American Society of Human Genetics 62nd Annual Meeting
November 6–10, 2012 San Francisco, California

PLATFORM ABSTRACTS

Tuesday, November 6

		Abstract Numbers
2. **4:30–6:30pm:** Plenary Abstract Presentations	Hall D	**#1–#6**

Wednesday, November 7
10:30am–12:45 pm: Concurrent Platform Session A (11–19):

11. Genetics of Autism Spectrum Disorders	Hall D	**#7–#15**
12. New Methods for Big Data	Ballroom 103	**#16–#24**
13. Cancer Genetics I: Rare Variants	Room 135	**#25–#33**
14. Quantitation and Measurement of Regulatory Oversight by the Cell	Room 134	**#34–#42**
15. New Loci for Obesity, Diabetes, and Related Traits	Ballroom 104	**#43–#51**
16. Neuromuscular Disease and Deafness	Room 124	**#52–#60**
17. Chromosomes and Disease	Room 132	**#61–#69**
18. Prenatal and Perinatal Genetics	Room 130	**#70–#78**
19. Vascular and Congenital Heart Disease	Room 123	**#79–#87**

Thursday, November 8
10:30am–12:45pm: Concurrent Platform Session B (29–37):

29. Next-Generation Sequencing: Methods and Applications	Hall D	**#88–#96**
30. Genetics and Intellectual Disability	Ballroom 103	**#97–#105**
31. GWAS from Head to Toe	Room 135	**#106–#114**
32. Cardiovascular Genetics: GWAS and Beyond	Room 134	**#115–#123**
33. Clinical Genetics: Mutations, Mutations and Syndromes	**Ballroom 104**	**#124–#132**
34. Cancer Genetics II: Clinical Translation	Room 124	**#133–#141**
35. Ethical, Legal, Social and Policy Issues	Room 132	**#142–#150**
36. Chipping Away at Autoimmune Disease	Room 130	**#151–#159**
37. Metabolic Disease Discoveries	Room 123	**#160–#168**

Thursday, November 8
4:30pm–6:45pm: Concurrent Platform Session C (38–46):

38. A Sequencing Jamboree: Exomes to Genomes	Hall D	**#169–#177**
39. Admixture and Demography	Ballroom 103	**#178–#186**
40. Analysis of Multilocus Systems	Room 135	**#187–#195**
41. Genes Underlying Neurological Disease	Room 134	**#196–#204**
42. Cancer Genetics III: Common Variants	Ballroom 104	**#205–#213**
43. Genetics of Craniofacial and Musculoskeletal Disorders	Room 124	**#214–#222**
44. Tools for Phenotype Analysis	Room 132	**#223–#231**
45. Therapy of Genetic Disorders	Room 130	**#232–#240**
46. Pharmacogenetics: From Discovery to Implementation	Room 123	**#241–#249**

Friday, November 9
8:00am–10:15am: Concurrent Platform Session D (47–55):

47. Structural and Regulatory Genomic Variation	Hall D	**#250–#258**
48. Neuropsychiatric Disorders	Ballroom 103	**#259–#267**
49. Common Variants, Rare Variants, and Everything in-Between	Room 135	**#268–#276**
50. Population Genetics Genome-Wide	Room 134	**#277–#285**
51. Endless Forms Most Beautiful: Variant Discovery in Genomic Data	Ballroom 104	**#286–#294**
52. Clinical Genetics: Complex Mechanisms and Exome-Discovery	Room 124	**#295–#303**
53. From SNP to Function in Complex Traits	Room 132	**#304–#312**
54. Genetic Counseling and Clinical Testing	Room 130	**#313–#321**
55. Mitochondrial Disorders and Ciliopathies	Room 123	**#322–#330**

Friday, November 9
4:30pm–6:45pm: Concurrent Platform Session E (59–67):

59. Genome Structure and Variation	Hall D	**#331–#339**
60. Advances in Neurodegenerative Disease	Ballroom 103	**#340–#348**
61. Missing Heritability, Interactions and Sequencing	Room 135	**#349–#357**
62. Exome Sequencing Uncovers Etiology of Mendelian Disease	Room 134	**#358–#366**
63. Transcriptional Regulation, Variation and Complexity	Ballroom 104	**#367–#375**
64. Epigenetics	Room 124	**#376–#384**
65. Advances in Ocular Genetics	Room 132	**#385–#393**
66. Cancer Genetics: Somatic Variants	Room 130	**#394–#402**
67. Developmental Insights into Human Malformations	Room 123	**#403–#411**

1

A novel molecular and functional mechanism predisposing to ototoxicity. *B. Wollnik[1], E. Pohl[1], N. Offenhäuser[2], A. Uzumcu[3], F.J. Kersten[4], A.K. Rzadzinska[5], O. Uyguner[3], B. Lorente[5], G. Nürnberg[6], M. Emiroglu[7], H. Kayserili[3], I. del Castillo[8], P. Nürnberg[6], T. Moser[9], C. Kubisch[10], K.P. Steel[5], P.P. Di Fiore[2], H. Kremer[4], Y. Li[1].* 1) Institute of Human Genetics, University of Cologne, Cologne, Germany; 2) IFOM, Fondazione Istituto FIRC di Oncologia Molecolar, Milan, Italy; 3) Department of Medical Genetics, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey; 4) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Welcome Trust Sanger Institute, Cambridge, Hinxton, UK; 6) Cologne Centre for Genomics, University of Cologne, Cologne, Germany; 7) Department of Otolaryngology, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey; 8) Unidad de Genética Molecular, Hospital Ramón y Cajal, Madrid, Spain; 9) InnerEarLab, Department of Otolaryngology, Göttingen University Medical School, Göttingen, Germany; 10) Institute of Human Genetics, Ulm University, Ulm, Germany.

While our knowledge about molecular mechanisms underlying Mendelian forms of hearing loss tremendously increased over the last years, the genetic basis and pathogenesis for drug induced hearing impairment remains unclear. Aminoglycosides are the most commonly used antibiotics worldwide. Although highly effective, their use is restricted by side-effects such as ototoxicity in a significant subset of patients. However, underlying pathogenesis and pharmacogenetic risk variants are largely unknown. Here we show that dysfunction of an actin remodeling protein (named here ARP) can result in a drug-inducible disturbance of actin dynamics and an irreversible hearing impairment in humans. By positional cloning, we identified a homozygous missense variant, p.L329P, in ARP as a cause of aminoglycoside-induced hearing impairment in a large consanguineous family from Turkey with 4 affected individuals. Complete ARP loss in knock out mice leads to hearing loss associated with shortened stereocilia. We demonstrate that the protein is a component of the tip complex that regulates stereocilia length and that it interacts with whirlin. The mutation severely impairs this interaction in vitro. Extensive biochemical studies showed that myosinXVa can stabilize the ARP-whirlin interaction complex, and we show for the first time that kanamycin has a negative effect on this complex formation, which is even more pronounced in mutant complexes, thereby explaining the development of hearing loss in affected individuals after aminoglycoside treatment. Taken together, we link ototoxicity after aminoglycoside treatment to actin dynamics and this finding will help in devising strategies to counteract this severe side-effect of aminoglycosides.

2

Genome-wide comparison of genetic and epigenetic regulatory mechanisms in primates. *Y. Gilad, A. Pai, R. Pique-Regi, C. Cain, J. Degner, N. Lewellen, K. Michelini, J. Pritchard.* The University of Chicago, Human Genetics, Chicago, IL.

Changes in gene regulation are thought to play an important role in adaptation and speciation, notably in primates. However, the extent to which changes in different regulatory mechanisms underlie gene expression evolution is not yet known. To address this gap, we comparatively characterized gene expression (using RNA sequencing) and genetic and epigenetic regulatory mechanisms in humans, chimpanzees, and rhesus macaques, using LCLs from 8 individuals from each species. Specifically, we used ChIP-seq to obtain genome-wide profiles of H3K4me3, H3K4me1, H3K27me3 and H3K27ac histone modifications, as well as binding of RNA polymerase II. We also collected DNaseI-sequencing from the same LCLs, and by using the CENTIPEDE algorithm we measured the strength of transcription factor binding for over 200 transcription factors in all three species. These data allowed us to identify both conserved and species-specific enhancer and repressor regulatory elements, as well as characterize similarities and differences across species in transcription factor binding to these regulatory elements. We found that that transcription factor binding and histone modifications in more than 67% of regulatory elements in putative promoter regions is conserved across the three species. In turn, by considering sequence conservation at genomic locations that showed differences in regulatory mechanisms across species we were able to better understand the extent to which changes in transcription factor binding are due to either cis- or trans- differences across species. Finally, we analyzed correlations between inter-species differences in the genetic and epigenetic regulatory mechanisms and variation in gene expression levels across species using a system of logistic regression models. Assuming that these correlations do imply a causal regulatory relationship, we estimate that up to 70% of inter-species gene expression differences can be accounted for by corresponding changes in transcription factor binding and/or the presence of histone modification marks.

3

Multidisciplinary and Translational Task Force for Neonatal Genomics. *E.E. Davis[1,2], A. Sabo[3], N.C. Oien[1], S.H. Katsanis[4], H. Cope[5], K. Sheets[6], A. Sadeghpour[1], K. McDonald[5], M. Kousi[1], J.R. Willer[1], J. Kim[4], S. Dugan-Rocha[3], D.M. Muzny[3], A. Ashley-Koch[5], E. Hauser[5], M. Hauser[5], J. Sun[2], J. Kurtzberg[2,7], A. Murtha[8], B. Boyd[8], W.B. Gallentine[9], R. Goldberg[2,10], M.T. McDonald[6], R.A. Gibbs[3], M. Angrist[4], C.M. Cotten[2,10], N. Katsanis[1,2].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) Department of Pediatrics, Duke University Medical Center, Durham, NC, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 4) Institute for Genome Sciences & Policy, Duke University, Durham, NC, USA; 5) Center for Human Genetics, Department of Medicine, Duke University Medical Center, Durham, NC, USA; 6) Division of Medical Genetics, Duke University Medical Center, Durham, NC, USA; 7) Department of Pathology, Duke University Medical Center, Durham, NC, USA; 8) Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA; 9) Division of Pediatric Neurology, Duke University Medical Center, Durham, NC, USA; 10) Neonatal and Perinatal Research Institute, Duke University Medical Center, Durham, NC, USA.

The accelerated implementation of genome-wide sequencing data as a first-pass diagnostic test is emerging as a powerful tool to secure a molecular diagnosis in the clinic. This has heightened the need for interpretive assays to determine the pathogenic potential of genetic variation. To address these challenges, and to capitalize on the opportunity to shorten the time to secure molecular diagnoses, we have created the Task Force for Neonatal Genomics at Duke University. The Task Force targets a uniquely vulnerable patient cohort: neonates, whose developmental anomalies are more likely to be within a timeframe for treatment. Anatomical defects are also amenable to functional analyses of multiple organ systems including central nervous system, renal, cardiac, craniofacial, skeletal, vascular, and skeletal muscle abnormalities. Our efforts target patients either prenatally in the Duke Fetal Diagnostic Center, or postnatally in the Duke Intensive Care Nursery and Pediatric specialty clinics and harness the full spectrum of clinical, genetics and cellular biology expertise, including the use of transient model organisms (primarily zebrafish). Here we report on 1) the interdisciplinary nature of our efforts; 2) our methodology for recruitment, data generation and analysis, and communication strategies between researchers and clinicians; 3) our analysis progress to date; and 4) our evolving approach to returning primary and secondary molecular findings to clinicians and family members. In our first year, we have screened > 50 referrals, enrolled 20 families, and have developed the capacity to enroll 50–100 families per year. In phenotype-appropriate patients, we couple whole exome sequencing of trios, a multi-tiered bioinformatic prioritization strategy, and functional modeling of novel variants of unknown significance in physiologically relevant vertebrate and cell-based models to inform allele pathogenicity. Successful examples include diagnosis of a channelopathy caused by a novel de novo SCN2A mutation in a severe epilepsy case, an early molecular diagnosis of Phelan-McDermid Syndrome, a novel mutation in X-linked spondyloepiphyseal dysplasia, and a proposed complex genetic mechanism for a hitherto unknown neonatal seizure syndrome. This initiative provides an unprecedented model for communication across an interdisciplinary research/clinical team with the ultimate goal of responsible and timely integration of new genetic technologies into clinical care.

4

Genome-wide Identification and Functional Analysis of Distant-Acting Craniofacial Enhancers. *C. Attanasio[1], Y. Zhu[1], M.J. Blow[2], A.S. Nord[1], V. Afzal[1], B. Hallgrimsson[3], D. FitzPatrick[4], H. Morrison[4], E.M. Rubin[1,2], L.A. Pennacchio[1,2], A. Visel[1,2].* 1) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA; 2) U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598 USA; 3) Department of Cell Biology & Anatomy, Faculty of Medicine, University of Calgary, Alberta, Canada; 4) MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Edinburgh.

The shape of the human face and skull is highly heritable, but the genetic factors that contribute to normal variation in craniofacial morphology remain poorly defined. Currently available data support a model in which a core set of craniofacial genes is controlled by arrays of non-coding elements that fine-tune their expression patterns. Sequence variation in these regulatory elements may contribute to normal phenotypic variation and represent risk factors for craniofacial birth defects. To determine the genomic location of craniofacial regulatory sequences and thereby enable their systematic exploration, we performed ChIP-seq analysis from embryonic day 11.5 mouse craniofacial tissues with the enhancer-associated protein p300, which identified 4,500 putative enhancers genome-wide. Approximately 100 of these sequences were analyzed in transgenic mouse reporter assays to validate their *in vivo* activities and to determine their precise spatial activity patterns at high resolution. Three-dimensional activity analysis of reporter staining by optical projection tomography (OPT) revealed a remarkable diversity in the developing craniofacial structures reproducibly targeted by different enhancers, consistent with a role in fine-tuning the expression patterns of their respective target genes. These ChIP-seq and transgenic enhancer catalogues are publicly available through FaceBase (www.facebase.org) and the Vista Enhancer Browser (enhancer.lbl.gov). To further evaluate the role of distant-acting enhancers in the development of craniofacial morphology, we selected three craniofacial enhancers, located near different genes involved in craniofacial development (*Msx1, Snai2, Isl1*), for deletion studies in mice. While none of the three enhancer deletions resulted in overt pathological phenotypes, normal mRNA expression of the respective target genes in mouse embryos was diminished in a spatial pattern consistent with enhancer reporter staining. Furthermore, micro-computed tomography (micro-CT) analysis of adult skulls demonstrated that all three enhancer deletions resulted in significant differences in skull morphology compared to wild-type controls. Taken together, these results indicate that thousands of distant-acting enhancers orchestrate gene expression during craniofacial development and are consistent with the notion that enhancers may contribute substantially to normal phenotypic variation and pathological aberrations of craniofacial morphology.

5

Translational cis-regulation of gene expression in human genome: the effect of human single nucleotide polymorphisms. *Q. Li[1], A. Makri[1], Y. Lu[1,2], L. Marchand[1], R. Grabs[1], M. Rousseau[1], H. Ounissi-Benkalha1[1], H. Qu[2], C. Polychronakos[1].* 1) Endocrine Genetics Lab, The Research Institute of the McGill University Health Center (Children's Hospital site) Montreal Qc, Canada; 2) The University of Texas School of Public Health, Brownsville Campus, Brownsville, TX.

Background: The variation in gene expression is heritable and can be mapped by quantitative trait locus analysis. Although methods exist for detecting the effect of single-nucleotide polymorphisms (SNP) on steady-state RNA (i.e. transcription or RNA stability), effects on translation have been studied only in a few isolated cases. We have developed a novel high throughput method that uses ribosomal association as proxy for translational efficiency of polymorphic mRNAs. As efficiently translated transcripts associate with multiple ribosomes while less active ones with fewer or none, we hypothesized that functionally allelic transcripts would show a detectable shift in this distribution. **Methods:** Cell lysates of lymphoblastoid cell lines (LCL) from 38 parents of the European HapMap set (CEU) were subjected to polyribosomal fractionation to separate soluble form polyribosomal RNA. Based on the poly-A RNA level, equal amounts of mRNA from nonpolysomal RNA vs. polysomal RNA of each LCL were evaluated for genome-wide expression profiling. The ratio of polysomal/nonpolysomal mRNA level was taken as the quantitative trait and tested for association with SNPs on the transcript. **Results:** A total of 98,363 exonic SNPs in 17,495 autosomal genes were evaluated, setting a Bonferroni threshold of 5×10^{-7}. By Spearman rank correlation, this threshold was met by 12 SNPs in 6 genes, while 71 SNPs in 27 genes met the false discovery rate Q-value<0.1. Among them, 5'UTR SNPs were enriched >2x (18%; vs. 8&%;). The top hit was rs1131017, in the 5'UTR of *RPS26*. Levels of its protein product, measured by quantitative Western blot and normalized for mRNA on the same sample, showed a significant (P=0.005) correlation with the number of G alleles in the 38 LCLs, validating this effect at the protein level. The same effect was also seen with in vitro translation of epitope-tagged allelic constructs (p= 0.002). **Conclusion:** Despite the small sample size, this proof-of-principle pilot has clearly shown that allelic effects on translation (validated,in one case,at the protein level) can be detected at a transcriptome-wide scale. We propose that this result represents the “tip of the iceberg” and that our novel method can be applied to larger sample sizes to cover an aspect of functional genomics that has not received much attention. It will add an important tool in the evaluation of genetic loci associated with complex disorders.

6

Lessons Learned from the NHLBI-Exome Sequencing Project. *S.M. Leal on behalf of the NHLBI-Exome Sequencing Project.* Molec/Human Gen, Baylor Col Med, Houston, TX.

The NHLBI-Exome Sequencing Project (ESP) was conceived to identify rare, putatively functional coding variants associated with heart, lung and blood related complex traits. To this end exome sequence data have been generated on >6700 individuals (4420 European Americans and 2312 African Americans) who are participants in several large cohort studies. Sequenced individuals included a deeply phenotyped random sample, quantitative trait extremes of interesting phenotypes (such as low-density lipoprotein (LDL), blood pressure, and body mass index) and disease cases (such as individuals with early onset myocardial infarction or stroke). Over 80 heart, lung, and blood phenotypes were available for analysis. Protein coding regions were sequenced to an average depth of 87X. Approximately 1.2 million coding variants were uncovered; of these ~1.1 million are rare (minor allele frequency<1%) and ~675,000 are rare and nonsynonymous. Given this is one of the largest medical sequencing studies ever undertaken, it was necessary to develop efficient methods and pipelines for variant calling, data quality control, analysis and interpretation of millions of single nucleotide variants from thousands of samples across multiple phenotypes. These methods allowed us to identify novel significant associations. For instance, we discovered and validated an association between variants in the *DCTN4* gene and risk of pseudomonas infection in individuals with cystic fibrosis (n= 91, $p<2.5\times10^{-6}$). We were also able to replicate several known associations across a number of traits [e.g., variants in the *APOB*gene are associated with LDL ($p<2.5 \times 10^{-6}$, n=3342) and the *LEPR* gene is associated with C-reactive protein ($p<2.5\times10^{-6}$, n=1791)]. Furthermore, we found multiple associations that are specific to a particular ethnicity [e.g., variants in the *DARC* gene are associated with white blood count ($p<10^{-13}$, n=1405 in African Americans) and the *F7* gene is associated with factor VII levels ($p<10^{-18}$, n=958 in European Americans)]. The ability to detect significant associations varied greatly by trait, depending not only on sample size but also on genetic architecture. The NHLBI-ESP provides novel information on the genetic etiology of several heart, lung and blood related traits, and guidance about using exome sequencing in population based cohorts to identify rare variants associated with complex traits.

7

Genetic Epidemiology of Autism Spectrum Disorder in a Cohort of over 11,000 Affected Sibships and Half-Sibships: Evidence of Genetic and Maternal Environmental Effects. *N. Risch[1,3], T.J. Hoffmann[1], M. Anderson[2], L.A. Croen[3], J. Grether[2], G. Windham[2]*. 1) Inst Human Gen, Univ California, San Francisco, San Francisco, CA; 2) CA Dept of Public Health, Richmond, CA; 3) Division of Research, Kaiser Permanente, Oakland, CA.

To examine genetic and environmental factors in autism, we rigorously linked California Department of Developmental Services records with state birth certificates to identify all sibs and half-sibs of individuals affected with autism spectrum disorder (ASD) born between 1990 and 2003. A total of 11,056 case sibships were identified, including 25,575 full sibs. Because more complete information was available for mothers, we were able to identify more maternal half sibs (3,276) than paternal half sibs (697). Using the birth certificate records to identify controls based on two-to-one matching to cases, we similarly identified a total of 22,891 control sibships. Using the singles method, the sib recurrence of ASD was 8.9% overall, and higher for male (13.7%) compared to female sibs (4.0%). We also noted a higher overall sib recurrence rate when the index case was female (12.1%) versus male (8.3%); these results are consistent with a multifactorial threshold model, with a higher threshold for females. For maternal half sibs, the overall recurrence (3.3%) was higher than for paternal half sibs (2.3%). We investigated stoppage by comparing reproductive behavior after the birth of an affected versus unaffected child. For the first few years after the birth of an ASD case, reproductive behavior was similar to controls; after 3 years, the probability of having another child dropped to 70%, and for subsequent children to 50% of the control rate. Hence, a more accurate recurrence estimate is derived by taking birth order into account, looking at sibs born sequentially after an affected. In that analysis, the overall sib recurrence was 10.1%; we observed that the risk to second born children is far higher (11.5%) than to later born sibs (7.3%); a similar observation was made in maternal half sibs (6.5% for second born compared to 3.0% for later born sibs; 4.8% overall). We also found a very significant effect of pregnancy interval, where the risk climbs to 17% for sibs born within one year of a prior affected sib, compared to 7% for sibs born at least four years after. An identical phenomenon was observed in the maternal half sibs. These results indicate a significant effect of maternal environment that was highest for second born children with a short inter-pregnancy interval. These results are very consistent with prior recent results for DZ twins, whose high recurrence is likely due to both genetic and shared maternal environmental factors.

8

Identifying inherited autism mutations using whole exome sequencing. *T.W. Yu[1], M.H. Chahrour[1], M.E. Coulter[1], S. Jiralerspong[2], K. Okamura-Ikeda[3], K. Schmitz-Abe[1], G.H. Mochida[1], J.N. Partlow[1], R.S. Hill[1], M. Al-Saffar[1], N.M. Mukaddes[4], A. Hashmi[5], S. Balkhy[6], G.G. Gascon[7], O. Oner[8], S. Al-Saad[9], T. Ben-Omran[10], L. Al-Gazali[11], V. Eapen[12], C. Stevens[13], S. Gabriel[13], K. Markianos[1], H. Taniguchi[3], N.E. Braverman[2], E.M. Morrow[14], C.A. Walsh[1]*. 1) Division of Genetics, Children's Hospital Boston, Boston, MA; 2) Department of Human Genetics and Pediatrics, McGill University, Montreal Children's Hospital Research Institute, Montreal, Quebec, Canada; 3) Institute for Enzyme Research, The University of Tokushima, Tokushima, Japan; 4) Istanbul Faculty of Medicine, Department of Child Psychology, Istanbul University, Istanbul, Turkey; 5) Armed Forces Hospital, King Abdulaziz Naval Base, Jubail, Kingdom of Saudi Arabia; 6) Department of Neurosciences and Pediatrics, King Faisal Specialist Hospital and Research Center, Jeddah, Kingdom of Saudi Arabia; 7) Clinical Neurosciences and Pediatrics, Brown University School of Medicine, Providence, Rhode Island, 02912; 8) Department of Child and Adolescent Psychiatry, Dr Sami Ulus Childrens' Hospital, Telsizler, Ankara, Turkey; 9) Kuwait Autism Center for Autism, Kuwait City, Kuwait; 10) Section of Clinical and Metabolic Genetics, Department of Pediatrics, Hamad Medical Corporation, Doha, Qatar; 11) Department of Paediatrics, Faculty of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates; 12) Infant, Child and Adolescent Psychiatry, University of New South Wales, Sydney, New South Wales, Australia; 13) Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA, 02142; 14) Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, Rhode Island, 02912.

To identify mutations that can account for the high heritability of autism spectrum disorders (ASDs), we applied high throughput sequencing to cases of familial ASD from consanguineous, multiplex pedigrees. In several cases we demonstrate a causative role for homozygous mutations in genes classically thought to cause neurometabolic disease and intellectual disability. For example, a Saudi family with three affected children was found to carry a homozygous, linked I308F missense change in AMT, a component of the glycine cleavage system. Mutations in AMT cause nonketotic hyperglycinemia(NKH), an autosomally inherited, neonatal-onset disorder associated with prominent seizures and early mortality, although rare, atypical forms of NKH have also been described with later onset and milder symptoms. One of the affected children in this family had a history of severe seizures as a neonate, but the remaining two had much milder symptoms with predominantly ASD as the clinical presentation. We prove biochemically that I308F is a hypomorphic allele that retains more residual activity than classic NKH-associated AMT mutations. We suggest that, in this manner, autism can be an unexpected manifestation of partial loss-of-function mutations in genes with more classically syndromic associations.

Extending these results, we analyzed whole exome sequencing data from a larger ASD cohort comprising >200 consanguineous families, screening them for point mutations in established human disease genes with known neurodevelopmental consequences. We uncover evidence for additional and clinically unexpected recessive mutations in genes known to cause neurometabolic disease (AMT, PEX7, MTRR, PAH), microcephaly (COH1, CEP152) and intellectual disability (HIST3H3). In some cases, phenotypic hallmarks of the monogenic condition could be recognized after variant identification by sequencing. These results underscore the power of whole exome sequencing in identifying clinically undiagnosed genetic conditions underlying a complex and heterogeneous disorder.

9

Identical by descent filtering in extended families reveals novel autism genes detected by exome sequencing. *H.N. Cukier[1], N.D. Dueker[1], S.H. Slifer[1], J.M. Lee[1], P.L. Whitehead[1], E. Lalanne[1], N. Leyva[1], I. Konidari[1], R.C. Gentry[1], W.F. Hulme[1], D. Van Booven[1], D.J. Hedges[1,2], V. Mayo[1], S.S. Ramsook[1], B.A. Barrionuevo[1], J.M. Jaworski[1], M.A. Schmidt[1,2], J.L. Haines[3], M.L. Cuccaro[1,2], J.R. Gilbert[1,2], M.A. Pericak-Vance[1,2].* 1) John P. Hussman Institute for Human Genetics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Autism spectrum disorders (ASDs) encompass a constellation of devastating neurodevelopmental conditions and studies to date demonstrate that the underlying etiology is extremely heterogeneous. Given this genetic complexity, we utilized extended families with multiple, distantly related, affected individuals which are likely to carry highly heritable risk factors to identify new ASD loci. We performed whole exome sequencing on at least two affected cousins across 40 multiplex ASD families to identify rare, segregating mutations that are incompletely penetrant. This includes 4 families with 5 ASD individuals, 7 families with 4 ASD individuals, and 8 families with 3 ASD individuals. A total of 164 individuals were captured with the Agilent SureSelect Human All Exon kit, sequenced on the Illumina HiSeq 2000, and the resulting data processed and annotated with BWA, GATK, and SeattleSeq. Each family had approximately 90,000 changes. Variants were filtered to those in identical by descent (IBD) regions delineated by SNP genotyping data. Initial analyses focused on novel and rare (MAF < 0.05) variants predicted to be detrimental, either by altering amino acids or splicing patterns. Validation was performed by Sanger sequencing and genotyping on the Infinium HumanExome BeadChip. In accordance with a dominant model of inheritance, this analysis identified 423 heterozygous changes, including those within known or suspected ASD genes (*AGAP1*, *CDH9*, *DLGAP2*, *NF1*, *NRCAM*, and *STXBP5*) and genes related to other neurological disorders including epilepsy (*CLCN2*), mental retardation (*CEP290*), and schizophrenia (*CSMD1*). One family carries 2 missense changes in *PRICKLE1*, a gene involved in neurite outgrowth. Twenty-two genes have variants in more than one family including *CIC*, a HMG-box protein previously implicated in neurodegeneration. Furthermore, three genes, *CXorf59*, *GLUD2*, and *SYN1*, had detrimental variants that followed an X-linked pattern of inheritance. The glutamate receptor *GLUD2* plays a key role in synaptic plasticity and the synaptic vesicle phosphoprotein *SYN1* has been previously linked to autism. In summary, we identified damaging variants in 10 genes known to be related to neurodevelopmental and neuropsychiatric disorders, and found novel putative ASD genes. By studying these unique pedigrees, we demonstrate that whole exome sequencing in extended families is a powerful source for the identification and verification of ASD genes.

10

The Discovery and Validation of Genes Recurrently Disrupted in Autism Spectrum Disorders. *B.J. O'Roak[1], L. Vives[1], A. Kumar[1], I.B. Stanaway[1], J. Egertson[1], E. Turner[1], C. Lee[1], G.L. Carvil[2], I.G. Phelps[2], D.R. O'Day[2], W. Fu[1], J. Hiatt[1], B. Martin[1], N. Krumm[1], B.P. Coe[1], R. Levy[1], E. Borenstein[1,3,4], D.A. Nickerson[1], H.C. Mefford[2], D.A. Doherty[2], J.M. Akey[1], R. Bernier[5], E.E. Eichler[1,6], J. Shendure[1].* 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 2) Department of Pediatrics, University of Washington School of Medicine, Seattle, WA; 3) Department of Computer Science and Engineering, University of Washington, Seattle, WA; 4) Santa Fe Institute, Santa Fe, NM; 5) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 6) Howard Hughes Medical Institute, Seattle, WA.

Autism spectrum disorders (ASD) have a strong genetic component, the etiology of which has yet to be fully elucidated due to extensive locus and allelic heterogeneity. It is widely speculated that sporadic (*de novo*) variants contribute significantly to the genetic basis of ASD. We are completing exome sequence of 1,658 individuals, including 134 trios (proband and parents) and 314 quads (an additional unaffected sibling) from 448 total simplex families. We have discovered over 100 ASD candidate genes based on *de novo* disruptions of loci unique to probands; however, most genes have only a single observed mutation. To prove pathogenicity, we developed a modified molecular inversion probe (MIP) based method enabling the ultra-low-cost resequencing of candidate loci in extremely large cohorts (amortized reagent costs <$1/sample/gene) with excellent sensitivity (99%) and positive predictive value (98%) at captured bases. We applied this method to 2,364 ASD probands from the Simons Simplex Collection (SSC), targeting 44 candidate genes (145 kilobases/sample). In the SSC cohort, we identified additional *de novo* truncating/splice-site disrupting events in *CHD8*, *DYRK1A*, *TBR1*, *GRIN2B*, *PTEN*, *ARID1B*, *ADNP*, *TBL1XR1*, and *CTNNB1*. Of note, 19 of the 44 genes screened intersect the previously reported 49-member protein-protein interaction (PPI) beta-catenin/chromatin remodeling network, which shows enrichment for autism candidate genes. Four additional genes intersect with an expanded PPI network of 74 genes, formed from over 400 SSC probands (O'Roak *et al.* 2012, Sanders *et al.* 2012). With the exception of *PTEN*, the genes with additional *de novo* disruptive events are all members of the expanded PPI network (binomial p<0.001). The most mutated locus was *CHD8*, a chromodomain helicase DNA binding protein linked to beta-catenin/Wnt regulation and neurodevelopment. In the combined MIP-exome dataset, 8/2,573 (~0.3%) probands carry a *de novo* nonsense, frameshift, or splice-site mutation in *CHD8*. Probands were notable for increased head circumference (6/8 with z-score >2, overall mean 2.3) and accounted for 2% of the children with ASD and macrocephaly—identifying a potential subphenotype for ASD classification. Our data strongly support a role for these nine genes in ~1% of sporadic ASD and implicate the beta-catenin/chromatin remodeling network in its etiology.

11

Rare complete human knockouts: population distribution and significant role in autism spectrum disorders. *E.T. Lim[1,2,3], M.J. Daly[1,2,3], ARRA Autism Sequencing Consortium.* 1) Analytic and Translational Genetics Unit, Mass General Hospital, MA; 2) Harvard Medical School, Boston, MA; 3) Broad Institute, Cambridge, MA.

It has been shown that an average healthy individual has ~20 genes that are inactivated by homozygous loss-of-function (LoF) variants. However, nearly all of these LoF variants are common variants (>5% allele frequency), and are found in a small subset of 100-200 apparently inessential genes, many of which are involved in chemosensation. In this study, we characterized the population distribution of low-frequency (≤5% allele frequency) homozygous and compound heterozygous (2-hit) LoF variants from exome sequence data in ~1000 cases with autism spectrum disorders (ASD) and ~1000 control exomes that were matched for ancestry.

We discovered a 2-fold enrichment of low-frequency 2-hit LoFs on the autosomes in cases versus controls and estimate that ~3% of cases may have complete LoF of a relevant autosomal gene contributing to their ASD phenotype. We further observed a similar 2-fold enrichment of rare hemizygous LoF variants on the X-chromosome in males with ASD compared to control males and estimate that ~2% of males with ASD have a hemizygous LoF in a relevant gene. This provides support for the role of rare complete knockouts to the genetic architecture of the disorder and provides insight into the genetic basis of the male gender bias observed in ASD.

12

Exome-based discovery of CNVs in Simplex Autism Families. *N. Krumm[1], B. Nelson[1], S. Girirajan[1], M. Dennis[1], C. Baker[1], M. Malig[1], NHLBI. Exome Sequencing Project[2], A. Quinlan[3], D.A. Nickerson[1], E.E. Eichler[1,4]*. 1) Dept. of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 2) National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD; 3) Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 4) Howard Hughes Medical Institute, Seattle, WA.

Autism Spectrum Disorder (ASD) is a common neurodevelopmental disorder with a strong genetic component, but individual genetic variants account for only a small fraction of cases. To enrich for pathogenic variants, recent studies have focused on “simplex” (i.e., sporadic) families from the Simons Simplex Collection, in which only one offspring has an ASD diagnosis. These studies have demonstrated increased de novo mutations in probands in comparison to their unaffected siblings; however, these de novo mutations are thought to explain only 10–15% of ASD cases. We used exome data from 342 ASD quads consisting of mother, father, proband and unaffected siblings and searched for disruptive genic CNVs using CoNIFER (Copy Number Inference From Exome Reads), a method for detecting CNVs from exome read-depth data (Krumm et al, 2012). Frequency for rare CNVs was estimated using 2,972 control exomes from the NHLBI ESP-GO project, also analyzed using CoNIFER. We focused on rare (<0.5% frequency) inherited events which were inherited by one offspring only. There were 242 of such singly-inherited rare CNVs, of which 79 (33%) were deletions and 137 (56%) disrupted genes (i.e., a breakpoint falling within the gene's CDS). Owing to the targeted nature of the exome-capture reaction, CoNIFER has nine-fold power over standard Illumina 1M SNP microarray platforms for the detection of genic events less than 10kbp in size. We exploited this increased sensitivity for small events to discover 26 CNVs which were previously missed either by Illumina SNP microarray assay or a custom Agilent array-CGH platform (with ~350k probes). The median size of these variants was 6kbp, and included a median of 4 exons. These CNVs disrupted several potentially pathogenic genes in probands, such as a 8kbp/5 exon disruptive duplication in *SNAP23*, involved in vesicle exocytosis, and 3kbp/3exon deletion in *ORC3*, a protein known to be involved in neural proliferation and memory in Drosophila. Our preliminary analysis suggests that inherited private CNVs play an important role in sporadic autism and emphasize the importance of finding smaller gene-disruptive events not routinely discovered using standard approaches.

13

Delta Catenin (CTNND2): genetics and function of a novel autism gene. *T. Turner[1,9], E. Oh[2], Y. Liu[2], M.X. Sosa[1,9], S. Sanders[3,9], K. Sharma[4], D. Moreno-De-Luca[5,9], T. Plona[6], K. Pike[6], D. Soppet[6], M.W. Smith[6], M. State[3,9], S.W. Cheung[7], C. Lese Martin[5,9], R. Huganir[4], E. Cook[8], N. Katsanis[2], A. Chakravarti[1,9]*. 1) Center for Complex Disease Genomics, Johns Hopkins University School of Medicine; 2) Center for Human Disease Modeling, Duke University; 3) Yale University; 4) Johns Hopkins University School of Medicine; 5) Emory University; 6) SAIC-Frederick, Frederick National Laboratory for Cancer Research; 7) Baylor College of Medicine; 8) University of Illinois at Chicago; 9) ACE Genetics Consortium.

We focused our autism gene discovery on individuals with the highest genetic liability, female-enriched multiplex families (FEMFs), rare autism families with at least 2 affected females. By whole exome sequencing of 10 unrelated FEMFs, who were the most severe in their family, we identified two rare, conserved to zebrafish, deleterious, heterozygous missense mutations in the delta-catenin gene (*CTNND2*). These mutations, G34S and R713C, were found in 2 of 10 individuals and are predicted to create a new phosphorylation site and affect protein-protein binding, respectively. Subsequently, we sought additional missense/nonsense changes by sequencing 22 RefSeq and 7 additional Ensembl exons in 361 affected females and 10 HapMap controls: we identified 5 new missense variants, 3 of which (G275C, R454H, T862M) are very highly conserved (S=PhyloP score >4), 1 (P189L) is moderately conserved (S>2), and 1 (P224L) is not highly conserved (S<2) but resides in a polyproline tract where FYN (Src receptor kinase) binds CTNND2. Exome sequencing in the Simons Collection yielded another change (Q507P: S>4) in a male with autism. We also identified 26 CNVs in various patient collections with 19 (73%) overlapping at least 1 *CTNND2* exon as compared to controls from DGV where there were 51 CNVs with none overlapping exons (P<2×10-12). To further explore the role of *CTNND2* in autism, we performed expression and functional studies. By examining 16 adult and 8 fetal human tissues we identified the highest expression to be in fetal brain indicating its importance in development. In addition, by looking at expression data across human development (Allen Brain Atlas), we discovered that genes positively correlated with *CTNND2* are enriched for cell morphogenesis whereas those negatively correlated are enriched for cell cycle function. We performed morpholino-based knock-down of ctnnd2 in zebrafish at 1 day post-fertilization and observed a significant convergence-extension phenotype that could be rescued by wild type human *CTNND2* but not by the autism G34S variant. Additional functional analysis by over-expression of wild type and G34S and R713C *CTNND2* alleles in rat primary hippocampal neurons and testing for effects on spine density and VGLUT1+ puncta are currently underway. Our results strongly suggest that *CTNND2* is a novel autism gene. (This work is funded by the Simons Foundation and NIMH grant R01MH081754).

14

Novel hotspots of recurrent copy number variation associated with autism spectrum disorder. *S. Girirajan[1], M.Y. Dennis[1], C. Baker[1], M.M. Malig[1], B.P. Coe[1], C.D. Campbell[1], K. Mark[1], T. Vu[1], C. Alkan[1], Z. Cheng[1], R. Bernier[2], E.E. Eichler[1,3]*. 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Psychiatry, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute, University of Washington, Seattle, WA.

Repeat architecture of the human genome predisposes certain regions to non-allelic homologous recombination (NAHR) resulting in copy number variants (CNVs). Over the last six years, our group and others have identified more than a dozen hotspots of recurrent microdeletion and microduplication associated with intellectual disability, autism, epilepsy, and schizophrenia. Most CNVs are large and mediated by large (>10 kbp) segmental duplications (SDs). We sought to explore the role that smaller CNVs (>5 kbp) flanked by smaller repeats might play in the pathogenicity of autism. We designed a custom targeted, high-density (a probe every 50 bp-1 kbp) microarray to identify CNVs in a total of 1,362 hotspot regions including 115 SD-mediated hotspots, 253 mini hotspots (flanked smaller SD blocks), 410 micro hotspots, and 584 Alu-mediated hotspots. We analyzed 2,684 individuals with sporadic or familial autism and 580 control individuals and compared our results with 19,747 population controls assessed using SNP microarrays. We discovered 34/115 SD-mediated hotspots with events, some of which were previously known to be associated with disease (e.g., 16p11.2, 1q21.1, and 17q12) as well as novel variants (e.g., 2q11.2, 12p11.23, 5p14.33, and 19p13.2), including a significant enrichment of duplications in several disease-associated regions including 1q21.1, 16p11.2, and the Williams-Beuren syndrome regions when compared to a developmental delay cohort. Among non-SD-mediated hotspots, we identified a total of 61/1,247 (5%) regions with rare CNVs in 78/2,588 (0.3%) of our cases. Of those events tested for inheritance, we find that only a small fraction are de novo (8/56) but the majority disrupt genes, including *CTNNA1*, *KDM4DL*, and *YWHAE*— known candidates for developmental delay. We also identified recurrent CNVs outside of hotspot regions that implicate interesting genes including three CNVs involving *HYDIN*— associated with hydrocephalus in mouse, and three CNVs involving *MBD5*— previously implicated with developmental delay and epilepsy. Our deep survey of regions of genetic instability has identified new candidates and recurrent rearrangements that confer risk to autism as well as other neurodevelopmental disorders.

15

Cluster Analysis Defines Subgroups of Phenotypic Expression for Autism Spectrum Disorders. *O.J. Veatch[1], B. Yaspan[1], N. Schnetz-Boutaud[1], M.A. Pericak-Vance[2], J.L. Haines[1]*. 1) Ctr Human Gen, Vanderbilt Med Ctr, Nashville, TN; 2) J.P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

Previous studies established a strong influence of genomic variation in the etiology of Autism Spectrum Disorder (ASD). However, effect sizes for associated regions are small and evidence from many analyses does not explain the heritability. Phenotypic heterogeneity is a likely reason for trouble identifying genetic factors. Our hypothesis is that subgrouping cases using phenotype data will increase the power to detect genes involved in ASD. We included Autism Diagnostic Interview-Revised (ADI), Autism Diagnostic Observation Schedule (ADOS) and Vineland Adaptive Behavior Scale (VABS) domain scores and ages, and head circumference (HC) measures for cases with an ASD diagnosis on both the ADI and ADOS. Spearman's correlation coefficients were calculated and variables were weighted according to the dataset structure. Principle Components Analysis predicts there are 15 data components. All variables, except HC and ADI RRB, have near equal contributions to component 1, explaining 26% of the phenotype variation. HC and ADI RRB contribute majorly to components 2 and 3, explaining another 26% of the variation. Using the agglomerative hierarchical method, the data clustered into 2 groups, one cluster representing cases with more severe ASD phenotypes. Kruskal-Wallis tests show that all variables, except HC and ADI RRB, are significantly different between the 2 clusters ($p<0.0001$). These clusters further divided into 6 subgroups; the more severe group into 2 and less severe into 4. Kruskal-Wallis tests show that the previously nonsignificant HC is very different between the 2 clusters in the severe group, while ADI RRB is very different between the 4 clusters in the less severe ($p<0.0001$). These data demonstrate that the importance of some variables is context-dependent. We calculated the odds of cases being assigned to the same cluster given a genetic relationship ($OR\approx1.50$; $p<0.00001$). We also estimated genetic relationships and compared relatedness within versus across clusters ($FST\approx0.15\pm0.26$). These data strongly suggest that the phenotypic clusters recapitulate genetic etiology. We validated cluster results via the Adjusted Hubert-Arabie Rand Index (AHARI) comparing the real data clusters to 1,000 data permutations (AHARI= 0.000027). All of the results for the initial dataset were replicated in an independent dataset. Our results suggest that more effective methods of phenotype definition will increase power to detect genetic factors influencing risk for ASD.

16

The detection, structure and uses of haplotype identity in population genetic data. *D. Xifara[1,2], I. Mathieson[1], G. McVean[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Statistics, University of Oxford, Oxford OX1 3TG, United Kingdom.

Extended regions of haplotype identity are the signature of recent common ancestry and are thus highly informative about diverse processes such as demographic history, the structure of recombination and the sharing of rare (untyped) variants. However, the identification of extended shared haplotypes is not straightforward, due to low levels of genotyping errors, particularly in sequence data. Moreover, current algorithms for estimating recent shared ancestry typically conflate the problem with haplotype estimation and are thus limited by the quality of phasing. Here, we introduce a model-based method for detecting extended haplotype sharing within a sample by identifying each individual's genealogical nearest-neighbours (NNs) at a given position along a chromosome. Our approach, which is related to the long-range phasing method of Kong et al. (2008), is robust to genotyping errors, and works directly from genotype data, thus avoiding the impact of phasing errors. We construct pseudo-trios at each locus along a chromosome, capturing local relatedness structure and identifying an individual's NNs as its surrogate parents. We describe two implementations of our model: a Hidden Markov Model approach which uses the Viterbi algorithm and incorporates LD information for the region, and an equivalent graph method that can be applied to large-scale data sets consisting of thousands of samples. The method leads naturally to an approach for statistical haplotype estimation, with performance comparable in accuracy to current methods. By applying the method to genome-wide SNP data from over 7,000 samples from the UK we show that the median length of haplotype identity is 1.8 cM, more than an order of magnitiude less than predicted by simple population genetic models, likely reflecting strong and recent population growth. Finally we show that individuals sharing maximal haplotype identity are actually not genealogical NNs more than 20% of the time, which has implications for the ability to impute rare variants.

17

Inferring and sequencing the founding bottleneck of Ashkenazim. *I. Pe'er[1], S. Carmi[1], S. Mukherjee[2], N. Parlamee[3], M. Bowen[4], K. Hui[4], V. Joseph[5], P.F. Palamara[1], L. Ozelius[6], I. Peter[6], A. Darvasi[7], K. Offit[5], H. Ostrer[8], J. Cho[4], L. Clark[3], G. Atzmon[8], T. Lencz[2], The Ashkenazi Genome Consortium.* 1) Dept Computer Sci, Columbia Univ, New York, NY; 2) Dept. of Psychiatry Research, Feinstein Institute of Medical Research, Manhasset, NY; 3) Dept of Pathology, Columbia Univ, New York, NY; 4) Dept. of Genetics, Yale University, New Haven, CT; 5) Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY; 6) Dept. of Genetics and Genomics Sciences, Mt. Sinai School of Medicine, New York, NY; 7) Dept. of Genetics, Hebrew University, Jerusalem, Israel; 8) Dept. of Genetics, Albert Einstein College of Medicine, Bronx, NY.

The Ashkenazi Jewish (AJ) population, currently including ~10 million individuals, has long been recognized as genetically isolated and therefore advantageous for genetic studies. Recent availability of GWAS data on thousands of AJ samples allows quantification of the isolation of this group, evaluation of its utility for sequencing studies, followed by pursuit of such WGS efforts. To perform such evaluation, we have developed novel methodology for inference of population genetic history based on the distribution of length and recent mutations in segments that are identical by descent (IBD), as observed by sequencing and SNP array data. We show such methodology to be uniquely effective in reconstructing recent demography, compared to previous methods more focused at pre-historic times. Applying this methodology to data from self-identified AJ samples, we show 85-90% of them belong to a genetic isolate related to other Mid-Eastern populations. This group has experienced an extreme bottleneck 30-35 generations ago, with subsequent expansion greatly exceeding the growth rate across all humans. Data are consistent with bottleneck size of merely 400 founders. This means that AJs are a relatively large group that is tractable for current sequencing throughput, with favorable study-size economics compared to other populations: several hundred individuals sequenced are expected to provide IBD segments to impute all common and rare variants in millions personal AJ genome, save the hundreds in each personal genome that are due to mutations in modern times. The Ashkenazi Genome Consortium (TAGC) has taken on this task with Phase I of the project now in progress, including 137 complete genomes of multiple-disease controls. Pilot TAGC samples show favorable QC measures (Ti/Tv=2.15±.003). We observe the total number of variants to be consistent with other European populations sequenced using the same platform and pipeline (SNV heterozygosity of 7.1×10^{-4}), but a significant increase in the fraction of novel heterozygote variants observed for SNVs (24% increase, $p<0.0017$) and other variants, as expected for a rapidly expanding isolated population, underrepresented in SNP databases. Variants detected in AJ samples, absent in a same-size group of European samples on the same platform tend to be shared across AJ samples, compared to European-detected variants, absent in AJ, consistent with a demographic effect of bottleneck in AJs, rather than with sequencing artifacts.

18

Statistical Methods for Association Test of Rare Variants Using Summarized Data without Individual-subject Information. *Q. Zhang, I. Borecki, M.A. Province.* Division of Statistical Genomics, Washington University School of Medicine, St Louis, MO.

Summarized data (usually in the form of contingency tables, such as case-control pooled sequence data and cohort allele frequency data from public databases) with no individual-subject information have been broadly used and reported in association studies of rare variants (RV). Most existing RV burden test methods, however, require individual-level data and therefore are inapplicable to summarized data. Methodological researches on the RV data with no individual information are very limited, leaving nearly untouched the question of how to perform a powerful collective test on a group of RVs based on summarized data. In this study we extensively investigate statistical properties of a few existing summarized-data-based RV test methods (EFT, CAST, C-ALPHA) and propose two novel methods (LRT, EPT). We demonstrate that the most commonly used EFT (exclusive frequency test) method produces incorrect p-values and can significantly diminish power. CAST produces correct p-values but cannot handle bidirectional-effect problem, whereas C-ALPHA considers bidirectional effects but tends to be greatly inflated with false positives. We compare statistical properties of these existing methods with our methods, a likelihood ratio test (LRT) method and an exact probability test (EPT) method. Extensive simulation shows that the EFT produces correct p-values and substantially increases power when there are bi-directional effects in data; LRT is slightly inflated when sample size is small (<1000), but outperforms most methods for large sample size (>3000). In terms of calculating correct p-value, EPT is the best one. Since EFT is time-consuming, we develop a fast version which can be used for the analysis of large scale data. Finally, we demonstrate the properties of these methods using real data. We also compare the power of these summarized-data based methods with individual-level-data based methods (CMC, SKAT), and discuss the gain and loss of power. Our study provides an informative guide as well as novel tools for RV association analysis using summarized data.

19

Testing for rare variant associations in the presence of missing data. *P. Livermore Auer[1], S. Leal[2], G. Wang[2], NHLBI Exome Sequencing Project.* 1) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

For studies of genetically complex diseases, many association methods have been developed to specifically analyze rare variants. These methods analyze rare variants in aggregate across a region which is usually a gene or transcript. When variant calls are missing, naïve implementation of rare variant association (RVA) methods can lead to inflated false positive rates as well as a reduction in power. For case-control data, only power will be reduced when the rate of missingness is equivalent between cases and controls; however, if the rates are differential between cases and controls there can be an increase in the false positive rate. For next generation sequencing data, it is inevitable that variant calls will be unavailable in specific genomic regions and sets of samples. Differential missingness can be caused when cases and controls are sequenced in batches that are subject to different experimental conditions (e.g., different capture arrays or sequencing machine) or if convenience controls are used from public repositories. Using sequence data from the NHLBI-Exome Sequencing Project we demonstrate that differential missingness can cause substantial increases in type I error, e.g. when 20% of the controls are missing variant calls and 0% for cases for an α=0.05 when analysis is performed using RVA methods such as the CMC (Li & Leal 2008), BRV (Morris & Zeggini 2010), WSS (Madsen & Browning 2009) and VT (Price et al. 2010) the type I error rates are 0.10, 0.10, 0.24 and 0.27 respectively. We developed extensions for four commonly used RVA tests (CMC-M, BRV-M, WSS-M and VT-M) and show that they control false positive rates without a reduction in power. Additionally, power to detect an association can be substantially improved by using the extended methods compared to removing variant sites with missing calls. When 1,000 cases and 1,000 controls are analyzed and 10% of the variant calls are missing for the simulated ALK gene which has 54 rare variant sites of which 75% are deemed to be causal with an odd ratio of 3.0 the power to detect an association (α=0.05) when analysis is performed using CMC-M, BRV-M, WSS-M and VT-M is 67%, 65%, 96% and 97%, respectively, however if those variant sites variant calls are removed and analysis is performed the power is ~0% for all methods. In order to maintain proper control of type I error without sacrificing power the extended RVA methods should be implemented when analyzing sequence or exome chip data.

20

Quantitative trait locus analysis for next-generation sequencing with the functional linear models. *M. Xiong[1], L. Luo[2], Y. Zhu[1]*. 1) Dept Biostatistics, Univ Texas Hlth Sci, Houston, TX; 2) Division of Epidemiology, Biostatistics and Preventive Medicine, University of New Mexico.

Next Generation Sequencing (NGS) technologies with faster and cheaper sequencing will generate so densely distributed data that genetic variants can be considered as observations varying over a continuum. This will dramatically facilitate the association studies of the entire allelic spectrum of genetic variation. Although in the past few years we have witnessed the rapid development of novel statistical methods for association studies of qualitative traits using NGS data, only a few statistics are proposed for testing the association of rare variants with quantitative traits. The QTL analysis of rare variants remains challenging. Analysis from low dimensional data to high dimensional sequence-based genomic data demands changes in statistical methods from multivariate data analysis to functional data analysis. In this report, we propose a functional linear model (FLM) as a general principle for developing novel and powerful QTL analysis methods designed for resequencing data. We have undertaken simulations to calculate the type I error rates and evaluate the power of the FLM and other eight existing statistical methods even in the presence of both positive and negative signs of effects. Since the FLM retains all of the genetic information in the data and explores the merits of both variant-by-variant and collective analysis and overcomes their limitation, the FLM has a much higher power than other existing statistics in all the scenarios considered. To further evaluate its performance, the FLM is applied to association analysis of six quantitative traits (QTL) in the Dallas Heart Study, and RNA-seq eQTL analysis with genetic variation in the low coverage resequencing data of the 1000 Genomes Project and gene expressions acquired by RNA sequencing (RNA-seq). The results of real data analysis shows that the FLM has much smaller P-values to identify rare variants associated with quantitative traits than other existing methods.

21

A rapid and powerful method for protein-protein interaction- and pathway-based association analysis in genome-wide association studies. *M. Li[1,2,3], S. Kwan[1,4], H. Gui[1], P. Sham[1,2,3]*. 1) Department of Psychiatry, The University of Hong Kong, Hong Kong, HK, Hong Kong; 2) Centre for Reproduction, Development and Growth, The University of Hong Kong, Hong Kong, HK, Hong Kong; 3) Centre for Genomics Sciences, The University of Hong Kong, Hong Kong, HK, Hong Kong; 4) Department of Medicine, The University of Hong Kong, Hong Kong, HK, Hong Kong.

It is well-known that the thousands of disease-susceptibility loci reported only explain a small proportion of heritability of their respective diseases. Recently quite a few studies have estimated that there remain a lot of unrevealed common variants which could account for much of the “missing heritability” but their effect size is often too small to pass the genome-wide significance level in conventional statistical analysis even in a reasonably large sample (say, more than 10,000 subjects). Properly introducing biological knowledge into the statistical analysis is a promising way to find out those small-effect genetic loci. Here, we developed a rapid and powerful method for protein-protein interaction (PPI)- and pathway-based association analysis in genome-wide association studies (GWAS). It seamlessly merged two existing set-based tests, extended Simes' test (known as GATES) and scaled chi-square test, as one independent statistical test to boost the power for the PPI- and pathway-based association analysis, and thus has the name of a HYbrid Set-based Test (HYST). Its `parental', GATES and scaled chi-square test adopt different strategies to combine association P values and can outperform each other when the number and linkage disequilibrium of SNPs vary. Computer simulation demonstrated that HYST had a reasonable type 1 error rate and was generally more powerful than its “parents” and other alternative tests (including VEGAS, PlinkSet and Logistic Kernel Machine Test) to detect a gene-set in which more than one gene is associated with the disease. We applied the method to three real complex disease datasets about Crohn's Disease, Rheumatoid arthritis and type-2 diabetes in the public domains. It took only ~10 to 30 min. to do a PPI- or pathway-based whole-genome association scan. It detected a number of highly-connected significant PPIs and pathways involving multiple confirmed disease-susceptibility genes not found in the conventional SNP- and gene-based association analyses. These results indicates that HYST can be effectively used to examine a collection of predefined SNP sets based on prior biological knowledge for revealing additional disease-predisposing genes of modest effects in GWAS. HYST has been implemented in a user-friendly software tool, named KGG (http://bioinfo.hku.hk/kggweb/), with integrated PPI and biological pathway resources for geneticists to insightfully explore the `missed' susceptibility genes of complex diseases.

22

Statistics for X-chromosome association. *U. Ozbek[1], D.E. Weeks[1, 2], W. Chen[3], J. Shaffer[2], S.M. Purcell[4, 5, 6], E. Feingold[1, 2]*. 1) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA; 4) Division of Psychiatric Genomics, Mount Sinai School of Medicine, New York, NY; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 6) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA.

Association between genotype and phenotype at autosomal loci is generally tested by chi-squared tests, or by regression models if there are covariates. For X-chromosome loci, it is not immediately obvious what the analogous statistics are. X chromosome data are often excluded from published analyses. A survey by Anastasia Wise found that only 32% of GWA papers from January 2010 through March 2011 analyzed the X chromosome (Wise AL 2011). Failure to analyze X data at all is obviously less than ideal, and can lead to missed discoveries. Even when the X chromosome is analyzed, it is usually done with suboptimal statistics. Several mathematically sensible statistics for X-chromosome association have recently been proposed. The optimality of these statistics, however, is based on very specific simple genetic models. In addition, while the simulation studies have been very informative, they have focused on single-marker tests and have not considered the types of error that occur when an entire chromosome is scanned. In this study, we comprehensively test the most promising X-chromosome association statistics using simulation studies that account for errors introduced by chromosome-wide testing. We also consider a wider range of models for sex differences and phenotypic effects of X inactivation. We found that many of the best statistics perform well even when there are variance differences between the sexes or small sex differences in allele frequency, but that unbalanced data can cause problems.

23

Joint association analysis of pleiotropy SNPs using GWAS summary statistics. *R.M. Salem[1,2,3], J.N. Hirschhorn[1,2,3]*. 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Department of Endocrinology, Boston Children's Hospital, Boston, MA.

Genome-wide association studies (GWAS) have been performed for a large number of traits and diseases. The majority of GWAS have focused on a single trait or disease, and if multiple phenotypes were analyzed, rarely are the phenotypes analyzed simultaneously. However, associated variants may have pleiotropic effects on multiple phenotypes, and pleiotropy is not considered in GWAS of a single phenotype. Pleiotropy could be important both to understand genetic architecture of complex traits and also to characterize more completely the underlying biology and phenotypic consequences of associated loci. We propose a method to systematically test and identify pleiotropic loci using GWAS summary statistics for the phenotypes of interest. The method builds on recently developed methods to jointly test multiple SNPs for association with a single phenotype, and allows approximation of joint association of multiple phenotypes with each potentially pleiotropic locus. The method uses GWAS summary-level statistics and estimated correlation of phenotypes of interest, to detect pleiotropic loci. In contrast to alternative methods, our approach explicitly considers the estimated regression coefficients from GWAS to detect pleiotropic loci, rather than relying solely on p-values. Our method is computationally fast and was applied to simulated GWAS data for both quantitative and dichotomous (case-control) phenotypes under several different models of genetic architecture with varying amount of pleiotropy. This new method can be used to leverage existing single phenotype GWAS data to study the role of pleiotropy in the genetic architecture of complex diseases.

24

Multivariate regression-based analysis of relative abundance data in metagenomics. *O. Libiger, N.J. Schork.* The Scripps Research Institute and Scripps Translational Science Institute, La Jolla, CA.

Recent advances in high-throughput genome sequencing have enabled the genomic characterization of entire microbial communities. By assessing the similarity among the sampled DNA sequences against previously identified sequences, the relative abundance of species within a community can be estimated. One can then address questions regarding e.g., the impact of microbial communities on human health by comparing abundance profiles obtained from several communities sampled at various locations or different times. While many methods exist for classification and clustering of metagenomic data, multivariate statistical methods for testing hypotheses regarding the differences in abundances have received little attention. The nature of the data poses several challenges: the abundance of different species may be correlated; while few species are usually common, a large number of them are often quite rare; and the quantity of different species often exceeds the number of samples. Multivariate regression-based approaches are well suited for this type of analyses. They can be used to quantify the collective effect of many species' abundances, they work within an established statistical framework that allows one to determine statistical significance of the results, and they are able to incorporate metadata as covariates in the model. We have assessed the utility of several multivariate regression-based techniques including distance-based regression, principal component regression, and regularized regression through simulation studies, and showcase their application in analyzing results of empirical studies including studies of microbial communities present on the palms of human hands.

25

Exome sequencing of more than 6,700 samples and the study of genetic susceptibility to common cancer. *A. Kiezun, A. McKenna, G. Kryukov, G. Getz.* Cancer Genome, Broad Institute, Cambridge, MA.

Despite tremendous progress, much of the genetic susceptibility to common cancer remains unknown. Previous work uncovered the complexity of the genetic architecture of common cancer: from rare variants of large effects to common variants of small effects. To date, genome-wide association studies have focused on common variants. In contrast, candidate gene studies have indicated that low frequency (1–5%) and rare (<1%) variants may have substantial impact on cancer risk. Lately, exome sequencing has emerged as a cost-effective way of discovering all coding variation but, in complex diseases, the study of rare variants is severely limited by statistical power and estimates show that over 10,000 samples will be required for robust association. Until recently, such large sample sizes were unattainable due to high cost.

In this study, we analysed whole-exome sequences of more than 6,700 samples (including more than 2,700 cancer cases from multiple tumor types) for association of low-frequency and rare coding variants with the cancer phenotype. For high data quality, we devised a specialized pipeline of variant calling and quality control. We called variants in all samples simultaneously using the BWA/Picard/GATK pipeline and performed stringent sample and variant QC. We analyzed (a) single-site associations of low-frequency variants, and (b) gene-based rare-variant burden using recently developed tests (Morris-Zeggini, Madsen-Browning, Variable-Threshold, Sequence Kernel Association Test). We controlled for population stratification and differential genotyping rate between cases and controls. We enriched the proportion of functional variants by PolyPhen2.

We present this ongoing study, its design, methods and preliminary results. Our findings demonstrate vast abundance of rare coding variation, significant inter-population differences in rare variants and the importance of carefully selecting variants for analysis. We also investigate the empirical distribution of effect sizes across the exome. We demonstrate the opportunities and challenges of exome sequencing for common cancer types. We show how coding variants can be analyzed to dissect genetic predisposition to cancer and we emphasize that many thousands of samples are required for robust detection of underlying genes.

26

Exome sequencing of families severely affected with breast cancer suggests eight new candidate genes: *ATR, BAP1, CHEK1, GEN1, KANK4, OBSL1, RAD51B* and *TP53BP1*. *C.H. Spurrell, A.M. Thornton, M.K. Lee, S. Casadei, S. Ng, T. Walsh, J. Shendure, MC. King.* University of Washington, Seattle, WA.

Families severely affected with breast cancer, but with wildtype sequences for all known breast cancer genes, were evaluated by exome sequencing of genomic DNA from 2 or 3 affected sisters or cousins per family, then all rare truncating mutations tested for co-segregation with breast cancer in the family as a whole. Analysis of the first 45 families reveals rare or private truncating mutations in eight genes. Six of these genes — *ATR, BAP1, CHEK1, GEN1, RAD51B,* and *TP53BP1* — function in biological pathways related to homologous recombination repair, as do *BRCA1* and *BRCA2*. *RAD51B* is a likely candidate as a breast cancer gene, given that inherited mutations in *RAD50*, *RAD51C*, and *RAD51D* have all been observed in breast or ovarian cancer. Similarly, *CHEK1* is a likely candidate, given the role of *CHEK2*. *BAP1* has an important role in inherited predisposition to melanoma; *BAP1* nonsense mutations in two breast cancer families are similar to the mutations seen in melanoma families. *GEN1*, a key Holliday junction resolvase, has been discounted as a breast cancer gene, because the critical truncating mutation, near the C-terminus, was equally common in cases and controls. However, we have identified a truncation much nearer the N-terminus co-segregating with breast cancer in a large family. Two other genes are new to cancer genetics. *KANK4* encodes an ankyrin repeat protein that may function as a membrane metallopeptidase. *OBSL1* encodes a cytoskeletal adaptor protein highly expressed in mammary glad and ovary. Recessive loss-of-function mutations in *OBSL1* cause 3M, a primordial growth disorder syndrome. In order to test whether other familial breast cancer patients carry mutations in these genes, we are employing a high-throughput Molecular Inversion Probe (MIP) approach. With MIPs, we capture all exons and splice junctions of each gene, then multiplex 384 DNA samples per lane for sequencing on our HiSeq platform. This approach allows us to efficiently investigate all candidate genes and identify those in which multiple rare mutations of functional significance co-segregate with breast cancer in multiple high-incidence families.

27

Rare variants in XRCC2 as breast cancer susceptibility alleles. *F.S. Hilbers[1], M.C. Völker-Albert[1], W.W. Wiegant[2], M.P.G. Vreeswijk[1,2], N. Hoogerbrugge[3], J.C. Oosterwijk[4], J.M. Collee[5], M.C. Southey[6], P. Peterlongo[7], P. Radice[7], F.J. Couch[8], K. Offit[9], I.G. Campbell[10], J. Benitez[11], C.J. van Asperen[12], H. van Attikum[2], P. Devilee[1].* 1) Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands; 2) Toxicogenetics, Leiden University Medical Centre, Leiden, The Netherlands; 3) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 4) Department of Clinical Genetics, University Medical Center Groningen, Groningen, The Netherlands; 5) Department of Clinical Genetics, Erasmus MC, University Medical Center, Rotterdam, The Netherlands; 6) Department of Pathology, The University of Melbourne, Melbourne, Australia; 7) Fondazione Istituto FIRC di Oncologia Molecolare (IFOM), Milan, Italy; 8) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, USA; 9) Cancer Genetics Laboratory, Memorial Sloane Kettering Cancer Center, New York, USA; 10) VBCRC Cancer Genetics Laboratory, Peter MacCallum Cancer Centre, Melbourne, Australia; 11) Human Genetics Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; 12) Department of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands.

Germline mutations in XRCC2 have recently been identified in a small number of breast cancer cases with a positive family history for the disease. Since XRCC2 is involved in DNA repair through homologous recombination and was recently identified as a Fanconi anemia gene, a role as a risk factor for familial breast cancer seems plausible. We analyzed the coding regions of XRCC2 in an international cohort of 3548 non-BRCA1/2 familial breast cancer cases and 1435 controls. Rare variants in XRCC2 were detected in 20 familial cases (0.6%) and 10 controls (0.7%). The only truncating variant was detected in a control. Polyphen and other in silico prediction tools predict that of the 17 missense variants, 9 were possibly or probably damaging, while the remaining 8 were benign. These categories were evenly distributed between cases and controls. To examine the effect of the detected variants on XRCC2 cellular function, human XRCC2 cDNA-constructs with these variants were transfected into the XRCC2-deficient hamster cell line IRS1. The degree of complementation was studied by scoring RAD51 foci formation after mitomycin C treatment, which is an indirect measurement of homologous recombination. Preliminary results show a correlation with functionality as predicted by Polyphen. More definitive data of this and other assays of homologous recombination will be presented at the meeting. Although the genetic association results suggest that XRCC2 variants do not confer risk of breast cancer, functional analyses may still reveal these extremely rare variants to impact XRCC2 function, and hence to play a role in breast cancer development.

28

HOXB13 is a susceptibility gene for prostate cancer: Results from the International Consortium for Prostate Cancer Genetics (ICPCG). *K. Cooney[1], W. Isaacs[2], J. Xu[3], E. Lange[4], L. Lu[3], S. Zheng[3], Z. Wong[3], L. Cannon-Albright[5], J. Stanford[6], E.A. Ostrander[7], C. Maier[8], J. Schleutker[9,10], D. Schaid[11], S. Thibodeau[12], G. Cancel-Tassin[13], F. Wiklund[14], R. Eeles[15], D. Easton[16], A. Wittemore[17], G. Giles[18], W. Catalona[19], D. Mandal[20], W. Foulkes[21], J. Carpten[22], D. Seminara[23] on behalf of The International Consortium for Prostate Cancer Genetics*. 1) Depts Internal Med & Urology, Univ Michigan, Ann Arbor, MI; 2) Dept Urology, Johns Hopkins Univ, Baltimore, MD; 3) Center for Cancer Genomics, Wake Forest Univ School of Med, Winston-Salem, NC; 4) Depts Genetics and Biostatistics, Univ North Carolina, Chapel Hill, NC; 5) Dept Med, Univ Utah School of Med, Salt Lake City, UT; 6) Div Public Health Sci, Fred Hutchinson Cancer Research Center, Seattle, WA; 7) Natl Human Genome Research Inst, NIH, Bethesda, MD; 8) Inst Hum Gen, Univ Ulm, Ulm, Germany; 9) Inst Biomed Tech, Univ Tampere, Tampere, Finland; 10) Dept Med Biochem and Gen, Univ Turku, Turku, Finland; 11) Dept Health Sciences Research, Mayo Clinic, Rochester, MN; 12) Dept Lab Med and Path, Mayo Clinic, Rochester, MN; 13) UPMC Univ, Paris France; 14) Depts Med Epid and Biostat, Karolinska Inst, Stockholm, Sweden; 15) Inst Cancer Research and Royal Marsden NHS Foundation Trust, Surrey, UK; 16) Centre Cancer Genetic Epid, Dept Oncology, Univ Cambridge, Stangeways Lab, Cambridge, UK; 17) Dept Health Research Policy, Stanford Comp Cancer Center, Stanford School Med, Stanford, CA; 18) Centre Mol, Env, Gen, Anal Epid, Univ Melbourne, Melbourne, Australia; 19) Northwestern Univ Feinberg School Med, Chicago, IL; 20) Dept Genetics, Louisiana State Univ Health Sciences Center, New Orleans, LA; 21) Program in Cancer Genetics, McGill Univ, Montreal, Canada; 22) Translational Genomics Research Inst, Phoenix, AZ; 23) Nat Cancer Inst, Nat Inst Health, Bethesda, MD.

A rare but recurrent mutation (G84E) in *HOXB13* was recently reported to be associated with prostate cancer risk. To examine this finding in a large international sample of prostate cancer families, we genotyped this mutation and 14 other SNPs in or flanking *HOXB13* in 2,443 prostate cancer families recruited by 15 groups of the International Consortium for Prostate Cancer Genetics (ICPCG). At least one mutation carrier was found in 112 families (4.6%), all of European descent. The proportion of families with subjects carrying the G84E mutation differed significantly across the ICPCG groups ranging from 0% to 22.4%. Within the carrier families, the G84E mutation was more common in men with a diagnosis of prostate cancer (194 of 382, 51%) than those without (42 of 137, 30%), P=9.9×10-8. The odds ratio for prostate cancer was 4.42 (95% confidence interval=2.56–7.64). A family-based association test found that the G84E mutation was significantly over-transmitted from parents to affected offspring, P=6.5×10-6. The prostate cancers in patients who carried the mutation had a wide spectrum of clinical characteristics including high risk disease. Analysis of flanking markers revealed that the G84E mutation resides in a rare haplotype in 95% of carriers suggesting a founder effect. Overall, these findings confirm that *HOXB13* G84E is a rare mutation found in approximately 5% of hereditary prostate cancer families in different, mainly European, populations where it is associated with prostate cancer risk. Future studies of prostate cancer in families harboring this rare allele will be needed to define the clinical utility of this observation.

29

Parkinson disease and melanoma: a common genetic pathway linked to PARKIN inactivation. *H. Hu[1,2,3], N. Dumaz[1,3], S. Lesage[4], L. Michel[1,3], V. Descamps[1,3,5], S. Mourah[3,6], C. Lebbé[1,3,7], N. Basset-Seguin[1,3,7], M. Bagot[1,3,7], A. Bensussan[1,3], L. Deschamps[3,8], MT. Leccia[3,9], A. Tsalamlal[2,3], P. Sivaramakrishna Rachakonda[10], S. Klebe[11], K. Rajive[10], C. Kannengiesser[2,3], A. Couvelard[3,8], B. Grandchamp[2,3], L. Thomas[12], A. Brice[4,13], N. Soufir[1,2,3]*. 1) Inserm U976, Centre de Recherche sur la Peau, Hôpital Saint Louis, AP-HP, Paris, France; 2) Department of Genetic, Hôpital Bichat Claude Bernard, AP-HP, Paris, France; 3) Université Paris 7, Paris, France; 4) Université Paris 6, Inserm UMR_S975, Centre de Recherche de l'Institut du Cerveau et de la Moelle Epinière, Hôpital Pitié-Salpêtrière, AP-HP, Paris, France; 5) Dermatology Department, Hôpital Bichat Claude Bernard, AP-HP, Paris, France; 6) Inserm U940, Laboratoire de Pharmacologie, Hôpital Saint Louis, AP-HP, Paris, France; 7) Dermatology Department, Hôpital Saint Louis, AP-HP, Paris, France; 8) Inserm CRB3, Département de Pathologie, Hôpital Beaujon, AP-HP, Paris, France; 9) Dermatology Department, CHU Grenoble, Grenoble, France; 10) Division of Molecular Genetic Epidemiology, German Cancer Research Centre, Heidelberg, Germany; 11) Department of Neurology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany; 12) Dermatology Department Hôpital de l'Hôtel-Dieu, Lyon, France; 13) Département de Génétique, Cytogénétique et Embryologie, Groupe Pitié-Salpêtrière, AP-HP, Paris, France.

Background: Parkinson's disease (PD), a neurodegenerative disease characterized by loss of melanin positive dopaminergic neurons is epidemiologically linked to cutaneous melanoma (CM). PARKIN (PARK2), the most important gene implicated in genetic predisposition to PD, encodes an E3 ubiquitin ligase involved in cyclin E degradation that has also been shown to be a tumor suppressor gene. We therefore investigated the role of PARK2 on melanoma susceptibility and oncogenesis. Patients and methods: The whole PARK2 coding region was sequenced in 500 CM patients recruited from oncogenetic survey and that matched one of following criteria: familial CM, or multiple CM or CM <25 years old. In addition, the presence of CNVs in PARK2 was investigated by MLPA, qPCR, and CGH. The frequencies of point mutations and CNVs in PARK2 was investigated in 320 matched healthy controls, and obtained for 2060 additional controls derived from ten publications in which PARK2 has been exhaustively studied (sequencing and CNV analysis) and were used as controls. Statistical analysis was carried out by comparing PARK2 abnormalities between patients and controls. Results: We identified 15 different PARK2 inactivated alleles (2 splicing, one frameshift, and twelve exonic CNVs in exons 2, 4, 8, 9, 10, 11 and 12 -all leading to a premature stop codon) in 16 melanoma patients (3.2%), whereas only such abnormalities were only present in 0.6% controls of both groups (P<0.0001 OR=5.56 [2.49–12.55]). In addition, PARK2 CNVs were present in 60% of melanoma cell lines and in 50% of primary tumors, often associated with BRAF or NRAS mutations, and LOH was detected in 50% melanoma cell lines. Western Blot analysis showed that PARKIN protein was absent in 90% of melanoma cell lines whereas present in melanocytic cells. Finally, transfection of wild type PARK2 cDNA markedly decreases cell proliferation of three melanoma cell lines. Conclusion: Here, we point out a common genetic pathway that could explain the epidemiological association between PD and CM, by showing that PARK2 inactivation plays an important role both in melanoma predisposition and oncogenesis. This provides new insights in CM oncogenesis that could be helpful for targeted therapy design, and may have clinical implications in genetic forms of PD (skin examination and photo-protection).

30

Exome Sequencing in Families at High Risk for Lymphoid Malignancies. *L.R. Goldin[1], M.L. McMaster[1], M. Rotunno[1], K.B. Jacobs[2], L. Burdette[2], M. Malasky[2], A. Hutchinson[2], M. Cullen[2], J. Boland[2], M. Yeager[2], M.A. Tucker[3], S.J. Chanock[2], N.E. Caporaso[1].* 1) Genetic Epidemiology Branch, DCEG/NCI, Bethesda, MD; 2) Core Genotyping Facility, DCEG/NCI, Bethesda, MD; 3) Human Genetics Program, DCEG/NCI, Bethesda, MD.

Single high penetrance genes that account for for familial hematologic malignancies have not been identified. Our group has accrued biospecimens and medical data for families at high risk for chronic lymphocytic leukemia (CLL), Hodgkin lymphoma (HL), and non-Hodgkin lymphomas (NHL). Although these conditions show strong familial aggregation, previous linkage studies did not identify significant loci causing susceptibility. We have conducted exome sequencing in 50 cases or obligate carriers from 17 families. We have sequenced two or more individuals per family, choosing the most distantly related cases as possible. Nimblegen v2 was used for library capture and sequencing was performed with the Illumina Hi Seq 2000. We identified high quality, non-synonymous variants shared by all cases in a family, filtering out those with population frequencies > 0.01. This screening identified an average of 48 variants per family. Further annotation was conducted and priority was also given to variants that are predicted to have a functional consequence and high conservation. After also selecting genes likely to be involved in immune function or carcinogenesis, 10–15 genes per family remained as candidates for further follow-up. In general, there were very few genes that were identified in more than one family suggesting that heterogeneity is extensive. In one CLL family, a non-synonymous SNP (rs 56400844) in the *CXCR4* (chemokine receptor 4) was shared by patients. This gene is in a linkage region identified in a previous study and is known to be differentially expressed in CLL tumors. Rare germ line variants in this gene have been previously described in CLL patients. Additional annotation shows that the identified variant is rare (0.06%) and is in a conserved, regulatory region. Further follow-up will be conducted to determine how much this gene accounts for familial CLL or other lymphoid malignancies. Analogous follow up procedures are in progress for the other identified variants.

31

Rare allelic forms of *PRDM9* associated with childhood leukemia. *J. Hussin[1,2], D. Sinnett[1,2], F. Casals[2], Y. Idaghdour[2], V. Bruat[2], V. Saillour[2], J-C. Grenier[2], J. Healy[2], T. de Malliard[2], J-F. Spinella[1,2], M. Lariviere[2], S. Busche[3], G. Gibson[4], A. Andersson[5], L. Holmfeldt[5], J. Ma[5], L. Wei[5], J. Zhang[5], G. Andelfinger[1,2], J.R. Downing[5], C.M. Mullighan[5], P. Awadalla[1,2].* 1) University of Montreal, Montreal, Canada; 2) Ste-Justine Hospital Research Centre, Montreal, Canada; 3) Department of Human Genetics, McGill University, Montreal, Canada; 4) Center for Integrative Genomics, School of Biology, Georgia Institute of Technology, Atlanta, Georgia; 5) St. Jude Children's Research Hospital, Memphis, Tennessee.

One of the most rapidly evolving genes in humans, PRDM9, is a key determinant of the distribution of meiotic recombination events. Mutations in this gene have previously been associated with male infertility and recent studies suggest that PRDM9 might be involved in pathological genomic rearrangements. As part of a genomic study of 111 families with children affected by acute lymphoblastic leukemia (ALL), we characterize the meiotic recombination patterns within a family with two siblings having hyperdiploid childhood ALL and observed unusual localization of maternal recombination events. The mother of the family carries a rare PRDM9 allele (the C allele), explaining the unusual patterns found. We further utilized exome sequencing data to discover a substantial excess of rare allelic forms of PRDM9 in 44 additional parents of children affected with ALL. The rare PRDM9 alleles are transmitted to the affected child in half of the cases, nonetheless there remains a signficant excess of rare alleles among patients relative to controls. We successfully replicated this latter observation in an independent cohort of 50 children with B-ALL, where we found an excess of rare PRDM9 alleles in aneuploid and infant B-ALL patients. As PRDM9 variability in humans has been suggested to influence genomic instability, PRDM9 histone H3K4 methyltransferase activity in the parental germline potentially modifies the risk of acquiring aneuploidies or genomic rearrangements associated with childhood leukemogenesis.

32

***De novo* mutation of the TGF beta family in early-onset ovarian cancer.** *I. Tournier[1], F. Charbonnier[1,2], S. Coutant[1], K. Walton[3], R. Marlin[1], M. Vezain[1], J. Tinat[1,2], E. Angot[4], R. Sesboué[1], J-C. Sabourin[1,4], D. Vaur[1,5], C. Harrison[3], T. Frebourg[1,2].* 1) Inserm U1079, University of Rouen, Institute for Research and Innovation in Biomedicine, Rouen, Normandy, France; 2) Department of Genetics, University Hospital, Rouen, Normandy, France; 3) Growth Factor Signalling Laboratory, Prince Henry's Institute, Clayton, Australia; 4) Department of Pathology, University Hospital, Rouen, Normandy, France; 5) Department of Genetics, Comprehensive Cancer Centre, Caen, Normandy, France.

In cancer genetics, the general dogma is that sporadic cancers result from the accumulation of somatic mutations whereas familial cancers are due to germline mutations followed by somatic mutations. The underestimated *de novo* mutation rate in humans recently revealed by exome analyses prompted us to investigate the contribution of *de novo* mutations in sporadic early-onset cancer. We used as a paradigm ovarian cancer without detectable mutation in known genes. We performed a comparative exome analysis in a young patient who developed, at 21 years of age, a sporadic early-onset ovarian adenocarcinoma with peritoneal metastases. The exome from the non tumoral ovarian tissue was captured and analysed on an Illumina platform. The 15715 detected SNVs and small Indels were then filtered using the EVA (Exome Variation Analyzer) software developed by our team to exclude non genic or intronic variants and variants present in the 131dbSNP database, the 1000 genome project and the 5379 exomes established by the NHLBI Grand Opportunity Exome Sequencing Project (ESP). The inherited genetic variations were then subtracted from the remaining mutations by comparing the exome of the index case with that of her parents in order to identify the potential deleterious *de novo* mutation. After this filtering scheme, the only remaining mutation with a high quality score was a mutation affecting one of the TGF beta family members expressed by the granulosa cells and known to be involved in ovarian development. Targeted Sanger sequencing confirmed the presence of the mutation in the ovarian tissue and blood lymphocytes of the index case and the absence of the mutation in the parents indicating that this *de novo* mutation had most likely occurred at the pre-zygotic level. Sequencing of this gene in 15 other cases of early-onset ovarian adenocarcinoma (19–30 years), without detectable *BRCA* mutation, most of them fitting with a sporadic presentation revealed another genetic variant in one patient who had developed an ovarian adenocarcinoma at 19 years of age. This variant has already been reported in the ESP, but at a low allelic frequency (0.002). Experiments are underway to determine whether the detected mutations abnormally activate the TGFbeta signalling. Our results provide new insights into the genetic determinism of ovarian cancer and provide a further example of an inherited form of cancer due to aberrant signals from stromal cells to epithelial cells.

33

Somatic activating mutations in PIK3CA cause progressive segmental overgrowth. *M.J. Lindhurst[1], V.E.R. Parker[2], F. Payne[3], J.C. Sapp[1], S. Rudge[4], J. Harris[2], A.M. Witkowski[1], Q. Zhang[4], M.P. Groeneveld[2], C.E. Scott[3], A. Daly[3], S.M. Huson[5], L.L. Tosi[6], M.L. Cunningham[7], T.N. Darling[8], J. Geer[9], Z. Gucev[10], P.A. Kreiger[11], V.R. Sutton[12], M.M. Thacker[13], C. Tziotzios[14], A.K. Dixon[15], T. Helliwell[16], S. O'Rahilly[2,17], D.B. Savage[2,17], M.J.O. Wakelam[4], R.K. Semple[2,17], I. Barroso[2,3], L.G. Biesecker[1].* 1) The National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) The University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Cambridge, UK; 3) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK; 4) The Babraham Institute, Babraham Research Campus, Cambridge UK; 5) Genetics Unit, Manchester Academic Health Science Centre, Manchester, UK; 6) Division of Orthopaedics, Children's National Medical Center, Washington DC; 7) Division of Craniofacial Medicine, University of Washington School of Medicine, Seattle, WA; 8) Department of Dermatology, Uniformed Services University of the Health Sciences, Bethesda, MD; 9) Greenwood Genetics Center, Greenwood, SC; 10) Department of Endocrinology and Genetics, Skopje Medical Faculty, Skopje, Macedonia; 11) Department of Pathology, A.I. duPont Hospital for Children, Wilmington, DE; 12) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 13) Department of Orthopaedic Surgery, A.I. duPont Hospital for Children, Wilmington, DE; 14) Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK; 15) School of Clinical Medicine, University of Cambridge, Cambridge, UK; 16) Liverpool Cancer Research UK Center, University of Liverpool, Liverpool, UK; 17) The National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge, UK.

The phosphatidylinositol-3-kinase (PI3K)/AKT signaling pathway is critical for cellular growth and metabolism. Mutations in key genes have been identified in numerous tumor samples, while loss of PTEN function or activation of AKT1, AKT2 or AKT3 have been implicated in disorders that feature overgrowth and/or hypoglycemia. We performed exome sequencing of DNA from affected and unaffected skin fibroblasts from a patient (C1) with unclassified severe overgrowth of the lower extremities and identified a cancer-associated variant in *PIK3CA* in DNA from the affected sample that was not present in the unaffected sample. This variant, c.3140A>T which predicts p.His1047Leu, was also found in DNA isolated from other affected tissues from patient C1 including muscle, bone, fibrous and adipose tissue. We then screened affected cells and tissues from 15 patients with clinical features that overlapped with patient C1 and found the p.His1047Leu variant in two patients and a second variant, p.His1047Arg, in eight patients. The mutation burden amongst the samples harboring these variants ranged from <1% to 50% but neither variant was found in any blood samples from these patients. The predominant finding in this cohort was segmental progressive overgrowth of subcutaneous, muscular and visceral fibroadipose tissue with skeletal overgrowth that was sometimes, but not always, distorting. The severity varied remarkably and ranged from massive overgrowth of both legs to overgrowth limited to two rays of one foot. Despite having mosaic overgrowth that is both progressive and sporadic, these patients did not meet the clinical criteria for Proteus syndrome. Their features more closely match those of CLOVES syndrome, however, these patients lacked the complex truncal vascular malformations that are commonly found in patients with CLOVES. We tested affected tissue from two patients that met the criteria for CLOVES syndrome and found a *PIK3CA* p.Glu542Lys variant in one patient and *PIK3CA* p.Glu545Lys in the other. These variants are also activating mutations commonly found in tumor samples. Finally, in a patient with an isolated congenital linear verrucous epidermal nevus, we found *PIK3CA* p.Glu545Lys in keratinocytes but not fibroblasts isolated from the lesion. These findings expand the spectrum of phenotypes associated with somatic activation of PI3K signaling and suggest multiple therapeutic targets for patients with progressive segmental overgrowth.

34

ChipEnrich: gene set enrichment testing for ChIP-seq data. *R.P. Welch[1], C. Lee[1], L.J. Scott[2], R.A. Smith[1], P. Imbriano[2], M.A. Sartor[1]*. 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI USA; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI USA.

Gene set enrichment testing is a method to identify pre-defined sets of genes that contain more experimentally relevant genes than would be expected by chance. This methodology was originally developed for the analysis of gene expression data, and has been adapted to new types of genome-wide data. Here we investigate the application of gene set enrichment testing to ChIP-seq data, specifically the locations of peaks called from piled up next-generation sequencing reads. There exist a number of challenges in applying gene set enrichment methods to this type of data. ChIP-seq peaks must be assigned to a gene, and given that no exhaustive database of gene regulatory domains exists, we must use a heuristic approach of assigning peaks to the nearest gene, the nearest TSS, or other locus definition. We define a gene locus as the region of the genome in which a peak would be assigned to a given gene. The length of a gene locus acts as a confounder, in that genes with longer locus lengths are more likely to have peaks assigned to them by chance, and therefore gene sets with longer gene loci on average will be detected as enriched. A proper test of gene set enrichment must adjust for gene locus length, as well as other potential confounders such as the mappability of the sequence in the locus. We developed a method called Chip-Enrich that empirically corrects for locus length and optionally mappability using a logistic regression model with smoothing spline terms for each covariate. We compare our method to two existing methods, Fisher's exact test (FET) and GREAT, on a number of experimental ChIP-seq datasets from the literature. We illustrate a number of issues in using these existing methods, and show that our method properly corrects for the bias introduced by locus length and mappability regardless of the transcription factor binding profile. We also confirm that Chip-Enrich correctly identifies the known biology of each transcription factor, and in some cases, is able to do so where other methods cannot. Chip-Enrich will be available as a Bioconductor R package that provides the user with: 1) the ability to test their data using Fisher's Exact test, ChIP-Enrich, or the binomial test used by GREAT, 2) 15 different annotation databases containing over 20,000 gene sets, 3) multiple methods of assigning peaks to genes, and 4) visualizations of various aspects of the ChIP-seq data profiles.

35

Enhanced exome sequencing to capture genome-wide common variants. *I.C.R.M. Kolder[1], K.I. Morley[1], E. Birney[2], I. Dunham[2], J.C. Barrett[1]*. 1) Dep. Human genetics, Wellcome trust sanger institute, Hinxton, Cambridgeshire, United Kingdom; 2) European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom.

Because it is not yet economically feasible to sequence high depth complete genomes in 1000s of individuals as part of complex disease studies, exome sequencing (focusing on only the coding part of the genome) has become a widely adopted alternative. Exome sequencing first uses a “pull-down” reagent to enrich a DNA library for only those sequences in target areas, followed by high depth genome sequencing. Recent studies have also shown that by using both the high depth areas in the target regions, as well as low depth “off target” areas, genotype imputation can be be used to provide reasonably high quality genotype calls at common variants across the genome in addition to the complete sequence data in the target areas.

We evaluate several extensions to this approach, including: (i) potential larger pull down targets which would include both coding exons and prioritized regulatory regions, such as Transcription Factor binding sites and selected DNaseI hypersensitivity sites; (ii) a similar functionally focused pull down which also prioritizes regions not in LD with previously selected regions in order to maximize genomic coverage; (iii) a mixture of pull down and shotgun reads from the whole genome. We have performed power simulations comparing these approaches to traditional exomes, whole genomes and GWAS, considering both genic and non-genic causal alleles. We will discuss the possibility of replacing traditional GWAS chips with these reagents.

36

Complete HIV-1 Genomes from Sequencing Single Molecules: Simple and Complex Samples. *M.P.S. Brown[1], M. Schaefer[2], Y. Gao[1], W. Kilembe[3], S. Allen[2], E. Hunter[2], E.E. Paxinos[1]*. 1) Pacific Biosciences, Menlo Park, CA, CA; 2) Emory University, Atlanta, GA; 3) Zambia Emory HIV Research Project, Lusaka, Zambia.

We sequenced complete HIV-1 genomes from single molecules using Single Molecule, Real Time (SMRT®) sequencing and derive de-novo full length genome sequences. SMRT® sequencing yields long-read sequencing results from individual DNA molecules with a rapid time-to-result. These attributes make it a useful tool for continuous monitoring of viral populations. The single molecule nature of the sequencing method allows us to estimate variant subspecies and relative abundances by counting methods. We detail mathematical techniques used in viral variant subspecies identification including clustering distance metrics and mutual information. Sequencing was performed in order to better understand the relationships between the specific sequences of transmitted viruses in linked transmission pairs. Samples representing HIV transmission pairs were selected from the Zambia Emory HIV Research Project (Lusaka, Zambia) and sequenced. We examine Single Genome Amplification (SGA) prepped samples and samples containing complex mixtures of genomes. Whole genome consensus estimates for each of the samples were made. Genome reads were clustered using a simple distance metric on aligned reads. Appropriate thresholds were chosen to yield distinct clusters of HIV genomes within samples. Mutual information between columns in the genome alignments was used to measure dependence. In silico mixtures of reads from the SGA samples were made to simulate samples containing exactly controlled complex mixtures of genomes and our clustering methods were applied to these complex mixtures. SMRT® sequencing data contained multiple full-length (greater than 9Kbp) continuous reads for each sample. Simple whole genome consensus estimates easily identified transmission pairs. The clustering of the genome reads showed diversity differences between the samples, allowing us to characterize the diversity of the individual quasi-species comprising the patient viral populations across the full genome. Mutual information identified possible dependencies of different positions across the full HIV-1 genome. The SGA consensus genomes agreed with prior Sanger sequencing. Our clustering methods correctly segregated reads to their correct originating genome for the synthetic SGA mixtures. The results open up the potential for reference-agnostic and cost effective full genome sequencing of HIV-1.

37

DeTCT pipeline: a software pipeline for the analysis of transcript count data. *J.A Morris, J.E Collins, I. Sealy, N. Wali, E. Busch-Netwich, R. White, D.L Stemple, J.C Barrett.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

We have developed a software pipeline for the analysis of data produced using DeTCT (Differential expression transcript counting technique). DeTCT is a high-throughput sequence-based approach for generating transcriptome profiles from total RNA samples. DeTCT uses polyA pull down to enrich for the 3' end of fragmented transcripts and Illumina sequencing to produce transcript count data.

The DeTCT pipeline first identifies the set of transcript 3' ends it will use to measure expression across samples; this step is performed using the mapped sequence data alone with no other annotation information. This feature of the pipeline is very useful for organisms that do not have a mature reference sequence, as it minimises the loss of sequencing reads which map outside of known transcripts. The transcript 3' ends are identified using a modified version of HPeak, a ChIP-Seq peak finding hidden Markov model based algorithm. The pipeline next processes the input data to generate counts at each transcript 3' end for each sample. At this stage the pipeline is capable of removing reads that are likely the result of a PCR duplication event, thus reducing the number of false positives in the final results. A sequence of random bases incorporated in each read pair allows us to identify candidate PCR duplicate reads which are then tested against the background distribution of the random base sequences from the entire experiment to produce a statistic that can be used as a cut off to remove likely PCR duplicates. Differential expression analysis is performed using the Bioconductor package DESeq with the final analysis results exported in a simple text format. As the pipeline is designed in a modular manner it is trivial to replace the methods used in any step with your favoured approach, such as in the differential expression analysis where implementing another Bioconductor package to analyse the count data would be simple and straightforward.

Where possible, steps in the pipeline have been coded in a manner that makes it simple to parallelise allowing users to easily exploit multiple computers and CPUs wherever possible in order to reduce analysis time. The DeTCT pipeline is open source software and is available to download from http://sourceforge.net/projects/detct/files/.

38

Fast Genome-Wide QTL Association Mapping with Pedigrees. *H. Zhou[1], E.M. Sobel[2], K. Lange[2,3]*. 1) Dept of Statistics, North Carolina State University, Raleigh, NC; 2) Dept of Human Genetics, UCLA, Los Angeles, CA; 3) Dept of Biomathematics, UCLA, Los Angeles, CA.

Genome-wide association studies (GWAS) have identified many common genetic variants associated with complex diseases and traits. However, it is widely accepted that these common variants explain at most a fraction of the genetic variation of most complex traits. This has resulted in a renewed interest in linkage-type analysis to detect rare variants. Family designs also allow for control of population stratification and study of parent-of-origin effects. The burgeoning next-generation sequencing (NGS) tools will increase the power of family-based and rare-variant approaches. However, pedigree likelihoods are notoriously hard to compute and current software for association mapping in pedigrees is prohibitively slow for dense maps.

We propose two ultra fast score tests for association mapping with pedigree-based GWAS or NGS data. Our two models differ by how association is parameterized. In the mean effect model, the two alleles of a SNP shift trait means. There is no confounding of association and linkage. In the variance effect model, correlations between trait values of different people are driven in part by local SNP sharing. The degree of sharing is measured through an identity by state (IBS) covariance matrix computed from the observed SNPs within a SNP group (usually a gene). Linkage and association are confounded in the variance component model. Within a pedigree, linkage is predominantly detected. Across pedigrees, association comes to the fore. The confounding of the two forces allows either to implicate the gene or SNP group.

We demonstrate with real and simulated data how fast and well both models perform. The mean component model has more power if a single common SNP explains trait variation. The variance component model is more appropriate when numerous rare SNPs in a gene impact trait values. Both models assume Gaussian variation of the trait. This framework carries with it several advantages. First, it works for random sample data, pedigree data, or a mix of both. Second, it enables covariate adjustment and correction for population stratification. Third, it accommodates both univariate and multivariate traits. Fourth, it fosters both likelihood ratio tests and score tests. Both models are now implemented in our comprehensive software package **Mendel** for easy use by the genetics research community.

39

Discovering SNPs Regulating Human Gene Expression Using Allele Specific Expression from RNA-Seq data. *E. Eskin[1], E. Kang[1], B. Han[1], A.J. Lusis[2], L. Martin[2], S. Shiffman[3]*. 1) Dept Computer Science, Univ California, Los Angeles, Los Angeles, CA; 2) Dept Human Genetics, Univ California, Los Angeles, Los Angeles, CA; 3) Genetics, Hebrew University, Jerusalem, Israel.

The study of the genetics of gene expression is of considerable importance to understand the nature of common, complex diseases. The main approach to understand the relationship between genetic variation and gene expression are the expression quantitative trait loci (eQTL) studies. Here we propose an alternative and complementary approach to identifying cis-acting regulatory variants by analyzing al- lele specific expression (ASE) in the human genome. Utilizing ASE has several advantages over traditional eQTL studies such that the measurements of ASE are more resistant to the confounding factors such as environmental or trans-acting factors which alter overall levels of expression. In this paper, we propose a novel analytical approach for identifying cis-acting variants which affect gene expression by ana- lyzing ASE. The proposed ASE mapping approach aims for the ASE mapping utilizing the ASE measurements from the RNA-seq data. We make two fundamental assumptions for our ASE mapping method. First, we can measure allele specific expression (ASE) accurately. Sec- ond, we can make a discrete ASE call for each individual(ASE or no ASE). In principle, these assumption is very realistic, since we can col- lect the large amount of data using RNA-seq. The proposed methodology based on these reasonable assumptions search for the candidate causal SNP set explaining the observed discrete ASE callings. We also propose the natural way to characterize the efficacy of the proposed ASE mapping called “reduction rate”. The key concept is that the efficacy of the ASE mapping depends not only the number of individuals but the number of allele specific expressed individuals. This method is easily and widely applicable to any RNA-seq data. For evaluation of the method, we used a RNA sequencing data of 54 lymphoblastoid cell lines derived from unrelated Nigerian individuals which we have the phased SNP data through HapMap release 22 data. We consider 2267 number of human ensembl genes which has at least one ASE individual. For 614 (27%) transcripts out of 2267, we found that either one variant or multiple variants in perfect linkage disequilibrium can explain the ASE pattern of the individuals.

40

Association of genetic variation affecting exon skipping to disease susceptibility. *Y. Lee[1], H. Im[2], W. Hernandez[3], NJ. Cox[1,4]*. 1) Department of Medicine, University of Chicago, Chicago, IL, 60637, USA; 2) Department of Health Studies, University of Chicago, Chicago, IL, 60637, USA; 3) Division of Biological Sciences, University of Chicago, Chicago, IL, 60637, USA; 4) Department of Human Genetics, University of Chicago, Chicago, IL, 60637, USA.

It is estimated that 15–50% of all human heritable diseases are affected by variants in canonical splice sites or splicing regulatory elements. Such variants have been shown to contribute to Mendelian diseases as well as tumor progression and susceptibility to cancer. For example, splice-changing intronic variants have been identified to be associated with lung adenoma prognosis, increased prostate cancer risk, and retinoblastoma. However, most studies have considered only the genetic determinants of AS. In this study, we examined intronic variants in splicing regulatory elements (SREs) as a mechanism to understand how intronic variants contribute to disease. We identified candidate SNPs for affecting exon skipping by analyzing sequence-driven AS models and intronic splicing enhancer (ISE) elements in the genome by scanning the genome for regions with putative ISE motifs having intronic SNPs within the elements. To experimentally validate our candidate ISE SNPs, we utilized exon-level expression microarrays of 176 HapMap lymphoblastoid cell lines. We then investigated the association of ISE-changing candidate SNPs within known cancer genes. We found 78 SNPs in ISEs that were associated with skipped exons in 40 known cancer genes, which is a statistically significant enrichment among SNPs associated with cancer in the NHGRI catalog (empirical p=0.031, odds ratio=1.54). One of the alleles at each of seven ISE SNPs within the cancer genes *NFAT5*, *GNA12*, and *DC53* effectively neutralize ISE motifs and were also associated with a moderate decrease in exon expression (linear regression, p<0.001, R^2>0.06). Furthermore, one of the ISE SNPs in *NFAT5* is in complete linkage disequilibrium (LD) with a SNP previously associated with early age of menarche (age at onset, LD r^2=1), a risk factor for breast and endometrial cancers. Three ISE SNPs in *GNA12* were also found to be in LD with SNPs associated in the risk of ulcerative colitis (LD r^2=0.954), a condition known to increase the risk of developing colorectal cancer. GWAS have accelerated the discovery of SNPs associated with risk of complex diseases, with a large number of these SNPs found in intronic regions. Our study identified several intronic SNPS that may regulate AS influence thereby increasing risk of common diseases.

41

Haplotype-based variant detection and interpretation enables the population-scale analysis of multi-nucleotide sequence variants. *E. Garrison[1], J.A. Rosenfeld[2], D. MacArthur[4, 5], Y. Xue[3], Z. Iqbal[6], S. Balasubramanian[7, 8], L. Habegger[8], R. Poplin[5], M. DePristo[5], G. Marth[1], M.B. Gerstein[7, 8, 9], C. Tyler-Smith[3], The 1000 Genomes Project.* 1) Biology, Boston College, Chestnut Hill, MA 02467; 2) IST/High Performance and Research Computing, University of Medicine and Dentistry of New Jersey (UMDNJ); 3) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs. CB10 1SA, UK; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston MA 02114; 5) Broad Institute of Harvard and MIT, Boston MA 02142; 6) The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK; 7) Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA; 8) Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520, USA; 9) Department of Computer Science, Yale University, New Haven, Connecticut 06520, USA.

Fully understanding the functional impact of genetic variation requires a determination of whether nearby variants are found on the same parental chromosome, also known as &ldquo∞ phase.” A pair of variants discovered in a single exon, transcription factor binding site, or splice junction may have dramatically different effects if they are in phase than if they are found as separately segregating alleles.

Current high-performance sequence variant detection methods apply a single-locus-based detection model to short-read sequencing alignment data. This approach to primary variant detection requires the reconstruction of phasing information post hoc using computationally expensive statistical approaches that depend on information combined across many samples from the same population. However, short-read sequencing data itself actually contains information that can be used to directly assess the co-localization of nearby sequence variants on the same haplotype.

We have developed methods to determine local phase between variants directly from short-read sequencing data. We applied these detection methods to investigate block substitutions of multiple consecutive bases, which we term multi-nucleotide polymorphisms (MNPs), in the 1094 low-coverage human genomes in Phase I of the 1000 Genomes Project. On average, we detect ~18,000 MNPs per genome. Most MNPs are short, with a rapid falloff in frequency with increasing length. Most MNPs in coding sequence have functional impact, which we have assessed by extending functional annotation methods to MNPs.

The generation of this large set of MNP calls has enabled the observation of mutational pathways that generate multi-base substitution effects. In many cases, we observe intermediate SNP alleles between the ancestral and derived MNP alleles. Our observations of these intermediate alleles and their frequencies indicate that the most common mutational pathway for 2bp MNPs (CA to TG via CG) is driven by cytosine deamination. In addition, we discuss the identification of other mutational mechanisms driving this class of polymorphism.

42

eQTL analysis identifies novel associations between genotype and gene expression in the human intestine. *B. Kabakchiev[1,2], NIDDK IBD Genetics Consortium[3], M. S. Silverberg[1,2].* 1) SLRI, Mount Sinai Hospital, Toronto, ON, Canada; 2) University of Toronto, Toronto, ON, Canada; 3) Yale University, New Haven, CT, USA.

Background: Genome-wide association studies have been pivotal to increasing our understanding of intestinal disease. However, the mode by which genetic variation results in phenotypic change remains largely unknown, with many associated polymorphisms likely to modulate gene expression. Here we report a comprehensive expression quantitative trait (eQTL) scan of intestinal tissue. Methods: Endoscopically and histologically normal tissue biopsies from the ileum of patients recruited at Mount Sinai Hospital in Toronto were obtained and preserved in RNAlater. Total RNA was extracted with the QIAGEN miRNeasy Kit and mRNA analysis was performed on Affymetrix Human Gene 1.0 ST arrays. DNA from the same individuals was genotyped using the Illumina Human OmniExpress platform. Cis-eQTL analysis (±50Kb around each gene) was carried out on 173 subjects encompassing the expression levels of 19,047 unique autosomal genes listed in the NCBI database and over 600K dbSNPs. This work was done in a custom software pipeline and the Kruskal-Wallis test was used to compare expression values across different genotypes. False discovery rate correction for multiple testing was applied at an alpha level of 5%. Results: The presence of 15,091 statistically significant cis-eQTLs associated with 2,629 genes was detected with this analysis. eQTLs associated with the same expression trait were in high linkage disequilibrium. Comparative analysis with other eQTL studies showed that 30.2%, 36.9%, 30.4%, 29.3% and 32.8% of genes identified as eQTLs in monocytes, hepatic tissue, lymphoblastoid cell lines, T-cells and fibroblasts respectively are also eQTLs in ileal tissue. Conversely, the majority of significant cis-eQTLs are novel and possibly tissue specific. These pertain to many aspects of cellular function from division to antigen processing and presentation. Our analysis confirmed previously published cis-eQTLs that are also inflammatory bowel disease associated SNPs: rs2298428 / UBE2L3, rs1050152 / SLC22A4 and SLC22A5. Conclusion: eQTL analysis of intestinal tissue substantiates reports in the literature that some eQTLs remain stable across cell types while many others are specific to the sampled location. Our findings not only confirm, but also significantly expand the number of known genotypes associated with expression and could help elucidate the mechanisms of intestinal disease.

43

A genome-wide association analysis of early-onset severe obesity: the SCOOP project. *E. Wheeler[1], N. Huang[1], E. Bochukova[2], S. Lindsay[1], J. Keogh[2], R.J.F. Loos[3], N.J. Wareham[3], S. O'Rahilly[2], M. Hurles[1], I. Barroso[1,2], I.S. Farooqi[2,4].* 1) Wellcome Trust Sanger Institute, Cambridge, UK; 2) University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 3) MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 4) National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK.

Obesity is a major public health problem with substantial impacts on morbidity and mortality. Considerable progress has been made in the discovery of multiple novel monogenic forms of this disorder, and more recently, common genetic variants have been reproducibly associated with body mass index (BMI) and obesity. However, these genetic variants only account for <5% of the variance in BMI. We performed a genome-wide association analysis in a cohort of 1,509 children with severe (>3 standard deviations of the BMI distribution), early-onset (onset below 10yrs) obesity and 5,380 publicly available WTCCC2 controls (http://www.wtccc.org.uk/ccc2/), all of UK Caucasian ancestry. Those SNPs with supporting evidence for association (N=29) were further evaluated in an additional 971 severely obese children and 1990 adult controls. A meta analysis of the discovery and follow-up stages identified four new loci with common and low-frequency alleles associated with severe obesity (LEPR, PRKCH, PACS1, RMST) and confirmed others identified previously (FTO, MC4R, TMEM18 and NEGR1). The association signal at LEPR, a known monogenic obesity locus, supports the idea that both common and rare variants can be involved in the pathogenesis of obesity at certain loci. We also show that the previously reported common, large (43kb) deletion at the NEGR1 locus, significantly associated with obesity risk, was entirely driven by a flanking smaller (8kb) deletion, where absence of the deletion increases risk of obesity. The smaller deletion at this locus removes a conserved transcription factor binding site for NKX6.1, involved in neuronal development in the mid and hindbrain. Comparison of our results with those already published for BMI, suggest that the genetic architecture of severe obesity does differ from that of more common obesity.

44

Mapping obesity traits using an integrated `omics' approach in adipose tissue from female twins. *A.K. Hedman[1], J.K. Sandling[2], E. Grundberg[2,3], K.S. Small[2,3], E. Meduri[2,3], S. Keildson[1], A. Nica[4], A. Buil[4], J.T. Bell[3], J. Nisbet[2], M. Sekowska[2], A. Wilk[2], A. Barrett[5], N. Hassanali[5], T-P. Yang[2], D. Glass[3], S-Y. Shin[2], L. Parts[2], N. Soranzo[2,3], R. Durbin[2], K. Ahmadi[3], K.T. Zondervan[1], C.M. Lindgren[1], T.D. Spector[3], E.T. Dermitzakis[4], M.I. McCarthy[1,5,6], P. Deloukas[2] for the MuTHER Consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 4) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 5) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Churchill Hospital, Oxford, UK; 6) Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK.

Epigenetic patterns and transcription are influenced by genetic and environmental factors, and both vary with obesity. We explored the associations between epigenetics, transcription and obesity-related traits utilizing mRNA expression (N = 825) and genome-wide DNA methylation (DNAm) profiles assayed on HumanMethylation450 BeadChips (N=648) from subcutaneous abdominal adipose tissue taken from well phenotyped female twins from the MuTHER study. We integrated such obesity-related patterns with genotypic information and disease loci to separate genetic and environmental contributions to the trait.

We focused analyses on the 34,685 (10%) most variable methylated sites and 23,596 mapped expression probes and found 34% (P<1.44E-06) of DNAm sites and 18.8% (P<2.12E-06) of transcripts to be associated with obesity-related traits. The largest number of associations, 29.7% of DNAm loci and 16.1% of transcripts, were for fat distribution trait DEXA-derived percentage trunk fat mass (PTFM). 28% of PTFM-associated DNAm were annotated to PTFM-associated genes, and for 348 of these, DNAm and gene expression were also directly associated (FDR 1%), suggesting that the effect may be mediated through changes in gene expression. Notably, PTFM-associated genes annotated to top PTFM-associated DNAm sites were primarily involved in metabolic processes, and their expression previously suggested to vary in obesity and under diet-induced stimuli.

We found PTFM-associated transcripts, but not DNAm, to be enriched (P<5E-08) for genetic control (higher heritability or direct control by SNPs in *cis*) compared to all profiled transcripts. Restricting analysis to PTFM-associated DNAm annotated to PTFM-associated genes, we found 14 SNPs involved in both transcript and DNAm variability. We sought to separate causal and reactive effects by integrating the genetic variants regulating PTFM-associated transcripts or DNAm in *cis* (eSNPs) with loci associated with relevant obesity traits, BMI or BMI-adjusted waist-hip-ratio (WHRadjBMI, another fat distribution trait), in recent meta-analyses efforts. We observed an enrichment of loci moderately associated with WHRadjBMI (P<0.05) and found a four-fold higher number of eSNPs overlapping with significant GWAS loci for WHRadjBMI, than for BMI. Analyses are underway to further explore the relationship and direction of effect between SNP, DNAm, gene expression and obesity traits.

45

Whole Exome Sequencing Identifies Candidate Causal Genes For Severe Insulin Resistance. *F. Payne[1], A. Daly[1], W. Bottomley[1], E. Raffan[2], E. Goncalves Serra[1], A. Thompson[1]. The UK10K Consortium., D.B. Savage[2], R.K. Semple[2], S. O'Rahilly[2], I. Barroso[1,2].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA. United Kingdom; 2) University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Cambridge CB2 0QQ. United Kingdom.

Genetic disorders of insulin action are highly heterogeneous, encompassing both severe insulin resistance (SIR) syndromes (a collection of rare disorders of extreme resistance to the glucose-lowering effects of insulin) and unrestrained metabolic or mitogenic insulin-like activity in the absence of insulin. Several exhibit a Mendelian inheritance pattern and causative variants in genes including *INSR*, *PPARG* and *AKT2* have been successfully identified by candidate gene re-sequencing. However, many cases of likely genetic aetiology are sporadic and their molecular basis unknown. To identify novel causal variants, we have undertaken whole exome sequencing (WES) in 74 patients with severely deranged insulin action and, where available, their family members. Some of these were sequenced as part of the UK10K Rare Disease project (http://www.uk10k.org/). Agilent baits were used to capture NCBI consensus coding sequence and Illumina paired-end reads analysed. Reads were mapped to the NCBI human genome reference sequence and single nucleotide variants and insertion/deletions called. An average of 9.6Gb of sequence was produced per patient, at a mean depth of 95 with 87% high quality coverage (mapping quality ≥ 30 in MAQ) at a minimum depth of 10, detecting an average of 72,497 variants. Families and individuals were analysed independently or grouped based on phenotype and likely mode of inheritance and DeNovoGear (http://sourceforge.net/projects/denovogear/) was used to identify variants with a high probability of being de novo. To prioritise putative causative variants, we filtered using the 1000 Genomes (http://www.1000genomes.org/) and NHLBI Exome Sequencing Project (ESP: http://evs.gs.washington.edu/EVS/) to exclude known variants and focused on those predicted to alter protein sequence (potentially functional). Variants were also prioritised based on presence in known candidate genes or genes within relevant pathways and on calling software quality measures. Potential candidates were confirmed by capillary sequencing and co-segregation with available family members examined. Using this approach, disease causing mutations were found in ~12 index cases to date. These include mutations within known genes presenting early or atypically (*WRN*, *PPARG*, *AGPAT2*, *PCNT*); in strong candidates with an established role in insulin action (*AKT2*, *PIK3R1*) and in genes not previously implicated in insulin action, but within pathways relevant to phenotypic features.

46

Exome analysis in 8,232 Finnish men identifies novel loci and low-frequency variants for insulin processing and secretion. *J.R. Huyghe[1], A.U. Jackson[1], M.P. Fogarty[2], A. Stančáková[3], H.M. Stringham[1], M.L. Buchkovich[2], C. Fuchsberger[1], J. Paananen[3], P.S. Chines[4], T.M. Teslovich[1], J.M. Romm[5], H. Ling[5], I. McMullen[5], R. Ingersoll[5], E.W. Pugh[5], K.F. Doheny[5], J. Kuusisto[4], L.J. Scott[1], F.S. Collins[4], G.R. Abecasis[1], R.M. Watanabe[6], M. Boehnke[1], M. Laakso[3], K.L Mohlke[2].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 3) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 4) Genome Technology Branch, National Human Genome Research Institute, Bethesda, MD, USA; 5) The Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD, USA; 6) Department of Preventive Medicine, Keck School of Medicine of USC, CA, USA.

Insulin secretion plays a critical role in maintenance of blood glucose levels, and failure to secrete sufficient insulin is a hallmark of type 2 diabetes. Genome-wide association studies (GWAS) have identified loci involved in these processes; however, many genetic contributions remain undefined. Until now it has not been possible to study the role of low-frequency (minor allele frequency (MAF) < 5%) nonsynonymous variants in complex traits exome-wide and systematically in large samples. To determine the role of low-frequency nonsynonymous variants in insulin processing and secretion, we designed an exome array based on variants discovered by sequencing > 10,000 subjects and genotyped ~242,000 SNPs on the Illumina Infinium HumanExome Beadchip in 8,232 non-diabetic Finnish men from the population-based Metabolic Syndrome in Men (METSIM) study. We identified two novel genes harboring low-frequency variants associated with insulin processing and secretion: TBC1D30 (2.0% MAF) and KANK1 (3.0% MAF), both of which function in G protein signaling pathways. We also identified coding low-frequency variants at two known loci. A nonsynonymous variant in SGSM2 (1.4% MAF) was independent of the GWAS-identified common signal. A nonsense variant in MADD (3.7% MAF) occurred on a haplotype containing the most associated SNP of two independent common GWAS signals at this locus. The nonsense allele, associated with decreased insulin secretion, always occurred with the GWAS allele associated with decreased secretion, and adjusting for one variant in a conditional analysis decreased but ccccccdid not eliminate association for the other variant. Finally, we observed that linkage disequilibrium with nonsynonymous variants in this region can extend up to 1 Mb. The interpretation of both single-variant and gene-based tests therefore needs to consider the effects of distant common SNPs, an especially important consideration when exome sequence data are analyzed in the absence of data on the surrounding noncoding regions. In conclusion, although sequencing will still be needed for a fully comprehensive analysis of variants, this study provides proof of principle that exome array genotyping is a valuable approach for identifying low-frequency functional variants, and for fine-mapping of GWAS-identified loci, in complex traits.

47

Global genomic and transcriptomic variation in human pancreatic islets reveals novel loci associated with type 2 diabetes and related traits. *J. Fadista, P. Vikman, I. Mollet, E.O. Laakso, U. Krus, O. Hansson, L. Groop.* Lund University Diabetes Centre, Malmö, Sweden.

The advent of exome and RNA sequencing technologies has provided the possibility to probe the genomic and transcriptomic landscape with unprecedented resolution. To study the interaction between genomic (exomeSeq) and transcriptomic (RNAseq) variation, we developed a pipeline for the discovery of global allelic imbalance of expression, RNA editing, and cis eQTLs for exome SNPs vs. gene expression in 42 human pancreatic islets (18 normal glycemic, 18 hyperglycemic and 6 diabetic). The availability of glycemic and insulin secretion data from these islets also allowed us to further interpret the complex trait of diabetes from a genotype → gene expression → phenotype perspective. By looking at a stringent set found in at least 2 individuals, we detected 639 RNA editing sites (62% novel SNPs). From these loci, 41% were in intergenic regions, 22% in 3'UTRs, 14% in introns, 12% in non-synonymous coding regions (including a variant predicted to be deleterious in the diabetic related gene VAMP8), 8% in synonymous coding regions, 2% in 5'UTRs, and 1% in splice-sites. Allelic imbalance of expression seen in at least 2 individuals were detected in 1287 loci (49% novel SNPs). From these loci, 37% were in intergenic regions, 25% in 3'UTRs, 9% in introns, 19% in non-synonymous coding regions (including variants predicted to be deleterious in diabetic related genes such as LAMC1, PPM1A, ATP1A1 and SGK1), 9% in synonymous coding regions, and 1% in 5'UTRs. Both RNA editing and allelic imbalance loci preferential target pathways related to metabolism and cell adhesion. Since many of these loci are also within 3'UTRs, interaction with microRNAs target sites could be expected, leading to an increase in post-translation complexity. Cis eQTLs were computed to find associations between exome SNPs vs. gene expression, and then look for common genes nominally associated (at p-value < 0.01) with HbA1c and insulin secretion. We found 10492 cis eQTLs (158 at FDR < 5%), 265 of whom contained 134 genes that were also associated with HbA1c. These included PDX1, GPR65, PTPN3, IL18R1, CPXM2, and SAMSN1 at FDR < 5%. In addition, we also found 82 eQTLs that contained 32 genes associated with insulin secretion, including KRTAP3-1, LOC100130238 and LOC285205 at FDR< 5%. The present data demonstrate the value of different layers of omics information in the same individuals to identify genes and pathways associated with type 2 diabetes.

48

Identification of a novel genome-wide significant association with type 2 diabetes risk in Mexican and Mexican Americans. *A.L. Williams[1,2], H. García-Ortiz[3], M.J. Gómez-Vázquez[4,5], C.A. Haiman[6], A. Huerta-Chagoya[4,7], A.K. Manning[1,2,8], C. Márquez-Luna[9], H. Moreno-Macías[4,10] for the SIGMA Type 2 Diabetes Consortium.* 1) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 2) Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; 3) Laboratory of Immunogenomics and Metabolic Diseases, National Institute of Genomic Medicine, Mexico City, México; 4) Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, México; 5) Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, México; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA; 7) Universidad Nacional Autonóma de México, Mexico City, México; 8) Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA; 9) Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, México; 10) Universidad Autónoma Metropolitana, México City, México.

Type 2 diabetes (T2D) afflicts 300M people worldwide, with disproportionate impact on populations of Mexican and Latin American ancestry. The SIGMA Diabetes Consortium was formed to systematically characterize the role of genetic variation in type 2 diabetes in Mexican and Mexican American populations. In particular, we set out to explore the hypothesis that there exist genetic variants that are common in Latinos but rare in other parts of the world, such that they might contribute strongly in this population, and might have been undetected by other studies.

We assembled a collection of 9,225 well characterized individuals, roughly half collected in Mexico, and half Latino samples from the Multiethnic Cohort Study. Each individual has been genotyped for 2.45M polymorphic SNPs using the Illumina OMNI 2.5M platform, and are being genotyped for 250,000 nonsynonymous SNPs using the Exome Chip.

Using these data we conducted a genome-wide association study (GWAS) of over 18 billion genotypes—to our knowledge, the largest GWAS performed using Latino samples to date. After data cleaning, association analysis was performed correcting for principle components of ancestry, age, and BMI using LTSOFT. We replicated multiple previously identified T2D associations including *TCF7L2* (p=2.0e-18) and *KCNQ1* (p=1.8e-14).

In addition, we identified a novel, genome-wide significant association of a missense variant on chromosome 17 (p=1.3e-13). The estimated odds ratio of the novel association is 1.24, making it one of the strongest risk alleles yet identified for T2D. This genetic variant is extremely common in individuals of Native American descent (frequencies of 40–60%), but rare in Europeans (<2%), which may explain why it was not identified previously. The variant is associated with lower age at onset of T2D, and more strongly with T2D risk in lean as compared to obese individuals.

This novel association may contribute to the larger burden of T2D in Latinos and suggests that the higher T2D prevalence in Latinos is not solely explained by non-genetic factors. Our study also demonstrates the value of performing disease studies in Mexico and other understudied populations, as this may reveal associations that may be undetected in studies to date.

49

Discovery and fine-mapping of type 2 diabetes susceptibility loci through trans-ethnic meta-analysis. *A. Mahajan[1], J.E. Below[2], W. Zhang[3], M.J. Go[4], E. Parra[5], A.P. Morris[1], AGEN-T2D, DIAGRAM, MA-T2D and SA-T2D Consortia.* 1) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, UK; 2) Departments of Human Genetics and Medicine, The University of Chicago, Chicago, Illinois, USA; 3) Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, St Mary's Campus, London, UK; 4) Division of Structural and Functional Genomics, Center for Genome Science, National Institute of Health, Chungcheongbuk-do, Korea; 5) Department of Anthropology, University of Toronto at Mississauga, 3359 Mississauga Road North, Mississauga, ON, Canada.

Recent meta-analyses of genome-wide association studies (GWAS) of type 2 diabetes (T2D) in European, Mexican American, South Asian and East Asian ancestry populations have demonstrated substantial overlap of loci contributing effects to susceptibility. We have performed GWAS trans-ethnic fixed-effects and MANTRA meta-analyses in 26,488 T2D cases and 83,964 controls in up to 2.5 million autosomal SNPs to: (i) identify novel susceptibility loci; (ii) assess the evidence of heterogeneity in allelic effects between ancestry groups; and (iii) improve fine-mapping resolution through differences in linkage disequilibrium (LD) structure between diverse populations. We identified 1 novel T2D susceptibility locus at genome-wide significance ($p<5\times10^{-8}$), mapping to TMEM154 ($p=4.2\times10^{-9}$). The lead SNP from the MANTRA analysis this locus demonstrated homogeneous allelic effects across ancestry groups. Furthermore, we observed nominal evidence of heterogeneity (Cochran's Q-statistic p<10-3) at lead SNPs at just 3 of the 55 established autosomal T2D susceptibility loci: KLF14 ($p=1.4\times10$-5), HNF4A ($p=4.0\times10$-5) and PEPD ($p=7.1\times10$-4). At all 3 of these loci, the East Asian ancestry group formed an outlier in allelic effects. We constructed credible sets of SNPs that encompass 95% of the posterior probability of being causal (or tagging an unobserved causal variant) across 30 T2D susceptibility loci previously identified in European ancestry GWAS. We compared genomic interval covered by the credible set in the trans-ethnic and the European ancestry only (12,171 cases and 56,862 controls) MANTRA meta-analyses. Fine-mapping resolution was improved by the addition of non-European ancestry GWAS at 23 of the loci, most notably at KCNJ11, where the credible set of SNPs was reduced from 134 (959kb) to just 1. This SNP, rs5215, is in strong LD with the previously implicated E23K variant, not reported in the European ancestry meta-analysis. However, the trans-ethnic credible set excludes another implicated variant at this locus, A1369S, in the ABCC8 gene. Our results clearly highlight the benefits of trans-ethnic meta-analysis of GWAS from multiple ancestry groups for discovery and fine-mapping of complex trait loci. Furthermore, allelic effects at GWAS loci are predominantly homogenous across ancestry groups, suggesting that these signals are unlikely to be driven by synthetic association with rare variation.

50

***TCF7L2* genetic variation is associated with impaired incretin effect and lower glucagon.** *B. Chamarthi[1,4], K.R. Littleton[2], V. Kaur[2], L. Chen[2], A.K. Manning[5], M.K. Thomas[6], M.S. Hudson[1,4], J.C. Florez[2,3,4,5].* 1) Division of Endocrinology, Diabetes & Hypertension, Brigham and Women's Hospital Boston, MA., Select a Country; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 3) Diabetes Unit, Massachusetts General Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) Broad Institute of Harvard and MIT, Cambridge, MA; 6) Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN.

The T allele at the intronic single nucleotide polymorphism rs7903146 in the transcription factor 7-like 2 (*TCF7L2*) gene is strongly associated with type 2 diabetes (T2D). The mechanisms for this increased risk remain uncertain. The objective of this study was to evaluate physiologic and hormonal effects of *TCF7L2* genotype before and after interventions that influence glucose physiology.

We genotyped rs7903146 in 165 participants (82 CC, 64 CT, 19 TT) in an ongoing pharmacogenetics study. Plasma glucose and insulin were drawn before and after 1) single dose of glipizide 5 mg (Visit (V) 1) and 2) 75-g oral glucose tolerance test (OGTT) after 2 days of metformin 500 mg BID (V2). Glucagon-like peptide 1 (GLP1), glucose-dependent insulinotropic polypeptide (GIP), proinsulin and glucagon were measured. We used linear regression models adjusted for age, BMI, gender, race, time (for repeated measures) and glucose (for hormone assessments) for statistical analyses.

A linear relationship was seen across genotypes for fasting glucose (V1 β 7.5, P<0.001; V2 β 5.1, P<0.01) with highest levels in TT homozygotes. Insulin, proinsulin and HOMA-IR were similar across genotypes. TT homozygotes compared to CT and CC genotypes, had similar glucose decline after glipizide and a greater decrease in fasting glucose from V1 to V2 (after metformin), but this was not significant on adjusting for baseline glucose, indicating that the drug response is driven by the higher baseline glucose and not genotype. V1 glucagon adjusted for glucose showed a linear relationship across genotypes with lower level in TT homozygotes (P<0.01). TT homozygotes had higher baseline GIP (P<0.01). Baseline GLP1 was similar but GLP1 response during OGTT showed a linear relationship with genotype with a steeper slope and greater early rise in TT homozygotes (P<0.001).

Our findings implicate altered GLP-1 and GIP signaling as contributory mechanisms for increased TT-associated T2D risk. Reduction in glucagon further suggests primary or adaptive changes in pancreatic alpha cell function in response to *TCF7L2* genetic variation.

51

Novel locus including *FGF21* is associated with dietary macronutrient intake. *A.Y. Chu[1], T. Workalemahu[2,3], N.P. Paynter[1], L.M. Rose[1], F. Giulianini[1], CHARGE. Nutrition Working Group[4], Q. Qi[2,3], G.C. Curhan[2], E.B. Rimm[2,3], D.J. Hunter[2,3], L.R. Pasquale[5,6], P.M Ridker[1,7], F.B. Hu[2,3], D.I. Chasman[1,8], L. Qi[2,3] on behalf of the DietGen Consortium.* 1) The Division of Preventive Medicine, Brigham & Women's Hosp, Boston, MA; 2) The Channing Laboratory and Department of Medicine, Brigham and Women's Hosp and Harvard Medical School, Boston, MA; 3) The Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, MA; 4) Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium; 5) The Department of Ophthalmology, Brigham and Women's Hosp and Harvard Medical; 6) The Mass Eye and Ear Infirmary, Boston, MA; 7) The Division of Cardiology, Brigham and Women's Hosp and Harvard Medical School, Boston, MA; 8) The Division of Genetics, Brigham and Women's Hosp and Harvard Medical School, Boston, MA.

Dietary intake assessed by food frequency questionnaire (FFQ) has been associated with obesity, diabetes, and other chronic conditions related to poor health. Genetic factors contribute to inter-individual variation in macronutrient intake (carbohydrates, protein, and fats), with heritability ranging between 0.11-0.65. To investigate the genetics of dietary intake, we assembled the DietGen Consortium consisting of 3 large prospective cohorts (Health Professionals Follow-up Study, Nurse's Health Study, and Women's Genome Health Study, n=33,388 with European ancestry), all employing similar FFQs and genotype data imputed genome-wide to the HapMap II reference panel. In DietGen, we conducted a genome-wide, fixed effects inverse-variance weighted meta-analysis of percentage of energy intake due to carbohydrate, protein, or fat. Analysis was performed both with and without adjustment for BMI. A total of 22 loci was association with at least one of the macronutrients at a level of significance (P<1E-5). The most significant SNP at each locus was evaluated for replication in a parallel genome-wide meta-analysis from the CHARGE Consortium (n=38,355 with European ancestry). A synonymous SNP in *FGF21* associated with percentage energy intake from protein adjusted for BMI was replicated (P=7.2E-4) by CHARGE and attained genome-wide significance in joint meta-analysis with DietGen (β[se]=-0.11%[0.02], P=7.9E-9). *FGF21* encodes fibroblast growth factor 21, a cytokine that influences cellular metabolism in animal models. In humans, elevated serum levels of the FGF21 protein have been linked to obesity and diabetes. Near the *TANK* gene, encoding a protein thought to interact with TRAF family members, a variant associated with carbohydrate intake replicated in CHARGE at nominal significance after Bonferroni correction (P=2.9E-3) and was just below genome-wide significance in joint analysis (β[se]=0.23%[0.04], P=9.6E-8). At the *FTO* locus, a variant not in LD (r^2=0.07) with the canonical BMI-associated *FTO* SNP (rs1558902) was associated with carbohydrate intake in DietGen at genome-wide significance (P=7.4E-9). However, this variant was not replicated in CHARGE (P=0.34) and joint meta-analysis no longer reached genome-wide significance (P=2.2E-4). Taken together, our results highlight *FGF21* and related pathways in macronutrient intake, suggest a role for the locus including the *TANK* gene, and reinforce the contribution of genetic analysis to understanding dietary intake.

52

The TRK-fused gene is mutated in hereditary motor and sensory neuropathy with proximal dominant involvement (HMSN-P). *H. Ishiura[1], W. Sako[3], M. Yoshida[4], T. Kawarai[3], O. Tanabe[3,5], J. Goto[1], Y. Takahashi[1], H. Date[1], J. Mitsui[1], B. Ahsan[1], Y. Ichikawa[1], A. Iwata[1], H. Yoshino[6], Y. Izumi[3], K. Fujita[3], K. Maeda[3], S. Goto[3], H. Koizumi[3], R. Morigaki[3], M. Ikemura[7], N. Yamauchi[7], S. Murayama[8], G. Nicholson[9], H. Ito[10], G. Sobue[11], M. Nakagawa[12], R. Kaji[3], S. Tsuji[1,2,13]*. 1) Department of Neurology, The University of Tokyo, Tokyo, Japan; 2) Medical Genome Center, The University of Tokyo, Tokyo, Japan; 3) Department of Clinical Neuroscience, The Tokushima University Graduate School of Medicine, Tokushima, Japan; 4) Department of Neuropathology, Institute for Medical Science of Aging, Aichi Medical University, Aichi, Japan; 5) Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI; 6) Yoshino Neurology Clinic, Chiba, Japan; 7) Department of Pathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 8) Department of Neuropathology and the Brain Bank for Aging Research, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan; 9) University of Sydney, Molecular Medicine Laboratory and University of Sydney, ANZAC Research Institute, Australia; 10) Department of Neurology, Kyoto University Graduate School of Medicine, Kyoto, Japan; 11) Department of Neurology, Nagoya University Graduate School of Medicine, Aichi, Japan; 12) Department of Neurology and Gerontology, Kyoto Prefectural University Graduate School of Medicine, Kyoto, Japan; 13) Division of Applied Genetics, National Institute of Genetics, Shizuoka, Japan.

[Background] Hereditary motor and sensory neuropathy with proximal dominant involvement (HMSN-P) is an autosomal dominant neurodegenerative disorder characterized by proximal-predominant muscle weakness and atrophy with widespread fasciculations and painful muscle cramps resembling amyotrophic lateral sclerosis (ALS) followed by distal sensory involvement. To date, large families with HMSN-P have been reported from two different regions in Japan (Kansai and Okinawa). Previous studies on families from the both regions showed a linkage to 3q. [Method] Linkage and haplotype analyses of two previously reported families from Kansai and two new families from Okinawa were conducted using SNP arrays. Enrichment of exonic sequences of an Okinawan patient was performed using SeqCap Human Exome 2.1 M array (NimbleGen) and the captured library was subjected to massively parallel sequencing by GAIIx and SOLiD4. Immunohistochemical studies were conducted in autopsied samples of a patient. [Results] The genome-wide linkage study revealed only one region on chromosome 3 showed a cumulative parametric multipoint LOD scores exceeding 4.0. In the Okinawan families, an extended disease-associated haplotype spanning 3.3 Mb within the candidate region was shared among the affected individuals, thus defining the minimum candidate region. GAIIx and SOLiD4 yielded average coverages of 29.0X and 26.8X, respectively. Exome analysis revealed the only novel nonsynonymous variant p.Pro285Leu in *TRK-fused gene* (*TFG*) in the minimum candidate region. The identical mutation was also found in Kansai families. The mutation cosegregated with the disease in all the families and was neither found in 964 Japanese control chromosomes nor in public databases. Furthermore, direct nucleotide sequence analysis of low coverage exons (<=10X) did not reveal any other novel nonsynonymous variants. Detailed haplotype analysis showed only 49 kb of shared segment at most between Kansai and Okinawan families, suggesting the mutations are of independent origins. Taken together, we concluded that *TFG* is the causative gene for HMSN-P. Neuropathological study of an autopsied patient revealed abnormal TFG/ubiquitin-positive cytoplasmic inclusion bodies in motor neurons and the dorsal root ganglia. [Conclusions] *TFG* is the causative gene for HMSN-P. HMSN-P is a novel proteinopathy showing TFG/ubiquitin-positive cytoplasmic inclusions, which highlights a new pathway leading to motor neuron degeneration.

53

Mutation in the Autophagy-related TECPR2 Gene Causes Hereditary Spastic Paraparesis. *D. Oz-Levi[1], B. Ben-Zeev[4,6], E. Ruzzo[3], Y. Hitomi[3], A. Gelman[2], K. Pelak[3], Y. Anikster[4,6], H. Reznik-Wolf[5,6], I. Bar-Joseph[4,5,6], T. Olender[1], A. Alkelai[1], M. Weiss[1], E. Ben-Asher[1], D. Ge[3], K. Shianna[3], Z. Elazar[2], D. Goldstein[3], E. Pras[5,6], D. Lancet[1]*. 1) molecular genetics, weizmann institute of science, Rehovot, Israel; 2) Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel; 3) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina, USA; 4) Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Israel; 5) The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Israel; 6) The Sackler School of Medicine, Tel Aviv University, Israel.

Hereditary spastic paraparesis (HSP, SPG) comprise a diverse group of neurodegenerative disorders characterized by progressive spasticity and hyperreflexia of the lower limbs estimated to affect 10 individuals per 100,000. Pure HSP is often dominant, whereas complicated HSP tends to be recessive and is accompanied by other neurological symptoms. The underlying genes and mutations have been discovered for only 27 of the 48 reported SPG types. We describe five patients from three unrelated Jewish Bukharian families who presented with apparently autosomal recessive HSP of the complicated form. Our patients are characterized by hypotonia evolving into spasticity with areflexia and parkinsonism, moderate to severe mental retardation and dismorphic features. Due to the incomplete resemblance to any other known form of HSP, we termed the present disease SPG49. Exome sequencing identified only one homozygous variant shared among all patients and absent in dbSNP and in 2,007 unaffected controls of non- Jewish-Bukharian origin. This is a one basepair frameshifting deletion in the tectonin beta propeller repeat containing 2 (TECPR2) gene, leading to a premature stop codon and significant degradation of the protein. TECPR2 has been recently reported as positive regulator of autophagosome accumulation. We therefore examined the autophagy-related fate of the SQSTM1 (p62) and MAP1LC3B (LC3) proteins in skin fibroblasts of an affected individual as compared to healthy control, and found that both protein levels were decreased, with a more pronounced decrement in the lipidated form of LC3 (LC3II). siRNA knockdown of TECPR2 in HeLa cells showed similar changes, consistent with aberrant autophagy. Our results are strengthened by the fact that autophagy dysfunction has been implicated in a number of other neurodegenerative diseases. Furthermore, SPG20, the phenotypically most similar HSP type, involves mutation in spartin, which mediates endosome formation, an allied membrane trafficking pathway. The discovered TECPR2 variation implicates autophagy for the first time in spastic paraparesis, and could offer novel genetic diagnostics and potentially also therapeutic insights.

54

Spinal muscular atrophy associated with progressive myoclonic epilepsy is caused by mutations in ASAH1. *J. Melki[1], J. Zhou[1], M. Tawk[2], FD. Tiziano[2], J. Veillet[1], M. Bayes[3], F. Nolent[1], V. Garcia[4], S. Servidei[5], E. Bertini[6], F. Castro-Giner[3], Y. Renda[7], S. Carpentier[4], N. Andrieu-Abadie[4], I. Gut[3], T. Levade[4], H. Topaloglu[7]*. 1) UMR-788, Inserm and University Paris 11, Biomedical Research Insitute, Kremlin Bicetre, France; 2) Istituto di Genetica Medica, Università Cattolica del Sacro Cuore, Roma, 00168, Italy; 3) Centro Nacional de Análisis Genómico, University of Barcelona, Barcelona, 080028, Spain; 4) Institut National de la Santé et de la Recherche Médicale UMR1037, Centre de Recherches sur le Cancer de Toulouse, Université Paul Sabatier, Toulouse, 31432, France; 5) Istituto di Neurologia, Università Cattolica del Sacro Cuore, Roma, 00168, Italy; 6) Unit of Molecular Medicine, Dept. Neurosciences, Ospedale Bambino Gesù Research Institute, Roma, 00165, Italy; 7) Child Neurology Unit, Department of Pediatrics, Hacettepe University, Ankara, 06100, Turkey.

Spinal muscular atrophy (SMA) is a clinically and genetically heterogeneous disease characterized by degeneration of lower motor neurons. The most frequent form is linked to mutations of the SMN1. Childhood SMA associated with progressive myoclonic epilepsy (SMA-PME) has been reported as a rare autosomal recessive condition unlinked to SMN1 mutation. Through linkage analysis, homozygosity mapping and exome sequencing in three unrelated SMA-PME families, we identified a homozygous missense mutation (c.125C>T; p.Thr42Met) in the ASAH1 exon 2 in affected children of two families and the same mutation associated with a deletion of the whole gene in the third one. Expression studies of the c.125C>T mutant cDNA in Farber fibroblasts showed an acid ceramidase activity deficiency of 32% of that generated by normal cDNA. This reduced activity was able to normalize the ceramide level in Farber cells raising the question of the pathogenic mechanism underlying the central nervous system (CNS) involvement in deficient cells. Morpholino knockdown of ASAH1 ortholog in zebrafish led to a marked loss of motor neuron axonal branching associated with increased apoptosis in the spinal cord. Our results reveal a wide phenotypic spectrum associated with ASAH1 mutations. An acid ceramidase activity below 10% results in Farber disease, an early onset disease starting with subcutaneous lipogranulomata, joint pains and hoarseness of the voice whereas a higher residual activity may be responsible for SMA-PME, a later onset phenotype restricted to the CNS and starting with lower motor neuron disease.

55

Genetic Variants in Chromatin Modifying Genes cause D4Z4 Hypomethylation, *DUX4* Expression, and Contraction-independent Facioscapulohumeral Muscular Dystrophy (FSHD2). *D.G. Miller[1], R.J.L.F. Lemmers[2], L.M. Petek[1], J. Balog[2], P.J. van der Vliet[2], G.J. Block[1], J.W. Lim[4], G.N. Filippova[4], A.M. Amell[1], G.W.E. Santen[2], B. Bakker[2], M.J. Bamshad[1], S.J. Tapscott[4], R. Tawil[3], S.M. van der Maarel[2]*. 1) Pediatrics, Genome Sciences, University of Washington, Seattle, WA, USA; 2) Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 3) Neurology, University of Rochester Medical Center, Rochester, NY, USA; 4) Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Facioscapulohumeral muscular dystrophy (FSHD) is an adult onset myopathy with epigenetic changes that most commonly occur as a consequence of reduction in the number of 3.3 kb D4Z4 units arrayed on chromosome 4 to 1–11 units. The contraction-induced epigenetic profile results in a variegated expression pattern of the normally repressed *DUX4* retrogene and produces myopathic changes that manifest as muscle weakness when contractions occur on arrays with permissive D4Z4 haplotypes. Five percent of FSHD-affected individuals develop FSHD despite having D4Z4 array lengths greater than the 10 unit threshold so FSHD occurring by this mechanism has been called FSHD2 or contraction-independent FSHD. D4Z4 arrays of individuals with FSHD2 show epigenetic de-repression similar to that seen by the more common mechanism but chromatin changes extend to similar arrays on chromosome 10, and the other chromosome 4 allele suggesting that contraction-independent array-de-repression may occur through the activity of genes involved in epigenetic modification of macrosatellite repeats. We performed exome sequencing of FSHD2-affected individuals and their unaffected family members to identify one of these loci. Rare variants present in 61% of FSHD2-affected individuals segregated independently of permissive haplotypes on chromosome 4 and only resulted in FSHD when a permissive haplotype and gene variant were present in the same individual. Consistent with our hypothesis, reduction of protein levels of this chromatin modifier results in D4Z4 array de-repression and a variegated pattern of *DUX4* expression in normal human myoblasts. Thus the non-Mendelian inheritance pattern seen in FSHD2 families is explained by digenic inheritance. Genes containing FSHD2-causing variants likely contribute to array de-repression in FSHD1 patients as well, making this discovery important for therapeutic strategies as well as diagnosis and counseling of FSHD-affected family members.

56

Unexpected extension of the phenotype spectrum associated with SMAD3 mutations in Aneurysms-Osteoarthritis Syndrome. *M. Aubart[1, 2], D. Gobert[2], N. Hanna[1, 3], C. Muti[2], J. Roume[2], V. Cusin[2], B. Grandchamp[2], L. Gouya[2, 3], D. Detaint[2], G. Jondeau[1, 2], C. Boileau[1, 2, 3]*. 1) INSERM U698, Paris, France; 2) Marfan and related disorders National Reference Center, Bichat University Hospital, Paris, France; 3) Molecular Genetics Laboratory, Ambroise Paré University Hospital, Boulogne, France.

Thoracic Aortic Aneurysms dissections (TAAD) are inherited in an autosomal dominant manner in about 20% of cases with less than 5% syndromic (Marfan, Ehler-Danlos, Loeys-Dietz). Genetic and clinical heterogeneity in isolated TAAD are important. Recently, mutations in the SMAD3 gene have been identified, associated with the presence of diffuse arterial aneurysms and tortuosity and early-onset osteoarthritis (Aneurysms osteoarthritis syndrome, AOS). To assess the prevalence of this new syndrome a cohort of 105 French patients with TAAD without mutations in other known genes (FBN1, TGFBR1, TGFBR2, ACTA2) was screened for mutations in SMAD3. 5 non-sense and 4 missense novel mutations were found. All missense mutations affected the active MH2-domain of the protein. After extensive familial screening, 51 subjects were considered as mutation-carriers either through direct molecular determination (34 subjects) or because they displayed a history of aortic surgery or sudden death (17 relatives). Among these 51 subjects, 27 had been evaluated for Marfan syndrome (MFS) rule out before molecular diagnosis. Subsequent extensive medical investigations could be performed in 21/27 SMAD3 mutation carriers (from 5 families). In these patients, a cardiovascular spectrum of abnormalities similar to Loeys-Dietz syndrome was observed. They involved TAAD in 66% (23/35), arterial tortuosity in 67% (10/15) and others aneurysms (including cerebral ones) in 60% (9/15). Skeletal anomalies usually associated with MFS were more frequent than in the general population, similar to TGFBR2 mutation carriers, but less frequent than in FBN1 mutation-carriers. No ophthalmologic anomalies were found. Osteoarthritis was present in all over-30-years old subjects, but without specific localization or gravity. Curiously, systematic neurological examinations revealed that 71% (15/22) of patients displayed neurological features such as muscle cramps, paresthesia, hypoesthesia, and gait disturbance. Among these, 9 patients (41 % with 3 different SMAD3 mutations) had an obvious peripheral neuropathy. Electromyography was performed in 5 patients and obviated the presence of an axonal peripheral neuropathy in 4. The association in these patients of this neuropathy with high arches and scoliosis enabled the diagnosis of type II Charcot-Marie-Tooth (CMT) disease. Thus surprinsingly, molecular AOS diagnosis seems to bridge two autosomal dominant and clinically heterogeneous diseases: TAAD and CMT-II.

57

Whole-exome sequencing for autosomal recessive non-syndromic deafness: 93% of known genes covered and OTOGL and SLITRK6 are novel genes. *M. Tekin[1,2], O. Diaz-Horta[1,2], D. Duman[3], J. Foster II[1,2], A. Sirmaci[1,2], M. Gonzalez[1,2], N. Mahdieh[4], M. Bonyadi[5], FB. Cengiz[3], R. Ulloa[1,2], S. Zuchner[1,2], S. Blanton[1,2]*. 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, USA; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami Miller School of Medicine, Miami, USA; 3) Division of Pediatric Genetics, Ankara University School of Medicine, Ankara, Turkey; 4) Kawsar Human Genetic Research Center, Tehran, Iran; 5) Faculty of Natural Sciences, Center of Excellence for Biodiversity, University of Tabriz, Tabriz, Iran.

Identification of the responsible mutation in autosomal recessive nonsyndromic hearing loss (ARNSHL) is difficult, since causative mutations in 39 different genes have so far been reported. After excluding GJB2 mutations, we performed whole-exome sequencing (WES) in 30 individuals from 20 unrelated multiplex consanguineous families with ARNSHL. Agilent SureSelect Human All Exon 50 Mb kit and an Illumina Hiseq2000 instrument were used. BWA was used to align sequence reads to the human genome (hg19) and variants were called using the GATK software package. An average of 93%, 84% and 73% of bases were covered to 1X, 10X and 20X within the ARNSHL-related coding RefSeq exons, respectively. Twelve homozygous mutations in known deafness genes, of which 9 are novel, were identified: MYO15A-p.Q1425X, -p.S1481P, -p.A1551N; LOXHD1-p.R383X, p.E677X; GIPC3-p.H170N; ILDR1-p.Q230X; MYO7A-p.G2163S; TECTA-p.Y1737C; TMC1-p.V529Lfs*2; TMPRSS3-p.F13Lfs*10; TRIOBP-p.R779Dfs*6. Sanger sequencing confirmed co-segregation of the mutation with deafness in each family. Uncovered regions with WES included those that are not targeted by the exome capture kit and regions with high GC content. Two novel deafness genes, OTOGL and SLITRK6 were identified. OTOGL encodes a protein that has structural similarities to epithelial-secreted mucin protein family that includes otogelin, which is a well-characterized inner ear protein of acellular membranes. Homozygous c.1430delT (p.V477Efs*25) mutation in OTOGL co-segregates with hearing loss in a consanguineous family with 4 affected and 3 unaffected children. SLITRK6 is a neurite-modulating activity protein member of the SLITRK family, highly expressed in the inner ear. A Slitrk6 KO mouse model is characterized by sensorineural hearing loss. Homozygous c.890C>A mutation introducing a stop codon (p.S297X) in SLITRK6 co-segregates with deafness in a consanguineous family with 3 affected and 3 unaffected children. Analysis of WES data shows at least two other excellent candidate genes for ARNSHL in the remaining families. This work was supported by NIH-NIDCD grant R01DC009645 to M.T.

58

Whole exome sequencing and more to unravel the genetics and genotype-phenotype correlations for deafness. *H. Kremer[1,2], M. Schraders[1], C. Zazo Seco[1], J. Oostrik[1], I. Feenstra[2], A.M.M. Oonk[1], E. van Beelen[1], M. van Bers[2], K. Neveling[2], J.A. Veltman[2], R.J.C. Admiraal[1], H.P.M. Kunst[1], R.J.E. Pennings[1], E.H. Hoefsloot[2].* 1) Dept Otorhinolaryngology, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 2) Dept Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands.

With the introduction of newborn hearing tests, early diagnosis of congenital hearing impairment is warranted in many countries worldwide. However, the underlying causes often remain elusive and therefore questions on prognosis and risks for further children in the family remain unanswered. We have employed a combination of strategies to identify genetic defects underlying autosomal recessive nonsyndromic hearing impairment (arNSHI) in a panel of about 150 Dutch individuals with presumed arNSHI which is the predominant type of prelingual hereditary HI. Two novel deafness genes have been identified by using high-density SNP arrays for homozygosity mapping and genotype sharing combined with candidate gene selection in small Dutch families with arNSHI. Defects in both genes cause moderate HI involving all frequencies which has been stable since diagnosis. SNP-genotyping combined with whole exome sequencing (WES) led to the identification of a third novel gene for arNSHI. The hearing loss in this family is highly progressive in early childhood and is initially only affecting the high and middle frequencies (ski-slope or Z-shape audiogram). WES was performed in 36 probands with presumed arNSHI and in 14 probands with presumed dominantly inherited HI (adNSHI). In the arNSHI cases, defects in GJB2 and deletions in GJB6 were excluded prior to WES. In 5 arNSHI cases underlying mutations were detected in known deafness genes (LOXHD1, TMC1, MYO15A, STRC) and for adNSHI mutations in 3 probands (MYO6, EYA4). Also, mutations in known genes for syndromic HI were identified in 3 probands with presumed arNSHI. Reevaluation of patients is being performed to confirm that HI is nonsyndromic in these cases. Novel candidate deafness genes emerged from WES which are being investigated further for confirmation. In conclusion, the presented results demonstrate that WES `only' is suitable for diagnosis in genetically highly heterogeneous disorders and that employing a combination of strategies facilitates the identification of novel deafness genes. WES so far resulted in the identification of the genetic defect in 25% of unrelated patients/families with presumed arNSHI and in 21% of those with presumed adNSHI. Genotype-phenotype correlations are emerging which are of utmost importance for prognosis, rehabilitation and genetic counseling in families.

59

A mutation in Ca2+ binding protein 2, expressed in cochlear inner hair cells, causes autosomal recessive hearing impairment. *I. Schrauwen[1], S. Helfmann[2], A. Inagaki[3], F. Wolk[2], M.A. Tabatabaiefar[4], M.M. Picher[2], M. Sommen[1], C. Zazo Seco[5,6,7], H. Kremer[5,6,7,8], A. Dheedene[9], C. Claes[1], E. Fransen[1], M.A. Chaleshtori[10], P. Coucke[9], A. Lee[3], T. Moser[2], G. Van Camp[1].* 1) Department of Medical Genetics, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium; 2) InnerEarLab, Department of Otolaryngology, Collaborative Research Center 889, University of Göttingen School of Medicine, Robert-Kochstr. 40, D-37075 Goettingen, Germany; 3) Depts. of Molecular Physiology & Biophysics, Otolaryngology-Head and Neck Surgery, and Neurology University of Iowa, USA; 4) Department of Medical Genetics, School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran; 5) Department of Otorhinolaryngology, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands; 6) Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands; 7) Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, 6500 HBNijmegen, The Netherlands; 8) Department of Human Genetics, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands; 9) Center for Medical Genetics, Ghent University, Ghent, Belgium; 10) Cellular and Molecular Research Center, School of Medicine, Shahrekord University of Medical Sciences, Shahrekord, Iran.

The encoding of sound in the human auditory system consists of a complex chain of events. Ca2+ ions play a crucial role in the signal transduction in inner hair cells (IHCs) of the cochlea. There the sound-evoked receptor potential activates voltage-gated CaV1.3 (L-type) Ca2+ channels that mediate Ca2+ induced exocytosis of glutamatergic synaptic vesicles. Ca2+ binding proteins (CABPs) expressed in IHCs may augment presynaptic Ca2+ influx by antagonizing calmodulin-mediated Ca2+ dependent inactivation (CDI) of CaV1.3 channels. In this study, we identified a splice site mutation (c.637+1G>T) in Ca2+ binding protein 2 (CABP2) in three consanguineous families from Iran with moderate-to-severe hearing loss. This mutation, most likely a founder mutation, probably leads to skipping of exon 6 and a premature truncation of the protein (p.F164SfsX4). Isothermal titration calorimetry revealed binding of Ca2+ by wild-type CaBP2 and the truncation, modeling exon skipping, altered Ca2+ binding. The truncated CaBP2 protein was also less potent in inhibiting CDI of CaV1.3 channels in HEK293T and caused less suppression of current density when compared to wild-type CaBP2. In conclusion, we show that genetic defects in CABP2 cause moderate-to-severe sensorineural hearing impairment. The mutation may cause a hypofunctional CaBP2 protein that alters Ca2+ signaling in hair cells.

60

Comprehensive diagnosis for hearing loss using personal genomics: The first 100 cases. *E. Shearer[1,2], E. A. Black-Ziegelbein[3], M. S. Hildebrand[1], A. P. DeLuca[3,4], R. W. Eppsteiner[1], S. E. Scherer[5], T. E. Scheetz[3,4,6], T. L. Casavant[3,4,7], R. J. H. Smith[1,2,7].* 1) Department of Otolaryngology - Head and Neck Surgery, University of Iowa, Iowa City, Iowa; 2) Department of Molecular Physiology & Biophysics, University of Iowa Carver College of Medicine, Iowa City, Iowa; 3) Center for Bioinformatics and Computational Biology, University of Iowa, Iowa City, Iowa; 4) Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 6) Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa; 7) Interdepartmental PhD Program in Genetics, University of Iowa, Iowa City, Iowa City, Iowa.

Background: Non-syndromic hearing loss (NSHL) is the most common human sensory impairment, with more than 65 genes and 1,200 mutations identified to date. This heterogeneity makes genetic diagnosis of NSHL using traditional methodologies expensive and time consuming, and limits research endeavors. To address the need for comprehensive genetic testing, we developed the OtoSCOPE® platform, which enables sequencing of all exons of all known genes involved in NSHL simultaneously. OtoSCOPE® offers high sensitivity, specificity and reproducibility and is now offered as a clinical diagnostic test for hearing loss. In this study, we used OtoSCOPE® to define the prevalence of disease-causing mutations in a large cohort of patients with idiopathic genetic hearing loss. **Methods**: Targeted sequence capture and massively parallel sequencing were performed on 100 persons with presumed genetic deafness, targeting the exons and flanking sequence of all 65 genes implicated in NSHL and Usher syndrome. Enriched libraries were sequenced on the Illumina HiSeq or MiSeq with sample pool sizes varying from 12–96. Data analysis was performed using a locally installed Galaxy framework with BWA for read mapping, GATK for variant calling, and a custom pipeline for variant annotation. We used a custom Perl script for copy number variation calling from our coverage data. All causative mutations were verified with Sanger sequencing. **Results**: 75 percent of patients had presumed autosomal recessive NSHL (ARNHSL); the remainder had presumed autosomal dominant NSHL (ADNSHL). Single nucleotide variations, indels and CNVs were identified in 14 genes. The overall solve rate for all patients was 58 percent;, with a solve rate of 60 percent; and 52 percent; for ARNSHL and ADNSHL, respectively. Of the causative mutations identified, a significant number were novel: 49 percent; and 67 percent; for ARNSHL and ADNSHL, respectively. **Conclusions**: These results provide the first comprehensive insight into genetic screening in a deaf population. Our data confirm the high degree of genetic heterogeneity underlying hearing loss and underscore the importance of a clinical diagnostic platform like OtoSCOPE®. In addition, the solve rate differences between ARNSHL and ADNSHL likely reflect genetic mapping data: there are fewer ARNSHL loci with unknown causative genes. Unmapped ARNSHL and ADNSHL families represent a valuable resource for novel gene discovery.

61

Characterization of *de novo* copy-number variations in two subjects with a constitutional "CNV mutator" phenotype. *P. Liu[1], K. Walter[4], K. Writzl[5], V. Gelowani[1], S. Lindsay[4], C.M.B. Carvalho[1], M. Withers[1], J. Wiszniewska[1], A. Patel[1], B. Rautenstrauss[6], M.E. Hurles[4], J.R. Lupski[1,2,3].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 3) Texas Children's Hospital, Houston, TX, USA; 4) Wellcome Trust Sanger Institute, Hinxton, UK; 5) Institute of Medical Genetics, UMC, Ljubljana, Slovenia; 6) Medical Genetics Center, Bayerstrasse 3-5, Munich, Germany.

De novo copy-number variation (CNV) can occur constitutionally in gametogenesis or in early development leading to sporadic genomic disorders. Such *de novo* CNVs appear to also be important in somatic mutagenesis relevant to cancer and population events important to species evolution. Since large pathological CNVs are rarely observed at more than one locus in a single patient, and are often *de novo*, current efforts in understanding their molecular features and underlying mechanisms have relied on comparing CNVs from different individuals. Therefore, knowledge regarding size, mechanism and spatial distribution of *de novo* genomic rearrangements in a single genetic background is lacking. We report two subjects with developmental problems and multiple congenital anomalies presenting a constitutional "CNV mutator" phenotype, in whom multiple *de novo* rearrangements are observed on different chromosomes. Such observations are distinguishable from the phenomenon of chromothripsis in which the multiple CNV changes concentrate on one chromosome [Liu *et al.* (2012) *Curr Opin in Genet Dev* PMID:22440479]. Array comparative genomic hybridization (aCGH) and SNP arrays were used to map large *de novo* CNVs in both subjects and determine parental origins whereas breakpoint sequencing was performed to glean insights into mechanism. Subject #1 carries 8 large (>100 kb) copy-number gains, ranging from 104 kb to 6.4 Mb. Subject #2 carries 11 large copy-number gains, ranging from 211 kb to 4.7 Mb. Whole genome sequencing was performed on subject #1 to characterize smaller sized CNVs not efficiently interrogated by aCGH. Breakpoint sequencing analysis showed that microhomologies and breakpoint complexities are the prevailing features left at rearrangement traces, suggesting that the rearrangements were likely produced by replication mechanisms such as fork stalling and template switching and/or microhomology mediated break-induced replication (FoSTeS/MMBIR). Haplotype analysis in subject #1 revealed that the duplicated or triplicated materials were derived from both the paternal chromosome and the maternal chromosome, suggesting a postzygotic timing of the mutations. Our results document a genome-wide spectrum of *de novo* CNVs in a "CNV mutator" phenotype background, and we suggest that errors in the cellular DNA replication machinery could lead to multiple independent *de novo* rearrangements. Our findings have important implications for genomic disorders, cancer and evolution.

62

Associations between gene expression and phenotypes in 16p11.2 rearrangements. *E. Migliavacca[1,2], K. Mannik[1], F. Zufferey[3], N.D. Beckmann[3], L. Harewood[1], A. Mace[4,2], Z. Kutalik[4,2], L. Hippolyte[3], A. Maillard[3], V. Siffredi[3], R.M. Witwicki[1], G. Didelot[1], S. Jacquemont[3], S. Bergmann[4,2], J.S. Beckmann[3,4], The 16p11.2 European Consortium collaborators.* 1) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, University of Lausanne, Lausanne; 3) Service de Genetique, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne; 4) Department of Medical Genetics, University of Lausanne, Lausanne.

The 600kb-long 16p11.2 deletion containing 28 genes and its reciprocal duplication are some of the most frequent causes of neurodevelopmental disorders with population prevalence of 1/2000. These recurrent rearrangements are associated with mirror phenotypes such as obesity and underweight, macro- and microcephaly, as well as autism spectrum disorder (ASD) and schizophrenia. The transcriptome alterations triggered by these imbalances and their possible associations with the aforementioned mirror phenotypes remain unexplored. Mutations in *PRRT2* were identified in patients with epilepsy and paroxysmal dyskinesia; phenotypes often reported in deletion patients. An atypical 118kb deletion from *MVP* to *KCTD13* segregated with ASD but not epilepsy, whereas a 136kb duplication encompassing *SPN* and *QPRT* was identified in an anorexic patient. Recent experiments pinpointed that expression levels of *KCTD13* determines the head size of zebrafish embryos. We generated gene expression profiles of lymphoblastoid cell lines derived from 50 deletion and 32 duplication carriers, as well as 50 controls. The level of expression of the 28 genes mapping to the 16p11.2 interval correlates to their gene dosage. We observe a greater correlation of the expression levels of *KCTD13*, *MVP* and *MAPK3* with each other within cohorts, three genes shown to have an epistatic effect on zebrafish head size. To investigate the functional relationship between gene-expression changes and clinical phenotypes we transformed anthropometric measurements into Z-scores using age, gender and geographically-matched reference populations. We observed that Z-scores computed for weight and BMI were associated with both *MAPK3* and *MVP* expression after adjustment for copy number status in a multivariate normal linear model for adults (weight: p=0.01, p=0.0005; BMI: p=0.02, p=0.0003, respectively). We are currently identifying the pathways perturbed in deletion and duplication patients with weighted correlation network analysis or the Iterative Signature Algorithm and the whole genome transcriptome data. The latter method allows reducing the complexity of large sets of expression data by decomposing it into transcription modules, subsets of genes that exhibit a coherent expression profile and are thus likely to play a role in the same biological process. Our study will improve the understanding of the molecular basis underlying the clinical phenotypes of 16p11.2 syndromes.

63

***De novo* triplication can arise from a duplication of the 17p12 region and confers a severe Charcot-Marie-Tooth, Type 1A phenotype.** *V. Gelowani[1], P. Liu[1], F. Zhang[4], S.B. Shachar[5], S.D. Batish[6], E. Roney[1], V. Drory[7,8], A. Orr-Urtreger[5,8], J.R. Lupski[1,2,3]*. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 3) Texas Children's Hospital, Houston, TX, USA; 4) State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China; 5) Genetic Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; 6) Athena Diagnositics, Inc., Worcester, MA, USA; 7) Department of Neurology, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; 8) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Charcot-Marie-Tooth (CMT) disease represents a clinically and genetically heterogeneous group of peripheral neuropathies. Its most common form, CMT1A, is predominantly caused by a common duplication of a 1.4-Mb region in human chromosome 17p12 as a result of nonallelic homologous recombination (NAHR). Homozygous CMT1A duplication has been previously reported to convey a more severe clinical phenotype than heterozygous duplication, consistent with a *PMP22* gene dosage effect. We identified a complex pedigree in which patients presented with different severity of CMT. Array comparative genomic hybridization (aCGH) and fluorescence *in situ* hybridization (FISH) were used to investigate the copy number status of the 17p12 region in each family member. Individuals affected with a milder form of CMT carry the common recurrent CMT1A duplication. Interestingly, a recurrent triplication of the 17p12 region was found in a family member diagnosed with a severe form of CMT, and this triplication likely arose as a *de novo* event from the CMT1A duplication on the maternally inherited chromosome. In addition to this family, we also identified six other anonymized unrelated CMT1A patients with the same apparent triplication. In order to understand the mechanism of how a further copy number increase can result from a duplication, we phased haplotypes of the duplicated/triplicated segment using genotyping information from polymorphic microsatellite markers. In aggregate, our data show that recurrent triplications can occur at the CMT1A locus. Our findings provide mechanistic insights of how recurrent triplications occur at a disease locus, further support the gene dosage hypothesis, and illustrate the concept of Clan Genomics [Lupski *et al.* (2011) *Cell* PMID: 21962505] wherein new mutations within a family or clan contribute to clinically relevant phenotypes that evolutionary forces have not yet had the time to select against.

64

A long, non-coding RNA from the Prader-Willi critical region forms a subnuclear cloud and recruits transcriptional activating complexes to the Snord116 locus in postnatal neurons. *W.T. Powell[1,2], R. Coulson[1,3], F. Crary[1], S. Wong[1], D.H. Yasui[1], J.M. LaSalle[1]*. 1) Medical Microbiology and Immunology, Genome Center, M.I.N.D. Institute, UC Davis School of Medicine, Davis, CA; 2) Physician-Scientist Training Program, Biochemistry, Molecular, Cellular, and Developmental Biology Graduate Group; 3) Genetics Graduate Group.

Prader-Willi syndrome (PWS) is one of the leading genetic causes of obesity in children and is characterized by intellectual disability, obsessive-compulsive tendencies, hyperphagia, and an increased risk for developing autism-spectrum disorders. PWS is an imprinted neurodevelopmental disorder caused by deletions of the epigenetically regulated 15q11-13 locus on the paternal allele, with the smallest deletions resulting in the loss of two types of non-coding RNAs (ncRNAs): *SNORD116* small nucleolar RNAs (snoRNAs) and the spliced exons of the host gene, *116HG*. While prior attention has focused on the functional relevance of the snoRNAs in PWS, the molecular function and potential relevance of *116HG* is unknown. Using RNA and DNA fluorescence in situ hybridization (FISH), we observed that *116HG* forms a subnuclear cloud that increases in size with postnatal neuronal maturation and precedes a paternal-specific, neuron-specific chromatin decondensation of the PWS locus. Furthermore, *116HG* remains tethered to its site of transcription and separates from the neighboring lncRNA *115HG* subnuclear cloud following chromatin decondensation during the first week of life. Using a novel antisense targeted RNA precipitation approach to isolate endogenous *116HG* with associated protein and chromatin from mouse brain, we identified SAF-A and RBBP5 as interacting proteins. As SAF-A has both RNA- and DNA-binding domains and is known to function as a tether for the lncRNA *XIST*, we hypothesize that SAF-A acts as a tether between the *116HG* lncRNA and the *Snord116* locus. RBBP5 is part of the MLL1 complex that methylates histone H3 lysine 4 to activate transcription. Investigations of additional genomic loci that associate with *116HG* by next-generation sequencing are in progress. Expression of *116HG* from an exogenous locus in a human cell line produced an RNA cloud similar to that formed by endogenous *116HG*. In order to test the functional relevance of *116HG* in the pathogenesis of PWS, transgenic mice expressing the complete locus including *Snord116* and *116HG* versus spliced *116HG* alone are under development. Together, these results reveal a novel function for a poorly understood lncRNA in the PWS critical region important to understanding the pathogenesis of PWS and related neurodevelopmental disorders.Supported by NIH-F31NS073164 and NIH-1R01NS076263.

65

Molecular analysis of patients whose clinical features overlap the 22q11.2 Deletion Syndrome. *S. Saitta[1], T. Busse[2], D. McDonald-McGinn[2], E. Zackai[2,4], S. Woyciechowski[3], J. Garbarini[3], E. Goldmuntz[3,4], C. Van Ravenswaaij[5], L. Hoefsloot[5], B. Emanuel[2,4], D. Driscoll[2,6]*. 1) Medical Genetics Institute, Cedars-Sinai Health System, Los Angeles, CA; 2) Division of Genetics, Children's Hospital of Philadelphia, PA; 3) Division of Cardiology, Children's Hospital of Philadelphia, PA; 4) Department of Pediatrics, Perelman Sch of Medicine, Univ of Penn, Philadelphia, PA; 5) Department of Genetics, University of Groningen, Netherlands; 6) Department of OB-GYN, Perelman Sch of Medicine, Univ of Penn, Philadelphia, PA.

Disruptions of related developmental processes may lead to genetically distinct disorders that present with clinical overlap. This is recognized in syndromes with variable phenotypic features such as DiGeorge/Velocardiofacial syndrome. Its associated defects are typically related to a deletion of chromosome 22q11.2 and include conotruncal cardiac defects, palate malformations, immune dysfunction and learning disabilities. In 10–15% of clinical cases no deletion by FISH or point mutations in the TBX1 gene are found. We identified a cohort of 100 such patients referred from clinicians nationwide between 1995–2005 and performed clinically-driven molecular studies in subsets of these patients. We first utilized an MLPA kit specific for 22q11.2 to screen the entire cohort and found 8 patients with deletions not detected by FISH. These deletions did not include TBX1. In 4/20 patients with midline defects (e.g. hypertelorism, cleft palate, cardiac defect) we found MID1 mutations associated with Opitz/GBBB syndrome. Pathogenic mutations in CHD7, seen in CHARGE syndrome were found in 5/20 patients from the cohort with ear anomalies, cardiac and immune defects. Two more patients from this group showed variants of unknown significance in CHD7. Ten patients with varied anomalies were analyzed by SNP array and 3/10 showed pathogenic copy number alterations of other chromosomes (1). Non-genetic factors like maternal diabetes are also associated with DiGeorge syndrome, and may be considered causative in patients with negative 22q11 FISH. Our pilot study suggests that patients with cardinal clinical features can have 22q11.2 deletions missed by standard FISH, or harbor a genomic disorder of another chromosome. Using chromosome microarray as an initial study will identify these patients and should be considered for those with VCFS phenotypes who were found negative by FISH. Single gene disorders can also present with malformations analogous to those of del22q11. While most cases involving MID1 and CHD7 are clinically distinct, our study shows that overlap can be significant. The future application of whole exome or whole genome sequence analysis should help identify genes affected in overlapping phenotypes that may function in common genetic pathways to those of the 22q11 deletion syndrome. 1. Busse et al, 2011, Hum Mut 32: 97.

66

Mouse model implicates *GNB3* copy number in a novel childhood obesity syndrome. *I.S. Goldlust[1], K.E. Hermetz[1], L.M. Catalano[1], R.A. Cozad[1], R.T. Barfield[2], K.N. Conneely[1,2], J.G. Mulle[1,3], S. Dharamrup[1], M. Hegde[1], K. Kim[4], B. Angle[4], A. Colley[5], A.E. Webb[6], E.C. Thorland[7], J. Ellison[8], J. Rosenfeld[8], B.C. Ballif[8], L.G. Shaffer[8], L.A. Demmer[9]. Unique Rare Chromosome Support[10], M.K. Rudd[1]*. 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA; 2) Department of Biostatistics and Bioinformatics, Emory University School of Public Health, Atlanta, GA 30322, USA; 3) Department of Epidemiology, Emory University School of Public Health, Atlanta, GA 30322, USA; 4) Division of Genetics, Northwestern University Feinberg School of Medicine, Chicago, IL 60614, USA; 5) Department of Clinical Genetics, South Western Sydney Local Health District, Liverpool, BC 1871, New South Wales, AU; 6) Amy E. Webb Pediatrics, Pismo Beach, CA 93449, USA; 7) Mayo Clinic, Rochester, MN 55905, USA; 8) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA 99207, USA; 9) Division of Genetics and Metabolism, Tufts University School of Medicine, Boston, MA 02111, USA; 10) Unique Rare Chromosome Support Group, Caterham, Surrey CR3 5GN, UK.

Copy number variation (CNV) is a recently recognized cause of early-onset obesity comorbid with intellectual disability (ID). Here we describe a recurrent unbalanced translocation that causes a new syndrome associated with ID, macrocephaly, eczema, seizures, and obesity. The translocation is mediated by non-allelic homologous recombination between ~280 kb of segmental duplications on chromosomes 8p23.1 and 12p13.31. We recruited seven unrelated subjects with the same unbalanced translocation der(8)t(8;12); however, we have not identified subjects with the reciprocal der(12)t(8;12). We analyzed the inheritance of the der(8)t(8;12) by FISH and found that four were *de novo* and three were inherited from a mother who carries the balanced translocation. SNP analysis of trios revealed that all four *de novo* translocations were maternal in origin.

The unbalanced translocation der(8)t(8;12) leads to duplication of over 100 genes on chromosome 12, including the G-protein beta 3 (*GNB3*) gene. *GNB3* was identified over a decade ago as an obesity candidate gene in association studies, but its functional role in weight gain has remained elusive. Previous studies focusing on SNPs in *GNB3* correlated haplotype with obesity, hypertension, depression, and drug response. Here, we generated a BAC transgenic mouse model carrying an extra copy of the risk haplotype of human *GNB3*. We used a linear mixed model to analyze the longitudinal effects of the presence of the BAC transgene on mouse weight. The weight of the mice is significantly associated with transgene genotype ($p=0.002$). On average, heterozygous male and female transgenic mice are 4.1–7.5% and 6.0–14.3% heavier, respectively, than their wild-type littermates from age 7 weeks. Using quantitative reverse-transcriptase PCR we detect higher expression of human *GNB3* in whole brain as compared to endogenous *Gnb3*. These data connect *GNB3* gene dosage and overexpression to elevated body mass index and provide evidence for a new genetic syndrome caused by a recurrent CNV.

67

Modelling neurogenesis impairment in Down syndrome using induced pluripotent stem cells from monozygotic twins discordant for trisomy 21. *Y. Hibaoui[1,2], I. Grad[2], S.R Sailani[1], A. Letourneau[1], S. Dahoun[1], S. Gimelli[1], M.F. Pelte[3], F. Bena[1], S.E. Antonarakis[1], A. Feki[2,4]*. 1) Department of Genetic Medicine and Development, University of Geneva Medical School and Geneva University Hospitals, 1 rue Michel-Servet, CH-1211 Geneva, Switzerland; 2) Stem Cell Research Laboratory, Department of Obstetrics and Gynecology, Geneva University Hospitals, 30 bd de la Cluse, CH-1211 Geneva, Switzerland; 3) Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, 1 rue Michel-Servet, CH-1211 Geneva, Switzerland; 4) Service de gynécologie obstétrique, HFR Fribourg - Hôpital cantonal, Chemin des Pensionnats 2-6, Case postale, 1708 Fribourg, Switzerland.

Down syndrome (DS), caused by trisomy 21, is the most common chromosomal disorder, with an incidence of 1 in 800 live births. Its phenotypic characteristics include intellectual impairment and several other developmental abnormalities, for the majority of which the pathogenetic mechanisms remain unknown. Here, we report the generation and the characterization of induced pluripotent stem cells (iPSCs) derived from monozygotic twins discordant for trisomy 21: Twin-N-iPSCs for the normal and Twin-DS-iPSCs for the DS-affected iPSCs. We hypothesize that these samples were ideal to study the effect of the supernumerary chromosome 21, since the rest of the genome is identical between the two samples. Karyotype and high-resolution array-based comparative genomic hybridization analysis, confirmed the chromosomal constitution of these iPSCs. Transcriptome analysis by mRNA-Seq showed alterations in the expression of genes that impact on DS features. In vivo differentiation of Twin-DS-iPSCs revealed an abnormal teratoma formation in NOD-SCID mice. In vitro, Twin-DS-iPSC-derived neurospheres showed a reduced number of neuroprogenitor cells (NPCs). This effect was associated with a decrease of cell proliferation and an increase of cell death. When NPCs were further induced to mature into neurons, we found structural changes in the architecture and density of neuron, glial cell and oligodendrocyte populations together with alterations in the expression of genes involved in lineage specification in neurogenesis and brain development. Furthermore, we provide novel evidence that the increased expression and activity of the dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A (DYRK1A) protein likely underlie these defects. Importantly, we found that targeting DYRK1A pharmacologically or by shRNA corrected these defects. In conclusion, these findings establish iPSCs generated from human monozygotic twins discordant for trisomy 21 as a unique model to recapitulate DS phenotype in vitro, study the detailed mechanisms involved in the pathogenesis of DS and design new therapies.

68

Discovery and interpretation of balanced chromosomal aberrations in neurodevelopmental abnormalities and prenatal diagnostics. *M.E. Talkowski[1,2,3], V. Pillalamarri[1], I. Blumenthal[1], C. Hanscom[1], Z. Ordulu[4], J. Rosenfeld[6], L.G. Shaffer[6], J.F. Gusella[1,2,3], C.C. Morton[3,4,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Departments of Neurology and Genetics, Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 4) Department of Obstetrics, Gynecology, and Reproductive Biology, Harvard Medical School, Boston, MA; 5) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 6) Signature Genomic Laboratories, PerkinElmer, Spokane, WA.

Whole-genome deep sequencing (WGS) has revolutionized genomic research but is not yet feasible for routine prenatal diagnostics. Large-scale copy number variant (CNV), whole-exome sequencing (WES), and low-coverage WGS studies have provided insight into annotation of the human genome, however balanced chromosomal abnormalities (BCAs) remain glaring blind spots in gene discovery and predictive diagnostics. We developed a convergent genomics approach that fills this void, sequencing structural alterations using customized large-insert jumping libraries at a research cost comparable to karyotyping. Our approach leverages CNV data from independent academic and commercial diagnostic cases and controls (current n > 47,000), as well as WES and WGS cohorts, for interpretation of pathogenicity from BCAs disruptions. This approach identified 22 novel candidates contributing to neurodevelopmental abnormalities, as well as later onset neuropsychiatric disorders, suggesting some shared genetic etiology. We have further identified cryptic inversions as a frequent source of complex BCAs, cryptic balanced chromothripsis in the germline, and mechanisms of BCA formation that are distinct from benign CNVs. More recently, we developed a rapid approach to prenatal sequencing using a 13-day research protocol. Our initial pilot study sequenced DNA extracted from amniocytes of a prenatal sample (DGAP239) with a de novo translocation, discovering direct disruption of CHD7, a causal locus in CHARGE syndrome and LMBRD1, a pathogenic locus in a metabolic syndrome; the neonate was ultimately diagnosed with CHARGE syndrome and expired following complications at age 10 days. A second study performed in real-time identified a translocation breakpoint within 14 days in a prenatal sample (DGAP247) enrolled at 16-weeks gestation that directly disrupted KHDRBS3, representing a variant of unknown clinical significance based on interpretation from our convergent analyses of BCAs, CNVs, WES, and WGS data. The phenotypic outcome from this variant will be assessed following birth. The former subject represents an interpretable example of the power of WGS in a prenatal diagnostic setting, however DGAP247 illustrates that ambiguities will often arise given our still immature annotation of the human genome. These data suggest jumping libraries represent a reflexive test for cytogenetic diagnostic studies that is tractable, interpretable, and immediately available to the diagnostic community.

69

Predisposition of acrocentric short arm fusions due to nuclear location, nucleolar disorganization, and telomere-induced DNA damage. *K.M. Stimpson[1,2], L.L. Sullivan[1], S. Chen[1], B.A. Sullivan[1,2].* 1) Institute for Genome Sciences & Policy, Duke University, Durham, NC; 2) Department of Molecular Genetics & Microbiology, Duke University, Durham, NC.

Nuclear organization, genomic features, chromatin landscape, transcription, and DNA repair are key factors in chromosome and genome stability. In a human cell culture system, we have shown that transient telomere dysfunction via expression of dominant-negative mutant telomere protein TRF2 (dnTRF2) results in non-random chromosome interactions and fusions. The rapidly fused regions are the short arms of the acrocentric chromosomes so that these *in vitro* engineered chromosomes are the same as Robertsonian translocations, the most common chromosomal rearrangement in humans. Acrocentric short arms are comprised of multiple repeat families, including ribosomal DNA (rDNA). The rDNA repeats, or nucleolar organizing regions (NORs), come into proximity and are the site of nucleolar assembly. Thus, acrocentric nuclear positioning during this process may contribute to their fast and frequent fusion. During telomere disruption by dnTRF2, we observed a clear change in nuclear organization including alterations in nucleolar architecture. Additionally, acrocentric short arm DNA regions become unstable, as denoted by the formation of extrachromosomal circular DNA (eccDNA), enrichment of DNA damage markers, and altered chromatin organization and compaction. DNA damage markers also accumulate at nucleoli but the damage did not appear to have a global impact on rDNA transcription. ChIP results suggest that the early acrocentric-specific response may be augmented as a result of non-telomeric functions for endogenous TRF2 at acrocentric short arms that are disrupted by dnTRF2. We propose that stability of the acrocentrics depends on proper nuclear and telomere protein function that prevents DNA damage and ensures proper chromatin compaction. This *in vitro* system suggests several novel mechanisms for acrocentric stability in mitosis that could be paralleled in meiosis and provide insight into the molecular basis of Robertsonian formation in humans.

70

Lessons learned from Next-Gen Cytogenetics: Whole genome sequence-based prenatal diagnosis of apparently balanced de novo chromosome rearrangements. *Z. Ordulu[1], M. Talkowski[2,3,4], V. Pillalamarri[3], S. Pereira[1], I. Blumenthal[3], C. Hanscom[3], A. Lindgren[1], N. Hussain[5], S. Connolly[6], L.E. Wilkins-Haug[1], J.F. Gusella[2,3,7], C.C. Morton[1,2,8]*. 1) Ob/Gyn, Brigham and Women's Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 3) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 4) Department of Neurology, Harvard Medical School, Boston, MA, USA; 5) Department of Pediatrics, University of Connecticut, CT, USA; 6) Department of Radiology, Children's Hospital Boston, Harvard Medical School, Boston, MA, USA; 7) Department of Genetics, Harvard Medical School, Boston, MA, USA; 8) Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

Apparently balanced de novo chromosome translocations and inversions detected prenatally are associated with risks of 6.1 and 9.4% respectively for congenital anomalies. Implementation of next-gen whole-genome sequencing in prenatal diagnosis for chromosomal rearrangements offers an unparalleled high resolution test for diagnosis and management of genetic disorders. We performed large-insert sequencing of custom jumping libraries of DNAs extracted from amniotic fluid cells oftwo pregnancies. DGAP239 presented as a third trimester pregnancy with multiple fetal anomaliesand 46,XY,t(6;8)(q13;q13)dn with unremarkable aCGH. Disruptions of CHD7 at 8q12.2, a causal locus in CHARGE syndrome, and LMBRD1 at 6q13, a pathogenic locus in a metabolic syndrome were detected in 13 days. DGAP247 presented as a 16-week pregnancy with advanced maternal age, normal ultrasound and 46,XY,inv(8)(q13q24.1)dn, and also with unremarkable aCGH. Breakpoints in a non-genic region in 8q11.21 and in KHDRBS3 at 8q24.23 were detected in 14 days. In DGAP239, sequencing provided precise breakpoints and diagnosis of CHARGE syndrome which was not possible based on prenatal imaging, cytogenetic and aCGH testing. Sequencing would have changed medical care from an initial plan to repair an isolated heart defect to management of a morbid condition requiring immediate assessment of breathing and feeding difficulties. In addition to breakpoint definition in DGAP247, sequencing reassured the parents that a known genomic syndrome is not associated with the single disrupted gene effectively reducing concern for an untoward outcome. While prenatal diagnostic tools of karyotyping and FISH can be enhanced by aCGH and SNP arrays, detection of precise breakpoints to nucleotide level is not feasible. This study illustrates the capabilities of next-gen sequencing with its superior power to yield invaluable diagnostic insights in pregnancies with de novo balanced rearrangements. Early detection of a genetic disorder is of significant importance for informing genetic counseling, and for managing the pregnancy and birth, in addition to further clinical follow-up. Despite potential complexities including disruptions of two genes with possibilities of a gene fusion, dysregulation, deletion/duplication at the breakpoints, the resolution is critical to realizingthe benefits. This study foretells an empowered prenatal diagnostic environment in which DNA sequencing becomes the standard of care.

71

The use of chromosome microarray analysis as a first-line test in pregnancies with a priori low risk for detection of submicroscopic chromosomal abnormalities. *F. Fiorentino[1], S. Napoletano[1], C. Caiazzo[1], M. Sessa[1], S. Bono[1], L. Spizzichino[1], S. Michiorri[1], A. Gordon[2], A. Nuccitelli[1], G. Rizzo[1], M. Baldi[1]*. 1) GENOMA- Molecular Genetics Laboratory, Via di Castel Giubileo, 11 00138 Rome - Italy; 2) Bluegnome Ltd, Cambridge CB22 5LD, UK.

Objectives: Large-scale prospective trials have shown that the use of chromosome microarray analysis (CMA) in prenatal diagnosis produces a substantial improvement of ~1–3% in detection rate of genetic aberrations compared to conventional karyotyping. However, at present, it is not clear yet what indication of prenatal diagnosis could benefit from CMA. In the absence of specific guidelines and large-scale studies for different categories of indications, it has been suggested that CMA should be offered only as an adjunction tool to selected groups of high risk pregnancies (e.g. abnormal ultrasound findings and a normal conventional karyotype). In this study we aimed to explore the utility of CMA in groups of pregnancies with a priori low risk for detection of submicroscopic chromosome abnormalities, usually not considered an indication for testing, in order to assess if CMA improves the detection rate of prenatal chromosomal aberrations. Method: A total of 3000 prenatal samples were processed in parallel using both whole-genome CMA and conventional karyotyping. The indications for prenatal testing included: advanced maternal age (AMA), maternal serum screening test abnormality (MSS), abnormal ultrasound findings (AUS), known abnormal fetal karyotype (AFK), parental anxiety (PA), family history of a genetic condition (FIS), cell culture failure (CCF). Results: The use of CMA resulted in an increased detection rate regardless of the indication for analysis (Table). This was evident in high risk groups (AUS-AFK), in which the percentage of detection was 5.8%(7/120), and also in low risk groups, such as AMA (6/1118, 0.5%) and PA (11/1674, 0.7%). A total of 24 (0.8%) fetal conditions would have remained undiagnosed if only a standard karyotype had been performed. Importantly, 17 (0.6%) of such findings would have otherwise been overlooked if CMA was only offered to high risk pregnancies. Conclusions: The results of this study suggest that, instead of restricting the use of CMA only to fetuses with ultrasound anomalies, more widespread CMA testing of fetuses would result in a higher detection of clinically relevant chromosome abnormalities, even in low risk pregnancies. Our findings provide substantial evidence for the introduction of CMA as a first-line diagnostic test for all pregnant women undergoing invasive prenatal testing, regardless of risk factors.

72

The challenge of preconceptional, preimplantation, and prenatal genetic diagnoses of mitochondrial DNA disorders. *J. Steffann[1], S. Monnot[1], N. Gigarel[1], P. Vachin[1], E. Herzog[1], P. Burlet[1], N. Frydman[2], A. Benachi[3], G. Chalouhi[4], Y. Ville[4], R. Frydman[2], AS. Lebre[1], A. Rotig[1], D.C. Samuels[5], C. Elie[6], A. Munnich[1], JP. Bonnefont[1]*. 1) Genetics Department, Necker Hospital, Paris, France; 2) Biology and Genetics of Reproduction, Antoine-Beclere Hospital, Clamart, France; 3) Obstetrics Department, Antoine-Beclere Hospital, Clamart, France; 4) Obstetrics Department, Necker Hospital, Paris, France; 5) Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, USA; 6) Medical Statistics Department, Necker Hospital, Paris, France.

Mitochondrial DNA (mtDNA) mutations cause a wide range of serious genetic diseases with high transmission risk, due to their maternal inheritance. Owing to the absence of efficient therapy for such disorders, &ldquō risk" couples often ask for preconceptional (PCD), preimplantation (PGD) or prenatal diagnosis (PND). All these procedures are based on mutant load (heteroplasmy) assessment in i) polar bodies (PCD), ii) blastomeres (PGD), or iii) fetal tissues (PND). Supposed variations of mutant loads across cells or tissues, and throughout the pregnancy, have however limited the use of such procedures so far. Here we report our 10-year experience in PCD, PGD and PND for pathogenic mtDNA mutations in MT-ATP6 (m.8993T>G, m.9185T>C, NARP/ Leigh), MT-ND3 (m.10197G>A), MT-ND4 (m.14459G>A), MT-TL1 (m.3243A>G, MELAS), and MT-TK (m.8344A>G, MERRF). Mutant loads were quantified in 46 polar bodies, 19 oocytes, 52 preimplantation embryos, 35 fetuses, and 11 whole placentas from carrier females. PCD was not a relevant approach, due to common mutant load discrepancies between first polar bodies and their counterparts. Conversely, mutant loads were constantly stable among various blastomeres from each of 42 embryos, making blastomere-based PGD an appropriate procedure. Analysis of 9 fetuses over 10 gestation weeks, after termination of pregnancy in the frame of PND, showed that mutant loads were stable across various tissues including amniocytes. However, analysis of multiple samples in 11 placentas carrying various mutations found intra-placental mutant load variations up to 55%. The highest variations were found in intermediate mutant load ranges, despite normalisation of the mutant load dispersion by the mean. Such results preclude the use of chorionic villous sampling for PND. Finally, no temporal variation of heteroplasmy was observed throughout pregnancy (14 fetuses). These data support analysis of a single amniocyte sample as the best PND approach. A major issue was the predictive value of the embryofetal mutant load for the postnatal outcome. 19 children were born from PGD/PND procedures and are healthy with a 4 months - 10 years follow-up. Preimplantation and prenatal heteroplasmy assessments are therefore presumably predictive of the postnatal prognosis, even though a longer clinical follow-up is required, and are thus valuable tools to prevent the recurrence of the disease in families where a mtDNA mutation segregates.

73

The Incidence and Spectrum of Genomic Imbalance in Products of Conception: New Insights From SNP Microarray Analysis of 2,400 Miscarriage Specimens. *B. Levy[1], S. Sigurjonsson[2], B.J. Pettersen[2], M.K. Maisenbacher[2], Z.P. Demko[2], R. Lathi[3], R. Tao[2], V. Aggarwal[1], M. Rabinowitz[2]*. 1) Dept Pathology, Columbia Univ, New York, NY; 2) Natera Inc., San Carlos, CA; 3) Stanford Fertility and Reproductive Medicine Center, Palo Alto, CA.

2,400 POC samples were analyzed using a SNP-based microarray platform. Results were successfully obtained on 99.9% of samples. Of the remaining 3,997 specimens, 862 were identified as male and 1,535 as female (m:f ratio = 0.56). Using the Parental SupportTM informatics technique, 529 specimens (22%) were identified to have significant maternal cell contamination and exclusion of these samples reveals a more balanced m:f ratio of 0.91. 1,868 specimens had negligible or no maternal cell contamination and were considered to be true fetal results. Of these, 59.3% showed classical cytogenetic abnormalities with aneuploidy accounting for 85.5%, triploidy for 10.3% and structural anomalies or tetraploidy for the remainder. 96.5% of trisomies were of maternal origin while the majority (57.6%) of monosomies were paternal in origin. Approximately two-thirds of triploids were digynic in origin. UPD was found in 7/752 cases that did not show gross cytogenetic abnormalities and 3 of these cases represented whole chromosome UPD (Androgenetic = 2 and Gynogenetic = 1). The remaining 4 UPD cases were all maternal UPD and possibly indicative of a trisomy rescue event. A separate evaluation of the samples for copy number changes (CNCs) below the resolution of routine cytogenetic analysis (<10 Mb) was also performed. Of the 2,397 specimens yielding a result, 752 did not show gross cytogenetic abnormalities and of these, 33 (4.4%) showed a CNC that ranged in size from 400 Kb - 9.5 Mb. 36.4% of these were clinically significant and the remaining were variants of unknown significance (VOUS). Twenty samples showed CNCs <10 Mb in addition to aneuploidy and 4 of these were classified as clinically significant while the remainder were considered VOUS. This is the largest POC study in the genomics era and our data supports the routine use of SNP microarrays for clinical cytogenomic evaluation of POC specimens. SNP microarray analysis yields a greater success rate (due to the ability to analyze non-viable tissue), a higher diagnostic return and most importantly, it allows for the reporting of "true" fetal results. The finding of a clinically significant CNC in approximately 1:63 samples with a "cytogenetically normal" result is consistent with the recent report by the NICHD prenatal microarray study group. The overall distribution of genomic imbalance in POC samples, including sub-microscopic imbalances and UPD will be discussed and compared to routine cytogenetic analysis.

74

Noninvasive Whole-Genome Sequencing of a Human Fetus. *J.O. Kitzman[1], M.W. Snyder[1], M. Ventura[1,2], A.P. Lewis[1], R. Qiu[1], L.E. Simmons[3], H.S. Gammill[3,4], C.E. Rubens[5,6], D.A. Santillan[7], M.K. Santillan[7], J.C. Murray[8], H.K. Tabor[5,9], M.J. Bamshad[1,5], E.E. Eichler[1,10], J. Shendure[1]*. 1) Dept of Genome Sciences, University of Washington, Seattle, WA; 2) Dept of Biology, University of Bari, Bari, Italy; 3) Dept of Obstetrics and Gynecology, University of Washington, Seattle, WA; 4) Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Dept of Pediatrics at University of Washington School of Medicine, Seattle, WA; 6) Global Alliance to Prevent Prematurity and Stillbirth, Seattle Children's Hospital, Seattle, WA; 7) Dept of Obstetrics and Gynecology, University of Iowa Hospitals and Clinics, Iowa City, IA; 8) Department of Pediatrics, University of Iowa, Iowa City, Iowa, USA; 9) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA; 10) 0Howard Hughes Medical Institute, Seattle, WA.

Analysis of cell-free fetal DNA in maternal plasma holds great promise for the development of non-invasive prenatal genetic diagnostics. However, previous studies have been restricted to detection of gross abnormalities such as trisomies or large-scale deletions and duplications, or to genotyping common polymorphisms using invasively sampled material. We combined genome sequencing of two parents, genome-wide maternal haplotyping, and deep sequencing of maternal plasma collected at 18.5 weeks gestation to demonstrate the first truly non-invasive whole genome sequencing of a human fetus. Inheritance at 1.1 million phased, maternally heterozygous sites was predicted with 99.3% accuracy, and inheritance at 1.1 million unphased, paternally heterozygous sites was predicted with 96.6% accuracy. Furthermore, 39 of 44 de novo point mutations in the fetal genome were detected, albeit with limited specificity. Subsampling these data and analyzing a second family trio by the same approach indicate that ~300 kilobase parental haplotype blocks combined with shallow sequencing of maternal plasma are sufficient to substantially determine the inherited complement of a fetal genome. However, ultra-deep sequencing of maternal plasma is necessary for the practical detection of fetal de novo mutations genome-wide. Although technical and analytical challenges remain, we anticipate that non-invasive analysis of inherited variation and de novo mutations in fetal genomes will facilitate the comprehensive prenatal diagnosis of the over 3,000 recessive and dominant Mendelian disorders with a known molecular basis.

75

Spina bifida risk is conferred by multiple polymorphisms in folate one-carbon pathway genes. *D. Gilbert[1], K. Lazaruk[1], J. Stein[1], J. Hardin[1], J. Witte[2], G. Shaw[3], E. Lammer[4], N. Marini[5], J. Rine[5]*. 1) VitaPath Genetics, Foster City, CA; 2) Department of Epidemiology and Biostatistics and Institute of Human Genetics, University of California San Francisco, San Francisco, CA; 3) Department of Pediatrics, School of Medicine, Stanford University, Stanford CA; 4) Children's Hospital Oakland Research Institute, Oakland, CA; 5) Department of Molecular and Cellular Biology, California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, CA.

Background: Pre-conception supplementation with folic acid clearly decreases the population risk of having an NTD-affected pregnancy. Genetic analysis between the folate one-carbon pathway genes and NTDs has detected few noteworthy associations, with limited replication success. We undertook a case-control study of the association between the most common NTD, spina bifida (SB) and variants across 37 genes in the folate pathway.

Methods: 949 Caucasian mothers who had given live birth to a SB affected child and 1,166 matched controls were analyzed. DNAs were genotyped for single nucleotide polymorphisms (SNPs) in 37 genes in the folate-homocysteine pathway. The potential association between each of these SNPs and SB live birth was investigated. Those SNPs showing an association under various genetic inheritance models were used to develop a polygenic risk score prediction model for SB.

Results: DNA sequencing of 37 folate pathway genes in 241 infants with SB identified 824 SNPs suitable for inclusion in the maternal case-control study. 949 case moms and 1,116 controls were genotyped with the 824 SNPs to define risk alleles and develop the prediction model. We found 58 single nucleotide polymorphisms (SNPs) with significant ($p < 0.05$) associations in mothers and SB-affected offspring, providing compelling evidence of a distributed genetic risk signature centered on the folate-one carbon pathway. From these associations, we have developed a risk prediction model. Risk alleles present in 43 SNPs can identify women at 5-fold risk increased for spina bifida affected pregnancy.

Conclusions: The polygenic risk score prediction model provides an important advance in identifying women who are at an increased risk of having a SB-affected child. These findings are under evaluation in an independent population, but strongly suggest women without history of a spina bifida affected pregnancy can be identified as high risk based on their genetic profile.

76

Bioinformatics approach for identifying the genetic contributions to preeclampsia. *A. Uzun[1,2], I. Kurihara[1], J. Tavormina[2], R. Cabezas[2], A. Laliberte[1], A. Dewan[3], E. Triche[2], J. Padbury[1, 2]*. 1) Pediatrics, Women and Infants Hospital, Providence, RI; 2) Brown University, Providence, RI; 3) Yale University, New Haven, CT.

Preeclampsia is recognized as a leading cause of maternal and fetal morbidity and mortality worldwide. It is a multi-system, hypertensive disorder of pregnancy that complicates 2–8 % of deliveries in US. We have developed a bioinformatics approach that uses a comprehensive set of sources to identify biologically relevant variants to investigate for association with preeclampsia. In order to describe the genetic architecture of preeclampsia, we developed a semantic data mining aggregation tool to assemble the primary literature related to preeclampsia and associated genes. This included published literature and publically available databases with expression arrays. We constructed a database of genes with significant biological information for their role in preeclampsia. During the curation process genes were classified according to their association with the phenotypic characteristics of preeclampsia, source of the tissue (maternal, fetal or both), presence of intrauterine growth restriction (IUGR) and/or gestational hypertension (GH) status. If the data were from experiments performed in species other than the human the species information was also noted. We curated more than 900 articles. We identified valid evidence for 1055 genes involved in preeclampsia and those genes were supported by 340 articles. We then performed a cluster analysis to determine whether the curated genes segregated by phenotype. Cluster analysis was performed with the entire set of curated genes and compared to genes identified solely by expression arrays or other agnostic techniques. We observed significant segregation of the curated genes into distinct preeclampsia phenotypes. The cluster analysis indicates that there is an unbiased grouping of genes around specific phenotypes. We believe the expression data to be more hypothesis free. We believe this is evidence that preeclampsia is multiple, distinct conditions or genetically unique subclasses of the same disorder. Using a unique set of curated information consisting of statistically validated genes, pathways and individual genetic variants allows more robust, prospective statistical testing of the genetic association(s) with disease. This innovative approach will help to deal with the challenges of genetic association studies in the context of massive amounts of genomics data.

77

Antenatal spectrum of CHARGE syndrome in 40 fetuses with *CHD7* mutations. *M. Legendre[1, 2], M. Gonzales[3, 4], G. Goudefroye[5], F. Bilan[2, 6], P. Parisot[7], M.-J. Perez[8], M. Bonnière[5], B. Bessières[5], J. Martinovic[5], A.-L. Delezoide[9, 10], F. Jossic[11], C. Fallet-Bianco[12, 13], M. Bucourt[14], J. Tantau[15], P. Loget[16], L. Loeuillet[17], N. Laurent[18], B. Leroy[17], H. Salhi[15], C. Rouleau[19], F. Guimiot[9], C. Chelin[3], A. Bazin[20], C. Alby[21], A. Kitzis[2, 6], Y. Ville[21, 22], F. Encha-Razavi[1, 5, 22], B. Gilbert-Dussardier[2], M. Vekemans[1, 5, 22], T. Attié-Bitach[1, 5, 22], SOFFOET.* 1) INSERM U-781, Hôpital Necker-Enfants Malades, Paris Cedex 15, France; 2) Service de Génétique, Centre Hospitalier Universitaire de Poitiers, Poitiers, France; 3) Service de Génétique et d'Embryologie Médicales, Hôpital Armand Trousseau, Assistance Publique Hôpitaux de Paris (APHP), Paris, France; 4) Université Pierre et Marie Curie, Paris 6, France; 5) Département de Génétique, Hôpital Necker-Enfants Malades, APHP, Paris, France; 6) UMR CNRS 6187, Université de Poitiers, France; 7) Service de Cardiologie Pédiatrique, Hôpital Necker-Enfants Malades, APHP, Paris, France; 8) Département de Génétique Médicale, Hôpital Arnaud de Villeneuve, Montpellier, France; 9) Service de Biologie du Développement, Hôpital Robert Debré, APHP, Paris, France; 10) Université Paris Diderot, Paris 7, France; 11) Département de Fœtopathologie, Hôpital de Nantes, Nantes, France; 12) Service de Neuropathologie, Hôpital Sainte-Anne, APHP, Paris, France; 13) Université Paris Descartes, Sorbonne Paris Cité, France; 14) Service d'Anatomie et de Cytologie Pathologiques, Hôpital Jean Verdier, APHP, Paris, France; 15) Service d'Anatomopathologie, Hôpital Saint-Vincent de Paul, APHP, Paris, France; 16) Service d'Anatomie et Cytologie Pathologiques, CHU Pontchaillou, Rennes, France; 17) Service d'Anatomie et de Cytologie Pathologiques, CHI Poissy, Saint Germain en Laye, France; 18) Service d'Anatomie Pathologique, CHU Dijon, France; 19) Département d'Anatomopathologie, Hopital Lapeyronie, Montpellier, France; 20) Département de Génétique, Laboratoire CERBA, Cergy Pontoise, France; 21) Service de Gynécologie Obstétrique, Hôpital Antoine-Béclère, Clamart, France; 22) Université Paris Descartes, Paris 5, France.

CHARGE syndrome is a rare usually sporadic multiple congenital anomalies disorder. Although the syndrome is well known in children, only one series of 10 fetuses with CHARGE syndrome has been reported to date. The diagnosis of CHARGE syndrome is increasingly realized antenatally, but remains challenging in many instances. Here we report the clinical analysis of 40 cases of CHARGE syndrome with a CHD7 mutation in which a fetal (38 cases) or neonatal (2) clinical, radiological and histopathological examination was performed. Strikingly, the rate of males is unexpectedly high in our series (sex ratio 2,6/1) as opposite to postnatal studies where CHARGE syndrome is described more frequently in females, suggesting a greater severity in males. Interestingly, some features never reported to date were recurrent in our series. Most cases would be classified as non-typical ones according to Verloes et al. diagnostic criteria as many features are not observed antenatally. This prompted us to propose diagnostic criteria more suitable for antenatal cases. Interestingly, one of the mutated fetuses was initially diagnosed as 3C syndrome reminding that the variable phenotypic spectrum of CHARGE syndrome shares features with many other syndromes that will be discussed. Finally, all but one mutations identified in our antenatal series are truncating suggesting a possible phenotype-genotype correlation. In conclusion, the clinical analysis of this series allowed us to refine the clinical description of CHARGE syndrome in fetuses, describe some new features, suspect a more severe phenotype in males and set up diagnostic criteria for the antenatal presentation in order to help the diagnosis of CHARGE syndrome during pregnancies and/or after termination of pregnancies following the detection of severe malformations.

78

Genetic normalization of day-3 embryos: Results from two independent preimplantation genetic screening (PGS) laboratories. *P. Brezina[1], E. Littman[2], Y. Sun[3], V. Phan[2], R. Anchan[4], A. Barker[5], M. Hughes[6], G.R. Cutting[7], W.G. Kearns[1, 8]*. 1) Gynecology and Obstetrics, Johns Hopkins Medical Institutions, Baltimore, MD, USA; 2) Red Rock Fertility Center, Las Vegas, NV, USA; 3) Reproductive Medical Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Peoples Republic of China; 4) Gynecology and Obstetrics, Harvard University, Boston, MA, USA; 5) Arizona Center for Fertility Studies, Phoeniz, AZ, USA; 6) Genesis Genetics, Detroit MI, USA; 7) Institute of Genetic Medicine, Johns Hopkins Medical Institutions, Baltimore, MD, USA; 8) The Center for Preimplantation Genetics, LabCorp, Rockville, MD, USA.

The most common cause of spontaneous miscarriage is aneuploidy. Three in vitro fertilization (IVF) clinics and two independent genetic laboratories determined the percentage of aneuploid day 3 embryos diagnosed by 23 chromosome single nucleotide polymorphism (SNP) or comparative genomic hybridization (CGH) microarray PGS that exhibit euploidy at the blastocyst stage. Samples were obtained from patients undergoing IVF and PGS. Single blastomeres from Day-3 embryos underwent DNA amplification and 23-chromosome microarray analysis. Aneuploid embryos that differentiated to the blastocyst stage either underwent surgery to separate the inner cell mass (ICM) from the trophectoderm (TE) or TE biopsy only. DNA amplification and SNP or CGH microarray analyses were performed on all cells. Informed consent was obtained for all studies. 84 patients were enrolled by three IVF clinics. A total of 672 embryos had Day-3 biopsy and PGS. Chromosomal content of 444 Day-3 embryos were determined by CGH microarray in one laboratory while the remaining 228 embryos were evaluated by SNP microarray in a second laboratory. Overall, 41% (269/661) of embryos were euploid and 59% (392/661) were aneuploid. Of Day-3 embryos with euploid karyotypes, 66% (177/269) developed to the blastocyst stage while only 26% (103/392) of the aneuploid embryos progressed to the blastocyst stage. Chromosomal content of 103 Day-5 blastocysts that developed from aneuploid Day-3 embryos were analyzed by microarrays and results were obtained for 79 blastocysts. CGH analyses of the TE from 37 embryos showed 41% (15/37) with a euploid karyotype and 59% (22/37) with an aneuploid karyotype. SNP analyses of the separate ICM and TE from 42 embryos showed 60% (25/42) with a euploid karyotype, while 31% (13/42) of embryos had an aneuploid karyotype. The combined results revealed that euploidy was observed in 51% (40/79) of blastocysts that developed from aneuploid Day-3 cleavage embryos. The results from two independent PGS laboratories indicate that genetic normalization of embryos can occur during early stages of development. Our data suggests that a possible mechanism for this normalization is the loss of aneuploid cells during differentiation.

79

Heterozygous germline mutations in a prototypical TGFβ repressor cause Shprintzen-Goldberg syndrome with aortic aneurysm. *A.J. Doyle[1,2,3], J.J. Doyle[1,2], M.E. Lindsay[1,2], S.L. Bessling[1,4], N. Huso[1,2], D. Schepers[5], L. Gillis[5], G. Mortier[5], L. Van Laer[5], D.W. Mohr[1], M.J. Caulfield[3], A.F. Scott[1], C.J. Curry[6], B.L. Loeys[5], A.S. McCallion[1,4], H.C. Dietz[1,2].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 2) Howard Hughes Medical Institute, Baltimore, Maryland, USA; 3) The William Harvey Research Institute, Queen Mary University of London, London, UK; 4) The Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 5) Center for Medical Genetics, Antwerp University Hospital and University of Antwerp, Antwerp, Belgium; 6) Genetic Medicine of Central California, University of California San Francisco, California, USA.

Recent work has definitively implicated TGFβ signaling dysregulation in the pathogenesis of multiple syndromic presentations of aneurysm, including Marfan syndrome (MFS), Loeys-Dietz syndrome (LDS) and LDS-like conditions. These disorders show a clear signature for increased TGFβ signaling in the aortic wall and therapeutic interventions that antagonize TGFβ in mouse models of MFS achieve phenotypic rescue. However, the location and character of many of the underlying mutations (i.e. heterozygous loss of function mutations in TGFβ receptor subunits (TβRI/II), signaling effectors (Smad3) or ligands (TGFβ2)) would intuitively infer decreased TGFβ signaling. In addition, some phenotypic features of LDS (i.e. craniosynostosis and cleft palate) have historically been attributed to low TGFβ signaling. Taken together, these data have engendered substantial controversy regarding the specific role of TGFβ in disease pathogenesis. Shprintzen-Goldberg syndrome (SGS) is a connective tissue disorder that includes virtually all craniofacial, skeletal and cardiovascular manifestations of MFS and LDS, with the additional findings of mental retardation and skeletal muscle hypoplasia. Using whole exome sequencing of a single affected child-unaffected parent trio, we identified *de novo* causative variation in the oncogene *SKI*. Subsequently, heterozygous *SKI* mutations were identified in 9 additional sporadic SGS patients. The ski oncoprotein negatively regulates TGFβ activity through inhibition of R-Smad phosphorylation by TGFβ receptors, nuclear translocation of the R-Smad/Co-Smad complex and expression of TGFβ target genes via recruitment of transcriptional repressors, such as histone deacetylases (HDACs). All of the mutations cluster in either the R-Smad binding site or the dachshund domain of the protein that coordinates HDAC recruitment. Cultured dermal SGS fibroblasts show a profound excess of TGFβ signaling and output gene expression at steady-state and a prolonged response upon stimulation with TGFβ. We also found that morpholino-induced silencing of *skia* and *skib* in zebrafish recapitulates many of the abnormalities seen in SGS, including altered cardiovascular, craniofacial and skeletal development. These data definitively implicate increased TGFβ signaling in the manifestations of MFS, LDS and SGS, challenge prevailing assumptions regarding the pathogenesis of craniofacial disorders and inform our understanding of 1p36 deletion syndrome (which encompasses *SKI*).

80

Loss-of-function mutations in TGFB2 cause Loeys-Dietz syndrome: towards solving the TGFβ paradox in aortic aneurysmal disease. *B. Loeys[1], M.E. Lindsay[2,3], D. Schepers[1], N. Ajit Bolar[1], J. Doyle[3], E. Gallo[3], J. Fert-Bober[4], M.J.E. Kempers[5], E. Fishman[6], Y. Chen[3], L. Myers[3], D. Bjeda[3], G. Oswald[3], B.M. Anderlid[7], M.Y. Yang[8], E.M.H.F Bongers[5], J. Timmermans[9], A.C. Braverman[10], N. Canham[11], G.R. Mortier[1], H.G. Brunner[5], P.H. Byers[8], J. Van Eyk[4], L. Van Laer[1], H. Dietz[3].* 1) Center for Medical Genetics, Antwerp University Hospital and University of Antwerp, Antwerpe, Belgium; 2) Department of Pediatrics, Division of Pediatric Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 3) Howard Hughes Medical Institute and McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 4) Department of Medicine, Division of Cardiology, Bayview Proteomics Center, Johns Hopkins University, Baltimore, MD, 21224, USA; 5) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 6) Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA; 7) Department of Clinical Genetics and Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden; 8) Departments of Pathology and Genetics, University of Washington School of Medicine, Seattle, WA 98195, USA; 9) Department of Cardiology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; 10) Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA; 11) Department of Clinical Genetics, North West Thames Regional Genetics Service, (Kennedy Galton Centre), Northwick Park Hospital, Watford Road, Harrow, HA1 3UJ, United Kingdom.

Loeys-Dietz syndrome (LDS) is an autosomal dominant connective tissue disorder with features that both overlap with and distinguish it from Marfan syndrome (MFS). In its most typical presentation, LDS is characterized by a triad of hypertelorism, bifid uvula/cleft palate and widespread aortic/arterial aneurysms and tortuosity. LDS is most often caused by heterozygous loss-of-function mutations in genes encoding positive effectors of transforming growth factor beta (TGFβ) signaling including either subunit of the TGFβ receptor (TGFβR1/2) and SMAD3. Nevertheless, in human aneurysmal aortic tissues a signature compatible with increased TGFβ signaling is consequently observed and in mouse models of MFS, antagonism of TGFβ signaling using either TGFβ neutralizing antibodies or angiotensin receptor blockers attenuates multisystem disease manifestations, including aortic aneurysm. These observations engender the controversy regarding the precise role of TGFβ in aortic aneurysmal disease. TGFβs comprise three multipotential cytokines, encoded by three separate genes, that regulate multiple aspects of cellular performance including, proliferation, differentiation, migration and synthetic repertoire. It has been our hypothesis that the apparent TGFβ paradox could be reconciled if compensatory paracrine or autocrine events in response to a relative loss of TGFβ signaling potential leads to functional overshoot. Here we describe identification and mechanistic characterization of a new gene for a syndromic thoracic aortic aneurysm presentation that lends validity to this pathogenic model. We report heterozygous deletions or loss-of-function mutations of the gene encoding the transforming growth factor beta 2 (TGFβ2) ligand in eight families characterized by a phenotype within the MFS/LDS spectrum and demonstrate upregulation of TGFβ signaling in aortic tissue from affected individuals. Furthermore, haploinsufficient Tgfb2+/− mice demonstrate aortic root aneurysm by 8 month of age and biochemical evidence of increased canonical and noncanonical TGFβ signaling. Mice that harbor a mutant MFS allele (Fbn1C1039G/+) in the context of Tgfb2 haploinsufficiency show a pronounced increase in TGFβ signaling and phenotypic worsening in association with normalization of TGFβ2 expression and excessive expression of TGFβ1. Taken together, these data implicate compensatory autocrine and/or paracrine events and excessive TGFβ signaling in the pathogenesis of TGFβ vasculopathies.

81

Genetic dissection of aortic disease in the Marfan syndrome. *F. Ramirez[1], L. Carta[1], M. Del Solar[1], M. Lindsay[2], H. Dietz[2], J.R. Cook[1]*. 1) Pharmacology & Systems Therapeutics, Mount Sinai School of Medicine, New York, NY; 2) Division of Cardiology, Johns Hopkins University, Baltimore, MD.

Thoracic aortic aneurysm (TAA) in Marfan syndrome (MFS) is associated with dilatation of the aortic root and ascending aorta, medial layer degeneration, and elevated TGFβ signaling secondary to structural or quantitative defects in fibrillin-1 (FBN1). While mouse models of MFS have underscored the importance of fibrillin microfibrils in aortic tissue maturation and homeostasis, the contributions of different cell lineages and developmental stages to TAA formation and progression are yet to be determined. To this end, we created an Fbn1 conditional null allele ($Fbn1^{Lox}$) and examined the phenotype of the relevant progenies from crosses between $Fbn1^{Lox/-}$ mice and transgenic mice harboring tissue-specific or inducible Cre-drivers. In contrast to the normal phenotype of mice with Fbn1 inactivation in endothelial cells (Cdh5-Cre), mice without fibrillin-1 microfibrils in the medial layer of either the aortic root and the entire aortic tree (SM22α-Cre) or only in the aortic arch (Wnt1-Cre) died from ruptured TAA. The genetic data therefore identify the medial layer of the asecending aorta as the predominant determinant of TAA formation. Fbn1 was also inactivated at different stages of postnatal life using the inducible R26Cre-ER^T transgene. Mice with Fbn1 inactivated soon after birth (P1) died from ruptured TAA within 30 days, whereas the average survival of those with Fbn1 inactivation at P4 was extended for an additional month. By demonstrating that fibrillin-1 microfibrils are mostly involved in aortic tissue growth and compliance, these findings implied that characterization of the early postnatal mechanisms promoting TAA formation might identify druggable targets with the result of mitigating disease progression in pediatric MFS patients.

82

Bicuspid aortic valve, aortic coarctation and patent ductus associated with *MATR3* disruption in human and mouse. *F. Quintero-Rivera[1,2,14], Q.J. Xi[3], K.M. Keppler-Noreuil[4], J.H. Lee[1], A.W. Higgins[13], R. Anchan[3], A.E. Roberts[6,14], G.A.P. Bruns[7,14], R. Berezney[8], B.D. Gelb[9], R.V. Lacro[6,14], D.J. Harris[7,14], A. Kamp[10], I.P. Moskowitz[10], W. Lu[11], C.C. Morton[5,12,14], J.F. Gusella[1,14], R.L. Maas[3,14]*. 1) Molecular Neurogenetics Unit and Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 2) Pathology & Laboratory Medicine School Medicine UCLA, Los Angeles,CA; 3) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston,MA; 4) University of Iowa Hospitals and Clinics, Iowa City, OH; 5) Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston,MA; 6) Department of Cardiology, Boston Children's Hospital, Boston, MA; 7) Division of Genetics, Boston Children's Hospital, Boston, MA; 8) Department of Biological Sciences, University at Buffalo, Buffalo, NY; 9) Pediatrics and Genetics & Genomic Sciences, Mount Sinai School of Medicine, New York, NY; 10) Departments of Pediatrics and Pathology, University of Chicago, Chicago, IL; 11) Renal Section, Department of Medicine, Boston University Medical Center, Boston, MA; 12) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 13) Department of Pathology, University of Massachusetts Medical School, Worcester, MA; 14) Harvard Medical School, Boston, MA.

Cardiac outflow tract malformations represent a structurally and genetically heterogeneous subset of congenital heart disease for which gene identification has been difficult. Balanced human chromosomal rearrangements associated with cardiac malformations provide a potential genetic entry point into this problem. We describe a 46,XY,t(1;5)(p36.11;q31.2)dn translocation carrier with a Noonan-like syndrome disorder, and bicuspid aortic valve (BAV), coarctation of the aorta (COA) and patent ductus arteriosus (PDA). Only a few human genes have previously been associated with BAV, COA and PDA. The 5q breakpoint disrupts *MATR3*, which encodes the nuclear matrix protein Matrin 3, while the 1p breakpoint disrupts *AHDC1,* which encodes AT hook DNA binding motif containing 1 protein. To determine whether either gene could account for the proband's cardiac phenotypes, we assayed the embryonic expression of mouse *Matr3* and *Ahdc1*. *Matr3* is strongly expressed in the developing heart and great vessels, whereas *Ahdc1* is not. Moreover, while homozygous *Matr3* loss-of-function mice are embryo lethal, heterozygous *Matr3* loss-of-function mice exhibit incompletely penetrant BAV, COA and PDA phenotypes, as well as ventricular septal defect (VSD). Our studies extend the range of *MATR3* mutant phenotypes in human, reveal a striking developmental sensitivity to *Matr3* gene dosage, and establish that *Matr3* loss-of-function in mouse causes aortic and aortic valve phenotypes strikingly similar to those associated with human *MATR3* disruption.

83

Identification of the cause of Blue Rubber Bleb Nevus Syndrome. *J. Soblet[1], N. Limaye[1], M. Cordisco[2], A. Dompmartin[3], O. Enjolras[4], S. Holden[5], A.D. Irvine[6], C. Labrèze[7], A. Lanoel[2], P.N. Rieu[8], S. Syed[9], C.J. van der Vleuten[10], R. Watson[11], S.J. Fishman[12], J.B. Mulliken[13], L.M. Boon[14], M. Vikkula[1]*. 1) Human Molecular Genetics, du Duve Institute, Université catholique de Louvain, Brussels, Belgium; 2) Department of Dermatology, Hospital Garrahan, Buenos Aires, Argentina; 3) Department of Dermatology, Université de Caen Basse Normandie, CHU Caen, France; 4) Consultation des Angiomes, Hôpital Lariboisière, Paris, France; 5) Department of Clinical Genetics, Guy's Hospital, London, United Kingdom; 6) National Children's Research Centre and Department of Paediatric Dermatology, Our Lady's Children's Hospital, and Trinity College, Dublin, Ireland; 7) Dermatology Department, Hôpital Pellegrin Enfants, Bordeaux, France; 8) Department of Pediatric Surgery, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; 9) Department of Dermatology, Great Ormond Street Hospital for Children, London, United Kingdom; 10) Department of Dermatology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; 11) Department of Paediatric Dermatology, Our Lady's Children's Hospital, Dublin, Ireland; 12) Department of Surgery, Vascular Anomalies Center, Children's Hospital Boston, Harvard Medical School, Boston, MA; 13) Department of Plastic and Oral Surgery, Children's Hospital Boston, Harvard School of Dental Medicine and Harvard Medical Schoool, Boston, MA; 14) Center for Vascular Anomalies, Division of Plastic Surgery, Cliniques universitaires Saint-Luc, UCL, Brussels, Belgium.

Blue rubber bleb nevus syndrome (BRBN) is a rare sporadic congenital disorder (OMIM # 112200) characterized by multiple venous malformations all over the skin, often on hands and feet. Patients can present with a few to hundreds of cutaneous and pathognomonic gastrointestinal lesions. These are most commonly in the small intestine, documented by endoscopy, colonoscopy, or magnetic resonance imaging (MRI). Although several case reports have been published, the etiopathology of BRBN is still unknown. Since inherited venous malformations (VMCMs) are caused by germline activating TIE2 mutations and common sporadic venous malformations (VMs) are due to strongly hyperphosphorylating somatic TIE2 mutations, we hypothesized that BRBN may also be part of the spectrum of TIE2-mediated phenotypes. To test this, we screened the coding region of TIE2 by direct sequencing of genomic blood DNA and cDNA from the resected lesions of 14 patients. In 16 tissues from 10 patients, we identified mutations leading to amino acid changes, absent in the blood DNA from patients as well as in cDNA from control tissues. These changes occur at highly conserved residues, and are not found in dbSNP. In contrast to VMCMs and VMs, BRBNs predominantly show double (cis) mutations, suggesting a phenotype-genotype correlation. They cause ligand-independent receptor hyperphosphorylation in vitro. These results unequivocally demonstrate that BRBNs are caused by post-zygotic activating TIE2 mutations.

84

Identifying genetic determinants of congenital heart defect in Down syndrome. *M.R. Sailani[1], P. Makrythanasis[1], S. Deutsch[1], A. Valsesia[2], E. Migliavacca[1], F. Santoni[1], A. Sharp[1], C. Serra-Juhe[4], S. Vicari[5], R. Rabionet[3], Y. Grattau[6], G. Dembour[7], A. Megarbane[6], R. Touraine[8], S. Kitsiou[9], H. Fryssira[9], C. Chatzisevastou-IouKidou[9], E. Kanavakis[9], G. Merla[10], L. Perez-Jurado[9], X. Estivill[3], J. Delabar[11], S.E. Antonarakis[1,12].* 1) University of Geneva, Geneva, Switzerland; 2) Swiss Institute of Bioinformatics, UNIL, Lausanne 1005, Switzerland; 3) Centre for Genomic Regulation, Biomedical Research Park (PRBB), Barcelona, Spain; 4) Universitat Pompeu Fabra, E-08003 Barcelona, Spain; 5) Department of Neuroscience, Children's Hospital Bambino Gesù, Rome, Italy; 6) Institut Jerome Lejeune, Paris, France; 7) Cliniques universitaires Saint-Luc, UCL, Bruxelles, Belgium; 8) CHU de Saint-Etienne, hôpital Nord, Saint-Etienne Cedex 2, France; 9) Department of Medical Genetics, University of Athens, Athens, Greece; 10) Medical Genetics Service, Hospital, IRCCS, San Giovanni Rotondo, Italy; 11) Unité de Biologie Fonctionnelle et Adaptative (BFA), CNRS EAC 4413, Université Paris Diderot, Sorbonne Paris Cité, 75205 Paris, France; 12) iGE3 institute of Genetics and Genomics of Geneva, Switzerland.

Congenital heart defect (CHD) is a common developmental defect of Down syndrome (DS) occurring in 40% of cases. While carrying three copies of genes or other functional genomic elements on chromosome 21 increases the risk for CHD, trisomy 21 itself may not be sufficient to cause CHD. Thus additional genetic variation and/or environmental factors could contribute to CHD risk. Here we use association studies to identify genomic variations that in concert with trisomy 21, determine the risk for CHD in DS. This case-control GWAS includes 187 DS with CHD (AVSD=69, ASD=53, VSD=65) as cases, and 151 DS without CHD as controls. Chromosome 21 specific association study revealed rs2832616 and rs1943950 (both cis-eQTLs for KRTAP7-1 gene) as CHD risk alleles (adjusted p-values < 0.05). Furthermore rs2183593 and rs7282991 (both cis-eQTLs for ADARB1 gene) were identified as risk factors for ASD. Since DS is likely to be a disorder of gene expression, 2-locus interaction was applied for whole genome eQTLs. A pair of interacting eQTL on chr2 (MAP4K4 gene) and chr11 (SPA17 gene) was identified. Furthermore, a search for chr21 risk CNVs for CHD was performed using a customized whole chr21 array of 135K probes in samples from 55 DS-AVSD and 53 DS controls without CHD. This analysis has revealed two CNV regions of 4.8 and 1.8 Kb near genes RIP4 and ZNF295 respectively (FDR=0.04); these regions are located in the previously identified DS-CHD region by studying partial trisomy 21 cases; and a third CNV region of 13.4 Kb (FDR=0.03) upstream of POFUT2 gene. The results of this study strongly suggest that the genetic architecture of CHD risk of DS is complex, and includes specific SNPs and CNVs variations on chr21, and interaction of non-chr 21 genomic variants.

85

Transcriptome-wide decreased variation in gene expression of Down Syndrome fibroblasts - selection or canalization? *K. Popadin, A. Letourneau, F. Santoni, S.E. Antonarakis.* Department of Genetic Medicine and Development,, University of Geneva Medical School, Geneva, Switzerland.

Down syndrome is characterized on one hand by extensive phenotypic variability, and on the other by a recognizable dysmorphic syndrome. Here we compare whole transcriptomes by RNA-Seq of primary fibroblasts from 8 Down Syndrome (DS) and 8 Normal (N) unrelated individuals age and sex matched in order to study variation in gene expression. Gene by gene comparison of the whole transcriptomes revealed a remarkable decreased variance in gene expression level in DS versus N (median of the distribution of ratios of variances in DS to N is 0.64, which is less than the expected 1; P value < $1*10^{-16}$, Wilcoxon signed-rank test). The decreased variance in DS could not be explained by differences in mean expression levels between DS and N. This effect is present for genes on each individual chromosome (except chromosome 21) and for different sets of genes studied. The effect is also robust to the number of passages of fibroblasts. The effect is stronger in young (<2 years) versus old (> 2 years) individuals, and for highly-expressed versus low-expressed genes. We propose two potential explanations for this observation. (I) Assuming that expression level variation is not neutral, we hypothesize that decreased variation in DS gene expression can result from purifying prenatal selection of DS individuals. Since DS individuals have decreased fitness due to extra-chromosome 21 they may not tolerate additional deleterious effects, associated to variation in gene expression levels. Therefore DS embryos may undergo prenatal selection, with survival of DS with the level of expression of the majority of genes close to optimal. Indeed, about 80% of DS are miscarried, and we hypothesize that the miscarried DS fetuses may have non optimal patterns of gene expression. (II) This observation could also be a result of canalization of gene expression, that aims to produce a similar phenotype regardless of the variability of the genotype. Although whole transcriptome comparison demonstrates on average decreased variance in DS, there is an outlier group of genes with the opposite trend, i.e genes with increased expression variation. This group is enriched with genes involved in metal binding. We suggest that increased variation in this set of genes may be associated with variable DS phenotypes. KP was supported by EMBO long-term fellowship program ALTF 527-2010.

86

Mutations in OLFML2B within the QT interval associated region 1q23.3 Disturb Cardiac Repolarization, Predispose to Long-QT Syndrome and to Sudden Infant Death (SIDS). *A. Pfeufer[3,4,5,24], C. Congiu[3], Z. Schäfer[4,5], H. Prucha[1,2,6], M. Vennemann[7], I. Sinicina[8], N. Strutz-Seebohm[9,10], H. Kartmann[11], M. Schell[11], E. Kremmer[12], E.R. Behr[13], N.H. Bishopric[14,15], R.J. Myerburg[15], L. Crotti[4,16], P.J. Schwartz[16,17,18], A.A. Hicks[3], P.P. Pramstaller[3], W. Rottbauer[19,20], S. Kääb[11], T. Meitinger[4,5], M. Näbauer[11], M. Cohen[21], M. Donner[22], D.T. Mage[23], H.W. Mewes[24,25], T. Bajanowski[26], G. Seebohm[9,10], M. Ueffing[1,2], C.J. Gloeckner[1,2].* 1) Department of Protein Science, Helmholtz Zentrum München, German National Research Center for Environmental Health, D-85764 Neuherberg, Germany; 2) Division of Experimental Ophthalmology, Institute for Ophthalmic Research, University of Tübingen, D-72076 Tübingen, Germany; 3) Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy - Affiliated Institute of the University Lübeck, Germany; 4) Institute of Human Genetics, Klinikum rechts der Isar der Technischen Universität München, D-81675 Munich Germany; 5) Institute of Human Genetics, Helmholtz Zentrum München, German National Research Center for Environmental Health, D-85764 Neuherberg, Germany; 6) Klinik für Dermatologie am Biederstein, Klinikum rechts der Isar der Technischen Universität München, D-81675 Munich, Germany; 7) Institute of Forensic Pathology, Klinikum der Universität Münster, D- 48149 Münster, Germany; 8) Institute of Forensic Pathology, Klinikum Innenstadt, Ludwig-Maximilians-University München, D-81777 Munich, Germany; 9) Institut für Physiologie, Universität Bochum, D-44780 Bochum, Germany; 10) Biochemie I, AG Kationenkanäle, Ruhr-Universität Bochum, D-44780 Bochum, Germany; 11) I. Medizinische Klinik, Klinikum der Universität München - Grosshadern, D-81777 Munich, Germany; 12) Institute of Immunology, Helmholtz Zentrum München, German National Research Center for Environmental Health, D-85764 Neuherberg, Germany; 13) Department of Cardiology, St George's University of London, London SW17 0RE, UK; 14) Departments of Molecular and Cellular Pharmacology and Pediatrics, University of Miami School of Medicine, Miami, FL 33136, USA; 15) Division of Cardiology, Department of Medicine, University of Miami School of Medicine, Miami, FL 33136, USA; 16) Department of Cardiology, University of Pavia and Fondazione IRCCS Policlinico San Matteo, 27100 Pavia, Italy; 17) Laboratory of Cardiovascular Genetics, IRCCS Istituto Auxologico Italiano, Milan, Italy; 18) Chair of Sudden Death, Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia; 19) Abteilung Innere Medizin 3, Universitätsklinikum Heidelberg, D-69120 Heidelberg, Germany; 20) Klinik für Innere Medizin II, Universitätsklinikum Ulm, D-89081 Ulm, Germany; 21) Department of Pathology, Sheffield Children's Hospital NHS Trust, Western Bank, Sheffield S10 2TH, UK; 22) Dupont Haskell Global Centers for Health & Environmental Sciences, Newark, DE 19711, USA; 23) Biomolecular Core Laboratory, AI duPont Hospital for Children, Wilmington, DE 19803, USA; 24) Institute Bioinformatics and Systems Biology IBIS, Helmholtz Zentrum München, German National Research Center for Environmental Health, D-85764 Neuherberg, Germany; 25) Chair of Genome Oriented Bioinformatics, Center of Life and Food Science, Technische Universität München, D-85350 Freising-Weihenstephan, Germany; 26) Institut für Rechtsmedizin, Klinikum der Universität Essen, D-45122 Essen, Germany.

Rationale: Several GWAS have mapped the strongest human cardiac repolarization and QT interval modifying QTL to Chr.1q23.3 near NOS1AP (CAPON). Predisposition to repolarization disturbances and its sequelae (e.g. sudden infant death syndrome, SIDS) can be caused by monogenic (Long-QT Syndrome, LQT) as well as complex etiologies (e.g. SCN5A-p.S1103Y). Aim: We investigated regional genes NOS1AP and OLFML2B for rare variants (mutations) with strong effects on cardiac repolarization to potentially cause monogenic LQT or SIDS under monogenic disease models. Results: Both genes are expressed in the human heart. Their knockdown in zebrafish induces cardiac dilatation and arrhythmia. Among 96 cases of LQT (without mutations in known genes) we identified two related heterozygous carriers of the OLFML2B missense mutation p.R527Q. In 513 SIDS cases we identified three heterozygous carriers of OLFML2B missense mutations p.P504L, p.G515E and p.Y557H. No mutations were observed in NOS1AP. All four OLFML2B mutations suppressed cellular export of the secreted glycoprotein. Expression of wildtype OLFML2B in Xenopus oocytes reduced current density of the voltage gated IKr potassium channel (KCNH2, Kv11.1) but not of other cardiac ion channels by 8±5% while mutations led to larger reductions (from -18% to -49%). Conclusions: Our findings suggest that OLFML2B mutations can predispose to LQT and to SIDS by decreasing myocardial repolarization reserve. We uncovered a novel monogenic disease gene illuminating its pathomechanism acting through IKr. In addition we identify novel roles and properties of the olfactomedin class of proteins providing an example how GWAS based QTL mapping may enhance our general understanding of pathophysiology.

87

The impact of inherited genetic variants associated with lipid profile, hypertension, and coronary artery disease on the risk of intracranial and abdominal aortic aneurysms. *F.N.G. van 't Hof[1,2], Y.M. Ruigrok[1,2], A.F. Baas[3], L.A.L.M. Kiemeney[4,5,6], S.H. Vermeulen[4], A.G. Uitterlinden[7,8], F. Rivadeneira[7], A. Hofman[8], G.J.E. Rinkel[1,2], P.I.W. de Bakker[3,9,10,11]*. 1) Department of Neurology, University Medical Center Utrecht, Utrecht, Netherlands; 2) Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands; 3) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Department of Epidemiology, Biostatistics and HTA, Radboud University Medical Centre, Nijmegen, The Netherlands; 5) Department of Urology, Radboud University Medical Centre, Nijmegen, The Netherlands; 6) Department of Cancer Registry and Research, Comprehensive Cancer Centre, Utrecht, The Netherlands; 7) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; 8) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 9) Department of Epidemiology, University Medical Center Utrecht, The Netherlands; 10) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; 11) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts.

Background: Epidemiological studies show that an unfavourable lipid profile and coronary artery disease (CAD) increase risk for abdominal aortic aneurysms (AAA) but not for intracranial aneurysms (IA), and that hypertension is a risk factor for IA but not for AAA. We evaluated these observations in a genetic approach. Specifically, we investigated single nucleotide polymorphisms (SNPs) associated with serum lipid levels, blood pressure and CAD, and tested their contribution to AAA and IA risk.

Methods: We defined sets of SNPs previously reported to be associated with serum lipid levels, CAD and blood pressure (BP). From previously collected genome-wide data, we extracted genotypes for these SNP sets in 709 IA cases and 2692 controls, and 807 AAA cases and 1905 controls (all of Dutch origin). For each trait SNP set, we computed a genetic score in each individual of the IA and AAA cohort by summing the observed number of risk alleles weighted by their published effect size. We tested the genetic scores for association to IA and AAA using logistic regression, with adjustment for sex, validated IA or AAA risk SNPs and principal components. As a negative control, we calculated a risk score based on 180 SNPs associated with human height.

Results: After QC, 709 cases and 2612 controls in the IA cohort, and 797 cases and 1866 controls in the AAA cohort were left for analysis. We found significant associations for genetic scores of total cholesterol (p=3.6×10-6), low-density lipoprotein cholesterol (p=5.7×10-7) and CAD (p=0.0014) with AAA, and for the BP score with IA (p=0.0037). A genetic score of height was not associated with IA (p=0.76) or AAA (p=0.16).

Conclusion: We demonstrate that genetic risk profiles of lipid factors and CAD are associated with AAA but not with IA, and the genetic risk profile of BP is associated with IA but not with AAA. This is consistent with epidemiological observations. These findings support the need for further studies to investigate whether some of these associated loci contribute directly to IA and AAA risk.

88

The value of population-specific reference panels for genotype imputation in the age of whole-genome sequencing. *C. Fuchsberger[1], B. Howie[2], M. Laakso[3], M. Boehnke[1], G. Abecasis[1] on behalf of the Genetics of Type-2 Diabetes (Go-T2D) Consortium.* 1) Center for Statistical Genetics, University of Michigan, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

Genotype imputation is a key step in the analysis of genome-wide association studies. Genotype imputation works by identifying haplotype segments shared between study participants, which are typically genotyped on commercial arrays, and a reference panel of more densely typed individuals, such as those provided by The 1000 Genomes Project. In this study we evaluate the value of population-specific reference panels for imputation analyses. Based on interim GoT2D low-pass (4x) whole genome sequence data, we constructed reference panels for 300 UK and 700 Finnish samples. Using a leave-one-sample-out approach, we masked genotypes on chromosome 20 at all sites except those included on Illumina 300k SNP arrays, then imputed the missing sites using the remaining reference haplotypes. For comparison, we also imputed those samples using the latest 1000 Genomes reference panel. We evaluated imputation accuracy by examining the squared correlation (r^2) between imputed dosages and masked genotypes. We found that subsets of 100 Finnish or 200 UK reference samples achieved similar imputation accuracy to the complete 1000 Genomes reference panel: for SNPs with MAF 1–3%, 3–5%, and >5%, the mean r^2 was 0.77, 0.84, 0.92 for Finnish and 0.70, 0.80, 0.92 for UK samples, respectively. Using the full population-specific panels, we observed increased imputation accuracy for all SNPs, especially those with MAF 1–3% (r^2 of 0.87 for Finnish, 0.78 for UK samples). To investigate the future gain in imputation accuracy, we repeated the experiment with 9533 Finnish samples from the METSIM Study genotyped on the ExomeChip. We restricted our analysis to the densely typed HLA. For rare SNPs with MAF <0.025% and 0.025–0.05% the mean r^2 was 0.68 and 0.76 using the 1000 Genomes reference panel, 0.85 and 0.95 using the Finnish reference panel, respectively. Furthermore, the Finnish panel allowed the imputation of 40% more rare variants with a mean r^2 of 0.74. Our data show that for samples from Finland and the UK, population-specific reference panels produce higher imputation accuracy and genomic coverage than do larger, publicly available but less specific reference panels. As expected, the gain is higher for the less diverse Finnish population, for which ~5000 samples capture well the investigated haplotype space. Our results suggest that upcoming population specific panels composed of 1000s of individuals will boost imputation quality substantially for rare variants.

89

Fast and accurate 1000 Genomes imputation using summary statistics or low-coverage sequencing data. *B. Pasaniuc[1,2], N. Zaitlen[1,2], G. Bhatia[1,2], A. Gusev[1,2], N. Patterson[2], AL. Price[1,2].* 1) Epidemiology, Harvard School of Public Health, Boston, MA; 2) Medical and Population Genetics, Broad Institute of Harvard and MIT.

Imputation of untyped genotypes using external reference panels (e.g. 1000 Genomes) is a widely used approach for increasing power in GWAS and meta-analysis. Current HMM-based imputation approaches require individual-level genotypes and are computationally intensive. An alternative is to employ Gaussian models that use linear predictors to infer missing data, an approach that has been previously proposed (Wen&Stephens 2010) but has not been tested on empirical 1000 Genomes data. This approach is computationally fast (~30 CPU hours for 1000 Genomes imputation of 3,000 target samples, vs. ~17,000 CPU hours for Beagle). We find that Gaussian imputation recovers 84% (53%) of the effective sample size (as quantified by average χ^2 association statistics) for common (>5%) and low-frequency (1–5%) variants, vs. 89% (58%) effective sample size recovered by Beagle imputation. Interestingly, even when only summary level data is available, Gaussian imputation of association statistics recovers 83% (51%) of the effective sample size at common (low-frequency) variants. We also quantify the reduction in effective sample size when reference panel does not match the target sample; when imputing into the 1000 Genomes UK sample using the rest of 1000 Genomes European data as reference, we find a reduction in effective sample size of 83%(53%) for summary-level imputation as compared to 88%(58%) for Beagle imputation.

In our recent work (Pasaniuc et al. Nat Genet 2012) we showed that GWAS based on extremely low-coverage sequencing can attain several times the effective sample size of array-based GWAS per unit cost. Here we extend Gaussian imputation to genotype calling from extremely low-coverage sequencing, again using 1000 Genomes reference panels. We show that our approach attains comparable accuracy to Beagle imputation at common SNPs (average r^2 accuracy at common variants of 0.76 for Gaussian imputation vs. 0.90 for Beagle in sequencing simulations at 0.5x coverage), with orders of magnitude reductions in runtime.

90

Accurate haplotype estimation using phase informative sequencing reads. *O. Delaneau[1], JF. Zagury[2], J. Marchini[1].* 1) Department of Statistics, University of Oxford, Oxford, United Kingdom; 2) Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, Paris, France.

Knowledge of haplotypes provides valuable insights into human disease and population genetics and has stimulated development of haplotype estimation methods in recent years. However, with increasing use of next-generation sequencing (NGS), further methodological developments are needed to exploit the phase information contained in sequencing reads spanning 2 or more heterozygous genotypes. In the 1000 Genomes Project data we have found that one third of heterozygous genotypes are covered by phase informative reads (PIRs). All the commonly used phasing methods ignore this source of information. This has motivated us to extend SHAPEIT to utilize both PIRs in each sample in addition to LD information between SNPs and across multiple samples. To evaluate how this method performs using data available from current technologies we applied it to two parents of an European trio deeply sequenced with 100bp reads and 300bp inserts, merged with 382 European individuals of the 1000 Genomes Project. We tested the effect of different sequence coverage in the trio parents, ranging from no reads available to 40x coverage. While Beagle provides error-free segments (EFS) of ~115kb length on average, SHAPEIT gives EFS of ~180kb without using any reads and jumps to ~200kb and ~210kb using respectively 5x and 10x coverage. Interestingly, there is no substantial improvement above 10x. We have also carried out a range of simulations to assess how performance changes with read length, insert size and error rate. We first simulated haplotypes using SFS_code to mimic realistic allele frequency spectra and demography and then simulated reads based on these haplotypes for comparison of phasing methods. First, we assessed the impact of the length of single end reads. 10x sequencing with reads of length of 200bp and 500bp results in EFS that are ~5% and ~28% longer on average than when not utilizing PIRs. Second, we assessed the impact of insert size in pair-end reads. 10x sequencing with inserts of size 500bp and 1kp results in EFS that are ~13% and ~20% longer. Finally, we assessed the impact of higher base error rates in long-range reads in an attempt to mimic newer sequencing technologies. We find that SHAPEIT also performs well in this setting, and EFS are ~45% longer with only 1x of 5kb reads with 4% of errors, and ~250% longer with 10x. Overall our results highlight the gains that can be achieved by using phase-informative reads when estimating haplotypes from NGS data.

91

An LD-based method for genotype calling and phasing using low-coverage sequencing reads and a haplotype scaffold. *A. Menelaou, J. Marchini.* Dept Statistics, Univ Oxford, Oxford, United Kingdom.

Following the sequencing of 2,500 individuals from the 1000 Genomes Project (1KGP), reference panels will be used via imputation for the detection of causal variants, replacing the HapMap reference panels used up to now. The quality of those reference panels, depends on the methods for genotype calling and phasing applied. It has been recognized that models that account for Linkage Disequilibrium (LD) greatly aid genotype calling. The majority of the methods employed are extensions of imputation and phasing algorithms, that use sequencing information in the form of genotype likelihoods and pool information across samples to estimate haplotypes and genotypes. We propose a novel method that assumes a study design where individuals are both genotyped and also sequenced at low-coverage, as is the case with the 1KGP. Using the genotyped SNPs, a haplotype scaffold is constructed using a phasing algorithm, and pedigree information is incorporated when available. We then phase each non-scaffold site one at a time using an MCMC algorithm that iteratively updates the unobserved alleles based on the genotype likelihoods at that site and local haplotype information. We use a simple multivariate normal model to capture both allele frequency and LD information around each site that results in very efficient MCMC updates. When sequencing data is available from trios Mendelian transmission constraints are easily accommodated into the updates. Since the method analyses one position at a time, it is highly parallelizable, it does not suffer from edge effects, and can be applied to a large sample. We applied the method to all variant sites on chromosome 20 of the phase 1 of the 1KGP. The resulting haplotypes were used as a reference panel on an imputation analysis on 16 individuals from Complete Genomics. Comparing the imputation accuracy when using the reference panel created by our methods versus other reference panels created from the same dataset, we achieve the highest imputation accuracy, especially at rare SNPs, for the African individuals, where the scaffold is mostly pedigree phased. The non-ref genotype concordance for SNPs with allele frequency <2% is 0.24 followed by 0.19 (SNPTools,Thunder), and 0.16 (Beagle). For the Americas and European individuals, we achieve competing imputation accuracies. Our method can handle indels and multi-allelic variants, and can adapt a reference panel to enrich the haplotype scaffold for future projects.

92

Mixed functional linear model for sequence-based quantitative trait association studies unifying population and family study designs. *M. Cao, Y. Zhu, M. Xiong.* Biostatistics, University of Texas School of Public Health, Houston, TX., Select a Country.

Population-based sample design is the current major study design for association studies. However, many rare variants are from recent mutations in pedigrees. The inability of common variants to account for most of the supposed heritability and the low power of population-based analysis tests for the association of rare variants have led to a renewed interest in family-based design with enrichment for risk alleles to detect the association of rare variants. An individual's disease risk is likely to come from the collected action of common variants segregating in the population and rare variants recently arising in extended pedigrees. Analyzing samples from populations and pedigrees separately is highly inefficient. It is natural to unify population and family study designs for association studies. Although in the past few years we have witnessed the rapid development of novel statistical methods for population-based association studies using next-generation sequencing (NGS) data, only a few methods have been developed for pedigree-based association studies with NGS data. It is now well documented that NGS can generate several millions or even dozens of millions of genetic variation data. As a consequence, these genetic variation data are so densely distributed across the genome that the genetic variation can be modeled as a function of genomic location. But, standard multivariate statistical analysis often fails with functional data. The emergence of NGS demands a paradigm shift in the analytic methods for QTL(eQTL) analysis from standard single-variate or multivariate data analysis to functional data analysis. In this report, we propose mixed functional linear models for sequence-based quantitative trait association studies to unify population and family study designs in which a continuous phenotype is taken as a scalar response, genetic variants across the genomic regions as functional predictors and additive genetic background effects due to population and family structure and cryptic relatedness as random effects. By intensive simulations we demonstrate that family-based association studies have much higher power than the population-based association studies and the mixed functional linear model substantially outperforms other statistics. The proposed statistics are applied to family-based whole genome sequencing SardiNIA/Proge NIA studies. The results show that the mixed functional linear model has a much smaller P-value than other statistics.

93

Rare Variant Extensions of the Transmission Disequilibrium Test Detects Associations with Autism Exome Sequence Data. *Z. He[1], B. O'Road[2], J. Smith[2], G. Wang[1], M. Kan[1], S. Hooker[1], B. Li[1], N. Krumm[2], D. Nickerson[2], E. Eichler[2], S. Leal[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA.

Many population-based rare variant association tests have been developed to analyze sequence data. They aggregate variants across a genetic region which is usually a gene. A drawback of these methods is that it is difficult to adequately control for population substructure/admixture and spurious associations can occur. For rare variants this problem can be substantial, because the spectrum of rare variation can differ greatly between populations. A solution is to perform analysis using the transmission disequilibrium test (TDT), which was developed to analyze trio data and is robust to population substructure/admixture. Sequence data is being generated on trios to detect de Novo events. These data are also useful to detect association with transmitted variants. We extended the TDT to test for rare variant (RV) associations using four commonly used methods: CMC (Li & Leal, 2008), WSS (Madsen & Browning 2009) a modified version of GRANVIL (Morris & Zeggini, 2010) and VT (Price et al. 2010). By reconstructing parental genotypes association testing can be performed when there is missing parental data. De Novo events can also be incorporated. P-value can be obtained analytically or empirically using either genotype of haplotype permutation. We demonstrate that for all RV-TDT tests type I error is well controlled even when there is a high level of population substructure. The power of the RV-TDT tests was evaluated using a number of population genetic and disease models. There is no one single most powerful RV-TDT test although in many situations the CMC and the modified version of GRANVIL outperform other RV-TDT methods, although the difference in power is small. The RV-TDT was used to analyze exome data from 199 Simon Collection autism trios. Using the RV-TDT methods to analyze rare [minor allele frequency (MAF) < 1%] and rare and low frequency variants (MAF 1–5%) several genes were found to be associated with autism. An association with rare variants within the ZAN gene (p=6×10-6) that lies in the Autism 1 region on chromosome 7q22.1 was detected. Associations were also observed with rare and low frequency variants located within the DCP1B, ABCA7 and PPP6R1 genes. Given the problem of adequately controlling for population substructure/admixture in rare variant association studies and the growing number of sequenced based trio studies the RV-TDT is extremely beneficial to elucidate the involvement of rare variants in the etiology of complex traits.

94

Methods for Noninvasive Prenatal Determination of Fetal Genomes. *M.W. Snyder[1], J.O. Kitzman[1], M. Ventura[1,2], A.P. Lewis[1], R. Qiu[1], L.E. Simmons[3], H.S. Gammill[3,4], C.E. Rubens[5,6], D.A. Santillan[7], M.K. Santillan[7], J.C. Murray[8], H.K. Tabor[5,9], M.J. Bamshad[1,5], E.E. Eichler[1,10], J. Shendure[1].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Biology, University of Bari, Bari 70126, Italy; 3) Department of Obstetrics and Gynecology, University of Washington, Seattle, WA; 4) Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Department of Pediatrics at University of Washington School of Medicine, Seattle, WA; 6) Global Alliance to Prevent Prematurity and Stillbirth, an initiative of Seattle Children's, Seattle, WA; 7) Department of Obstetrics and Gynecology, University of Iowa Hospitals and Clinics, Iowa City, IA; 8) Department of Pediatrics, University of Iowa, Iowa City, IA; 9) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA; 10) Howard Hughes Medical Institute, Seattle, WA.

The sequencing of cell-free DNA in maternal plasma is a promising approach to non-invasive prenatal diagnosis, as this pool consists of a mixture of maternal- and fetal-derived molecules. Deep sequencing of the maternal plasma, coupled with haplotype-resolved genome sequencing of the parents, can be used to prenatally predict inheritance of parental alleles and to identify de novo mutations, effectively allowing for the non-invasive determination of the fetal genome. To assess inherited variation, we developed a hidden Markov model (HMM) that exploits allelic imbalance across parental haplotype blocks to predict transmission based on sequence data from maternal plasma. To evaluate this HMM, we tested it in the context of a trio where we had genome sequenced both parents, had haplotype resolved the maternal genome, and had deeply sequenced the maternal plasma. Using this model, accuracy of the inferred inherited alleles at 1.1 × 10^6 phased, `maternal-only' heterozygous sites was 99.3%. Among the top 95% of sites ranked by haplotype block length, prediction accuracy rose to 99.7%, suggesting that remaining inaccuracies can be mitigated by improvements in haplotyping. We compared these results to those obtained when parental genomes are statistically phased with a reference panel of 379 individuals, and found that prediction accuracy decreases with population minor allele frequency. We also identified 72 blocks predicted by the HMM to have been partially transmitted and, by comparison to overlapping clones, determined the subset representing likely recombination events. We also predicted de novo mutations in the fetus by considering high-quality bases observed in the plasma but inconsistent with Mendelian inheritance. This approach was highly sensitive, detecting 89% of true events, but lacked specificity, generating 2.5 × 10^7 candidates. We developed a downstream filtering approach for controlling specificity. After filtering, we reduced the number of candidates to approximately 4,000, a more manageable set for triaging and targeted validation depending on predicted functional impact. We identified and validated 17 true de novo mutations from this subset, confirming a 2,800-fold increase in signal-to-noise ratio. Further algorithmic improvements to the prediction of de novo events in the fetal genome are ongoing.

95

Associating complex traits with rare variants identified by NGS: improving power by a position-dependent kernel approach. *U. Bodenhofer, S. Hochreiter.* Institute of Bioinformatics, Johannes Kepler University, Linz, Austria.

Current high-throughput sequencing technologies have allowed for an easy and cost-efficient identification of rare single-nucleotide variations (SNVs), many of which have already been proven to be associated with diseases or complex traits. Despite these successes, genome-wide association studies involving rare variants remain statistically challenging. Classical single-SNV association studies particularly suffer from poor statistical power, as the potentially large number of SNVs often leads to poor significance upon false discovery rate (FDR) correction. To overcome these difficulties, approaches have been proposed that do not consider all SNVs individually; instead, they group SNVs and perform tests on those groups. This can either be done by grouping SNVs that are in the same genomic region of interest (e.g. the same transcript or exon) or by windowing along each chromosome. The choice of the groups/windows is crucial: FDR correction does not pose a serious problem if there are large, and consequently fewer, windows, but the local tests have poor power for large windows. If smaller windows, and consequently a larger number thereof, are chosen, the local tests perform well, but FDR correction nullifies this advantage. The windowing approach is based on the implicit assumption that accumulations of SNVs that are associated with the outcome (disease or trait) are unlikely to be caused by chance. In other words, the closer two SNVs are on the genome, the more likely they have similar effects on the outcome. Following this line of thought, we propose the Position-Dependent Kernel Association Test (PODKAT). PODKAT employs ideas similar to the acclaimed Sequence Kernel Association Test (SKAT) by Wu et al. with the crucial difference that the pairwise genomic distances of SNVs are explicitly taken into account in a way similar to position-dependent sequence kernels. Thereby, PODKAT is potentially able to be used for larger window sizes without drastically sacrificing statistical power. This claim is validated using a large test bed of data sets. Currently, many studies are underway that will identify many more rare SNVs, thereby also making the aforementioned statistical difficulties more severe. We strongly believe that PODKAT will provide an even stronger advantage over competing methods for further increased numbers of SNVs.

96

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: Demographic and Behavioral Influences on Telomeres and Relationship with All-cause Mortality. *C. Schaefer[1], S. Sciortino[1], M. Kvale[2], K. Lapham[2], J. Lin[2], D. Ranatunga[1], S. Rowell[1], M. Sadler[1], S. Miles[1], W. McGuire[1], D. Ludwig[1], L. Walter[1], I. Listerman[2], S. Van Den Eeden[1], R. Whitmer[1], C. Quesenberry[1], N. Risch[2], E. Blackburn[2].* 1) Div Research, Kaiser Permanente, Oakland, CA; 2) Institute for Human Genetics, UCSF, San Francisco, CA.

Telomere shortening is a biomarker of aging, but it is still unclear whether it plays a direct causal role in aging-related health changes and mortality. Mendelian syndromes and heritability studies demonstrate the biological underpinnings of telomere length (TL), yet other studies show that environmental and behavioral factors also influence telomere lengths. The Genetic Epidemiology Research Study on Adult Health and Aging (GERA) multiethnic cohort (average age = 63 years) has measured TL from saliva samples on over 100,000 individuals with linked electronic medical records. A detailed survey of demographic and behavioral factors was conducted 2 years prior to saliva collection, providing a unique opportunity to address questions of telomere epidemiology and aging. With these data, we examine demographic relationships with TL, behavioral influences, and relationship of TL with all-cause mortality following sample collection. As expected, TL is inversely correlated with age, and women have longer telomeres than men except as young adults. All analyses controlled for age and gender. As seen in other studies, we find significantly longer TL among African Americans than other groups, but no significant difference between whites, Latinos and Asians. TL is positively correlated with level of education and body mass index (BMI), and negatively correlated with cigarette smoking and alcohol consumption. All the above factors were independently significant in multivariate as well as univariate analyses. No associations were found with physical activity. We also found no relationship with diagnosis of major depression or stress-related disorders, even when limited to recent episodes. We found that short TL was prospectively associated with mortality, although only those with the shortest TL were at increased risk; the association persisted even after adjustment of the demographic and behavioral factors such as age, sex, race, education, physical activity, BMI, smoking, and alcohol consumption. In summary, while we found a broad range of demographic and behavioral factors that influence TL, a significant relationship of TL with mortality persisted after adjustment for all these factors. While this could indicate a direct effect of TL on health, it will also be important to examine the extent of pre-existing morbidities in these individuals to understand their possible role in the pathway between TL and longevity.

97

Diagnostic Exome Sequencing in Patients with Intellectual Disability of Unknown Cause. *J. de Ligt, M.H. Willemsen, B.W.M. van Bon, T. Kleefstra, H.G. Yntema, T. Kroes, A.T. Vulto-van Silfhout, D.A. Koolen, P. de Vries, C. Gilissen, A. Hoischen, H. Scheffer, B.B.A. de Vries, H.G. Brunner, J.A. Veltman, L.E.L.M. Vissers.* Department of Human Genetics - 855, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.

Intellectual disability (ID) is a common condition that carries lifelong medical and social consequences. The causes of ID remain largely unknown due to its extensive clinical and genetic heterogeneity. De novo mutations may play an important role in ID as most individuals present as isolated cases without family history, as has recently been demonstrated in a small number of individuals with ID. Here we evaluate the diagnostic potential and role of de novo mutations in a cohort of 100 patients with ID of unknown cause using family-based exome sequencing. The coding regions of >21,000 genes were captured and sequenced in 100 ID patients and their unaffected parents in a diagnostic setting. All patients were counseled and consented by a clinical geneticist prior to inclusion and studied by SNP microarrays to exclude causal de novo CNVs. A systematic data analysis pipeline was developed to identify and clinically interpret de novo mutations and X-linked maternally-inherited mutations in male patients. In addition, a high-throughput re-sequencing strategy was set up to screen an additional cohort of over 750 ID patients for mutations in candidate ID genes. All de novo mutations identified in this study were interpreted in the context of the clinical diagnosis. A total of 79 unique coding de novo mutations were identified and validated in 53 patients. In 13 patients damaging de novo (n=10) as well as X-linked maternally-inherited (n=3) mutations were identified in known ID genes, resulting in a minimal diagnostic yield of 13% in this cohort. In addition, potentially causative de novo mutations were found in another 22 patients in genes not previously associated with ID. Screening of additional ID patients for mutations in five candidate genes revealed further de novo mutations in DYNC1H1, CTNNB1 and GATAD2B. More importantly, phenotypic comparison of patients with de novo mutations in the same gene showed clear overlapping phenotypes, thereby establishing pathogenicity for these three genes. The combined diagnostic yield of exome sequencing was therefore 16% in this series. Because the remaining 19 candidate ID genes represent excellent candidates genes for ID, the future diagnostic yield of this approach may be as high as 35%. De novo mutations represent a major cause of previously unexplained ID that is now accessible through large scale sequencing strategies. Exome sequencing is indicated as a diagnostic assay for patients with ID of unknown cause.

98

C-terminal deletions of the AUTS2 locus cause distinct syndromic features and cognitive impairment. *E. Voorhoeve[1], G. Beunders[1], C. Golzio[2], L. Pardo[1], J. Rosenfeld[3], M. Talkowski[4], I. Simonic[5], A. Lionel[6], S. Vergult[7], R. Pyatt[8], J. van de Kamp[1], A. Nieuwint[1], M. Weiss[1], P. Rizzu[1], D. Posthuma[1], L. Verwer[1], H. Meijers-Heijboer[1], B. Menten[7], G. Mortier[9], S. Scherer[6], E. Eichler[10], S. Girirajan[10], N. Katsanis[2], A. Groffen[1], E. Sistermans[1].* 1) Department of Clinical Genetics, VU medical Center, Amsterdam, Netherlands; 2) Center for Human Disease Modeling, Department of Cell Biology, Duke University Medical Center, Durham, USA; 3) Signature Genomics Laboratories, Perkin Elmer, Inc., Spokane, WA, USA; 4) Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Harvard University, Boston, USA; 5) East Anglian Medical Genetics Service, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; 6) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Center for Medical Genetics, University Hospital, Ghent, Belgium; 8) Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH, USA; 9) Department of Medical Genetics, Antwerp University, Edegem, Belgium; 10) Department of Genome Sciences and Howard Hughes Medical Institute, University of Washington, Seattle, USA.

Background Translocations involving 7q21.22 have been associated with autism and intellectual disability (ID). Discordant evidence has suggested that these clinical phenotypes might be driven by a number of genes (AUTS2, WBSCR17, CALN1). Here, we examined whether AUTS2 disruptions are causal for neurocognitive defects.

Methods An international cohort of ~50,000 patients and 16,000 controls was examined for CNVs in AUTS2 by array CGH and we performed detailed phenotypic analyses. 5'RACE experiments were performed to test for alternative transcripts in human brain. A zebrafish knock down model was generated to test the potential of different AUTS2 splice isoforms to induce some of the anatomical phenotypes seen in patients.

Results We found 44 AUTS2 deletions. Microdeletions disrupting the coding sequence were causal to a complex syndromic ID/autism phenotype including short stature, microcephaly, cerebral palsy and distinct facial dysmorphisms. The exonic AUTS2 deletions occur with a frequency comparable to NSD1 deletions causing Sotos syndrome. We also discovered a novel, highly conserved, C-terminal AUTS2 isoform in human brain. Clinical and functional studies showed this isoform to be a key contributor to the human phenotype because (a) patients with C-terminal deletions exhibited significantly more severe and pleiotropic aspects of AUTS2 syndrome, and (b) the significant microcephaly and craniofacial defects seen in auts2 knockdown zebrafish embryos were rescued by this C-terminal isoform.

Conclusions We identified a hitherto unappreciated syndromic phenotype caused by deletions in AUTS2, required for both brain and craniofacial development. The C-terminal portion of the gene contributes significantly to the phenotype, demonstrating how transcriptional complexity can underpin human pathology.

99

Autism Traits in the RASopathies. *I. Corbin[1], G. Desachy[1], K. Rauen[2], L. Weiss[1]*. 1) Dept. of Psychiatry and Inst. for Human Genetics, University of California San Francisco, San Francisco, CA; 2) Department of Pediatrics, University of California San Francisco, San Francisco, CA.

Autism spectrum disorders (ASDs) are developmental disorders with impairment in language, communication, social skills, broad interests and flexible behaviors. Recent advances in ASD research has implicated Ras signaling, studied extensively in cancer. Interestingly, it has long been suggested that neurofibromatosis type 1 (NF1) may be associated with ASDs, although conflicting data called these reports into question. NF1 is a RASopathy, a syndrome caused by germline mutations in genes encoding components of the Ras/mitogen-activated protein kinase signaling pathway. Other RASopathies include Noonan syndrome (NS), Costello syndrome (CS) and cardio-facio-cutaneous syndrome (CFC). Our study aims to assess autism traits in RASopathies. Measures used are the Social Responsiveness Scale (SRS) and the Social Communication Questionnaire (SCQ). The SRS is a quantitative measure with continuously distributed and highly heritable scores. The SCQ is a qualitative threshold-based screener, based on the Autism Diagnostic Interview, the gold standard diagnosis method. The Fisher Exact test was used on SCQ outcome to test association between RASopathies and ASD. The student's t-test was used to compare SRS scores between each RASopathy and unaffected siblings or idiopathic ASD. Results in four RASopathies (NF1 n=37, NS n=23, CS n=28) show that a fraction of subjects (NF1 5%, NS 35%, CS 18%, CFC 56%) have significant social impairment at the level of ASD by SCQ, suggesting variation across disorders and within disorder. In NS, CS and CFC a significantly higher rate of ASD by SCQ is seen when compared to unaffected sibling controls. NF1 subjects did not show a significant difference, although the odds ratio for NF1 is elevated (OR=2), suggesting low power. SRS scores show significantly lower social responsiveness comparing those with NF1, NS, CS, and CFC with controls (P=3.2 × 10-3, 6.6×10-6, 1.3×10-6, and 9.7×10-14 respectively), though SRS scores are significantly less severe when comparing NF1, NS, CS, and CFC to idiopathic ASD (P=3.4×10-19, 3.4×10-08, 2.2×10-16, and 6.4×10-03 respectively). In conclusion, we demonstrate that RASopathy subjects have significant social impairment compared with controls. The fraction of clinically significant impairment varies across disorder and social responsiveness varies within each RASopathy. This systematic study supports that autism may be associated with Ras pathway dysregulation in human development.

100

Identification of novel recessive mutations in genes for intellectual disability. *B. De Vries[1], J.H.M. Schuurs-Hoeijmakers[1], A.T. Vulto-van Silfhout[1], L.E.L.M. Vissers[1], J. de Ligt[1], C. Gilissen[1], I. van de Vondervoort[1], M.T. Greally[2], C.W. Ockeloen[1], M.H. Willemsen[1], E.M. Bongers[1], G. Hira[1], J.Y. Hehir-Kwa[1], R. Pfundt[1], T. Kleefstra[1], K. Neveling[1], S. Reinateo[3], A. Vitello[3], P. Failla[3], D. Greco[3], M. Fichera[3], O. Galesi[3], B.M.W. van Bon[1], J.A. Veltman[1], C. Romano[3], M.A. Willemsen[1], H.G. Brunner[1], H. van Bokhoven[1], A.P.M. de Brouwer[1]*. 1) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Donders Institute for Brain, Cognition and Behaviour, Department of Pediatric Neurology and department of neurology, Department of; 2) National Centre for Medical Genetics, Our Lady's Children's Hospital, Crumlin, Dublin; 3) Unit of Pediatrics and Medical Genetics, Unit of Neurology, Laboratory of Medical Genetics IRCCS Associazione Oasi Maria Santissima, Troina, Italy.

Recent studies have shown that intellectual disability (ID) is mostly monogenic, but highly heterogeneous in its origin. Mutations in more than 10% of all genes might be involved in this common disorder affecting 2–3% of the general population. So far, small sibling families with ID have received little attention from the scientific community due to technical limitations to identify the underlying genetic defect. We used an exome sequencing approach to systematically identify recessive pathogenic alleles in 20 sibling families with ID. Sixteen families had affected brother-sister or sister-sister pedigrees (all but one of non-consanguineous parents) and four families consisted of affected brother pairs. Exome sequencing was performed on genomic DNA of one individual in each family. Pathogenic mutations were identified in three genes. In *DDHD2*, compound heterozygous frameshift mutations segregated with a spastic paraplegia-ID-structural brain malformation phenotype within a brother-sister family. *DDHD2* has not previously been implicated in human disease. A hemizygous stop mutation in *SLC9A6* and a hemizygous codon deletion in *SLC6A8*, both on the X-chromosome, were identified in two brotherpair families. Furthermore, seven novel candidate genes for ID were identified. In *TDP2*, a homozygous splice donor site variation inevitably resulting in aberrant mRNA splicing segregated in a family with three affected brothers. Recessive missense variations affecting amino acids that are conserved amongst (at least) vertebrates and predicted to influence protein function were identified in *SYNE1*, *ZNF582*, *MCM3AP* and *PAMR1* (last two in the same family). All four genes are expressed in neuronal tissue and all have a functional link to protein networks implicated in ID. A hemizygous missense variation affecting a conserved amino acid in *BCORL1*, a transcriptional corepressor, was present in two affected brothers. And lastly, in *PTPRT*, a protein tyrosine phosphatase involved in signal transduction and cellular adhesion in the central nervous system, we identified a combination of a heterozygous intronic deletion with a heterozygous missense variation that segregated with a complex ID phenotype in a brother-sister family. In this study, we identified (potentially) pathogenic mutations in one novel, two known and seven candidate ID genes in nine out of 20 (45%) families, clearly demonstrating that novel recessive mutations can readily be identified in small sibling families.

101

Causal de novo SNVs, indels and CNVs in children with undiagnosed developmental disorders. *M. Hurles[1], M. van Kogelenberg[1], K. Morley[1], T. Fitzgerald[1], S. Gerety[1], A. Tivey[1], S. Gribble[1], S. Al-Turki[1], S. Clayton[1], C. Wright[1], J. Barrett[1], H. Firth[1,2], D. FitzPatrick[3], N. Carter[1] on behalf of the DDD project.* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 2) East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK; 3) MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Western General Hospital, Edinburgh, UK.

To delineate the genetic architecture of severe undiagnosed developmental disorders in UK children we have deeply phenotyped over 2,500 affected children and their parents through a nationwide network of clinical geneticists, and recruited the families into a genetic research study entitled the Deciphering Developmental Disorders project. Seventy-five percent of the families are simplex. We are interrogating the causal roles of coding and regulatory SNVs, indels and CNVs by applying exome-array comparative genomic hybridization (exome-aCGH) to detect deletions and duplications, and exome-sequencing to detect sequence variants, in all coding exons, known enhancers, and highly conserved elements. We have profiled over 1,500 probands using exome-aCGH and determined the inheritance status of potential causal CNVs using SNP chips and custom assays. Our results are consistent with previous studies suggesting causal large CNVs can be identified in approximately 15% of children who had not previously been screened on clinical microarrays. We are identifying causal smaller and intragenic CNVs in those children who have been previously screened on clinical microarrays, and thus evaluate the advantages of exome-aCGH. We will describe exome sequence analysis of 230 of the parent-offspring trios. In a preliminary analysis of 64 trios we have identified likely causal de novo mutations in known developmental disorder genes in 14% of the families and in a further ~20% of families we have identified de novo mutations in highly plausible candidate genes, including several genes recurrently mutated in patients with similar phenotypic features. We have modeled some of these plausible candidate genes in zebrafish and identified concordant developmental phenotypes in morphant zebrafish for a subset of these. We will describe the breakdown of these causal and putative causal variants by phenotype and family history.

102

Making Headway with the Molecular and Clinical Definition of Rare Genetic Disorders with Intellectual Disability. *M.H. Willemsen[1], W.M. Wissink-Lindhout[1], L.E.M. Vissers[1,3], A.P.M. de Brouwer[1], J.H.M. Rensen[2], N. de Leeuw[1], R. Pfundt[1], H.G. Yntema[1], J. de Ligt[1,3], J.A. Veltman[1,3], H.G. Brunner[1,3], H.M.J. Lantman - de Valk[4], B.C.J. Hamel[1], H. van Bokhoven[1,3,5], T. Kleefstra[1,3].* 1) Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 2) Pluryn, Care provider for people with intellectual disabilities, Oosterbeek, The Netherlands; 3) Nijmegen Centre for Molecular Life Sciences, Institute for Genetic and Metabolic Disease, Nijmegen, The Netherlands; 4) Department of primary and community care, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Department of Cognitive Neuroscience, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Background The cause of intellectual disability (ID) is unknown in at least 50%. Knowing the cause is of major importance in care and counseling of patients and families and provides insight in co-morbidity, associated behavioral problems, prognosis and recurrence risk. It precludes further unnecessary, often incriminating testing, and fruitless interventions, and instead enables specific anticipation on associated health and behavioral problems, since the extreme clinical and molecular heterogeneity goes along with diverse presentations. In recent years, advances in genetic technologies have provided great new diagnostic opportunities. We aimed to identify genetic causes of ID in a cohort of 253, mainly adult, patients with unexplained ID receiving services from Dutch residential settings.

Methods This study comprised two phases. Part one of our study included a multidisciplinary clinical evaluation, followed by specific genetic diagnostic tests if indicated, a standard genome-wide array analysis and a metabolic screen in all 253 individuals. Subsequently, 47 of 191 (24.6%) patients that remained undiagnosed were selected for further studies using next generation sequencing (NGS) approaches. In 40 patients with sporadic ID we used family-based whole exome sequencing (WES) to detect *de novo* mutations, and in 7 patients with familial ID we performed WES or X-exome sequencing.

Results Part one resulted in 18.3% genetic diagnoses comprising of 12% chromosomal abnormalities and 5% monogenetic defects, mostly fitting syndromes for which the causative genes were identified recently as *EHMT1, TCF4* and *CDKL5*. In 1.3% a primary metabolic cause was established. Part two resulted in another 36.2% (17/47) likely pathogenic genetic causes which suggests that the total yield of likely genetic diagnoses in the studied cohort is 54.5%. NGS revealed pertinent diagnoses including mutations in the known ID genes *PDHA, GRIN2A, LRP2* and *OPHN1*, mutations in novel ID genes, leading to the definition of novel ID syndromes, and mutations in plausible novel ID genes.

Conclusion By careful clinical evaluation and making use of current day technologies, we identified over 50% of likely genetic causes in patients selected for previously unexplained ID, which reveals a comprehensive and promising yield of the currently available advanced and rapidly developing genetic diagnostic repertoire and is of major importance for the care and counseling of patients and families.

103

***MBD5* dosage affects multiple neurodevelopmental pathways in common with other genetic syndromes.** *S.V. Mullegama[1], J.A. Rosenfeld[2], C. Orellana[3], B.WM. van Bon[4], E.A. Repnikova[5], L. Brick[6], L. Dupuis[7], D.J. Stavropoulos[8,9], D.L. Thrush[5], J.G. Foster[5], K. Manickam[5], A. Lin[10], J.C. Hodge[11], M.E. Talkowski[12,13,14], J.F. Gusella[10.12.14.15], S. Schwartz[16], S. Aradhya[17], R.E. Pyatt[5], B.BA. de Vries[4], R. Mendoza-Londono[7], L.G. Shaffer[2], S.H. Elsea[18]*. 1) Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA; 2) Signature Genomic Laboratories, PerkinElmer Inc., Spokane, WA, USA; 3) Prenatal Genetics and Diagnostics, University Hospital of La Fe, Valencia, Spain; 4) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH, USA; 6) Department of Pediatrics, Clinical Genetics Program, McMaster University Medical Center and McMaster Children's Hospital, Hamilton, ON, CA; 7) Department of Pediatrics, Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, ON, CA; 8) Department of Pediatric Laboratory Medicine, Cytogenetics Laboratory, Hospital for Sick Children, Toronto, ON, CA; 9) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON, CA; 10) Medical Genetics, MGH for Children, Boston, MA, USA; 11) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; 12) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 13) Department of Neurology, Harvard Medical School, Boston, MA, USA; 14) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 15) Autism Consortium of Boston, Boston, MA, USA; 16) Laboratory Corporation of America, Research Triangle Park, NC, USA; 17) GeneDx, Gaithersburg, MD, USA; 18) Department of Pediatrics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA.

Copy number variations (CNVs) that involve chromatin-modifying genes play an important role in the genetic etiology of many neurodevelopmental disorders (NDs), including intellectual disability (ID), epilepsy, and autism. Studies show that many NDs associated with CNVs emerge from abnormal dosage of chromatin modifying genes suggesting that strict regulation of gene dosage is required for proper neurodevelopment. We hypothesize that phenotypically similar but genetically distinct NDs involve overlapping genetic pathways affected by improper dosage of genes critical to neurodevelopment and behavior. We provide evidence that gene dosage is critical to the functioning of key developmental pathways through analysis of the chromatin modifying gene, methyl-CpG binding domain 5 (*MBD5*), in a cohort of >100 individuals with *MBD5* defects. We previously reported that haploinsufficiency of *MBD5* is the primary causal factor in 2q23.1 deletion syndrome and that mutations in *MBD5* are associated with autism. We now provide evidence that increased dosage of *MBD5* is equally detrimental. We report 24 unrelated dup 2q23.1 cases, establishing a duplication syndrome complementary to the microdeletion syndrome, with a phenotype that includes ID, motor delay, language and sleep impairments, behavioral problems, and autistic features. Duplications range from .068–34.98 Mb and include *MBD5*, which is over-expressed in patient cells. The phenotypes associated with deletion, mutation, or duplication of *MBD5* and the gene expression changes observed indicate *MBD5* is dosage sensitive and critical for normal development. Dup 2q23.1 causes a phenotype similar to del 2q23.1 and other NDs, like Smith-Magenis syndrome and autism, suggesting shared molecular pathways. We show in patient cells deleted or duplicated for *MBD5* and in *MBD5* RNAi knockdown cell lines that *RAI1* and *MBD1* have significantly altered expression when *MBD5* dosage is altered. Expression microarray data and confirmatory qPCR further show that *MBD5* plays a role in chromatin remodeling, affecting gene expression in circadian rhythm, neuronal development, and cell growth/survival pathways. Corroborating, we show that *NR1D2*, a circadian rhythm gene, has altered expression in *MBD5* cases, thus identifying a specific pathway affecting sleep. Overall, these findings support a dosage-specific role for *MBD5* in neuronal development by mediating key nuclear functions important for regulation of gene dosage and expression.

104

Exome sequencing in X-linked intellectual disability family assess the role of the KIAA2022 gene in the aetiology of intellectual disability. *M. Rio[1], S. Mouton[2], AC. Mazery[1], C. Bole-Feysot[3], P. Nitschke[4], N. Bahi-Buisson[5], A. Munnich[2], L. Colleaux[2]*. 1) Département de Génétique, Hôpital Necker-Enfants Malades, Assistance Publique-Hôpitaux de Paris, INSERM U781, Université Paris Descartes, Paris, France; 2) INSERM U781, Université Paris Descartes, Département de Génétique, Hôpital Necker-Enfants Malades, Assistance Publique-Hôpitaux de Paris, Paris, France; 3) Plateforme Génomique, Institut Imagine, Université Paris Descartes, Paris, France; 4) Plateforme Bioinformatique, Institut Imagine, Université Paris Descartes, Paris, France; 5) Département de Neuropédiatrie, Hôpital Necker-Enfants Malades, Assistance Publique-Hôpitaux de Paris, Paris, France.

X-linked intellectual disability (XLID) is a heterogeneous group of syndromic and non-syndromic disorders. Mutations in more than 90 other genes have been involved in XLID. Yet, each gene account for only a very small number of cases and linkage data or mutation screening of known XLID genes strongly suggest that many more remain to be discovered. More recently, the search for novel XLID genes has shifted toward high-throughput sequence analysis of all genes encoded by the X chromosome. Here we report on the clinical and molecular characterisation of an X linked non syndromic mental retardation family ascertained by next-generation sequencing (NGS) of most of the X-chromosome transcripts. Two brothers were referred for genetic work up because of intellectual disability. The youngest presented with severe intellectual disability, seizures and disordered hyperactivity. The elder brother was highly social, communicated verbally with good sentence structure and was able to read and write. His IQ was 50. Their 38 years old maternal uncle was unemployed, unable to read and write and his IQ was 46. All of them had normal neurological examination and no remarkable facial features were present. The two affected sibs were tested using the XSeq™ Research Screening Panel developped by Raindance Technologies. We generated coverage of at least 90 reads for more than 90% of the targeted sequences and identified 1259/1397 genetic variants per proband. We filtered variants to systematically identify common events in each proband and focused on protein-altering and splice-site variants. We identified 2 previously unreported variants in the PLP2 gene and in the KIAA2022 gene that were not present in an internal exome dataset of 125 individuals or among the 5,379 exomes available from the NHLBI Exome Sequencing Project. Additional analysis showed that only the KIAA2022 mutation segregated with known carrier status and in all affected individuals. Interestingly, disruption of this gene have previously been identified in two mentally boys and this gene is highly expressed in fetal brain and the adult cerebral cortex. Our data confirm the role of the KIAA2022 gene in the aetiology of intellectual disability and the interest of NGS in small XLMR families.

105

Biallelic mutations of a ubiquitin-ligase-encoding gene cause an Ohdo-like intellectual disability syndrome. *G. Borck[1], B. Dallapiccola[2], R. Ramirez-Solis[3], A. Segref[4, 5], H. Thiele[6], A. Edwards[7], M.J. Arends[8], X. Miro[9], J.K. White[3], J. Desir[10], M. Abramowicz[10], M.L. Dentici[2], K. Hofmann[4], A. Har-Zahav[11], E. Ryder[3], N.A. Karp[3], N.J. Ingham[3], G. Nuernberg[6], S. Abdelhak[12], M. Pasmanik-Chor[13], O. Konen[14], R.I. Kelley[15], M. Shohat[11,16], P. Nuernberg[5,6], J. Flint[7], K.P. Steel[3], T. Hoppe[4, 5], C. Kubisch[1], D.J. Adams[3], L. Basel-Vanagaite[11,16]*. 1) Institute of Human Genetics, University of Ulm, Ulm, Germany; 2) Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; 3) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; 4) Institute for Genetics, University of Cologne, Cologne, Germany; 5) CECAD, University of Cologne, Cologne, Germany; 6) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 7) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 8) University of Cambridge, Department of Pathology, Addenbrooke's Hospital, Cambridge, UK; 9) Institute of Molecular Psychiatry, University of Bonn, Bonn, Germany; 10) Department of Medical Genetics, Hôpital Erasme, Brussels, Belgium; 11) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 12) Molecular Investigation of Genetic Orphan Diseases, Pasteur Institute, Tunis, Tunisia; 13) Bioinformatics Unit, G.S.W. Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel; 14) Imaging Department, Schneider Children's Medical Center of Israel, Petah Tikva, Israel; 15) Kennedy Krieger Institute, Baltimore, MD, USA; 16) Raphael Recanati Genetics Institute, Rabin Medical Center, Beilinson Campus, Petah Tikva, Israel.

Ubiquitination plays a crucial role in neurodevelopment and genetic alterations of the ubiquitin ligase-encoding *UBE3A* gene contribute to Angelman syndrome, a common form of syndromic intellectual disability, and autism. While the function of UBE3A has been widely studied, little is known about its paralog UBE3B. Using exome and capillary sequencing, we here identify biallelic mutations of *UBE3B* in four patients from three unrelated families presenting an autosomal-recessive Ohdo-like intellectual disability disorder characterized by developmental delay, growth retardation with a small head circumference, facial dysmorphisms including blepharophimosis, ectodermal anomalies, and low cholesterol levels. *UBE3B* encodes an uncharacterized E3 ubiquitin ligase. The identified *UBE3B* variants include one frameshift and two splice-site mutations leading to aberrant splicing as well as a missense substitution affecting a highly conserved residue of the HECT domain. We show that the murine ortholog *Ube3b* is expressed in the central nervous system and in craniofacial structures, an expression pattern that is relevant to the phenotype observed in individuals with the human syndrome. Disruption of mouse *Ube3b* leads to reduced viability and recapitulates key aspects of the human disorder, such as reduced weight and brain size and a downregulation of cholesterol synthesis. We establish that the Caenorhabditis elegans ortholog of *UBE3B*, *oxi1*, functions in the ubiquitin/proteasome-system in vivo and is especially required under oxidative stress conditions. Together, our data reveal the pleiotropic effects of UBE3B deficiency and reinforce the physiological importance of ubiquitination in neuronal development and function in mammals.

106

Androgenetic alopecia: identification of four new genetic risk loci and evidence for the contribution of WNT-signaling to its etiology. *S. Heilmann[1,2], A.K. Kiefer[3], N. Kluck[1,2], D. Drichel[4], A.M. Hillmer[5], C. Herold[6], J.Y. Tung[3], N. Eriksson[3], S. Redler[1,2], R.C. Betz[1,2], R. Li[7], H. Stefansson[8], D.R. Nyholt[9], K. Song[10], S.H. Vermeulen[11,12], S. Kanoni[13], G. Dedoussis[14], N.G. Martin[9], L.A. Kiemeney[12,15,16], V. Mooser[10], K. Stefansson[8], J.B. Richards[7,17], T. Becker[4,6], F.F. Brockschmidt[1,2,18], D. Hinds[3,18], M.M. Nöthen[1,2,18].* 1) Insitute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics,Life & Brain Center, University of Bonn, Bonn, Germany; 3) 23andMe, Mountain View, CA, USA; 4) Institute of Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany; 5) Genome Technology and Biology Group, Genome Institute of Singapore, Singapore, Singapore; 6) German Center for Neurodegenerative Disease (DZNE), Bonn, Germany; 7) Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 8) deCODE genetics, Sturlugata 8 IS-101, Reykjavík, Iceland; 9) Queensland Institute of Medical Research, Brisbane, Australia; 10) Genetics Division, GlaxoSmithKline, King of Prussia, Pennsylvania, USA; 11) Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 12) Department of Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 13) Genetics of complex traits in humans, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 14) Department of Dietetics-Nutrition, Harokopio University, Athens, Greece; 15) Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 16) Comprehensive Cancer Centre of the Netherlands (IKNL), Nijmegen, The Netherlands; 17) Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 18) equally contributing senior authors.

Androgenetic alopecia (AGA, male pattern baldness) is the most common form of hair loss among humans. It is characterized by a progressive loss of hair from the scalp. Research has established that the pathogenesis of AGA is driven by androgens based on a genetic predisposition as the major precondition. During the past years, candidate gene and genome-wide association studies have identified single nucleotide polymorphisms (SNPs) at eight different genomic loci to be associated with AGA. Despite these recent breakthroughs in the understanding of the genetics of AGA, a significant fraction of the overall heritable risk still awaits identification. Furthermore, understanding of the pathophysiology of AGA is incomplete, and thus each newly associated locus may provide novel insights into the contributing biological pathways. The aim of the present study was to identify additional AGA risk loci by replicating SNPs at twelve genomic loci that show association with AGA in a recent meta-analysis (P-value < 1×10^{-5}) but fell below the threshold of genome-wide significance (P-value > 5×10^{-8}). We analyzed a total of 3,443 cases and 3,597 controls of European descent to confirm the association with AGA at these loci. The combined analysis of the replication and the meta-analysis data identified four additional genome-wide significant risk loci for AGA on chromosome (chr)2q35, chr3q25.1, chr5q33.3, and chr12p12.1. The strongest association signal was obtained for rs7349332 (P-value = 3.55×10^{-15}) on chr2q35 located intronically in *WNT10A*. *WNT10A* is a member of the family of WNT-genes that encode for small secreted signaling proteins which play important roles during development and tissue homeostasis. Interestingly, WNT-signaling has also been implicated in the regulation of hair development and hair cycling. This seems to be of particular interest, as changes in hair follicle dynamics are a key feature in AGA. Indeed, genotype-specific expression studies showed the AGA risk allele of rs7349332 to be associated with lower *WNT10A*-expression in human hair follicle. These changes in expression levels might be of functional relevance in the regulation of hair cycle dynamics. Our present study thus provides the first genetic evidence for an involvement of WNT-signaling to AGA-development. It is hoped that a deeper understanding of the role of WNT-signaling in AGA will provide the basis for the development of new therapeutic options in the future.

107

A polymorphism in human estrogen-related receptor beta (ESRRβ) is associated with early indications of hearing loss from acoustic overload in young adult musicians. *V.C. Henrich[1], S.L. Phillips[2], S.J. Richter[3], S. Teglas[4], R. Morehouse[5].* 1) Ctr Biotech, Genomics, and Health Research, Univ North Carolina,Greensboro, Greensboro, NC; 2) Department of Communication Sciences and Disorders, Univ North Carolina, Greensboro, Greensboro, NC; 3) Department of Mathematics and Statistics, Univ North Carolina, Greensboro, Greensboro, NC; 4) Music Research Institute, Univ North Carolina, Greensboro, Greensboro, NC; 5) Department of Communication Sciences and Disorders, Appalachian State Univ, Boone, NC.

Noise-induced hearing loss (NIHL) is a world-wide health problem, a growing concern among young people, and the most frequently observed disability among returning war veterans. Previous genetic association studies with adult factory workers have found specific genetic predispositions to NIHL. In this case-control study, we examined 271 SNPs in 52 candidate genes from 204 participants selected from 572 college-age musicians based on hearing status. A bilateral loss of hearing acuity at 4000 and/or 6000 Hz was hypothesized to be a phenotype arising from underlying genotypes. Loss of 15dB or more compared to hearing thresholds at lower and higher frequencies appears as a “notch” on audiograms. When subjects with no notches, a single notch, and bilateral notches were compared, a nonsynonymous SNP in the gene that encodes a nuclear receptor, the estrogen-related receptor (ESRRβ), was associated with bilateral notches. Heterozygous CT carriers of the variant (rs61742642; C→T, P386S) accounted for 26% (18/68) of all subjects showing bilateral notches (OR = 2.8, CI = 1.4-5.9, p= 0.003). Among the 204 case-control subjects, 29 (14.2%) carried the CT genotype and 18 of them had a notch in both ears. Other closely linked SNPs showed no evidence of association with bilateral notching. This SNP also does not show evidence of an association with high-frequency hearing loss (8000 Hz) suggesting that NIHL involves a different mechanistic impairment. The ESRRB gene is expressed in the stria vascularis, which is responsible for restoring ion balance in the cochlear endolymph following sound exposure. Certain familial mutations of ESRRB cause congenital deafness in homozygotes. The SNP results in a substitution of an evolutionarily conserved residue on the outer surface of the receptor, which itself has been implicated in regulation of cellular redox states. These data, therefore suggest an important role for this polymorphism of ESRRβ in susceptibility to noise-induced hearing loss.

108

Dissection of polygenic variation for human height into individual variants, specific loci and biological pathways from a GWAS meta-analysis of 250,000 individuals. *T. Esko[1], A.R. Wood[2], S. Vedantam[3,4,5], J. Yang[6], S. Gustaffsson[7], S.I. Berndt[8], J. Karjalainen[9], H.M. Kang[10], A.E. Locke[11], A. Scherag[12], D.C. Croteau-Chonka[13], F. Day[14], R. Magi[1], T. Ferreira[15], J. Randall[15], T.W. Winkler[16], T. Fall[7], Z. Kutalik[17], T. Workalemahu[18], G. Abecasis[10], M.E. Goddard[6], L. Franke[9], R.J.F. Loos[14,19], M.N Weedon[2], E. Ingelsson[7], P.M. Visscher[6], J.N. Hirschhorn[3,4,5], T.M. Frayling[2], GIANT Consortium.* 1) Estonian Genome Center, University of Tartu, Tartu, Tartumaa, Estonia; 2) Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK; 3) Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital, Boston, Massachusetts 02115, USA; 4) Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA; 5) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 6) University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland, Australia; 7) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden; 8) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892, USA; 9) Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 10) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA; 11) Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA; 12) Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Germany; 13) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA; 14) MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK; 15) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK; 16) Public Health and Gender Studies, Institute of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Regensburg, Germany; 17) Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland; 18) Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA; 19) Mount Sinai School of Medicine, New York, NY, USA.

Adult human height is a highly heritable polygenic trait. Previous genome-wide analyses have identified 180 independent loci explaining an estimated 1/8th of the heritable component (80%). Our aims were a) to increase the understanding of the role of common genetic variation in a model quantitative trait, and b) to help understand the biology of normal growth and development. Within the GIANT consortium, we performed a GWAS of ~250,000 individuals of European ancestry. We tested for the presence of multiple signals at individual loci using an approximate conditional and joint multiple SNP regression analysis. We identified 698 independent variants associated with height at p<5×10-8, which fell in 424 loci (+/-500kb from lead SNP) and altogether explained 1/4 of the inherited component in adult height. Half of the loci contained multiple signals of association. By applying a novel pathway analysis approach that uses co-expression data from 80,000 samples to predict the biological function of poorly annotated genes, we observed enrichment for novel and biologically relevant pathways in these loci. For example, for more than 10 % of the loci a gene was found in their vicinity with a predicted “regulation of ossification” function (GO:0030278, WMW P < 10-34), including newly identified genes such as PRRX1and SNAI1. Other genes and pathways newly highlighted by pathway analysis include WNT (WNT2B, WNT4, WNT7A) and FGF (FGF2, FGF18) signaling and osteoglycin. We also noted an excess of signals across the entire genome, with the median test statistic twice that expected under null (lambda = 2.0). This result is consistent with either a very deep polygenic component to height that covers most of the genome or population stratification contributing partly to the results, or a combination of the two. Encouragingly, initial results from family based analyses and mixed models that correct for distant relatedness across samples indicate that a large proportion of the discovered signals are genuine height-associated variants rather than confounded by stratification. In conclusion, data from 250,000 individuals show that adult height is highly polygenic with, typically, multiple signals of association per locus now accounting for 1\4 of heritability. Furthermore, these results suggest that increasing GWAS sample sizes can continue to uncover substantial new insights into the aetiological pathways involved in common human phenotypes.

109

Genome-Wide Association Studies (GWAS) meta-analysis for fracture risk points to loci related to hormonal and neurological pathways: the GEFOS Consortium. *L. Oei[1,2,3], H.F. Zheng[4], E.E. Ntzani[5], C.M. Nielson[6,7], U. Styrkarsdottir[8], P.M. Ridker[9], K.K. Tsilidis[5], K. Estrada[1,2,3], A. Enneman[1,3], A. Vernon-Smith[10], R.D. Jackson[11,12], S. Trompet[3,13], T. Lehtimäki[14], S. Kaptoge[15], T.B. Harris[16], J. Eriksson[17], N. Amin[2], A. Metspalu[18,19], P.C. Sham[20], E. Evangelou[5], J.P.A. Ioannidis[5,21], G. Thorleifsson[8], A.G. Uitterlinden[1,2,3], S.A. Cummings[22], T. Spector[23], D.P. Kiel[24,25], D. Chasman[9], E. Orwoll[7], J.B. Richards[4,23], F. Rivadeneira[1,2,3], GEFOS consortium.* 1) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands; 2) Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 3) Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), The Netherlands; 4) Department of Human genetics and Epidemiology and Biostatistics, McGill University, Montréal, QC, Canada; 5) Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece; 6) Public Health and Preventive Medicine, Oregon Health & Science University, Portland, OR, USA; 7) Bone and Mineral Research Unit, Department of Medicine, Oregon Health & Science University, Portland, OR, USA; 8) deCODE Genetics, Reykjavík, Iceland; 9) Brigham and Women's Hospital, Boston, MA, USA; 10) Icelandic Heart Association and University of Iceland, Kopavogur, Iceland; 11) Department of Internal Medicine, The Ohio State University, Columbus, Ohio, USA; 12) Center for Clinical and Translational Science, The Ohio State University, Columbus, Ohio, USA; 13) Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands; 14) University of Tampere and Tampere University Hospital, Tampere, Finland; 15) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 16) Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute of Aging, National Institutes of Health; 17) Center for Bone and Arthritis Research, Institute of Medicin, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 18) Estonian Genome Center, University of Tartu, Tartu, Estonia; 19) Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 20) Department of Medicine, The University of Hong Kong, Hong Kong, China; 21) Stanford Prevention Research Center, Stanford University, Stanford, California, USA; 22) Research Institute, California Pacific Medical Center, San Francisco, CA, USA; 23) Department of Twin Research and Genetic Epidemiology, King 's College London, London, UK; 24) Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts, USA; 25) Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.

Risk of osteoporotic fracture is heritable; by applying Genome-wide Complex Trait Analysis (GCTA) to the Rotterdam Study we derived a heritability estimate of 0.21 (95%CI: 0.05-0.37). This is the first large-scale fracture GWAS meta-analysis in the Genetic Factors of OSteoporosis (GEFOS) consortium. Cases were individuals (>18 years) with fractures confirmed by medical, radiological or questionnaire reports. The discovery phase comprised 21 GWAS cohorts (n=93,364, of which 16,542 were cases). Additive association was tested for imputed SNPs (HapMap CEU release 22, build 36) with logistic regression models adjusted for sex, age, height, weight. Results were meta-analyzed with inverse variance fixed-effects in METAL. GCTA joint-multiple SNP conditional analysis was used to detect secondary signals. The strongest signals associated at GWS ($P<5\times10^{-8}$) included SNPs in 7q21 (SHFM1) ($P=2.4\times10^{-10}$, odds ratio = 1.10, 95% CI: 1.07-1.13) and in 18p11 (FAM210C) ($P=1.4\ x10^{-8}$, odds ratio = 1.09, 1.06-1.12). These loci were recently identified as bone mineral density (BMD) associated loci by our parallel GEFOS effort. The large linkage disequilibrium block harbouring the 7q21 signal includes SHFM1; this region is characterized by genomic re-arrangements leading to deletion of DSS1, DLX5 and DLX6. The latter two code for members of the Wnt signalling pathway, and cause ectrodactyly when both are deleted or mutated. After conditional analysis two more signals reached GWS in 17q11 ($P=1.3\times10^{-8}$) close to SOST, and 3q2 ($P=3.7\times10^{-8}$) next to ZIC4 and ZIC1. Sclerostin, antagonizes bone formation by binding to LRP5/6 receptors and inhibiting Wnt signalling; while heterozygous deletion of ZIC1 and ZIC4 is involved in Dandy-Walker malformation, where affected individuals have motor deficits such as delayed development, hypotonia and ataxia. Suggestive SNPs ($P<5\times10^{-6}$) mapped in or near RSPO3 ($P=6.3\times10^{-7}$), LRP5 ($P=1.4\times10^{-6}$), CDC42EP3/FAM82A1 ($P=2.1\times10^{-6}$), ZFHX3 ($P=2.8\times10^{-6}$), NOVA1 ($P=4.1\times10^{-6}$), FBLN5 ($P=5.0\times10^{-6}$). Several of these loci have been found associated with bone mineral density, height, breast or prostate cancer; and mutations in some of these genes are known to cause skeletal abnormalities. These markers will undergo further in-silico and de-novo replication in >80,000 individuals (~30,000 cases). In conclusion, this large-scale GWAS meta-analysis for fracture risk pin-pointed loci involved in the regulation of bone mineral density, neurological and endocrine function.

110

Genetic landscape of the red blood cell. *J.C. Chambers[1], P. van der Harst[2], W. Zhang[1], I. Mateo Leach[2], J. Sehmi[1], N. Verweij[2], D.S. Paul[3], A. Rendon[4], U. Elling[5], H. Allayee[6], A. Radhakrishnan[3], J.S. Beckmann[7], G.V. Dedoussis[8], P. Deloukas[3], A.A. Hicks[9], S. Sanna[10], M. Uda[10], J. Penninger[5], C. Gieger[11], J.S. Kooner[1], W. Ouwehand[3], N. Soranzo[3], HaemGen RBC consortium.* 1) Imperial College London, London, London, United Kingdom; 2) University of Groningen, University Medical Center Groningen, The Netherlands; 3) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 4) Department of Haematology, University of Cambridge, Cambridge, UK; 5) Institute of Molecular Biotechnology of the Austrian Academy of Sciences, 1030 Vienna, Austria; 6) Department of Preventive Medicine, USC Keck School of Medicine, Los Angeles, CA 90033, USA; 7) Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 8) Nutrition and Dietetics, Harokopio University, Athens, Greece; 9) Center for Biomedicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy - Affiliated Institute of the University of Lübeck, Lübeck, Germany; 10) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy 09042; 11) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.

Red blood cells provide the primary mechanism for oxygen transport in the circulation. To refine our understanding of the genetic factors influencing red blood cell formation and function, we carried out a genome-wide association (N=71,861) and replication (N-63,506) study of haemoglobin and five related red blood cell phenotypes. Genome wide significance was set at P<1×10-8 allowing a Bonferroni correction for the 6 inter-related red blood cell phenotypes (Neff~5). We report 75 independent genetic loci associated with one or more red blood cell phenotypes at P<1×10-8, 43 of these loci are novel. Together these loci explain 5–10% of the phenotypic variance in each trait. To investigate underlying mechanisms, we searched for sequence variants at the red blood cell loci that might influence protein coding or gene regulation. We first identified 39 non-synonymous SNPs that are in high LD (r2>0.8) with sentinel SNPs at the red blood cell loci (~6-fold enrichment, P= 0.01). We then used formaldehyde-assisted isolation of regulatory elements followed by next-generation sequencing (FAIRE-seq) to identify nucleosome-depleted regions (NDRs) that may represent active regulatory elements. We find 60 SNPs located within one of these NDRs that are either: i. one of the 75 sentinel SNPs from the red blood cell GWAS, or ii. in high LD (r2>0.8) with a sentinel SNP. These SNPs and regions are strong candidates for functional genomic elements influencing red blood cell formation and function, and a priority set for experimental evaluation. Next we used complementary bioinformatic strategies (proximity, expression QTL, coding variant and GRAIL) to identify a core set of 121 genes as candidates for mediating the observed associations. We show that the candidates are enriched for genes known to be involved in haematological development and function (P=10-63), and are preferentially expressed in erythroblasts, red blood cell precursors (P=0.005). We find that 43 of these candidate genes have haematopoietic phenotypes in M. musculus or D melanogaster, providing strong functional support for a set of genes conserved across phyla that are important in blood cell formation or survival. Our findings provide extensive new insights into the genes and genetic variants that may influence haemoglobin levels and related red blood cell indices, and will provide the basis for a deeper knowledge of the biological mechanisms involved in haematopoiesis and red blood cell function.

111

Discovery and fine-mapping of serum protein loci through trans-ethnic meta-analysis. *A.P. Morris[1], Y. Okada[2,3], F.J.A. van Rooij[4,5], B.P. Prins[6], M.F. Feitosa[7], M. Karakas[8], J. Felix[4,5], B.Z. Alizadeh[6], L.A. Cupples[9,10], J.R.B. Perry[11], N. Franceschini[12], CHARGE Consortium Protein Working Group.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Center for Genomic Medicine, RIKEN, Yokohama, Japan; 3) Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 4) Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands; 5) Netherlands Consortium for Healthy Aging, The Netherlands Genomics Initiative, Leiden, The Netherlands; 6) University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 7) Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA; 8) Department of Internal Medicine II-Cardiology, University of Ulm Medical Center, Ulm, Germany; 9) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 10) National Heart, Lung, and Blood Institute's Framingham Heart Study, Division of Intramural Research, Bethesda, MD, USA; 11) Peninsula Medical School, University of Exeter, Exeter, UK; 12) Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA.

The serum concentration of proteins, such as albumin, are associated with a wide range of disorders including malnutrition, cancer, and cardiovascular, kidney, and inflammatory diseases. However, relatively little is known about the genetic architecture and biological mechanisms underlying these highly heritable traits.

We began by performing fixed-effects meta-analysis of genome-wide association studies (GWAS) of serum albumin and total protein concentrations in up to 53,190 individuals of European ancestry. We identified novel loci at genome-wide significance ($p<5\times10^{-8}$) for serum albumin at *HPN-SCN1B* ($p=3.3\times10^{-15}$), *SERPINF2-WDR81* ($p=6.8\times10^{-13}$), *TNFRSF11A-ZCCHC2* ($p=3.9\times10^{-9}$), and *FRMD5-WDR76* ($p=2.0\times10^{-8}$), and for total protein on chromosome 6q21.3 ($p=3.4\times10^{-9}$). We then combined the European ancestry data with up to 10,168 individuals from a GWAS of Japanese descent using trans-ethnic meta-analysis (MANTRA) to: (i) identify additional novel loci for serum protein concentrations; (ii) assess the evidence of heterogeneity in allelic effects between ancestry groups; and (iii) improve fine-mapping due to varying linkage disequilibrium structure between diverse populations.

The MANTRA analysis identified additional loci with strong evidence of association with serum protein concentrations (Bayes' factor >100,000), including *ELL2* which achieved genome-wide significance through trans-ethnic fixed-effects meta-analysis ($p=1.1\times10^{-8}$). There was little evidence of heterogeneity in allelic effects between ancestry groups at any of the identified serum protein loci, except at *ARID5B* for total protein. Here, the lead SNP (rs2675609) was strongly associated in the Japanese GWAS ($p=1.7\times10^{-6}$), but not in the European meta-analysis ($p=0.014$), and the allelic effects were in opposite directions in the two ancestry groups. We also observed substantial improvements in fine-mapping resolution after addition of the Japanese GWAS to the European ancestry meta-analysis at the 6q21.3 locus for total protein, where the interval covered by SNPs accounting for 99% of the probability of being causal was reduced from 346kb to just 37kb.

Our results highlight the advantages of trans-ethnic meta-analysis for the discovery and fine-mapping of complex trait loci, and provide initial insights into the underlying genetic architecture of serum protein concentrations.

112

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: A genome-wide association study of telomere length in a multi-ethnic cohort of 100,000 subjects. *M. Kvale[1], K. Lapham[1], T. Hoffmann[1], S. Sciortino[2], L. Walter[2], Y. Banda[1], I. Listerman[1], J. Lin[1], S. Hesselson[1], P. Kwok[1], E. Blackburn[1], C. Schaefer[2], N. Risch[1,2].* 1) Inst Human Genetics, Univ California, San Francisco, San Francisco, CA; 2) Kaiser Permanente Northern California Division of Research, Oakland, CA.

The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort contains data on telomere lengths and genotypes on over 675,000 SNPs for each of 100,000 subjects. The unprecedented number of subjects, along with data assayed in single experiments and consistent data processing, create unique opportunities for large scale genome wide association studies without the heterogeneity typically present in consortium meta-analyses. We conducted a genome-wide association study of telomere length in four major racial/ethnic groups (white, African American, Asian, Latino) comprising the GERA cohort. Data on telomere length were derived from relative qPCR assays using a novel standard curve analysis; genotypes were assayed on custom-designed Affymetrix Axiom arrays. Primary analysis was based on the white subjects (N≈70,000), with significant regions examined for replication in the other major race/ethnicity groups. We identified 63 genomic regions with genome-wide significance ($P<5\times10^{-8}$). Of these, 14 showed clear replication in at least one of the other major race/ethnicity groups. One of the identified locations was at the oligonucleotide/oligosaccharide-binding fold containing 1 (OBFC1) gene on chromosome 10 ($P<10^{-15}$); this association has been reported in some but not all previous GWA studies. Another prior candidate, the Telomerase RNA component (TERC) gene, also showed significant association in whites ($P<10^{-16}$), but only provided weak evidence in other ethnic groups. Both genes have a known role in telomere maintenance. Two SNPs showed extreme significance ($P<10^{-80}$), on chromosomes 1 and 12. These SNPs encoded amino acid substitutions, one for a G-coupled receptor, and the other for a Migration and Invasion Inhibitory Protein. Another 8 loci have P values less than 10^{-20}. The relationship between these loci and telomere maintenance with age has yet to be determined. In summary, the reliability of our association data has been demonstrated by the compelling replication of two known telomere maintenance genes. The sizable number of new highly significant locations that we identified provides an opportunity to better understand the genetic basis for telomere maintenance with age. The large amount of longitudinal clinical data available through EMRs on these subjects along with extensive environmental and behavioral data also makes it an invaluable resource to further understanding the role of telomeres in health and disease.

113

Heritability of the Variation in Aging in Two Longitudinal Family Cohort Studies: SardiNIA/Progenia Study and Framingham Heart Study. *J. Bragg-Gresham[1], S. Sanna[2], C. Sidore[1,2,3], A. Mulas[2], F. Busonero[2], A. Maschio[2], M. Urru[4], F. Reinier[4], R. Berutti[2,4], M. Marcelli[4], R. Cusano[2], M. Oppo[4], D. Pitzalis[2], M. Zoledziewska[4], A. Angius[4], C. Jones[4], A. Cao[2], M. Uda[2,3], S. Kardia[5], D. Schlessinger[6], F. Cucca[2,3], G. Abecasis[1].* 1) Biostatistics, Univ Michigan, Ann Arbor, MI; 2) Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Monserrato, 09042, Italy; 3) Università degli Studi di Sassari, Dip. Scienze Biomediche, Sassari, 07100, Italy; 4) CRS4, Laboratorio di Genomica, Parco tecnologico della Sardegna, Pula; 5) Michigan Center for Genomics & Public Health, Ann Arbor, University of Michigan, MI, USA; 6) Laboratory of Genetics, National Institute on Aging, Baltimore, MD, USA.

Normal aging is associated with diverse physiological changes in all organ systems. These changes do not occur at the same rate among individuals and some of the variation in the rate of aging between individuals is believed to be under genetic control. Starting with a set of >6,000 richly phenotyped individuals from the SardiNIA longitudinal study of aging, we first identified a set of physiological changes that were strongly associated with age. We found that a linear model including a measure of kidney function (estimated glomerular filtration rate), two measures of cardiovascular function (systolic blood pressure and intimal media thickness of the carotid artery) and waist circumference resulted in a strong predictor of chronological age (r2 = 0.81). We validated this set of predictors in the Framingham Heart Study and observed continued high ability to predict chronological age (r2 = 0.35). We used our model to assign a “physiological age” to each individual and reasoned that differences between this “physiological age” and each individual's chronological age would help quantify the rate of aging in each individual. In the SardiNIA sample, increases in “physiological age” were associated with an increase in all-cause mortality, even after adjusting for chronological age (a one standard deviation increase in the difference between physiological and chronological age corresponded to a 60% increase in mortality) and with significantly higher odds of a variety of aging related conditions (hypertension, metabolic disorder, kidney/ urologic disease, coronary artery disease, immunologic disorders, diabetes, and depression). Overall, we observed heritability of 30–40% in nuclear and extended pedigrees from the Sardinia and Framingham Heart Study samples. Heritability was highest among 44–63 year olds (peaking at h2 of ~0.60) and lowest among the youngest and the oldest individuals in each group. Our investigations result in quantitative measure of physiologic aging that is highly heritable and should be amenable for GWAS analysis. Genomewide association studies and replication are currently ongoing.

114

Over 250 novel associations with human morphological traits. *N. Eriksson, C.B. Do, J.Y. Tung, A.K. Kiefer, D.A. Hinds, J.L. Mountain, U. Francke.* 23andMe, Mountain View, CA.

External morphological features are by definition visible and are typically easy to measure. They also generally happen to be highly heritable. As such, they have played a fundamental role in the development of the field of genetics. As morphological traits have frequently been the target of natural selection, their genetics may also provide clues into our evolutionary history. Many rare diseases include dysmorphologic features among their symptoms. However, aside from height and BMI, currently little is known about the genetics of common variation in human morphology. Here we present a series of genome-wide association studies across 18 self-reported morphological traits in a total of over 55,000 people of European ancestry from the customer base of 23andMe. The phenotypes studied include hair traits (baldness, unibrow, hair curl, upper and lower back hair, widow's peak), as well as many soft tissue and skeletal traits (chin dimple, nose shape, dimples, earlobe attachment, nose-wiggling ability, the presence of a gap between the top incisors, joint hypermobility, finger and toe relative lengths, arch height, foot direction, height-normalized shoe size). Across the 18 phenotypes, we find a total of 281 genome-wide significant associations (including 53 for unibrow and 29 each for hair curl and chin dimple). Almost all of these associations are novel; we believe this is the largest set of novel associations ever described in a single report. Many of these SNPs show pleiotropic effects, e.g., a SNP near GDF5 is associated with hypermobility, arch height, relative toe length, shoe size, and foot direction; another near AUTS is associated with both back hair and baldness. Nearby genes are significantly enriched to be transcription factors (p<1e-14) and to be involved in rare disorders that cause cleft palate, ear, limb, or skull abnormalities (p<1e-7). A SNP near ZEB2 is associated with both widow's peak and chin dimple; mutations in ZEB2 cause Mowat-Wilson syndrome, which includes distinctive facial features such as a pronounced chin. Morphology-associated SNPs are also enriched within regions that have been identified as undergoing selection since the divergence from Neanderthals (18 associations in 11 regions, p = 4e-5). The abundance of these SNPs, which include the ZEB2 and GDF5 associations above, suggest that physical traits may have played a significant role in driving the natural selection processes that gave rise to modern humans.

115

Coronary artery disease loci identified in over 190,000 individuals implicate lipid metabolism and inflammation as key causal pathways; evidence for independent signals in many of the risk loci. *S. Kanoni[1], C. Willenborg[2,3], M. Farrall[4,5], T.L. Assimes[6], J.R. Thompson[7], E. Ingelsson[8], D. Saleheen[9-11], J. Erdmann[2,3], M.P. Reilly[12], R. Collins[13], S. Kathiresan[14,15], A. Hamsten[16,17], U. Thorsteinsdottir[18,19], J.S. Kooner[20], J. Danesh[10], C.N.A. Palmer[21], R. Roberts[22], H. Watkins[4,5], H. Schunkert[2,3], N.J. Samani[23], P. Deloukas[1] for the CARDIoGRAMplusC4D consortium.* 1) Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Cambridge, UK, Cambridgshire, United Kingdom; 2) Universität zu Lübeck, Medizinische Klinik II, Lübeck, Germany; 3) Deutsches Zentrum für Herz-Kreislauf-Forschung, Lübeck, Germany; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 5) Cardiovascular Medicine, University of Oxford, Oxford UK; 6) Department of Medicine, Stanford University School of Medicine; 7) Department of Health Sciences, University of Leicester, UK; 8) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 9) Center for Non-Communicable Diseases, Karachi, Pakistan; 10) Department of Public Health and Primary Care, University of Cambridge, UK; 11) Department of Medicine, University of Pennsylvania, USA; 12) Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, USA; 13) Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, UK; 14) Cardiology Division, Center for Human Genetic Research, and Cardiovascular Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, USA; 15) Broad Institute of Harvard/MIT, Cambridge, USA; 16) Atherosclerosis Res Unit, Department of Medicine, Karolinska Institutet; 17) Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; 18) deCODE genetics, Reykjavik, Iceland; 19) University of Iceland, Faculty of Medicine, Reykjavik, Iceland; 20) National Heart and Lung Institute, Imperial College London, Hammersmith Hospital, London, UK; 21) Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK; 22) Institute for Molecular Medicine FIMM, University of Helsinki and Public Health Genomics Unit, Helsinki, Finland; 23) Department of Cardiovascular Sciences, University of Leicester & National Institute for Health Research Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK.

Coronary artery disease (CAD) is the commonest cause of death in the world. We undertook a 2 stage meta-analysis and here we report the association analysis results in up to 63,746 CAD cases and 130,681 controls, identifying 16 loci reaching genome-wide significance for the first time. In total, 14 loci reached genome-wide significance, namely IL6R, APOB, VAMP5-VAMP8-GGCX, SLC22A4-SLC22A5, ZEB2-AC074093.1, GUCY1A3, KCNK5, LPL, PLG, TRIB1, ABCG5-ABCG8, FURIN-FES, FLT1 and AK097927 (in the young cases subgroup analysis only). Another 6 loci reaching p< 10-6 were further validated in 4 independent studies and 2 loci (EDNRA and HDAC9) reached genome-wide significance in a 3-stage combined meta-analysis. Our results are taking the number of such loci for CAD to 47, and a further 103 independent variants (r2< 0.2) strongly associated with CAD (false discovery rate 5%). Together with the genome-wide loci these variants explain approximately 10.6% of CAD heritability. In total 14 CAD loci (6 novel and 8 previously reported) harbor a gene for which a mouse knock out model has a relevant cardiovascular phenotype. Of the 47 genome-wide significant lead SNPs, 12 demonstrate a significant association with a lipid trait and 5 with blood pressure but none with diabetes. Network analysis with 233 candidate genes (extended set of associated loci at 10% FDR) generated 5 interaction networks comprising 85% of the putative CAD genes. The 4 most significant pathways mapping to these networks are linked to lipid metabolism and inflammation underscoring their causal role in the genetic aetiology of CAD. Furthermore, we implemented the GCTA tool to undertake fine mapping in 27 of the 47 CAD loci via an approximate conditional analysis method. Prior to the analysis, we validated the robustness of GCTA versus the exact conditional analysis in a subset of studies. We then used meta-analysis summary-level statistics from 27 Metabochip studies and a representative reference panel for linkage disequilibrium estimation. Our preliminary findings provide evidence of additional independent signals in at least half of the regions. To further investigate these findings we will perform a conditional analysis using our 2-stage meta-analysis results after performing imputation to 1,000 Genomes.

116

Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *D. Gu, X. Lu, L. Wang, S. Chen, J. Huang, H. Li, X. Yang, Y. Hao, China Atherosclerosis Genetics Consortium.* State Key Laboratory of Cardiovascular Disease, Fuwai Hospital,Chinese Academy of Medical Sciences, Beijing, Beijing, China.

Coronary artery disease (CAD) is leading cause of death and disability worldwide. Most of genome-wide association studies (GWAS) of CAD focused on samples of European origin, and the identified loci altogether explained only a small fraction of the risk for CAD. We carried out a two-stage GWAS study of CAD in a sample of ~33,000 Han Chinese. In the discovery stage, we performed a meta-analysis of the two GWAS studies comprising 1,515 CAD cases and 5,019 controls with approximately 2.2 million genotyped or imputed autosomal SNPs. In the replication stage, we genotyped top association SNPs in 3 replication samples comprising 15,460 cases and 11,472 controls. We successfully identified four new loci for CAD on chromosome 2, 4, 6, and 12 (*P* value ranging from 5.68×10^{-10} to 2.77×10^{-15}). We also replicated four loci previously identified in European populations (6p24.1 in *PHACTR1*, 6q23.2 near *TCF21*, 9p21.3 near *CDKN2A/B* and 12q24.13 near *C12orf51*). These findings provide new insights into biological pathways for the susceptibility of CAD in Han Chinese population. Further study and integration of multiethnic GWAS findings will surely promote a fuller and better understanding of the global genetic architecture of CAD.

117

Discovery of 63 Novel Loci and Refinement of Known Loci Associated with Lipid Levels. *C. Willer[1,2], G.R. Abecasis[3], M. Boehnke[3], L.A. Cupples[4], P. Deloukas[5], P.W. Franks[6], S. Gustafsson[7], E. Ingelsson[7], S. Kathiresan[8,9,10], K.L. Mohlke[11], G.M. Peloso[10], S.S. Rich[12], S. Ripatti[5,13,14], M. Sandhu[5,15], E.M. Schmidt[2], S. Sengupta[3], Global Lipids Genetics Consortium.* 1) Cardiovascular Medicine, University of Michigan, Ann Arbor, MI; 2) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 3) Dept of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 4) Biostatistics, Boston University, Boston, MA; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 6) Genetic and Molecular Epidemiology, Lund University, Malmö, Sweden; 7) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 8) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 9) Broad Institute, Cambridge, MA; 10) Center for Human Genetic Research, Massachusetts General Hosptial, Boston, MA; 11) Department of Genetics, University of North Carolina, Chapel Hill, NC; 12) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 13) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Finland; 14) Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland; 15) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

To identify novel loci associated with blood lipid levels, we performed a targeted follow-up of previous genome-wide association studies (GWAS) for fasting serum levels of LDL cholesterol, HDL cholesterol, triglyceride levels and total cholesterol, which are heritable, treatable risk factors for cardiovascular disease. Using the Illumina custom Metabochip array, we genotyped 65,345 genetic variants with preliminary evidence for association with these traits or other cardiovascular and metabolic traits. After meta-analysis with previous GWAS results, resulting in a total sample size of 180,141 to 188,578 individuals, we identified 167 genomic regions associated with lipid levels, 63 of which were novel genomic regions. Novel loci include strong biological candidate genes *DAGLB*, *APOH*, *VLDLR* and *LRPAP1*. Gene-set enrichment analyses identified genes near novel genome-wide significant index SNPs that appear in pathways with known lipid-related genes including *PPARA*, *ABCB11*, *UGT1A1* and *INSIG2*. Using dense genotyping in individuals of European, East Asian, South Asian, and African ancestry, we narrow association signals at 12 loci including *CETP*, *APOE*, *SORT1* and *PCSK9*. We find that loci associated with HDL and LDL cholesterol are often associated with other cardiovascular and metabolic traits, including coronary artery disease, type 2 diabetes, blood pressure, and body mass index. In addition to increasing the number of genetic loci associated with blood lipids, we narrow and restrict potential candidate genes in previously identified loci, and characterize associated biological pathways to guide future biological research.

118

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: Genome-wide Association Study of Plasma HDL and LDL and Treatment Response in over 100,000 Subjects. *T.J. Hoffmann[1,2], M.N. Kvale[2], Y. Banda[2], S.E. Hesselson[2], L. Walter[3], S. Sciortino[3], D.K. Ranatunga[3], D. Ludwig[3], C. Iribarren[3], R. Grant[3], P. Kwok[2,4], C. Schaefer[3], N. Risch[1,2,3].* 1) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 3) Kaiser Permanente Northern California Division of Research, Oakland, CA; 4) Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA.

Plasma lipid concentrations are an important heritable risk factor for cardiovascular diseases that are often targets for therapeutic interventions. The Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort contains extensive longitudinal lipid, prescription medication, and genotypes (over 675,000 markers) for nearly all of its 100,000 participants, composed of 81% White, 7.5% Asian, 7% Latino, and 3.5% African American race/ethnicities. Approximately 1 million HDL and LDL cholesterol measures from electronic health records were linked to genetic data for this analysis.

We conducted a primary genome-wide association (GWA) study among the white subjects using simple mean/medians for each individual. All analyses were adjusted for age, BMI and ancestry covariates. Follow up analysis was conducted in the other race/ethnicity groups. We compared the results of our association analysis of HDL measurements to previously reported GWAS hits ($p < 5\times10^{-8}$) from the large meta-analysis of Teslovich et al. (2010), and found that at a $p < 0.001$, 38 SNPs replicated, 8 did not replicate, and one was uninformative. Further, we identified 7 novel loci at genome-wide significance ($p < 5\times10–8$). To illustrate the power of this cohort, a previously known SNP in the CETP gene (rs3764261) had a $p < 10–265$ in whites, $< 10–72$ in Latinos, $< 10–41$ in Asians, and < 1017 in African Americans. Combined, these are the most significant P-values ever recorded in GWA studies. When comparing our LDL results to those of Teslovich et al. (2010), at a $p < 0.001$, 36 SNPs were replicated, and one was uninformative. Twelve novel loci for LDL were also identified at genome-wide significance. Findings generally replicated in the other race/ethnicity groups, although there was greater variability in results across groups for LDL than for HDL. We also compared the response of LDL to statin treatment (comparing lipid test results prior to versus on treatment) on nearly 28,000 individuals. Two of 4 previously reported GWAS hits were replicated, and 2 novel SNPs were identified.

This study demonstrates the remarkable power of this cohort for identifying genes underlying complex traits using high-quality longitudinal electronic health records linked to extensive GWA data.

119

Genome-wide screen with 1000 Genomes imputed data identifies 19 new lipid loci and new variants with stronger effects in previously known loci. *I. Surakka[1,2], A.-P. Sarin[1,2], R. Mägi[3], M. Horikoshi[4,5], S. Wiltshire[4,5], T. Esko[3,6], T. Ferreira[4], L. Marullo[4,7], G. Thorleifsson[8], A. Mahajan[4], S. Hägg[9], A. Isaacs[10,11], M. Beekman[12,13], J.S. Ried[14], T.W. Winkler[15], C.P. Nelson[16,17], C. Willenborg[18,19], A. Morris[4], M.I. McCarthy[4,5,20], I. Prokopenko[4,5], S. Ripatti[1,2,21] for the ENGAGE Consortium.* 1) Institute for Molecular Medicine Finland FIMM, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Estonian Genome Center, University of Tartu, Tartu, Estonia; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 5) Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, Oxford, United Kingdom; 6) The Institute of Molecular and Cellular Biology of the University of Tartu, Tartu, Estonia; 7) Department of Evolutionary Biology, Genetic Section, University of Ferrara, Ferrara, Italy; 8) deCODE Genetics, Reykjavik, Iceland; 9) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 10) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands; 11) Centre for Medical Systems Biology, Netherlands Genomics Initiative, Leiden, the Netherlands; 12) Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; 13) Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands; 14) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 15) Public Health and Gender Studies, Institute of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Regensburg, Germany; 16) Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; 17) National Institute for Health Research (NIHR) Leicester Cardiovascular Disease Biomedical Research Unit, Glenfield Hospital, Leicester, UK; 18) Universität zu Lübeck, Med. Klinik II, Lübeck, Germany; 19) Deutsches Zentrum für Herz-Kreislauf-Forschung e. V. (DZHK), Lübeck, Lübeck; 20) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom; 21) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom.

Despite the high number of already known genetic loci associated with blood lipids (high and low density lipoprotein cholesterols (LDL), total cholesterol and triglycerides), a substantial proportion of the genetic contribution to trait variances remains unexplained. In this study, we screened for novel associations with blood lipid levels by imputing the June 2011 release of 1000Genomes (2188 haplotypes) into seventeen population-based studies of European origin (up to 51,204 individuals) with genome-wide association (GWA) data. We anticipated that the denser genetic map and greater coverage of low frequency and rare variants would enhance the identification of associated and causal alleles at both known and novel loci. In a fixed-effects meta-analysis following imputation, we found 78 loci associated with at least one of the lipid measures with $P<5e-8$, 19 of these previously-unreported (>2Mb from known loci). For eight of these (near *ZDHHC18*, *TMEM48*, *PSG9*, *PRKAG3*, *ADH1A/B*, *OR4C13*, *ADAMTS3*, and *DUSP3*) the most strongly associated SNP had minor allele frequency (MAF) below 5%. For example, the lead SNP for LDL-cholesterol at the *PSG9* locus had a MAF of 1.4% and per-allele beta of 0.29 (SD). We could replicate previously-reported associations at 3 missense SNPs (*APOE*, *ANGPTL4* and *PCSK9*, with MAFs 16.5%, 3.0% and 1.9% respectively) identified from candidate-gene resequencing efforts, and 3 others (*APOB*, *GCKR* and *HNF4A*, with MAFs 23.8%, 35.9% and 4.1%) found in GWA studies. Initial fine-mapping analyses have identified at least two coding variants (*ABCA6/8* and *MOSC1*, with MAFs 1.9% and 28.0%) that are lead SNPs at those loci, and several other examples of low frequency lead SNPs. As an example, a SNP close to *LPA* had beta=0.24 for LDL, MAF=0.9% and distance of 173kb from the previously reported lead SNP. As a group, the most-associated variants at the 78 lipid loci were enriched for lower than expected associated p-values for four related traits for which data were available: body mass index, waist-to-hip ratio, fasting glucose and insulin. For example, glucose p-values for SNPs associated with total cholesterol deviated from the expected uniform distribution (P = 3e-14). Our results illustrate the utility of imputation strategies using 1000Genomes (and other reference sets) as a complement to genome-wide association and genome resequencing approach particularly for variants in the low frequency range.

120

Rare coding variation and risk for myocardial infarction: an exome chip study of ~6,000 cases and controls. *R. Do on behalf of the NHLBI Exome Sequencing Project - Early Onset Myocardial Infarction.* Center for Human Genetics, Massachusetts General Hospital, Cambridge, MA.

Myocardial infarction (MI) is the leading cause of death in the U.S. and is heritable. Through genome-wide association studies (GWAS), the role of common genetic variation in risk of MI has been thoroughly evaluated and more than 30 loci have been associated with MI. The extent to which rare variants contribute to MI risk is unknown. We tested the hypothesis that rare coding variation contributes to MI risk. We evaluated 1,989 cases with MI and 3,995 controls free of MI from two cohorts - the Ottawa Heart Study and the Women's Health Initiative - and genotyped these samples using a novel genotyping array ("exome chip") focused on genetic variation in the protein coding regions of the human genome. The "exome chip" contains ~250,000 coding variants discovered through exome sequencing in ~12,000 individuals and, in addition, includes all common GWAS variants previously associated with MI. Collectively, the array represents nearly all non-synonymous coding and splice-site variation with a >1:1000 allele frequency in the European population. Association testing for MI was performed using single-marker logistic regression correcting for ten principal components of ancestry. From this screen, the strongest association signal was seen at the common non-coding variant at chromosome 9p21 previously discovered by GWAS (55% allele frequency, OR 1.27, p=3×10-9). For rare coding variants (defined as an allele frequency < 5%), the strongest association signal was with LPA (I4399M, 2.3% allele frequency, OR 1.73, p=1×10-5). Of note, a well-studied variant in PCSK9 (R46L, 1.6% frequency, OR = 0.66) had a modest association (p=0.02). After excluding previously reported GWAS variants, there is an excess of rare coding variants with p<0.05 but which do not exceed chip-wide significance (p<2×10-7). Post-hoc power calculations show that large sample sizes will be required to confidently detect the signal from rare variants. For example, 13,500 cases and 13,500 controls would be required to detect a locus with the same frequency and effect as PCSK9 R46L at chip-wide significance. At present, we are attempting to replicate our top novel results in additional independent samples to provide more refined insights into the role of rare coding variation in risk for MI.

121

High Exome Mutational Burden in 58 African Americans with Persistent Extreme Blood Pressure. *KD.H. Nguyen[1], A.C. Morrison[2], A. Li[2], R. Gibbs[3], E. Boerwinkle[2], A. Chakravarti[1].* 1) Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 2) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.

High blood pressure (BP) is a major cardiovascular risk factor in African Americans (AA). Despite its modest heritability (35%), ~63 BP loci have been implicated by genome-wide association studies in European and African ancestry samples. We explored exome sequencing in 58 African Americans (AA) at the extremes of BP distribution across multiple visits in the Atherosclerosis Risk in Communities study (~1%tile and 99%tile residuals of the baseline age- and sex-corrected systolic BP) to demonstrate the enrichment of deleterious mutations genome-wide and to identify novel genes. We identified 67,298 high quality coding/splicing variants (≥10X coverage, ≥2 copies of the variant alleles, PHRED-like score ≥30, call rate ≥90%); each variant had a phyloP conservation score (S) and was classified as synonymous, mild missense (exon splice junction, non-NMD nonsense, nonsynonymous) or severe missense (intron splice junction, NMD nonsense). We assumed that the observed exomic mutation profile (kernel density of variants for each S value) from the 58 individuals was a mixture of two profiles, (1- β) of random subjects (107,727 variants in 61 AA individuals from the 1000G Project) and β of `true' mutations (70,393 Mendelian / disease causing mutations from the Human Genome Mutation Database), and estimated the mutational burden (β^) by least squares. This analysis estimated an overall β^= 6%, with values of 2%, 12% and 38% for the synonymous, mild missense and severe missense variants, respectively. Importantly, β^ increased with higher conservation scores to ~100%. Across each of the 3 mutation classes, β^ was slightly higher for variants observed exclusively in the top than the bottom BP group (14%/12%, 27%/25%, 60%/41%, for synonymous, mild and severe missense variants respectively). Conversely, we observed β^= 0 for variants that were present in both the top and bottom BP classes irrespective of mutation class. By considering only variants at class-specific phyloP thresholds, S≥5 and 4.5, for the mild and severe missense variants (β^=100%), we estimate that a minimum of 2,412 variants in 1,881 genes, or an average burden of ~42 mutations at ~32 genes per subject, are involved in BP. Consequently, our results showed that BP extreme subjects have distinct global mutational burden; there is a significant enrichment of deleterious coding mutations at highly conserved sites in these individuals; and the identified genes reveal new BP candidate genes.

122

Chipping a hole-in-one from the FAIRE way: Use of post-GWAS fine-mapping genotyping arrays for functional variant discovery. *A. J. P. Smith[1], F. Drenos[1], P. Howard[1], C. Giambartolomei[2], P. J. Talmud[1], V. Plagnol[2], S. E. Humphries[1].* 1) Cardiovascular Genetics, University College London, London, United Kingdom; 2) UCL Genetics Institute, University College London, London, United Kingdom.

Genome-wide association studies have identified many loci associated with cardiovascular disease and related risk factors. These studies commonly identify only proxies in strong linkage-disequilibrium with the causal genetic variant. Fine-mapping can further reduce the number of potential causal SNPs, but may not identify the functional variant. The majority of associations occur outside of coding regions and are likely to play a role in gene regulation. We developed a high-throughput method, termed FAIRE-gen, to localize potential causal variants using FAIRE-enrichment of open chromatin followed by allelic quantification using fine-mapping genotyping chips. Allelic imbalance in heterozygotes following FAIRE can identify SNPs that influence chromatin structure and therefore regulatory potential. To begin to address the localisation of functional regulatory SNPs on a large scale, FAIRE-gen was carried out with the Illumina IBC BeadChip; a 50K custom chip for SNPs focussed around 2,000 potential cardio-metabolic genes. An fQTL (FAIRE-QTL) was identified with rs7120118 at the *NR1H3* locus, the gene coding for LXR-α. This variant was also associated with HDL-C and *NR1H3* gene expression levels. Using the 200K Metabochip, containing fine-mapping GWAS loci for cardiovascular disease and related traits, this fQTL was replicated, and novel fQTLs identified for further HDL-C loci, including *LIPC*, *CETP* and *APOB*. Analysis of T2D GWAS loci with this chip also identified potentially causal variants, including SNPs at the *TCF7L2*, *CDKAL1*, *CDKN2A*loci. By identifying causal non-coding SNPs and their mechanism of action we will have a greater understanding of the pathways leading from genetic variant to disease pathology. FAIRE-gen may become a valuable tool to detect causal regulatory variants, particularly when used in conjunction with fine-mapping studies.

123

Strong association of one carbon metabolism genes with stroke and change in post-methionine load homocysteine levels in the Framingham Heart and Vitamin Intervention for Stroke Prevention Studies. *S.R. Williams[1,2], Q. Yang[3,4], F. Chen[1], X. Liu[3], K. Keen[1,5], P. Jacques[6], W.M Chen[1,5], G. Weinstein[7], F.C. Hsu[8], A. Beiser[3,7], L. Wang[9], K.F. Doheny[10], P.A. Wolf[7], M. Zilka[10], J. Selhub[6], B.B. Worrall[5,11], S. Seshadri[7], M.M. Sale[1,12,13], GARNET (The Genomics and Randomized Trials Network).* 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Cardiovascular Research Center, University of Virginia, Charlottesville, VA; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 4) The Framingham Heart Study, Framingham, MA; 5) Department of Public Health Sciences, University of Virginia, Charlottesville, VA; 6) Nutritional Epidemiology Program and Vitamin Metabolism Laboratory, Jean Mayer USDA Human Nutrition Research Center on Aging, and Freidman School of Nutritional Science and Policy, Tufts University, Boston, MA; 7) Department of Neurology, Boston University School of Medicine, Boston, MA; 8) Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC; 9) Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic College of Medicine, Rochester, MN; 10) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 11) Department of Neurology University of Virginia, Charlottesville, VA; 12) Department of Medicine, University of Virginia, Charlottesville, VA; 13) Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA.

Baseline homocysteine levels (tHCY), as well as change between pre- and post-methionine load test homocysteine levels (ΔΔPOST), have been associated with risk of incident stroke in several community-based cohorts; tHCY in the Framingham Heart Study (FHS) and Δ POST with recurrent stroke in the Vitamin Intervention for Stroke Prevention (VISP) study. We conducted genome-wide association (GWA) analyses of Δ POST in 2100 VISP participants (mean age 67 years, 36% female) and 4810 FHS offspring (mean age 59 years, 53% female) attending the sixth quadrennial examination (1994–1996), using 1000 Genomes imputation and linear regression analysis, adjusting for age, sex, principal components of ancestry (and familial relationships in FHS). VISP trial inclusion criteria included ischemic stroke within 120 days and homocysteine levels in the highest population quartile. Using a sample size weighted meta-analysis with the VISP and FHS cohorts, we were able to detect genome wide significant association for ΔPOST with SNPs within, or flanking, the glycine N-methyltransferase (*GNMT*) gene ($p=1.60\times10^{-63}$, rs9296404). GNMT is a regulatory enzyme whose function in the liver is to clear excess S-adenosyl-L-methionine through glycine methylation and the creation of S-adenosyl-L-homocysteine via the one carbon metabolism pathway. Meta-analysis also revealed 4 additional one carbon metabolism genes associated with ΔPOST, *ALDH1L1*($p= 7.3\times10^{-13}$), *CBS*($p= 3.15\times10^{-26}$), *CPS1*($p= 9.10\times10^{-13}$), and *PSPH*($p= 1.17\times10^{-16}$), indicating a strong genetic component to ΔPOST and this pathway. One of these genes, *ALDH1L1*, was also associated with incident ischemic stroke in FHS (n=168 events over 7 years of follow-up; p=0.02, rs2364368) presenting a novel association with stroke and this component of the one carbon metabolism pathway. Further, haplotype analysis in VISP indicate two major *GNMT* groups that correlate with high (19μM/L) and low (14μM/L) methionine metabolizers ($p\leq0.001$) inferring functional differences due to *GNMT* variants. Finally, functional analysis of the *GNMT* promoter within the high and low methionine metabolizers, demonstrate regulatory differences that may directly lead to the differences in ΔPOST. Taken together these data identify novel genetic variants in the one carbon metabolism pathway associated with methionine metabolism, homocysteine levels and stroke risk. Further exploration of these cohorts for possible pharmacogenomic interactions is underway.

124

Baraitser-Winter syndrome : delineation of the phenotypicspectrum in a large series of molecularly defined patients. *A. Verloes[1], O.A. Abdul-Rahman[2], J. Allanson[3], J.F. Atkin[4], M. Baraitser[22], H. Brunner[5], N. Chassaing[6], K. Devriendt[7], V. Drouin[8], A. Fry[9], J.P. Fryns[7], F. Giuliano[10], K.W. Gripp[11], D. Lacombe[12], A. Lin[13], G. Mancini[14], M. Marble[15], M. Nezarati[16], M. Nowaczyk[17], S. Osimani[1], M. Rossi[18], C. Rusu[19], Y. Sznajer[1], C. Van Ravenswaaij[20], J. Masliah[1], J.B. Rivière[21], B.W.M. van Bon[5], A. Hoischen[5], W. Dobyns[21], D. Pilz[9].* 1) Robert DEBRE University Hospital, Paris, France; 2) University of Mississippi Medical Center, Jackson, Mississippi, USA; 3) Children's Hospital of Eastern Ontario, Ottawa, Canada; 4) Ohio State University, Columbus, Ohio, USA; 5) Nijmegen Medical Centre, Nijmegen, The Netherlands; 6) Toulouse University Hospital, France; 7) University Hospital Leuven, Leuven, Belgium; 8) University Hospital, Rouen, France; 9) University Hospital of Wales, Cardiff, UK; 10) University Hospital, Nice, France; 11) duPont Hospital for Children, Wilmington, Delaware, USA; 12) University hospital, Bordeaux, France; 13) Massachusetts General Hospital, Boston, MA, USA; 14) Erasmus Medical Center, Rotterdam, The Netherlands; 15) Children's Hospital of New Orleans, New Orleans, Louisiana, USA; 16) North York General Hospital, Toronto, Ontario, Canada; 17) McMaster University, Hamilton, Ontario, Canada; 18) University Hospital, Lyon, France; 19) University Hospital, Iasi, Romania; 20) University of Groningen, Groningen, The Netherlands; 21) Seattle Children's Hospital, Seattle, Washington, USA; 22) London, UK.

Baraitser-Winter syndrome (BWS) is a dominant MCA disorder. it was shown by exome sequencing to result from heterozygous missense mutations in one of the two ubiquitous cytoplasmic actin-encoding genes ACTB and ACTG1 (Rivière, Nat Genet 2012,44:440). We present detailed phenotypic description and neuroimaging of approx. 30 patients with BWS, emphasizing the clinical variability of the syndrome, which also encompass Fryns-Aftimos syndrome. The major clinical anomalies are a striking facial dysmorphism (present in all cases) with hypertelorism, broad nose with large tip, congenital ptosis, ridged metopic suture, and highly arched eyebrows. Iris or retinal coloboma is present in many cases, as does deafness. Pachygyria with an anterio-posterior gradient is present in most cases. Progressive joint stiffness and postnatal microcephaly may develop with time. Intellectual disability and epilepsy are variable and correlate with CNS anomalies.

125

Three novel mutations in MED12 cause Ohdo syndrome Maat-Kievit-Brunner type. *A.T. Vulto-van Silfhout[1], A. Hoischen[1], B.W.M. van Bon[1], W.M. Nillesen[1], C. Gilissen[1], F. Gao[2], J.M. Spaeth[2], B.C. Hamel[1], T. Kleefstra[1], M.A.A.P. Willemsen[3], H. van Bokhoven[1], H.G. Yntema[1], B.B.A. de Vries[1], H.G. Brunner[1], T.G. Boyer[2], A.P.M. de Brouwer[1].* 1) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 2) Department of Molecular Medicine/Institute of Biotechnology, University of Texas Health Science Center, San Antonio, USA; 3) Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Ohdo syndrome [MIM 249620] is characterized by intellectual disability and typical facial features including blepharophimosis. Clinically the blepharophimosis-intellectual disability syndromes have been classified in five distinct subgroups: del(3)(pter) type, Ohdo type, Say-Barber-Biesecker-Young-Simpson (SBBYS) type, Verloes type, and Maat-Kievit-Brunner (MKB) type. Here, we performed exome sequencing in two families with two affected males with Ohdo syndrome MKB type, which is characterized by X-linked inheritance and facial coarsening at older age. Two novel missense mutations were identified in Mediator of RNA polymerase II transcription subunit 12 (MED12; NM_005120.2), p.(Arg1148His) and p.(Ser1165Pro), that segregated with the disease phenotype. Upon subsequent analysis of an additional cohort of nine single male patients with Ohdo syndrome, we detected one additional de novo missense change in MED12: p.(His1729Asn). This patient also presented with the clinical hallmarks of Ohdo syndrome MKB type. Previously, three other specific missense mutations in MED12 have been described in patients with FG syndrome and Lujan-Fryns syndrome. Our patients clearly differ from these two syndromes as they have the classical Ohdo features ptosis and blepharophimosis, while the tall forehead, macrocephaly, and broad thumbs and halluces as seen in Lujan-Fryns and FG syndrome and the typical small ears and hair whorls as seen in FG syndrome are less apparent. However, in adulthood the facial appearance of Ohdo patients type MKB becomes more coarse, and the distinction between these syndromes becomes less apparent. The occurrence of three different hemizygous missense mutations in three unrelated families with Ohdo syndrome MKB type, shows that MED12 is the causative gene for this Ohdo syndrome subtype. The identification of an X chromosomal gene in Ohdo syndrome has important implications for the recurrence risk in the families.

126

Heterogeneity of mutational mechanisms and modes of inheritance in auriculo-condylar syndrome. *C. Gordon[1], A. Vuillot[1], A. Omarjee[1], S. Park[2], J. Horst[3], R. McGowan[4], E. Tobias[5], S. García-Miñaúr[6], M. Bitner[7], L. Jakobsen[8], P. Kroisel[9], A. Stewart[10], R. Palmer[7], A. Munnich[1], M. Holder[11], A. Lin[12], A. Henderson[10], L. Basel-Vanagaite[13], E. Gerkes[14], L. Wilson[7], M. Cunningham[2], S. Marlin[15], S. Lyonnet[1], J. Amiel[1].* 1) INSERM U781, Université Paris Descartes - Sorbonne Paris Cité, Institut Imagine, Paris, France; 2) Center for Tissue and Cell Sciences, Seattle Children's Research Institute, Seattle, WA; 3) Division of Pediatric Dentistry, Department of Orofacial Sciences, UCSF, San Francisco, CA; 4) North Scotland Regional Genetics Service, Ashgrove House, Aberdeen, Scotland; 5) University of Glasgow, Yorkhill Hospital, Glasgow, Scotland; 6) Instituto de Genetica Medica y Molecular, Hospital Universitario La Paz, Madrid, Spain; 7) Great Ormond Street Hospital, London, UK; 8) Copenhagen University Hospital, Copenhagen, Denmark; 9) Medizinische Universität Graz, Austria; 10) Institute of Genetic Medicine, International Centre for Life, Newcastle upon Tyne, UK; 11) Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHRU Lille, France; 12) Massachusetts General Hospital, Boston, MA; 13) Schneider Children's Medical Center, Israel; 14) Universitair Medisch Centrum Groningen, The Netherlands; 15) Hôpital d'Enfants Armand-Trousseau, Paris, France.

Auriculo-condylar syndrome (ACS) is a rare craniofacial disorder consisting of micrognathia, hypoplasia of the mandibular condyle and a specific malformation of the ear at the junction between the lobe and helix. Missense mutations in the phospholipase C, beta 4 (PLCB4) and guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 3 (GNAI3) genes have recently been identified in ACS patients by exome sequencing. These genes are predicted to function within the G protein-coupled endothelin receptor pathway during craniofacial development. Here we report five additional cases with heterozygous PLCB4 missense mutations, and one novel heterozygous GNAI3 missense mutation. Analysis of familial cases indicates a high degree of intrafamilial phenotypic variability. For each protein, the mutated residues map to sites predicted to participate in catalysis. Our analysis of the mutational spectrum indicates that certain residues represent hotspots; of the total of ten ACS PLCB4 point mutations now described, four disrupt Arg621 and two disrupt Asp360. The narrow distribution of mutations within protein space suggests that the mutations may result in dominantly interfering proteins, rather than haploinsufficiency. Another argument for heterozygous loss-of-function not being sufficient comes from one ACS patient in whom we identified a homozygous deletion of four exons within PLCB4, predicted to result in the absence of PLCB4 protein; his consanguineous parents did not present the ACS phenotype. In addition to ACS, the homozygous deleted patient presented with central apnea, a phenotype that has not been previously reported in ACS patients, but which may have been underdiagnosed due to masking by obstructive apnea. These findings indicate that ACS can be either a dominant disorder with variable penetrance (catalytic domain missense mutations) or recessive (loss-of-function alleles). We have also examined the expression pattern of Plcb4 during craniofacial development, by in situ hybridisation in mouse embryos. A key question for future investigation is whether mutations in PLCB4 and GNAI3 (and other endothelin pathway members) may underlie related disorders of the first branchial arch, such as OAVS and Robin sequence.

127

Genetic heterogeneity of Myhre syndrome. *C. Le Goff[1], C. Michot[1], C. Mahaut[1], A. Abhyankar[2], W. le Goff[3], V. Serre[1], A. Afenjar[4], A. Brooks[5], N. Brunetti-Pieri[6], P. Campeau[7], A. Destrée[8], M. di rocco[9], D. Donnai[10], R. Hennekam[11], D. Heron[12], S. Jacquemont[13], S. Mansour[14], S. Marlin[15], R. McGowan[16], H. Murphy[10], M. Simon[17], I. Stolte- Dijkstra[18], J. Tolmie[16], R. Touraine[19], N. Van der Aa[20], T. Van Essen[21], A. Verloes[22], J.L. Casanova[2], A. Munnich[1], V. Cormier-Daire[1]*. 1) Dept Gen, Hopital Necker, INSERM U781, Université Paris DEscartes Sorbonne cité, Paris, France; 2) St. Giles Lab of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University,, New York, NY, United States; 3) INSERM, UMR_S939, Dyslipidemia, Inflammation and Atherosclerosis in Metabolic Diseases, Paris, France; 4) Service de neuropédiatrie, centre de référence anomalies du développement, Hôpital Armand Trousseau, Paris, France; 5) Department of Clinical Genetics, Sophia Children's Hospital, Erasmus MC, Rotterdam, Netherlands; 6) Department of Pediatrics, Federico II University of Naples Telethon Institute of Genetics and Medicine, Napoli, Italy; 7) Laboratory of Dr. Brendan H. Lee Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States; 8) Institut de Génétique Humaine IPG, Charleroi, Belgium,; 9) Unit of Rare Diseases, Department of Pediatrics, Gaslini Institute, Genoa, Italy; 10) Genetic Medicine, Manchester Academic Health Science Centre, University of Manchester and Central Manchester University Hospitals NHS Foundation Trust, St Mary's Hospital, Manchester, UK; 11) Department of Pediatrics H7-236 Academic Medical Center University of Amsterdam, Amsterdam, Netherlands; 12) Unité de Génetique Clinique, Hôpital La Pité Salpétrière, Paris, France,; 13) Service de génétique médicale CHUV, Lausanne, Switzerland,; 14) Clinical Genetics SW Thames Regional Genetics Service St George's, University of London, London, United Kingdom; 15) Unité de génétique Clinique, Hôpital Armand Trousseau, Paris, France,; 16) Department of Medical Genetics, FergusonSmith Centre, Yorkhill Hospital, Glasgow, United Kingdom; 17) Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, Netherlands,; 18) Department of Genetics, Clinical Genetics Section University Medical Center Groningen, Groningen, Netherlands; 19) service de génétique, CHU Hopital nord, St Etienne, France; 20) Clinical Geneticist Department of Medical Genetics University Hospital and University of Antwerp, Antwerp, Belgium; 21) department of genetics, University Medical Center Groningen, University of Groningen, Groningen, Netherlands,; 22) Département de Génétique, INSERM U676, Hôpital Robert Debré, Paris, France,.

Myhre syndrome (OMIM #139210, MS) is a developmental disorder characterized by pre and postnatal short stature, brachydactyly, facial dysmorphism (short palpebral fissures, maxillary hypoplasia, mandibular prognathism, short philtrum), muscular hypertrophy, joint limitation and deafness. Other features include intellectual disability behavioral disturbance, cardiac defects, and cryptorchidism. Skeletal manifestations include thickened calvarium, cone-shaped epiphyses, shortened tubular bones, hypoplastic iliac wings, broad ribs and large vertebrae with short and large pedicles. All reported MS cases are sporadic supporting de novo dominant mutations. Using exome sequencing in 2 MS cases, we selected mothers against DPP homolog 4 (SMAD4) as a candidate gene based on its pivotal role in BMP and TGFβ signalling. SMAD4 mutations were subsequently found in additional 9 affected individuals. We identified only 3 distinct heterozygous missense SMAD4 mutations all affecting Isoleucine 500 which in located in the Mad Homology 2 (MH2) domain, near the monoubiquitinated site Lys519. Following this initial study, we collected the samples of 16 additional MS families. We identified SMAD4 mutations in 10/16 patients. The absence of mutations in 6 patients prompted us to compare the group 1 with SMAD4 mutations and the group 2 without mutations. All 27 patients present with major signs of MS, namely facial dysmorphism, pseudomuscular hypertrophy, thickened skin, joint limitations and brachydactyly. Patients mostly have mild to moderate mental retardation and deafness. A distinction might be observed concerning the visceral malformations, patients in group 2 have no congenital heart malformations and intestinal and urinary tract malformations were only reported in group1. 17/21 patients in group 1 have pre- and post-natal growth deficiency (-2 to -5 DS) whereas only 1/6 in group 2. We further demonstrated a defect in SMAD4 ubiquitination in patient fibroblasts and increased level of SMAD4 suggesting a stabilization of SMAD4 protein in Myhre syndrome. We finally observed a decreased expression of downstream TGFβ target genes supporting impaired TGFβ driven transcriptional control in MS. We conclude that MS is a genetically heterogeneous condition. Ongoing studies will hopefully lead to the identification of another disease gene presumably also involved in TGFβ pathway.

128

Seven novel families with ADCL favor clinical and molecular homogeneity. *C. Bodemer[1], B. Callewaert[2], M. Kempers[3], E. Bourrat[4], M. Renard[2], O. Vanakker[2], F. Malfait[2], J. De Backer[2], PJ. Coucke[2], S. Hadj-Rabia[1], A. De Paepe[2]*. 1) Service de dermatologie - MAGEC, Hôpital Necker - Enfants Malades, Paris, France; 2) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 3) University Medical Center St. Radboud, Nijmegen, the Netherlands; 4) Service de dermatologie, Hopital Saint-Louis, Paris, France.

Elastin gene aberrations have been associated with a variety of phenotypes. Grossly, loss-of-function mutations result in arterial stenoses, while mutant protein expression induces cutis laxa. Autosomal dominant cutis laxa (ADCL)is a rare disorder that further presents with typical facial characteristics, inguinal hernias, aortic root dilatation and pulmonary emphysema. In most patients, frameshift mutations are found in the 3' region of the elastin gene (exons 30–34) and result in a C-terminally extended protein, but exceptions have been reported. We describe 21 patients from 7 families with ADCL, the largest cohort thus far reported, and found C-terminal frameshift mutations in the elastin gene in all probands. Our data favor homogeneity both on the clinical and molecular level. All mutations reside in exons 30, 32, 33 and 34. One mutational hotspot in exon 32 (c.2262delA) is documented. We confirm the previously reported findings comprising generalized skin involvement (97%), inguinal hernia (51%), aortic root dilatation (55%) and emphysema (37%), necessitating regular cardiovascular and pulmonary evaluations. We further evidence wide intra- and interfamilial variability and amelioration of the skin phenotype in ageing patients. We conclude that ADCL is a variable, but recognizable phenotype that is caused by comparable mutations in the elastin gene. Both first and both last authors contributed equally.

129

Comprehensive clinical and molecular analysis of 12 families with type I recessive cutis laxa. *B. Callewaert[1], C-T. Su[2], T. Van Damme[1], P. Vlummens[1], F. Malfait[1], O. Vanakker[1], B. Schulz[2], M. Mac Neal[2], EC. Davis[3], JGH. Lee[3], A. Salhi[4], S. Unger[5], K. Heimdal[6], S. De Almeida[7], U. Kornak[8], H. Gaspard[9], J-L. Bresson[10], K. Prescott[11], ME. Gosendi[12], S. Mansour[13], GE. Pierard[14], S. Madan-Khetarpal[15], FC. Sciurba[16], S. Symoens[1], PJ. Coucke[1], L. Van Maldergem[17], Z. Urban[2], A. De Paepe[1]*. 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 3) Department of Anatomy and Cell Biology, McGill University, Montreal, Canada; 4) Department of Dermatology, Faculty of Medicine, Alger, Algeria; 5) Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 6) Department of Medical Genetics, Oslo University Hospital, Olso, Norway; 7) Department of Medical Genetics, Hospitals Civis De Lisboa, Lisboa, Portugal; 8) Charité, Campus Virchow-Klinikum, Berlin, Germany; 9) Institute of Human Genetics, Heidelberg University, Heidelberg, Germany; 10) Service de Genetique, CHU Saint-Jacques, Besançon, France; 11) Department of Clinical Genetics, Chapel Allerton Hospital, Leeds, UK; 12) Department of Pediatrics, General Hospital of Llerenda, Badajoz, Spain; 13) Department of Clinical genetics, St George's University of London, London, UK; 14) Department of Dermatopathology University Hospital of Liège, Liège, Belgium; 15) Division of Medical Genetics, Department of Pediatrics, Children's Hospital of Pittsburgh, PA, USA; 16) Division of Pulmonary and Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; 17) Centre de Génétique Humaine, Université de Franche-Comté, Besançon, France.

Autosomal recessive cutis laxa type I (ARCL type I) is characterized by generalized cutis laxa with pulmonary emphysema and/or vascular complications. Rarely, mutations can be identified in FBLN4 or FBLN5. Recently, LTBP4 mutations have been implicated in a similar phenotype. Studying FBLN4, FBLN5 and LTBP4 in 12 families with ARCL type I, we found bi-allelic FBLN5 mutations in 2 probands, whereas 9 probands harbored biallelic mutations in LTBP4. In one patient, no mutations were found. FBLN5 and LTBP4 mutations cause a very similar phenotype associated with severe pulmonary emphysema, in the absence of vascular tortuosity or aneurysms. Gastro-intestinal and genitourinary tract involvement seems to be more severe in patients with LTBP4 mutations. Functional studies showed that most premature termination mutations in LTBP4 result in severely reduced mRNA and protein levels. This correlated with increased transforming growth factor beta (TGFβ) signaling. However, one mutation, c.4127dupC, escaped nonsense-mediated decay. The corresponding mutant protein (p.R1377fsX27) caused altered binding to fibrillin-1 and loss of binding to fibronectin, leading to an abnormal morphology of microfibrils in fibroblast cultures, while retaining normal TGFβ signaling. Therefore, LTBP4 mutations may cause disease through both loss of function and gain of function mechanisms.

130

M694V mutation in Armenian-Americans: a ten-year retrospective study of *MEFV* mutations testing for Familial Mediterranean Fever at UCLA. *F.S. Ong[1,2], H. Vakil[3], Y. Xue[2], K.H. Shah[1], J.Z. Kuo[1,2], K.E. Bernstein[1,4], D.L. Rimoin[2,5,6,7], J.I. Rotter[1,2,5,6,7], J.L. Deignan[3], K. Das[3], W.W. Grody[3,7]*. 1) Dept. of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048; 2) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048; 3) Dept. of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095; 4) Dept. of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048; 5) Dept. of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048; 6) Dept. of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095; 7) Dept of Pediatrics and Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095.

Familial Mediterranean fever (FMF), inherited in an autosomal recessive manner, is a systemic auto-inflammatory disorder characterized by recurrent attacks of fever with peritonitis, pleuritis, pericarditis, arthritis, and erysipelas-like skin lesions. The marenostrin-encoding fever gene (*MEFV*), located on chromosome 16p13.3, is the only gene in which mutations are currently known to cause FMF. We conducted a retrospective case series review of all the test results for all *MEFV* gene mutations testing completed at UCLA Clinical Molecular Diagnostic Laboratory between February 2002 and February 2012 and reviewed the charts of patients that tested positive. We found that the M694V mutation is linked to an adverse FMF clinical outcome in the Armenian population, manifested by earlier onset of disease, severity of disease, and renal amyloidosis sequela. A total of 476 cases for suspected FMF cases were tested. There were 69 double positive mutation cases (14.5%) and 83 single mutation cases (17.4%). Of the 69 cases of double positive mutations, 34 cases were of Armenian ethnicity (49.3%) and 35 cases of non-Armenian descent (50.7%). Of the 34 cases of Armenian ethnicity, 12 (35.3%) were M694V homozygotes and 17 (50%) of the remaining 22 were M694V compound heterozygotes. In contrast, out of the 35 cases of non-Armenian ethnicity, there was only one (2.9%) M694V homozygote, while 17 (48.6%) of the remaining 34 cases were M694V compound heterozygotes. The difference in the M694V homozygous mutation is highly significant between the two groups (p=0.0006 by Fisher's exact test). Of the 69 cases of double positive mutations, FMF Type I or Type II (amyloidosis) data was gathered for 30 cases (43.5%). There were only three cases of FMF Type II (10%). All three were Armenian — 2 M694V homozygotes, 1 M694V/M680I (G/C) compound heterozygote. Of the 27 FMF Type I, M694V homozygotes accounted for 6 cases (22.2%), all were Armenian. For early onset disease (defined as <18 years of age), data was available for 41 of the 69 double positive mutations cases (59.4%). Of these 41 cases, there were 31 cases of early onset disease (75.6%), 17 of whom were Armenians (54.8%). M694V homozygotes accounted for seven of the 31 cases (22.6%) where data was available, and all seven were Armenians. Finally, among the double positive mutations cases, there were four cases of documented renal amyloidosis. Three of these cases were M694V homozygotes and all three were of Armenian descent.

131

Clinical features of individuals with Floating-Harbor syndrome due to mutations in *SRCAP*. *S.M. Nikkel[1,2], A. Dauber[3,4], R.L. Hood[2], M. Feingold[3,4], M. Connolly[3,4], M.J.M. Nowaczyk[5], S.M. White[6], A. Afenjar[7], F. Brancati[8], I. Cordeiro[9], A. Destrée[10], F. Forzano[11], E.M. Honey[12], D. Héron[13], C.M. Jacob[13], S.G. Kant[14], U. Kini[15], E. Kirk[16], E. Lemos Silveira-Lucas[17], L. Silveira Lucas[18], L. Audi Delaney[18], B. Santos da Cunha[18], V. Mericq[19], K. Pope[6], S. Price[20], J.M. Wit[14], D.E. Bulman[1,2], K.M. Boycott[1,2] FORGE Canada Consortium.* 1) Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 2) University of Ottawa; 3) Children's Hospital Boston; 4) The Manton Center for Orphan Disease Research at Children's Hospital Boston; 5) McMaster University Medical Centre, Hamilton; 6) Royal Children's Hospital, Melbourne; 7) Hôpital Armand Trousseau, Paris; 8) Tor Vergata University Hospital, Rome; 9) Hospital de Santa Maria, Lisboa; 10) Institut de Pathologie et de Génétique, Belgium; 11) Medical Genetics Unit, Galliera Hospital, Genova; 12) University of Pretoria, Pretoria; 13) Groupe Hospitalier Pitié-Salpêtrière, Paris; 14) Leiden University Medical Center; 15) Oxford University Hospitals NHS Trust; 16) Sydney Children's Hospital; 17) Ambulatório de Genética Médica da Prefeitura de Porto Alegre, Brazil; 18) Faculdade de Medicina da Pontifícia Universidade Católica do Rio Grande do Sul, Brazil; 19) University of Chile, Santiago; 20) Northampton General Hospital NHS Trust.

Floating-Harbor syndrome (FHS) is a rare autosomal dominant disorder characterized by short stature, delayed bone age, expressive language delay and a characteristic facial appearance. Recently, mutations in the *SRCAP* gene, which clustered in the 34th exon, were found to be causal for FHS in 13 individuals. Expanding on this study we have performed a comprehensive genotype-phenotype correlation that now includes an additional 24 mutation-positive individuals. In our cohort of 37 patients with *SRCAP*-FHS we observed an almost equal number of affected males and females ranging in age from two to 44 years of age at the time of data collection. The p.Arg2444* and p.Arg2435* mutations were seen in 50% and 25% of the affected individuals, respectively. Mutations were identified in two mother-daughter pairs that define the boundaries of the critical region in exon 34 of *SRCAP*. Twenty individuals without *SRCAP* mutations in the critical region were reviewed prior to sequencing for likelihood of a diagnosis of FHS. Three individuals, who most closely resembled the FHS phenotype, had complete gene sequencing, but no mutations were identified. Clinical review of the 37 individuals with *SRCAP* mutations highlighted several recurrent findings including gastrointestinal issues (colonic stricture, constipation, and/or vomiting/reflux in 13 - only one individual with Celiac disease), ophthalmic findings (hyperopia, strabismus, nystagmus in 10), and ear issues (recurrent otitis, conductive hearing loss, cochlear abnormalities in 11). There were a number of clinical features, although not previously thought to be common in FHS, that appeared to be over-represented including genitourinary anomalies (renal agenesis, hydronephrosis, cryptorchidism, hypospadias in 11), clavicular anomalies (pseudarthroses or hypoplasia in 4), hip dysplasia (3), 11 pairs of ribs (2), seizure disorders (3) and cardiac malformations (5). Two individuals had hypothyroidism. Autism was not a feature of any mutation-positive individual. Comprehensive review of the clinical features of 37 individuals with mutations in *SRCAP* will be presented along with core clinical diagnostic criteria and suggestions for management of long-term complications.

132

A prospective natural history study of *DICER1*-related familial pleuropulmonary blastoma (PPB) syndrome shows incomplete penetrance, pleiotropy and variable expressivity. *D.R. Stewart[1], L. Doros[2], G. Glenn[1], A. Bauer[3], G. Williams[4], A. Carr[5], J. Ivanovich[6], R. Kase[5], L. Harney[5], K.A. Schultz[4], C.P. Kratz[7], L.P. Dehner[8], D.A. Hill[9], Y. Messinger[4].* 1) Clincal Genetics Branch, National Cancer Institute, Rockville, MD; 2) Division of Pediatric Hematology/Oncology, Children's National Medical Center, Washington DC; 3) Uniformed Services University of the Health Sciences, Bethesda, MD; 4) International PPB Registry, Children's Hospitals and Clinics of Minnesota, Minneapolis, MN; 5) Westat, Rockville, MD; 6) Hereditary Cancer Core, Washington University School of Medicine, St. Louis, MO; 7) Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany; 8) Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO; 9) Division of Pathology, Children's National Medical Center, Washington DC.

Background. *DICER1*-related familial PPB is an autosomal dominant tumor-predisposition syndrome that arises secondary to mutations in *DICER1*, a gene critical in microRNA biogenesis. Carriers of *DICER1* mutations (&ldquôfecteds”) are at an increased but as of yet unquantified risk to develop PPB, cystic nephroma, ovarian stromal cell tumors, as well as thyroid abnormalities and other tumors and dysplasia. In a prospective natural history study of the syndrome at the NIH, we comprehensively evaluated 5 families (total of 21 participants) with at least two members with a *DICER1* mutation. *Results*. All families had at least one affected parent and 1–3 affected children. Each family harbored a unique truncating mutation in *DICER1*: 3 families with nonsense mutations (in exons 10, 11 and 17) and 2 families with frameshift mutations (in exons 22 and 24). There were 15 individuals with *DICER1* mutations, and 6 individuals with proven or presumed wildtype *DICER1*. Of the 15 *DICER1* mutation carriers, 9 (60%) had known *DICER1*-related lesions in at least 1 organ system prior to our evaluation. The 6 (40%) asymptomatic *DICER1* mutation carriers ranged in age from 10 months to 49 years; 2/4 had multiple cysts by chest CT (suspected regressed PPB), 2/6 had unrecognized nodules and/or goiter on thyroid ultrasound (US), and 3/6 had kidney cysts on renal US (possibly cystic nephroma). In *DICER1* mutation carriers, 8/11 (73%) had cysts, suspected PPB or evidence of PPB resection on chest CT imaging, 7/15 (47%) had thyroid abnormalities on thyroid US and 3/15 (20%) had renal abnormalities on renal US. No testicular masses were observed on testicular US. 1/8 (12.5%) had a history of bilateral ovarian Sertoli-Leydig cell tumor, and the others had normal pelvic US. No significant dental abnormalities, eye or nasal tumors were found. No biochemical, hormonal, carcinoembryonic antigen (CEA) or alpha-fetoprotein (AFP) abnormalities were found. Phenotype in a parent did not predict phenotype in a child. One child (an ex-23-week premature infant) with a *DICER1* mutation had a history of non-verbal autism and a patent ductus arteriosus. *Conclusions*. *DICER1*-related familial PPB is incompletely penetrant, pleiotropic and features variable expressivity. Age-dependent penetrance for disease features and evidence-based screening recommendations need to be determined.

133

Clinical implementation of a cancer care model based on comprehensive molecular profiling of tumor-normal pairs. *J.C. Taylor[1], K. Kaur[1], S. Henderson[2], E. Domingo[3], A. Cutts[2], J. Woods[4], C. Motley[5], B. Dougherty[6], M. Middleton[7], B. Hassan[8], Y. Wang[4], E. Beasley[4], M. Naley[4], I. Tomlinson[1, 3], A. Schuh[2], TSB LifeTech Consortium.* 1) Oxford Biomedical Res Ctr, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 2) Department of Haematology, Oxford Radcliffe Hospitals NHS Trust, Oxford; 3) Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, Oxford; 4) Life Technologies, Foster City, California; 5) Johnson & Johnson, New Brunswick, New Jersey; 6) AstraZeneca, London, UK; 7) Department of Medical Oncology, Churchill Hospital, Oxford; 8) Weatherall Institute of Molecular Medicine, Oxford.

Molecular profiling of cancer samples within the clinical laboratory can have strong predictive value for patient prognosis, the development of drug resistance and can inform the selection of appropriate therapeutic options. In recent years there has been a considerable increase in the number of genes with a proven value for determining cancer treatment course however routine molecular profiling of tumors for somatic mutations is currently confined to a very limited selection of genes. Comprehensive and routine screening of multiple genes in parallel is likely to increase the impact of molecular profiling for clinical use by identifying and basing treatment options on those mutations that may influence response to therapy, tumor progression and the development of drug resistance, for each individual patient. As part of a Technology Strategy Board funded large scale academia-industry collaboration in partnership with Life-Technologies, Astra Zeneca and Johnson & Johnson, we are undertaking a prospective clinical study to evaluate the efficacy of implementing an innovative cancer care model based on a massively parallel sequencing test. This test will provide a comprehensive molecular profile of the tumor in a panel of 150 genes focused on drugs that have therapeutic or pharmacogenetic value and have been approved or are in clinical trials. Concomitant software development will convert sequence data to annotated mutations and provide clinical decision support. As a precursor to the clinical implementation of this test, we have evaluated the workflow, bioinformatics pipeline and accuracy of Ion AmpliSeq(TM) Cancer HotSpot Panel targeted enrichment and Ion Torrent PGM(TM) sequencing for routine use in a clinical molecular diagnostics environment as a method for introducing a molecular profiling strategy for routine clinical diagnostic use. We screened a cohort of 40 formalin fixed paraffin embedded samples with 54 known clinically relevant mutations. We identified 95% of the positive control mutations and developed an early stage pipeline for moving from sequence data to clinical relevance. 33 mutations of clinical interest were identified in addition to the positive controls; a subset of these mutations is being investigated further in the research laboratory. The assay has been translated to the Oxford University Hospital NHS Trust molecular diagnostics laboratory, where it is being implemented alongside existing molecular diagnostic testing methods.

134

Clinical Implementation of Single Nucleotide Polymorphism (SNP) Microarrays in Pediatric Cancer and Non-Malignant Hematologic Disorders. *X. Lu, Y. Zhao, S. Gurusiddappa, C. Lau, J. Shohet, P. Rao, K. Rabin, S. Plon.* Department of Pediatrics, Texas Children's Cancer/Hematology Ctr, Baylor College of Medicine, Houston, 77030.

Array-based technology has been showing great impact in clinical cancer Cytogenetics. This study was to test the feasibility of single nucleotide polymorphism (SNP) microarrays in the clinical diagnosis of pediatric cancers. A total of 80 cases including 40 acute leukemias (27 B-cell precursor Acute Lymphoblastic Leukemia/BCP-ALL, 6 T-cell lymphoblastic leukemia, 6 acute myeloid leukemia and one mix-phenotypic acute leukemia), 17 solid tumors, 5 Myelodysplastic disorders, 6 non-malignant hematologic disorders, 5 lymphomas, one chronic myeloid leukemia cell line, as well as 6 normal remission marrows were tested using Illumina Cyto 12 and/or 1M quad SNP arrays. Overall the SNP array data showed good concordance with cytogenetic/FISH data. SNP arrays revealed most cytogenetically unidentifiable abnormalities including copy neutral loss of heterozygosity (CN-LOH). In addition, SNP arrays provided prognostic information in 10 acute leukemia cases previously reported with normal cytogenetics and/or FISH and detected deletion of cancer or cancer related genes (STIL1, EVI1, IKZF1, PAX5, PTEN, MLL, BTG, ERG and VPREB1 etc.) at the exon level. Importantly, in one case with thrombocytopenia and a known constitutional RUNX1 gene deletion, SNP array revealed a low level (~20%) of mosaic CN-LOH for chromosome 21; and in two BCP-ALL cases reported as hyperdiploid by chromosome /FISH analyses, SNP array detected CN-LOH in all disomic chromosomes, indicating that the hyperdiploidy in fact resulted from doubling of a hypodiploid clone. Finally, SNP array detected three novel amplifications at 2p25.2 in two Neuroblastoma cases originally reported with MYCN amplification and, amplicons at 2p25.1~p24.3 encompassing 16 genes in a high-grade glioma originally reported with double minute chromosomes. Additional replication and dilution studies were also conducted and showed consistent reproducibility and established the sensitivity of the SNP array. In conclusion, SNP arrays not only could confirm most karyotypic/FISH data but could also detect additional genomic aberrations, some with significant clinical implications. This pilot study shows that SNP array has great potential as a diagnostic tool in pediatric cancer Cytogenetics and can be integrated into routine clinical cancer cytogenetic testing.

135

A Prospective Clinical Trial to Evaluate DNA Sequencing as a Diagnostic Tool to Guide Cancer Therapy. *A.M.K Brown[1], P.L. Bedard[2,3], B. Tran[2,3], J. Dancey[1,7], E. Winquist[4], S.J. Hotte[5], G. Goss[6], S. Welch[4], T. Zhang[2,3], L. Stein[1], V. Ferretti[1], S. Watt[1], W. Jiao[1], K. Ng[1], P. Shaw[1], B.G. Neel[2,3], T.J. Hudson[1], J.D. McPherson[1,3], S. Kamel-Reid[2,3], L. Siu[2,3].* 1) Ontario Institute for Cancer Research,Toronto Ontario, Canada; 2) Princess Margaret Hospital-University Health Network,Toronto Ontario, Canada; 3) University of Toronto,Toronto Ontario, Canada; 4) London Health Sciences Centre, University of Western Ontario, London Ontario, Canada; 5) Juravinski Cancer Centre, Hamilton Health Sciences Centre, Hamilton Ontario, Canada; 6) Ottawa Hospital Cancer Centre, Ottawa Hospital Research Institute, Ottawa Ontario, Canada; 7) NCIC Clinical Trials Group, Kingston, Ontario, Canada.

BACKGROUND: We are conducting a multicenter clinical trial to evaluate the feasibility of including next-generation sequencing in routine clinical care. The goals of the study are to determine patient acceptance of research biopsies for genomic sequencing, optimal methods and procedures for sample collection, DNA extraction for successful analysis, review and reporting of mutations back to clinicians and patients with three weeks. METHODS: Patients with metastatic solid tumors potentially eligible for early clinical trials are recruited from five cancer centers in Ontario. After informed consent, all patients undergo a fresh biopsy of their tumor, provide a blood sample and permission to retrieve their archived tumor specimens. DNA derived from FFPE and blood are analyzed using a targeted gene sequencing approache in parallel with genotyping using the Sequenom MassArray. All mutations detected are validated by Sanger Sequencing or other methods in a CAP/CLIA certified laboratory. Results are reviewed by an expert panel to determine whether results are actionable and reportable. RESULTS: In the first year of the trial, we have recruited >90 patients who are potential candidates for early phase clinical trials of targeted agents. The median number of previous treatments the patients have received is 3. Median time with metastatic disease is 17 months. Over 90% of patients approached have consented to take part in the study. Over 30% of patients have had somatic mutations determined to be actionable by an expert panel of clinicians and scientists. In 14%, the sequencing of the tumor has impacted treatment decisions by matching a targeted therapy to the genetic profile of the tumor with treatment benefit observed in 4 cases so far (1 PR in ovarian cancer, 1 SD in breast cancer, 2 clinical benefits in thyroid and unknown primary squamous cell cancers). In four patients, the added benefit of sequencing the entire exon was demonstrated in comparison to genotyping by identifying novel mutations not present on the genotyping panels. The time from initial consent to the delivery of a clinical report has been 21 days or fewer in 62% of patients. Bioinformatics tools have been developed to assist with logistics of sample handling, validation of the pipeline and reporting mechanisms and are being optimized for eventual routine inclusion into the clinical environment.

136

Whole genome sequencing of a highly aggressive melanoma identified BRAF L597 mutants associated with sensitivity to MEK inhibitors. *Z. Zhao[1,2,3], K.B. Dahlman[1,2], J. Xia[3], H. Hutchinson[2], C. Ng[4], D. Hucks[1], P. Jia[3], M. Atefi[4], Z. Su[1], S. Branch[4], P. Lyle[5], D.J. Hicks[1], V. Bozon[6], J.A. Glaspy[4,7], J.L. Netterville[1,8], C.L. Vnencak-Jones[5,9], J. Sosman[1,10], A. Ribas[4,7], W. Pao[1,10].* 1) Vanderbilt-Ingram Cancer Center, Vanderbilt Univ, Nashville, TN; 2) Dept Cancer Biology, Vanderbilt Univ, Nashville, TN; 3) Dept Biomedical Informatics, Vanderbilt Univ, Nashville, TN; 4) Dept Medicine, Division of Hematology-Oncology, UCLA, Los Angeles, CA; 5) Depts of Pathology, Microbiology, and Immunology, Vanderbilt Univ, Nashville, TN; 6) Millenium Pharmaceuticals, Inc., Cambridge, MA; 7) Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA; 8) Dept Otolaryngology, Vanderbilt Univ, Nashville, TN; 9) Dept Pediatrics, Vanderbilt Univ, Nashville, TN; 10) Dept Medicine/Division of Hematology-Oncology, Vanderbilt Univ, Nashville, TN.

Recurrent driver mutations in BRAF (V600), NRAS (G12/13, Q61), KIT (W557, V559, L576, K642, D816), GNAQ (Q209) and GNA11 (Q209) have previously been identified in malignant melanomas, but over one-third of cases have no known potentially actionable targets. To identify novel drivers, we performed whole-genome sequencing (WGS) of a tumor-normal pair from a patient with an aggressive, chemotherapy-naïve metastatic melanoma which was pan-negative for the mutations listed above. Our data analysis pipeline revealed the landscape of somatic single nucleotide variants (339,057 novel SNVs), insertions and deletions (46,145 novel indels), copy number variations (269 novel CNVs), and structural variants (75 novel SVs) in this tumor genome. A subset of the nonsynonymous SNVs was validated by direct sequencing (127 SNVs, 93% validation rate) resulting in the identification of a somatic BRAF L597R mutation. Analysis of this BRAF mutation in 49 tumors negative for BRAF V600 mutations as well as driver mutations in KIT, NRAS, GNAQ, and GNA11 showed that 2 (4%) harbored L597 mutations and another 2 involved BRAF D594 and K601 mutations. In vitro signaling induced by the L597R mutant was suppressed by MEK inhibition. Collectively, these data demonstrate the utility of WGS in identifying actionable mutations in tumors and potential therapeutic implications of BRAF L597 mutations in melanoma.

137

Identification of novel mechanisms of drug resistance in BRCA1-deficient cancer by exome and RNA sequencing. *K.K. Dhillon[1], T. Taniguchi[1,2].* 1) Human Biology and Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Howard Hughes Medical Institute.

Acquired resistance to cisplatin therapy is an obstacle for effective treatment of ovarian and breast cancers. BRCA1 or BRCA2 loss is commonly observed in these cancers and cisplatin is initially effective for the treatment of BRCA-deficient cancers. We have shown that re-expression of BRCA1/2 due to secondary BRCA1/2 mutations in recurrent ovarian cancers is associated with cisplatin resistance. However, restoration of BRCA1/2 does not account for all occurrences of cisplatin resistance. Therefore, we hypothesize that restoration of DNA repair even in the absence of functional BRCA1/2 may lead to cisplatin resistance in cancer cells. To test this hypothesis, we developed an in vitro model of cisplatin resistance using the BRCA1-mutated breast cancer cell line, HCC1937. We generated cisplatin-resistant clones by culturing cells in cisplatin. Surprisingly, none of the resistant clones showed BRCA1 re-expression or secondary BRCA1 mutations. However, consistent with our hypothesis, a subset of clones restored DNA damage-induced foci formation of CtIP, RAD51 and FANCD2, which are required for DNA repair and normally require functional BRCA1. To identify mechanisms of cisplatin resistance, we performed exome and RNA-sequencing of parental HCC1937 cells and four cisplatin-resistant clones. Data from cisplatin-sensitive HCC1937 parental cells were used as a baseline to identify genetic and expression variations unique to cisplatin-resistant clones. Of particular interest, we found that FANCI, a component of the Fanconi anemia-BRCA pathway, was up-regulated in the three cisplatin-resistant clones that showed restoration of DNA repair foci. Importantly, depletion of FANCI in the cisplatin-resistant clones rendered these cells sensitive to cisplatin again and significantly reduced CtIP and FANCD2 foci formation. Overexpression of FANCI in cisplatin-sensitive parental HCC1937 cells resulted in increased CtIP and FANCD2 foci formation although it did not lead to increased resistance to cisplatin. This suggests that FANCI overexpression in resistant clones is a mechanism of cisplatin resistance in conjunction with additional, yet to be identified, changes. In summary, our studies showed that overexpression of FANCI with additional changes, may lead to restoration of DNA repair and cisplatin resistance in BRCA1-deficient cancer cells.

138

BRCA1 and BRCA2 mutational spectrum in a normal population - implications for clinical diagnostics and incidental findings. *E. Ruark, K.W. Lau, A. Renwick, E. Ramsay, F. Zhang, S. Seal, N. Rahman.* Institute of Cancer Research, Sutton, United Kingdom.

As sequencing costs fall ever lower and large-scale gene, exome and genome sequencing in patients and healthy individuals comes ever closer, the correct interpretation of the clinical implications of sequence variation becomes an ever increasing and challenging imperative. The BRCA1 and BRCA2 genes are often promoted as exemplars for the value of population sequencing and of the importance of reporting incidental findings. However, although it is over 15 years since these genes were identified many hundreds of sequence changes are still reported as `variants of unknown significance'. Moreover, although available evidence indicates the great majority of such variants are not associated with high risks of disease; the management of individuals with these variants is hugely variable, with some women being recommended mastectomy and/or oophorectomy. Multiple approaches have been used to try to predict the clinical consequences of BRCA1 and BRCA2 variants, but the majority of rare variants remain in the `unknown' category. A major limitation in evaluating variants has been the absence of mutational information of these genes from the normal population. To address this we are sequencing the BRCA1 and BRCA2 genes in 1000 population controls from the UK through exome sequencing and validation of all rare variants by Sanger sequencing. We have completed 500 samples and the results of the full 1000 will be presented. We identified 4 clearly pathogenic, truncating mutations in BRCA2 and none in BRCA1 in 500 individuals. By contrast, we identified 48 missense variants in BRCA2 and 28 in BRCA1 of which 40 and 15 respectively had frequencies <1%. We evaluated the 55 rare missense variants with in silico prediction programmes commonly used in diagnostics: Polyphen, SIFT and AlignGVD. 33 were predicted to be potentially deleterious by at least one in silico prediction programme and would have been reported as `variants of unknown significance'. Thus >6% of normal individuals would receive a result that could lead to considerable anxiety and clinical uncertainty if such results were fed back. The results also suggest that the frequency of BRCA2 mutations is higher than previously thought, which has implications for cancer risk estimation. The full spectrum of variation in BRCA1 and BRCA2 will be presented together with the implications for current variant prediction and risk estimation programmes and our revised process for management of BRCA variants in diagnostics.

139

Targeted re-sequencing of 10 ovarian cancer candidate genes in 2,240 cases and 355 controls. *H. Song[1], M. Cieck[2], J. Cunningham[2], B. Fridley[2], E. Dicks[1], P. Harrington[1], S. Ramus[3], S. Gayther[3], E. Goode[2], P. Pharoah[1]*. 1) Dept Oncology, Univ Cambridge, Cambridge, United Kingdom; 2) Mayo Clinic, Rochester, Minnesota, USA; 3) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, California, USA.

Background: Several genes with rare, deleterious alleles are associated with susceptibility to epithelial ovarian cancer (EOC). The aim of this study was to estimate the contribution of deleterious mutations in ten candidate genes (*BRCA1*, *BRCA2*, *MSH6*, *MLH1*, *MSH2*, *PMS2*, *RAD51B*, *RAD51C*, *RAD51D*, and *TIPARP*) to EOC in the population. **Methods**: We used multiplexed 48.48 Fluidigm access arrays for targeted sequence library preparation. A total of 421 primer pairs were designed to cover the coding sequences and splice sites. We used next-generation sequencing technology to sequence lymphocyte DNA for 2,240 cases of EOC and 355 controls from SEARCH and Mayo Clinic studies. We included 44 pair of duplicates and 33 known *BRCA*1/2 mutation carriers for quality control. GATK was used for variants detection and Annovar was used for variants annotation. Amplicons with a read depth of <15 were excluded from the analysis. Alternate alleles were called if > 40 percent of reads differed from the consensus sequence for read depth 15 to 29 or if > 30 percent of reads differed from the consensus sequence for read depth >29. **Results**: The median read depth was 104 (IQR 85–122). The concordance rate for duplicates was 98%. We detected 29 of the known *BRCA1/2* mutations (88%). Four were missed because the read depth of the relevant amplicon was <15. 168 of cases (7.2%) carried deleterious variants in one of the genes: 62 (2.7%) in BRCA1, 80 (3.5%) in BRCA2, 9 (0.4%) in *MSH6*, 3 (0.13%) in *RAD51C*, 7 (0.31%) in *RAD51D*, and one in each of *MSH2*, *PMS2* and *TIPARP*. No deleterious variants were identified in *RAD51B* gene. Homologous recombination deficiency is specifically associated with high-grade serous EOC. The estimated prevalence of *BRCA1* and *BRCA2* mutations in these cases was 5.3% and 6.4% after adjusting for the fact that NGS will not detect large genomic rearrangements - approx. 10% mutations in *BRCA1* or *BRCA2* - and we estimate another 15% of mutations were missed due to inadequate coverage. Conclusions: BRCA1 and BRCA2 together account for about 1 in 12 of all EOC cases and 1 in 9 cases of high-grade serous EOC. Mutations in *MSH6*, *MLH1*, *MSH2*, *PMS2*, *RAD51B*, *RAD51C*, *RAD51D*, and *TIPARP* together account about 1% of EOC. It will be feasible to use a targeted sequencing approach for clinical testing of multiple genes in a single assay. Large-scale targeted resequencing may also be a useful approach to identifying novel, rare ovarian cancer susceptibility alleles.

140

Enhanced detection of low-level mosaic mutations in RB1 gene in Sporadic Unilateral RB by Ion Torrent semiconductor sequencing: Risk of second cancer. *Z. Chen[1], S. Walther[1], K. Moran[1], D. Gerhart[2], T. Ganguly[2], A. Ganguly[1]*. 1) Genetic Diagnostic Laboratory, Department of Genetics, University of Pennsylvania, Philadelphia, PA; 2) DNA Sequencing Facility, Department of Genetics, Perelman school of Medicine, University of Pennsylvania.

BACKGROUND: Retinoblastoma (RB) is the most common intraocular cancer in children. Sporadic RB is caused by de novo germline mutations in RB1 gene in ~80% of newly diagnosed bilateral cases and ~15% of unilateral cases. De novo germline mutation may be the result of either a pre-conception mutation in a parental germ cell, or an early post-zygotic mutation giving rise to somatic and/or germline mosaicism. Detection of a mosaic mutation, which is challenging, alters the future risk of second cancer for the affected individual. Deep sequencing methods are significantly more efficient in detecting low-level mosaic mutations missed by traditional sequencing. METHODS: We performed deep sequencing of RB1 gene on DNA isolated from blood of 69 RB probands with unilateral disease, with 93 somatic mutations identified in tumors, but without any germline mutation. The individual exons of RB1 gene with known somatic mutations were amplified using DNA of the respective individual. The amplicons were pooled in equimolar concentrations, barcoded libraries prepared and sequenced on the Ion PGM using 316 chip and 100b sequencing kit. Sequencing was performed in duplicate. To estimate the level of background sequencing error, all the tested exons were amplified using DNA from three healthy individuals, pooled, and sequenced following the same procedure. Sequence data was analyzed by Torrent Suite (Life Technologies) and NextGENe software (Softgenetics). RESULTS: About 250 Mb data were generated on a single 316 chip run and the average coverage for each exon was above 10,000. Five low-level mosaic mutations were identified. Sanger sequencing missed three of these mutations and two were detectable upon re-examination of the chromatograms. These variants were called with high statistical significance ($p < 0.0001$). Thus, the incidence of mosaic germline mutation is estimated to be 7.2% in unilateral RB probands. DISCUSSION: Use of deep sequencing platform, Ion PGM, yielded highly sensitive detection of low-level mosaic mutations in RB1 gene. While previous methods using allele specific PCR had predicted presence of ~3% mosaic mutations in unilateral RB, the current platform indicates that the rate can be as high as 7%. Considering, 15% as the standard germline mutation frequency in sporadic unilateral RB, this is a significant increase. This finding changes the genetic counseling for risk of second cancer in the proband and for future affected offsprings.

141

Risk of colorectal cancer for monoallelic and biallelic MUTYH mutation carriers. *A.K. Win[1], S.P. Cleary[2,3], J.G. Dowty[1], D.D. Buchanan[4], J.P. Young[4], N.M. Lindor[5], R.W. Haile[6], P.A. Newcomb[7], L. Marchand[8], J.L. Hopper[1], S. Gallinger[2,3], M.A. Jenkins[1], the Colon Cancer Family Registry.*
1) School of Population Health, The Univ Melbourne, The Univ Melbourne, VIC, Australia; 2) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 3) Cancer Care Ontario, Toronto, Ontario, Canada; 4) Cancer and Population Studies Group, Queensland Institute of Medical Research, Bancroft Centre, Herston, Queensland, Australia; 5) Department of Health Science Research, Mayo Clinic Arizona, Scottsdale, Arizona, USA; 6) Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA; 7) Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 8) University of Hawaii Cancer Center, Honolulu, Hawaii, USA.

Background: Germline mutations of the DNA base excision repair gene MUTYH are associated with an increased risk of colorectal cancer. Due to the rarity of these mutations, previous studies have been underpowered to provide precise estimates of age- and gender-specific cancer risks for monoallelic and biallelic mutation carriers. Methods: We identified 269 families of monoallelic (n=227) and biallelic (n=42) mutation carriers (probands) from the Colon Cancer Family Registry (47 from Australasia, 138 from USA and 84 from Canada). These carrier families had been ascertained either because they had a family history of cancer (n=33) or from sampling population cancer registries independently of family history (n=236). We estimated the hazard ratios (HR) of colorectal cancer incidence for mutation carriers relative to the general population (based on age-, sex- and country-specific cancer incidences), and hence the age-specific cumulative risks (penetrance) using a modified segregation analysis that incorporated both genotyped and ungenotyped relatives and conditioned on ascertainment to produce unbiased estimates. Results: We observed a total of 413 colorectal cancers with median age at diagnosis of 58 years (range 23–93) in the families. The HR for monoallelic mutation carriers was 1.85 (95% confidence interval, 1.28–2.68) [2.39 (1.55–3.67) for males and 1.28 (0.64–2.57) for females; p=0.13]. The HR for biallelic mutation carriers was 69.27 (28.10–170.75) [214.65 (82.59–557.86) for males and 19.59 (3.19–120.37) for females; p=0.02]. The cumulative risks to age 70 yrs were estimated to be: monoallelic mutation, 51% (41–81%) for males and 4% (3–5%) for females; and biallelic mutation, 87% (56–99%) for males and 76% (44–97%) for females. Conclusions: This international study, the largest one to date, provides more precise and accurate estimates of both absolute and relative colorectal cancer risks for MUTYH mutation carriers. Men carrying a biallelic mutation in MUTYH have a higher risk of colorectal cancer than women.

142

Newborn screening for cystic fibrosis: Preliminary results on the false positive experience. *C.J. Barg[1], F.A. Miller[1], R.Z. Hayeems[1,2], P. Durie[3], J.C. Carroll[4], P. Chakraborty[5,6], B.K. Potter[7], Y. Bombard[8,9], K. Tam[10], L. Taylor[11], E. Kerr[12], C. Davies[5], J. Milburn[5], F. Ratjen[11], A. Guttmann[1,2,13]*. 1) Institute of Health Policy, Management & Evaluation, University of Toronto, Toronto, Canada; 2) Institute for Clinical Evaluative Sciences, Toronto, Canada; 3) The Research Institute and Department of Pediatrics, The Hospital for Sick Children, Toronto, Canada; 4) Department of Family and Community Medicine, Mount Sinai Hospital, University of Toronto, Canada; 5) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, Canada; 6) Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, Canada; 7) Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada; 8) Yale University, Department of Epidemiology and Public Health, New Haven, Connecticut, USA; 9) Center for Health Policy and Outcomes, Memorial Sloan Kettering Cancer Center, New York, USA; 10) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada; 11) Division of Respiratory Medicine, Department of Pediatrics, The Hospital for Sick Children, Toronto, Canada; 12) Department of Psychology, The Hospital for Sick Children, Toronto, Canada; 13) Division of Paediatric Medicine, Department of Pediatrics, The Hospital for Sick Children, Toronto, Canada.

Introduction: Many newborn screening (NBS) programs have expanded to test for additional conditions, often including cystic fibrosis (CF). The clinical benefits of early detection for infants with CF are established. However, early studies identified psychosocial harm in families of infants with false positive (FP) NBS results for CF. Whether well designed programs of infant retrieval and confirmatory testing can minimize harms for families receiving FP results remains unknown. **Methods:** We used a mixed-methods cohort design to obtain prospective self-report data from all mothers of infants with FP CF NBS results 1–3 months after confirmatory testing at Ontario's largest NBS treatment centre, and from a randomly selected control sample of mothers of similarly aged screen negative infants from the same region. Mothers completed a questionnaire assessing their experience, attitudes, and psychosocial response. Willing mothers of FP infants also completed a qualitative interview. **Results:** To date, 88 mothers of FP infants (response rate = 61%) and 254 controls (response rate = 45%) have completed questionnaires; 13 FP mothers were interviewed. Preliminary data show no significant differences between mothers of FP and control infants on measures of psychosocial response (*ps* > .05). Qualitatively, mothers of FP infants reported that the time between notification of the positive NBS result and confirmatory testing was highly stressful, but identified personal, familial, and clinical sources of support. Groups did not differ on marital status and income (*ps* > .05), though control mothers reported higher educational attainment. Both samples had a significantly higher household income and were more likely to be married or living common-law compared to the Ontario population (*ps* < .05). **Conclusions:** Preliminary findings suggest that mothers of infants with FP CF NBS results do not experience measurable psychosocial harm in the early postpartum months. The NBS treatment centre from which mothers were recruited has implemented recommendations to minimize delay between initial notification and confirmatory testing and to provide accurate and comprehensible risk information to parents at initial notification. These factors may explain the absence of measurable psychosocial burden, despite mothers' qualitative acknowledgment of initial distress. Recruitment biases limit generalizability to the full population, which is likely to include more vulnerable families.

143

Conflicting views on newborn and infant genetic screening: perspectives of relatives of children with genetic conditions causing developmental delay and parents of healthy children. *S.A. Metcalfe[1, 2], A.D. Archibald[1, 3], A.L. Atkinson[1, 2], C. Hickerton[1], S. Lawton[1, 2], B.J. McClaren[1], S.H. Wong[1, 2]*. 1) Genetics Education & Health Research, Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, Australia; 2) Dept Paediatrics, The University of Melbourne, Melbourne, Australia; 3) Victorian Clinical Genetics Service, Royal Children's Hospital, Melbourne, Australia.

The rapid pace of technological advances has engendered vigorous debate about expanding newborn screening (NBS) to include genetic conditions for which there are no curative treatments. It is recognized that early testing will reduce the `diagnostic odyssey', earlier interventions may improve quality of life and information from screening can be valuable for future reproductive decision-making. One important consideration for policy makers is the views of parents, as stakeholders, about acceptability of such programs. We surveyed 62 parents attending child immunization clinics regarding their interest in genetic screening for conditions causing developmental delay, with 49 parents also providing comments to open-ended questions. We conducted semi-structured in-depth interviews with a subgroup of 14 parents of 26 healthy children (each parent having at least one child ≤2yr), as well as interviewing a further 15 parents of boys with Duchenne muscular dystrophy (DMD) and 10 relatives of individuals with spinal muscular atrophy types 2 and 3 (SMA2/3). Interviews were digitally-recorded and transcribed. Thematic analysis of interviews and survey text responses was performed, with independent coding by at least 2 researchers. 79% of those surveyed expressed an interest in screening for conditions causing developmental delay, with the most common reason (22 responses) being to allow early intervention/management, and 9 comments relating to help plan and prepare for the future. Most parents interviewed also supported such screening provided it is voluntary. However, while parents of healthy children generally thought that NBS should be offered, perceiving this as a convenient time, by contrast parents of children with DMD or SMA2/3 were overall much more supportive of screening at a later stage (1–2 years of age), as they valued having a period of `normality'. They perceived that receiving a diagnosis from NBS would have impacted negatively on the relationship with their child. Those favoring infant screening suggested it could be targeted to children whose parents are concerned about their child's development, while parents of healthy children often worried that targeted screening might reduce equity of access. These divergent views probably reflect the contrasting experiences participants had with their children's development. These findings highlight the ethical complexities of offering population screening for conditions causing developmental delay.

144

Do Research Participants Really Want to Know? The Seattle Colorectal Cancer Family Registry Experience on the Return of Research Genetic Test Results. *M. Laurino[1,2,3], D. Fisher[1], W. Grady[1,3,4], P. Newcomb[1,5].* 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) University of Washington School of Public Health Institute for Public Health Genetics, Seattle, WA; 3) Seattle Cancer Care Alliance, Seattle, WA; 4) University of Washington School of Medicine, Seattle, WA; 5) University of Washington School of Public Health Department of Epidemiology, Seattle, WA.

The Seattle Colorectal Cancer Family Registry (CCFR) is part of six collaborating international registries funded by the National Cancer Institute to specifically provide a resource for interdisciplinary colorectal cancer (CRC) research. The informed consent process during the population and clinic based recruitment of individuals with CRC and members of their family from 1998–2007 did not specifically discuss the possibility of receiving individual result disclosure. Since several professional advisory groups recently emphasize the clinical validity and utility of the genetic research finding influencing the proposed guidelines about when to disclose research test results, researchers at the Seattle CCFR decided to offer genetic counseling (GC) to their registry participants tested for deleterious mismatched repair (MMR) gene mutations. This decision is based on the increasing evidence that supports the effectiveness of cancer screening for early detection and cancer prevention in individuals with Lynch syndrome. Considering there is still limited information on participant uptake of the offer to receive genetic research results, we propose to share our experience following implementation of the “Do You Want to Know (DYWTK)” GC protocol. Since February 2011, we initiated re-contact of eligible registry participants and offered GC to discuss the opportunity to receive individual results. To date, 53/ 72 (73.6 %) of eligible participants expressed an interest to receive their genetic research results and 14/72 (19.4%) participants declined this offered opportunity. For those interested, 25/53 (47.2 %) completed their GC 1 session and 22/25 (88 %) completed their GC 2 session and received their genetic research results. Non-responders (i.e., participants who did not return their DYWTK response form or follow-up phone call) were 11/72 (15.3%) following DYWTK approach letter, 19/53 (35.8 %) to schedule for GC 1 session, and 3/25 (12%) to complete their GC 2 session. We are currently conducting GC assessment surveys (e.g., decliner, baseline GC 1 session, and post 2-month and 12-month GC 2 session) to explore the reasons why some participants accept and others decline GC and the opportunity to know their genetic test results offered in the research setting. Outcomes from these GC surveys provide valuable insights on how findings from genetic and genomic research are utilized for personal and family health promotion.

145

The Student-Athletes' Knowledge of Sickle Cell Trait and the Impact of Mandatory Genetic Testing. *N. Lovick[1], L. Mar[1], M. Treadwell[2], J. Youngblom[1], C. Hartshorne[2].* 1) California State University, Stanislaus, Turlock, CA; 2) children's Hospital & Research Center Oakland, Oakland, CA.

Sickle cell trait (SCT), generally considered to be a benign condition, has been associated with several case reports of exercise-related sudden death. In 2010, the National Collegiate Athletic Association (NCAA) approved mandatory SCT testing of all Division I student athletes. This paper investigated three main questions: Are student-athletes' receiving any pre-test counseling or education; What is the student-athletes' understanding of SCT and the associated health implications; For students identified as having SCT, have their training habits, performance, or participation in sport been impacted? A survey was developed and distributed to student-athletes participating in an NCAA division I sports team in California. Responses from 233 student-athletes (57.3% female) indicated that the majority of student-athletes are not receiving any pre-test counseling and/or educational materials (63.3%). The majority of participants (up to 64.3%) were unable to answer basic, but important, questions that are often discussed by genetics providers prior to offering genetic testing. Those who received pre-test counseling were not able to correctly answer these questions at a significantly higher frequency than those who did not (p=0.39). Four participants were identified as having SCT. Only two of them received post-test education/counseling about being a carrier. Two indicated that their practice habits have somewhat changed, including drinking more water and taking more rests to avoid exhaustion. One person indicated that their coach treats them differently by providing them with more water breaks then everyone else on the team. No one indicated that their performance in sport has changed since finding out they were a carrier. These results indicated that no major ramifications resulted from testing positive, but that more thorough genetic counseling prior to testing is needed to ensure fully informed consent and avoid any confusion surrounding SCT and the associated health implications.

146

Impact of Direct-to-Consumer Pharmacogenomic Testing. *C.S. Bloss[1], N.J. Schork[1, 2], E.J. Topol[1, 2, 3].* 1) Scripps Translational Science Institute and Scripps Health, La Jolla, CA; 2) 2Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA; 3) Scripps Clinic Medical Group, La Jolla, CA.

Aim of Investigation: To evaluate the psychological, behavioral and clinical impacts of direct-to-consumer (DTC) pharmacogenomic (PGx) testing. Methods: We performed a nested case-control analysis of data from the Scripps Genomic Health Initiative, a longitudinal cohort study originally designed to assess the impact of DTC genomic testing for common disease in a large sample of adults. Analyses presented are based on web-based assessments performed at baseline and long-term follow-up, at which point a subset of participants had received PGx testing for 12 medications or drug reactions, including succinylcholine, 5FU/capecitabine, abacavir, azathiopurine/6-MP, beta blockers, carbamazepine, floxacillin, irinotecan, Plavix, statin induced myopathy, statin response, and warfarin. Psychological, behavioral, and clinical outcomes at long-term follow-up were compared between PGx recipients and non-recipients. Secondary analyses within PGx recipients were also performed to evaluate the impact and perceptions of the test as a function of current or previous use of medications on the PGx panel. Results: Out of 1325 individuals who completed long-term follow-up, 481 (36.3%) received PGx testing. Psychological and behavioral (diet or exercise) outcomes were not statistically different between individuals who received PGx testing versus those who did not. Receipt of PGx testing was, however, associated with higher overall screening test completion at follow-up ($p<.05$). Among PGx test recipients, 8.8% reported sharing their PGx test results with their physician, and 18.3% self-reported current or previous use of at least one medication on the PGx panel. The majority of test recipients perceived the test to have somewhat (34.9%) or definite (58.0%) personal utility, Conclusions: PGx testing was associated with higher overall screening test completion, and the majority of PGx test recipients perceived the test to have personal utility even though most did not report current or previous use of any of the medications on the panel.

147

Impact of Genomic Risk for Type 2 Diabetes on Health Behaviors. *S.B. Haga[1], W. Barry[2], R. Mills[1], J. Sullivan[3], H.F. Willard[1], L.P. Svetkey[4], G.S. Ginsburg[1].* 1) Inst Genome Sci & Policy, Duke Univ, Durham, NC; 2) Department of Biostatistics and Bioinformatics, Duke Univ Medical Center, Durham, NC; 3) Department of Pediatrics, Div Medical Genetics, Duke University Medical Center, Durham, NC; 4) Department of Medicine, Div Nephrology, Duke University Medical Center, Durham, NC.

We are conducting a clinical study to assess the impact of health and genetic literacy on risk comprehension, risk perceptions, and health behavior of individual genomic risk information for Type 2 diabetes mellitus. A total of 300 individuals were randomly recruited from the general public: 62% self-identified as White, 26% as African-American, 72% female, and 67% have a college degree or higher. At baseline, we assessed health literacy, genetic literacy, and perceptions and attitudes about genetic testing. After delivery of test results, we assessed the intention to adopt or actual adoption of healthy lifestyle changes with respect to diet and exercise in 240 of the participants at 3 months. Overall, 47% of participants increased their exercise activity, 57% reduced their fat intake, and 36% increased their fruit/vegetable/fiber intake after receiving their genetic risk for T2DM compared to baseline. Participants at increased genetic risk were no more likely to change their diet and/or exercise behavior compared to those at decreased or same as population risk. In addition, we did not observe any association with improved diet or exercise activity with respect to health literacy, gender or income level. Analysis of 6-month follow-up data is pending as well as additional sub-group analysis to identify factors that may be associated with change of health behaviors. In conclusion, while a significant number of participants improved their diet and exercise habits, we find no specific effect due to knowledge of genetic risk for T2DM, confirming other reported findings.

148

African American attitudes toward exome and whole genome sequencing. *J. Yu[1], J. Crouch[2], S.M. Jamal[1], H.K. Tabor[1,2], M.J. Bamshad[1,3].* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Seattle Children's Hospital and Research Institute, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA.

Exome sequencing and whole genome sequencing (ES/WGS) present patients and research participants the opportunity to receive multiple genetic results of varying clinical and personal utility. Yet, this potential for direct benefit also risks disenfranchising populations such as African Americans (AA) that are already underrepresented in genetic research. Lower rates of genetic test utilization by AA as compared to European Americans (EA) may be attributed, in part, to differences in genetic awareness and knowledge, misperceptions about genetic risk, and differences in orientations to family, time, and spirituality. Given these barriers, understanding AA perspectives on participating in ES/WGS research and the prospect of receiving genetic results from ES/WGS will be critical to ensuring parity in genomic healthcare and research. We conducted a series of eight focus groups (n=60) to investigate if and how attitudes toward participation in ES/WGS research participation and return of results from ES/WGS differ between AA and EA. We found that perspectives on participating in ES/WGS research and receiving ES/WGS results differed in several ways between AA and EA. AA expressed a need for collective decision-making about ES/WGS participation. In contrast to EA, most AA were unwilling to enroll their children in ES/WGS research. While almost all were willing to participate in ES/WGS research for altruistic reasons, AA questioned the value of receiving individual genetic results. For instance, they were skeptical that receiving genetic results would impact their reproductive decisions and that genetic results on complex conditions would motivate disease prevention behaviors. Those who wanted to receive results wanted them in multiple formats but most AA insisted on receiving a copy of their genome if sequenced. AA related these themes to an expressed deep mistrust of biomedical research and negative healthcare experiences. While AA and EA in this study were interested in and willing to participate in ES/WGS research, they differed in their perspectives on ES/WGS research participation and the value of receiving individual genetic results. These preliminary findings highlight the need to investigate differences among racial ethnic groups with respect to the ethical issues raised by ES/WGS, and they further suggest the need to develop and test culturally tailored strategies for consenting and returning results to AA.

149

Personalized health literacy in the age of personalized medicine: results from a deliberative public engagement exercise. *B.J. Wilson[1], J.C. Carroll[2], S.G. Nicholls[1], S.M. Craigie[1], H. Etchegary[3], D. Castle[4], B.K. Potter[1], J. Little[1,5], L. Lemyre[6] on behalf of the Emerging Team in Genomics in Screening.* 1) Epidemiology & Community Med, Univ Ottawa, Ottawa, ON, Canada; 2) Department of Family Medicine, Mount Sinai Hospital, University of Toronto, Toronto; 3) Clinical Epidemiology, Memorial University, Newfoundland,; 4) ESRC Innogen Centre, University of Edinburgh, UK; 5) Canada Research Chair in Human Genome Epidemiology; 6) School of Psychology & Institute of Population Health, University of Ottawa.

The era of personalized medicine means that decision-making using genomic information will no longer be confined to specialist settings with supportive genetic counseling resources, or to only populations affected by rare diseases. To maximize the utility gained from genomic tests, it is important to consider how best to promote genomic or `personalized health literacy' - “sufficient knowledge and appreciation of genomics principles to allow informed decision making for personal well-being and effective participation in social decisions on genetic issues.” - among target populations: Aim To examine baseline levels of personalized health literacy in a general population with no special experience of genetic tests. Methods We conducted a public engagement exercise to explore the idea of incorporating a hypothetical genomic profiling test into routine screening for colorectal cancer (CRC). Within a workshop format, we applied stepped information sets, in neutral language, describing the application and purposes of this approach to screening, and possible implications from personal, familial, health system and societal perspectives. We recorded group discussions using audio and/or field notes, and encouraged free association written responses, supplemented by a structured attitude survey at the end of the workshop. Qualitative data were transcribed and standard content analysis approaches, using Roger's knowledge framework as an initial template (`awareness', `how-to', and `principles' knowledge). The research was conducted in Ontario and Newfoundland, Canada. Results Five workshops were conducted involving 120 individuals, aged 51–88 years (70% females). Participants readily understood the application of genomic profiling in the CRC screening context, were positive about personal knowledge as a way to promote and protect health, and most would consider asking for such a test, if it existed. They stressed the need for adequate understanding as a basis for making well-informed personal health care decisions. While the data suggested gaps in `principles' and `how-to' knowledge in respect of technical aspects of genomics, there was nevertheless a clear grasp of salient issues pertaining to the need for validity in genomic tests, the necessity of effectiveness evidence for preventive interventions to underpin the utility of personalized tests, and the broader set of `principles' issues relating to health system and societal aspects of test implementation.

150

Dynamics, definitions and discrepancies: public perspectives on the systematic collection and use of family health history in routine health care. *H. Etchegary[1], B.J. Wilson[2], S.M. Craigie[2], S.G. Nicholls[2], D. Castle[3], J.C. Carroll[4], J. Allanson[5], B.K. Potter[2], P. Chakraborty[6,7] on behalf of the CIHR Emerging Team in Genomics in Screening.* 1) Clinical Epidemiology, Memorial University, St. John's, Newfoundland, Canada; 2) Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada; 3) ESRC Innogen Centre, University of Edinburgh, UK; 4) Department of Family Medicine, Mount Sinai Hospital, University of Toronto, Toronto; 5) Children's Hospital of Eastern Ontario; 6) Newborn Screening Ontario, Children's Hospital of Eastern Ontario; 7) Department of Pediatrics, University of Ottawa, Ottawa.

The systematic collection of family health history (FHH) information may contribute to personalized health care. A recent consensus statement raised numerous research questions about the routine use of FHH in health care. Despite an incomplete evidence base, many health care organizations have implemented FHH initiatives and encourage the integration their use across the spectrum of patient interactions. We designed a project to explore public and patient perspectives on the systematic collection and use of FHH in routine health care settings. Aim: To identify issues relevant to the routine collection and use of FHH in routine health assessment in two contexts: chronic disease risk assessment in adults, and general health assessment in the early childhood period. Methods: We conducted semi-structured workshops in two targets populations (recruited from the community), healthy adults aged 50 years and over, and parents of children aged less than 5 years. In a focus group format, we presented stepped, neutrally framed information sets, outlining: FHH as a potential tool in health care; any empirical evidence on its utility; and (hypothetical) potential issues which might arise from a personal, family, and socio-legal perspective. Qualitative discussion data were supplemented by structured attitude questions. Results: While generally positive towards the idea of FHH as a readily available, non-invasive tool in health care, many issues emerged which are relevant to its systematic implementation in health care. Prominent among these were issues related to the definition of “family” and what happens when people have no access to their own FHH; many concerns about the legitimacy of expecting individuals to seek out FHH information when family relationships and dynamics many be complex; concerns about access to, or protection of, FHH information contained in databases; and concerns about the idea that FHH data could be linked across family members under the care of the same health care organization. Two particular issues emerged specific to FHH as a tool in the early childhood period. The first was the implications of, in effect, screening parents and their families for unsuspected conditions through the act of assessing a child's FHH. The second was the potential for conflict between disclosing complete FHH in the interests of the child and protecting the privacy of a parent whose preference is not to disclose personal health information.

151

Immunochip: Redefining the genetic architecture of multiple sclerosis. *J. McCauley, International Multiple Sclerosis Genetics Consortium.* Hussman Institute for Human Genomics, University of Miami, Miami, FL.

Multiple sclerosis (MS) is a chronic debilitating autoimmune disease of the central nervous system that affects approximately 2.5 million people worldwide. We have employed the Immunochip genotyping array to fine-map regions previously established as associated with susceptibility to MS, to interrogate regions of established relevance in other autoimmune diseases and to screen for additional novel MS susceptibility loci. The final manufactured Illumina iSelect custom beadchip platform contained 196,524 variants nominated for inclusion by multiple disease consortia. In total the chip enables dense genotyping in 186 regions of established relevance in at least one of 12 primary autoimmune diseases, including 38 regions that are already established as being associated with MS susceptibility. After comprehensive quality control, we performed a preliminary analysis of 160,640 SNPs (consisting of 107,136 variants across the 186 fine-mapping intervals and 53,504 additional variants) in a dataset consisting of 17,097 cases and 20,055 controls from twelve countries. Eighteen of the fine-mapped intervals (not previously established in MS) show evidence of association with genome-wide significance ($P < 5 \times 10^{-8}$), including for example Chr2q32 in the region of the STAT4 gene (rs9967792, $p = 6.08 \times 10^{-9}$). Of note, six of these 18 regions are known to be associated with inflammatory bowel disease (IBD). In addition a further 28 independent loci outside of these fine-mapping regions also show evidence of association with genome-wide significance. Many of these loci were previously suggested in our second generation genome-wide association study and meta-analyses. Our results clearly demonstrate the utility of the Immunochip and are set to nearly double the list of associated loci. Ongoing analyses are focused on fine-mapping efforts across these loci.

152

Dense genotyping of candidate genes indentifies 16 new susceptibility loci in ankylosing spondylitis. *A. Cortes[1], P.C. Robinson[1], P. Leo[1], D.M. David[2], M.A. Brown[1], International Genetics of Ankylosing Spondylitis and Wellcome Trust Case Control Consortium.* 1) The University of Queensland Diamantina Institute, Brisbane, QLD, Australia; 2) School of Social and Community Medicine, Bristol University, Bristol, UK.

Introduction/Aim: Ankylosing spondylitis (AS) is a highly heritable inflammatory arthritis common in both Asian and European populations. Thus far genes identified include the HLA-B*27 allele, and 13 non-MHC loci identified in European populations. In this study we aimed to better characterize the genetic architecture of AS and to fine-map known susceptibility loci. **Materials and Methods:** We successfully genotyped 129,030 polymorphic SNPs in 10,624 AS affected and 15,174 healthy individuals of European and Asian descent using the Illumina Immunochip microarray, which was designed for immunogenetic studies. **Results:** In this study we identified 16 new AS risk loci reaching genome-wide significance ($P < 5 \times 10^{-8}$), bringing the number of known non-MHC loci to 27. We found multiple independent association signals in 8 of these loci, caused by both common and low frequency variants, suggesting that multiple genetic variants within a gene can affect disease susceptibility. A second MHC association with the classical HLA-A*0201 was observed in both HLA-B*27 positive and negative disease (OR= 1.2; $P= 4.5 \times 10^{-9}$). European and Asian specific signals were observed in *IL23R* and *PTGER4*. **Discussion:** This study has replicated all attempted genome-wide significant loci reported in European populations and identified 16 novel susceptibility loci. Identified loci implicate microbial sensing (*NOS2*, *NKX2-3*, *SH2B3* and *ICOSLG*), intracellular antigenic peptide handling (*ERAP1*, *ERAP2*, *LNPEP* and *NPEPPS*) and CD8+ T cells (*EOMES* and *IL7R*) pathways as important in AS etiology as well as increase the number of susceptibility genes in the TH17 pathway (*TYK2* and *IL6R*). **Conclusion:** This increased characterization of the genetic architecture of AS aids greatly in explaining the currently poorly understood high observed heritability and familiality in AS. This data also guides functional studies towards uncovering how these genes cause disease and in the development of new therapeutics.

153

Dense fine-mapping study identifies novel disease loci and implicates coding and non-coding variation in primary biliary cirrhosis risk. *J.Z. Liu[1], M.A. Almarri[1], D.J. Gaffney[1], G.F. Mells[2], L. Jostins[1], H.J. Cordell[3], S. Ducker[4], D. Day[2], M.A. Heneghan[5], J.M. Neuberger[6], P.T. Donaldson[4], A. Bathgate[7], A. Burroughs[8], M. Davies[9], D.E. Jones[4], G.J. Alexander[10], J.C. Barrett[1], R.N. Sandford[2], C.A. Anderson[1], The UK PBC Consortium and The Wellcome Trust Case Control Consortium 3.* 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, UK; 2) Department of Medical Genetics, Cambridge University, Cambridge, UK; 3) Institute of Genetic Medicine, Newcastle University, Newcastle, UK; 4) Institute of Cellular Medicine, Newcastle University, Newcastle, UK; 5) Institute of Liver Studies, King's College Hospital NHS Foundation Trust, London, UK; 6) The Liver Unit, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; 7) Scottish Liver Transplant Unit, Royal Infirmary of Edinburgh, Edinburgh, UK; 8) The Sheila Sherlock Liver Centre, The Royal Free Hospital, London, UK; 9) The Liver Unit, St James University Hospital, Leeds, UK; 10) Department of Hepatology, Cambridge University Hospitals NHS Foundation Trust, UK.

We genotyped 2,861 cases from the UK PBC consortium and 8,514 UK population controls across 196,524 variants within 186 known autoimmune risk loci using the Immunochip array. We identified three novel genome-wide significant loci (P<5×10-8), increasing the number of known loci to 25. The most associated variant at 19p12 is a low-frequency non-synonymous SNP in TYK2, further implicating JAK/STAT and cytokine signaling in disease pathogenesis. A further five loci contain non-synonymous variants in high linkage disequilibrium (LD) (r2>0.8) with the most associated variant in the locus. For seven of the loci the lead SNP is in high LD with a known eQTL in B-lymphoblastoid cells. We found multiple independent common, low-frequency and rare variant association signals at five loci. Imputation of the classic human leukocyte antigen (HLA) alleles identified four independent HLA-allele associations, replicating DQA1*0401, DQB1*0602 and DQB1*0301 and newly implicating DRB1*0404 in disease pathogenesis. We integrated our association results with ENCODE DNase-seq data across 16 cell lines to see if, for a given tissue, our association result were enriched within open-chromatin (OC) regions. An enrichment score (E) was calculated by comparing the number of OC regions among candidate causal variants (r2>0.8 to the most associated SNP in each locus) against the total number of OC regions across all Immunochip loci. For each cell line, the null distribution and an empirical P-value of E were calculated by performing 1000 permutations, and repeating the enrichment analysis for each permutation. We tested for differences in E between cell lines by comparing E in one cell line to that using all other cell lines. We observed a modest enrichment of OC regions among genome-wide significant loci in B-lymphoblastoid cells when compared to the other 15 cell lines (P=0.068). Indeed, of the 26 independent non-HLA signals tagged on Immunochip, 15 have SNPs in B-lymphoblastoid OC regions in high LD (r2>0.8) with the most associated variant. This study demonstrates how dense fine-mapping arrays coupled with functional genomic data can be utilized to identify candidate causal variants for functional follow-up, and suggest relevant tissues for such studies. Furthermore, our enrichment analysis protocol can be applied to other functional annotations and phenotypes to identify the cell types in which disease associated variants are likely to be acting.

154

Fifteen novel psoriasis susceptibility loci: disease-specific signals highlight the role of innate immunity. *L.C. Tsoi[1], S.L. Spain[2], J. Knight[3,4], E. Ellinghaus[5], P.E. Stuart[6], F. Capon[2], J. Ding[1], Y. Li[1], T. Tejasvi[6], J.E. Gudjonsson[6], H.M. Kang[1], A.M. Bowcock[7], U. Mrowietz[8], S. Koks[9], T. Esko[10], J.J. Voorhees[6], M. Weichenthal[8], P. Rahman[11], D. Gladman[12], C.EM Griffiths[13], A. Reis[14], J. Kere[15], R.P. Nair[6], A. Franke[5], J.NWN Barker[2,16], G.R. Abecasis[1], J.T. Elder[6,17], R.C. Trembath[2,18], Genetic Analysis of Psoriasis Consortium, Wellcome Trust Case Control Consortium 2.* 1) Biostatistics, University of Michigan, Ann Arbor, MI. 48109, USA; 2) Division of Genetics and Molecular Medicine, King's College London, London, UK; 3) Neuroscience Research, Centre for Addiction and Mental Health, Toronto, ON, Canada M5T 1R8; 4) National Institute for Health Research (NIHR), Biomedical Research Centre, Guy's and St. Thomas' NHS Foundation Trust; 5) Institute of Clinical Molecular Biology, Christian-Albrechts-University, 24105 Kiel, Germany; 6) Department of Dermatology, University of Michigan Ann Arbor, MI 48109, USA; 7) Division of Human Genetics, Department of Genetics, Washington University at St. Louis, St. Louis, MO; 8) Department of Dermatology, University Hospital, Schleswig-Holstein, Christian-Albrechts-University, 24105 Kiel, Germany; 9) Department of Physiology, Centre of Translational Medicine and Centre for Translational Genomics, University of Tartu, 50409 Tartu, Estonia; 10) Estonian Genome Center, University of Tartu, 51010 Tartu, Estonia; 11) Department of Medicine, Memorial University, St. John's, Newfoundland A1C 5B8, Canada; 12) Department of Medicine, Division of Rheumatology, University of Toronto, Toronto Western Hospital, Toronto, Ontario M5T 2S8, Canada; 13) Dermatological Sciences, Salford Royal NHS Foundation Trust, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; 14) Institute of Human Genetics, University of Erlangen-Nuremberg, Erlangen, Germany; 15) Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden, and Folkhälsan Institute of Genetics, Helsinki, Finland, and Department of Medical Genetics, University of Helsinki, Finland; 16) St John's Institute of Dermatology, King's College London, London, UK; 17) Ann Arbor Veterans Affairs Hospital, Ann Arbor, MI, 48105, USA; 18) Queen Mary University of London, Barts and the London School of Medicine and Dentistry, London, UK.

Psoriasis is a chronic, potentially disfiguring, immune-mediated inflammatory disease of the skin with a prevalence of 0.2 to 2%. To gain further insight into the genetic architecture of psoriasis, we conducted a meta-analysis of three genome-wide association studies (GWAS) and two independent datasets genotyped on the Immunochip, involving 10,588 cases and 22,806 controls in total. Nineteen out of the 21 known psoriasis loci achieve genomewide signfiicance in our meta-analysis, and we identified 15 novel disease susceptibility regions, increasing the number of psoriasis-associated loci to 36 for Caucasians. Conditional analyses identified five independent signals within previously known loci. The 39 independent signals detected in the current study collectively account for 14.3% of the total psoriasis liability variance, or approximately 22% of its estimated heritability. We identified two significant pairwise interactions: *HLA-C* (rs4406273)-*LCE* (rs6677595) and *HLA-C* (rs4406273)-*ERAP1* (rs27432), as well as one additional pair showing supportive evidence of interaction: *IL23R* (rs9988642)-*TYK2* (rs34536443). The two significant interactions confirmed or extended results of previous studies, and the *IL23R-TYK2* genetic interaction is consistent with the known biological interactions of their encoded proteins. Novel shared disease regions encompassed a number of genes whose products regulate T-cell function (e.g. *RUNX3*, *TAGAP* and *STAT3*), and which overlap most often with intestinal diseases (Crohn's disease and celiac disease). Novel psoriasis-specific susceptibility intervals were notable for candidate genes whose products are involved in innate host defense, encoding interferon-mediated antiviral responses (*DDX58*), macrophage activation (*ZC3H12C*), and NF-κB signaling (*CARD14*, *CARM1*). These results portend a better understanding of shared and distinctive genetic determinants of immune-mediated inflammatory disorders and emphasize the importance of the skin in innate and acquired host defense.

155

MHC fine-mapping in celiac disease reveals structural basis of HLA-gluten interaction. *J. Gutierrez-Achury[1], G. Trynka[1], K.A. Hunt[2], J. Romanos[1], D. van Heel[2], C. Wijmenga[1], P.I.W. de Bakker[3,4,5].* 1) Department of Genetics, University Medical Center of Groningen, Groningen, The Netherlands; 2) Blizard Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, London E1 2AT, United Kingdom; 3) Brigham and Women's Hospital, Boston, MA; 4) Broad Institute of Harvard and MIT, Cambridge, MA; 5) University Medical Center, Utrecht, The Netherlands.

Background: The major histocompatibility complex (MHC) harbors risk for inflammatory and infectious diseases. However, the dissection of the causative alleles is challenging because of the high degree of linkage disequilibrium and its characteristic genetic structure. To refine the MHC association in celiac disease, and to identify independent risk variants we genotyped > 10,000 SNPs across the MHC region in 12,016 European celiac cases and 11,920 matched controls using a custom genotyping array (Immunochip). Methods: Using an european ancestry reference panel, we imputed the classical alleles of the HLA-A, B, C, DQA1, DQB1, DRB1, DPA1 and DPB1 genes, and the polymorphic amino acid positions in the gene products in 5 cohorts (UK, Netherlands, Poland, Spain and Italy). We used stepwise logistic regression conditional analysis, with gender as covariates in the null model, to identify independent associations. Results: Of all variants tested (SNPs, HLA classical alleles and amino acids), the strongest association was at amino acid position 55 in the DQβ1 protomer near the binding groove (-log(p)=2067). After adjusting for position 55, we found associations at DQα1-position 44 (-log(p) = 162.2), DQβ1-position 57 (-log(p) = 93.6), DQβ1-position 9 (-log(p) = 12.0) and DQβ1-position 52 (-log(p) = 10.7). The haplotypes formed by these five positions recapitulate the known risk-conferring effects of the DQ2.5 (OR=10.4), DQ2.2 (OR=4.7), and DQ8 (OR=2.1) alleles. After controlling for the HLA-DQ effects, we identified 4 additional associated variants across the MHC, including HLA-A-position 101, *HLA-B*08* and HLA-DPB1-position 38. Conclusions: Our fine-mapping strategy confirms the role of HLA-DQ molecules in the presentation of gluten antigens, and provides important structural insights about the molecular interaction between HLA and gluten. We also demonstrate evidence for independent effects in HLA-A, -B, -DPB1 and -DPA1 that warrant further study.

156

Host-microbe interactions shape genetic risk for inflammatory bowel disease. *J. Barrett, International IBD Genetics Consortium.* Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

Crohn's disease (CD) and ulcerative colitis (UC) are the two common forms of inflammatory bowel disease. They are illustrative of the success of genome-wide association studies and collaborative meta-analyses, which have previously implicated many susceptibility genes for these disorders. To extend these successes, and develop an integrated view of the underlying biology of IBD, we performed an enlarged GWAS meta-analysis of CD and UC, followed by replication and fine-mapping using the Immunochip custom array, in a total sample size of >75,000 cases and controls.

First, we identify 71 new loci resulting in a total of 163 with genome-wide significance. Most loci contribute to both phenotypes, though often with different effect sizes; the genetic etiology of CD & UC is largely shared, but different pathways appear to play roles of differential importance. We apply literature and PPI network analyses as well as eQTL and GO enrichment to annotate the functional pathways underlying IBD risk. We observe striking overlap between susceptibility loci for inflammatory bowel disease and mycobacterial infection. Gene expression network analysis emphasizes this relationship, with pathways shared between host responses to mycobacteria and those predisposing to inflammatory bowel disease. We observe an enrichment of signals of both directional and balancing selection at IBD associated SNPs, with a particular enrichment of balancing selection at loci also associated with bacterial defense, indicating that host-microbe interactions play a key role in this archetypal complex disease.

Our bioinformatic techinques identified single genes supported by multiple lines of evidence in 39 loci. We have used the dense fine-mapping data from Immunochip, followed by 1000 Genomes imputation, to assess the number of independent effects at these loci, and attempt to characterize likely causal alleles. Preliminary analysis suggests that independent effects are widespread, including the extreme case of the *NOD2* gene, with 9 independent variants. We have integrated our genetic data with databases of gene expression and other functional annotations such as ChIP-seq from the ENCODE project. In total, this dataset presents a vivid summary of the current knowledge of genetic risk of a complex disease.

157

Sequencing-based and multiplatform Genome-Wide Association Study for Multiple Sclerosis and Type 1 Diabetes in Sardinians. *I. Zara[1,7], E. Porcu[2,3,7], M. Zoledziewska[2,7], M. Pitzalis[2,7], M. Valentini[1], A. Mulas[2], F. Busonero[2], R. Atzeni[1], M. Oppo[1], F. Reinier[1], R. Berutti[1,3], R. Pilu[1], F. Deidda[1], C. Sidore[2,3,4], R. Piras[2], A. Loi[2,3], Sonia. Sanna[2], E. Cocco[5], F. Poddie[3], G. Farina[6], G. Rosati[6], L. Lianas[1], G. Cuccuru[1], G. Zanetti[1], A. Angius[1,2], M.G. Marrosu[5], C.M. Jones[1], G.R. Abecasis[4], Serena. Sanna[2], F. Cucca[2]*. 1) CRS4, Center for Advanced Studies, Research and Development in Sardinia, Parco Scientifico e Tecnologico della Sardegna, Pula, Italy; 2) Istituto di Ricerca Genetica e Biomedica, ex INN, Consiglio Nazionale delle Ricerche, Monserrato, Italy; 3) Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy; 4) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA; 5) Centro Sclerosi Multipla, Dipartimento di Scienze Neurologiche e Cardiovascolari, Università di Cagliari, Cagliari, Italy; 6) Clinica Neurologica, Università di Sassari, Sassari, Italy; 7) These authors contributed equally to this work.

Multiple Sclerosis (MS) and Type 1 Diabetes (T1D) are extremely common in Sardinians, the incidence being among the highest worldwide. More than fifty loci known to increase susceptibility in Europeans for each of the two diseases have been reported, but are unable to explain the unusual frequency observed in Sardinia. To detect novel loci associated with both diseases, including those that are Sardinian specific, we genotyped ~10,000 individuals, including ~850 trios, with the Illumina Immunochip (IC) and a subset of ~6,000 individuals, including ~240 trios, with the Affymetrix 6.0 array (A6). After applying standard quality filters, we analyzed 3,994 unrelated controls, 2,934 MS and 1,662 T1D unrelated cases genotyped with the IC, and 2,149 unrelated controls, 2,273 MS and 1,376 T1D unrelated cases genotyped with the A6 array. Furthermore, to better explore the Sardinian specific genetic contribution, we imputed ~13 million variants using a reference panel of 1,656 haplotypes deriving from low-pass sequencing of 1,146 Sardinian samples (4x coverage, on average). For T1D, we observed genome-wide significant association at the HLA class II (p-value up to 9×10-272) and the INS gene (p-value up to 6.52×10–20), and suggestive evidence at IL2RA and CLEC16A genes (p-value < 5×10-5). Interestingly, at INS and IL2RA, the top variants resulting from the IC GWAS were unlinked to those detected in A6 GWAS and from published GWAS. For MS, we detected genome-wide significant associations at HLA (p-value up to 1.27×10–58), IL2RA (p-value up to 1.2×10-9) and CLEC16A (p-value up to 4.71×10–12). Excluding the SNP at CLEC16A, the best signals in our GWAS at known loci are independent (r2<0.2) from the best signals found in IMSGC (Nature, 2011). In addition to known loci, we observed two additional loci for which SNPs with borderline significance (p-value < 5×10-6) in the A6 GWAS data became genome-wide significant in IC GWAS, and another one with genome-wide significance in the latter but only nominal evidence in the former. Variants at these three potentially novel loci are now being followed up for validation in an independent case-control data set. Our study combines the standard GWAS analysis with sequence data coming from individuals of the same isolated population, increasing the power to detect Sardinian specific variation responsible for the high incidence of MS and T1D in Sardinia.

158

Admixture Mapping for Asthma in Latinos Identifies Additional Heritable Risk Factors from Genome-Wide Meta-Analysis Data. *C.R. Gignoux[1], D.G. Torgerson[1], J.M. Galanter[1], L.A. Roth[1], C. Eng[1], D. Hu[1], S. Huntsman[1], R.D. Hernandez[1], R.A. Mathias[2], S. Sen[1], K.C. Barnes[2], E.G. Burchard[1]*. 1) UCSF, San Francisco, CA; 2) Johns Hopkins University, Baltimore, MD.

As the field of human genetics moves increasingly towards next-generation sequencing, the wealth of currently available genome-wide association study data is in danger of being ignored. Here we show the power of admixture mapping to uncover novel risk factors within existing genome-wide association study data. We performed admixture mapping in 7,002 individuals of mixed ancestry (3,906 Hispanic/Latino and 3,102 African-American) from the EVE Asthma Genetics Consortium, testing whether ancestry, rather than genotype, at each locus was associated with disease status. We uncovered a novel Latino-specific genome-wide significant admixture mapping peak ($p \sim 6.8\times10\text{-}6$) in 18q12 centered on *SMAD2*, a member of the TGF-β pathway known to play an important physiological role in asthma but never before detected in genome-wide association studies. We replicated the association in the locus in an additional 3,774 Latino individuals in the GALA II study ($p \sim 5\times10^{-3}$). We then measured *SMAD2* expression in blood in a subset of GALA II individuals and find significantly lower expression in people with asthma than with healthy controls (25% relative decrease, $p \sim 1.2\times10^{-4}$). Remarkably, in the original GWAS, no p-value for genotypic association neighboring *SMAD2* was below 10-4. This suggests that rare, population-specific variants near *SMAD2* are not well tagged by existing arrays. Follow-up imputation using 1000 Genomes identifies a potential candidate SNP upstream of *SMAD2* with an elevated allele frequency in Native Americans with an OR of 1.67 (95% CI 1.32–2.1). The discovery of this ethnic-specific risk factor argues for the importance of sequencing diverse populations to determine a more complete picture of the genetics of asthma. In addition, as the number of genome-wide datasets continues to grow, we demonstrate the utility of alternative mapping strategies on existing GWAS datasets.

159

Deep exome sequencing of psoriasis identified new association signals contribute by INDELs, CNVs and rare SNPs. *X. Jin[1, 3], H. Tang[2], H. Jiang[1], D. Cao[1], H. Shao[1], Q. Li[1], J. Shen[1], L. Song[1], Y. Shi[1], J. Mei[1], X. Yang[1], L. Coin[4], Y. Li[1], X. Zhang[2], J. Wang[1]*. 1) BGI, Shenzhen, China; 2) Anhui Medical University, Hefei, Anhui 230022, China; 3) School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510641, China; 4) Department of Genomics of Common Disease, Imperial College, London, London W12 0NN, United Kingdom.

Psoriasis is a common chronic, autoimmune and hyper proliferative skin disease following a pattern of polygenetic or multifactorial inheritance. Although Genome-wide association studies (GWAS) have found several association signals for psoriasis, a large part of heritability for the disease still remains unexplained. Here we sequenced whole exome of 700 cases and 800 controls from Han Chinese population to ~30X. After careful quality control, 547,978 SNPs were identified, in which 372,289 (63.25%) with minor allele frequency (MAF) less than 1% (93.09% were novel) and 217,268 (39.65%) were singleton (95.89% were novel). Several gene-based association detection methods were used to address signals from rare variants (MAF<5%). Number of genes shown significant association signals were known to be important in immune system related pathways. Except SNPs, the data also allow us to detect association signals based on INDELs and CNVs. We use an newly development method to estimate allele frequency spectrum of INDELs and detect association signals. We also developed an exome sequencing analysis pipeline to identify disease associated CNVs, and to generate absolute copy number genotypes at putatively associated loci. The method successfully re-discovered the LCE3B_LCE3C CNV associated with Psoriasis (p-value = 5×10e-6) and identified number of new association regions. Significant signals found by SNPs, INDELs and CNVs were collected together and using custom designed target region sequencing strategy to validate in another 10,000 cases and 10,000 controls. The study contributes to our understanding of the genetic basis of psoriasis and highlights the power and importance of using next-generation sequencing technology to identify associations between phenotype and different kinds of genetic variants.

160

Mutations in *DDHD2* cause recessive spastic paraplegia with intellectual disability, thin corpus callosum and periventricular white matter hyperintensities. *A.P.M. de Brouwer[1], J.H.M. Schuurs-Hoeijmakers*[1], E.J. Kamsteeg*[1], S. Ben-Salem[2], S.T. de Bot[3], I. van de Vondervoort[1], S. Vermeer[1], J. Schwartzentruber[4], B.R. Ali[2], S.A. Al-Yahyaee[2], S. Tariq[2], T. Pramathan[2], R. Bayoumi[2], B.P. van de Warrenburg[3], W.M. van den Akker[1], C. Gilissen[1], J.A. Veltman[1], I.M. Janssen[1], A.T. Vulto-van Silfhout[1], S. van der Velde-Visser[1], A. Diekstra[1], C.E. Erasmus[3], M.A. Willemsen[3], L.E.L.M. Vissers[1], H. van Bokhoven[1], R.A. Wevers[5], L. Al-Gazali#[2], M.T. Geraghty#[4], B.B.A. de Vries#[1]*. 1) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Institute for Genetic and Metabolic Disorders, Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The N; 2) Department of Paediatrics, Faculty of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates; 3) Department of Pediatric Neurology and department of neurology, Department of Pediatrics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 4) Department of Pediatrics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 5) Department of Laboratory Medicine, Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Phospholipase A1 (PLA1) comprise a group of enzymes that hydrolyze the sn-1 ester bond of phospholipids, producing 2-acyl-lysophospholipids and fatty acids. None of the members of the PLA1 family have been implicated in human disease so far. We report on four families with a clinical presentation of complex hereditary spastic paraplegia (HSP) due to mutations in *DDHD2*, one of the three intracellular PLA1s. We used exome sequencing to study two sibling families with a complex and presumed autosomal recessive HSP, presenting as a combination of progressive spasticity with an onset before 2 years of age, intellectual disability (ID) and cerebral abnormalities consisting of a marked thin corpus callosum, periventricular white matter hyperintensities and an unusual peak on brain MRS imaging that we propose to represent an, as yet, unknown lipid molecule. In both families compound heterozygous mutations were identified in *DDHD2*: c.1804_1805insT in combination with c.2057delA, and c.1386dupC in combination with c.1978G>C. Sequencing of a follow-up cohort consisting of 55 individuals with presumed autosomal recessive HSP resulted in the identification of a homozygous frameshift, c.1546C>T, and a homozygous stop mutation, c.859C>T, within *DDHD2* in two more families. The first mutation was identified in a large consanguineous Omani family, with linkage to a 9cM interval on chromosome 8p encompassing *DDHD2*. The second family consisted of an affected male from a consanguineous Iranian family. The clinical presentation of these additional families was very similar to the two original families in our study with early onset HSP, ID, thin corpus callosum and white matter abnormalities. All mutations affected the DDHD domain of the protein that is essential for its phospholipase activity. qPCR analysis of mRNA expression levels of *DDHD2* in a panel of different human tissues shows highest expression in adult human brain tissue. The function of DDHD2 as a PLA1 suggests that the peak observed by cerebral MRS imaging might be caused by an ineffective hydrolysis of phospholipids. We show that mutations in *DDHD2* cause a specific HSP-ID-structural brain abnormalities phenotype, thereby linking a member of the phospholipase A1 family to neurodevelopmental disease. Moreover, we observed an unusual peak with MRS imaging, likely caused by ineffective phospholipid hydrolysis, that presents a useful diagnostic measure to distinguish the *DDHD2*-phenotype from other HSP phenotypes.

161

Lipidomics of Gaucher disease: Substrate composition and nature is dependent on tissue/region and acid β-glucosidase mutations - phenotypic implications. *Y. Sun[1,3], W. Zhang[2], Y. Xu[1,3], B. Quinn[1], N. Dasgupta[1], B. Liou[1], K.D.R. Setchell[2,3], G.A. Grabowski[1,3]*. 1) Div Human Gen, Cincinnati Children's Hosp; 2) Div Pathology and Lab Med, Cincinnati Children's Hosp; 3) Dept Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH.

Gaucher disease results from mutations in *GBA1* that leads to defective acid β-glucosidase (GCase)-mediated cleavage of glucosylceramide (GC) and glucosylsphingosine as well as the heterogeneous manifestations in the viscera and CNS. To investigate the substrate composition and nature in Gaucher disease, the mutation, tissue, and age dependent accumulations of different GC species were characterized by LC-MS/MS in the mice with *Gba1* missense mutations alone or with isolated saposin C deficiency (C*). Of those mutants, D409V mutation with one null allele (9V/null) led to GC species accumulation primarily in the visceral tissues with preferential accumulations of GC24:0 in the lung and GC16:0 in the liver, but no preferential splenic and brain storage. The progressive and preferential age-dependent accumulations of different GC species were found in D409V/null visceral tissues. In combination with C*, the V394L mutation (4L;C*) had significant defects of GC18:0 degradation in the brain, whereas the D409H mutation with C* (9H;C*) led to visceral substrate accumulation including all GC species. Glucosylsphingosine was poorly degraded by the V394L and D409H GCases in the brain, and by the D409V GCase in the visceral tissues. The N370S mutation did not cause significant substrate accumulation in the newborn mice. These results demonstrate age, tissue/regional, and mutation-specific quantitative differences in GC species and glucosylsphingosine accumulation. Such tissue differential GC species and glucosylsphingosine accumulations influence the mutational-dependent tissue/regional expression of Gaucher disease phenotypes.

162

Sterol precursors induce Niemann-Pick C disease phenotypes in Smith-Lemli-Opitz Syndrome causing defective LDL-cholesterol utilization that is corrected by imino-sugars. *C.A. Wassif[1,3], E. Lloyd-Evans[2], L.J. Haslett[2], I.M. Williams[3], C.L. Toth[1], F.D. Porter[1], F.M. Platt[3]*. 1) PDGEN, NICHD/HDB/NIH, Bethesda, MD; 2) Cardiff School of Biosciences, University of Cardiff, CF10 3AX; 3) Department of Pharmacology, University of Oxford, Oxford, OX1 3QT, UK.

Smith-Lemli-Opitz syndrome (SLOS), is an autosomal recessive multiple malformation, cognitive impairment disorder of cholesterol synthesis, caused by mutations in the 3β-hydroxysterol Δ7-reductase gene (DHCR7). DHCR7 catalyzes the conversion of 7-dehydrocholesterol (7-DHC) to cholesterol. This primary biochemical defect leads to a bona fide induction of a Niemann Pick Type C (NPC) disease phenotype, characterized by storage of unesterified cholesterol within the lysosome and failure to deliver LDL-cholesterol to the ER. The inhibition of the NPC pathway correlates with increasing concentrations of 7-DHC. Similar to NPC1 cells, sphingosine storage is an initiating factor in disrupting cholesterol transport in SLOS. Accumulation of sphingosine in turn inhibits lysosomal calcium entry, which disrupts NAADP signaling and endocytosis. Cellular expression of NPC1 increases as levels of 7DHC rise in what appears to be a compensatory attempt to restore intracellular trafficking. Human and mouse SLOS fibroblast alike demonstrated significant increases of glycosphingolipids (GSLs) and sphingosine. These findings are specific to SLOS, and are not observed in fibroblasts with cholesterol synthesis defects such as lathosterolosis and desmosterolosis. This inhibition of NPC1 function also appears *in vivo* as demonstrated by finding significantly elevated levels of sphingosine in both the liver and cortex of the null SLOS mouse model at E18.5 and the cortex of the hypomorphic T93M/null model at multiple ages. These finding also translate clinically, as SLOS patients have elevated sphingosine and GSL levels in their cerebrospinal fluid (CSF). This increase in CSF sphingosine correlates whith both residual DHCR7 activity and phenotypic severity. Thus, decreased bioavailability of cholesterol due to the NPC-like defect may contribute to SLOS pathology. This data suggested that miglustat a GSL synthase inhibitor that is known to cross the blood brain barrier and is used in treatment of NPC1 patients, may be of therapeutic benefit in SLOS patients. To test this hypothesis we treated fibroblasts with miglustat to observe any beneficial effect on the NPC phenotypes. Treatment of these cells with miglustat prevented the storage of GSLs and reduced the storage of sphingosine. Lysosomal accumulation of sphingomyelin in SLOS cells was redistributed back to the plasma membrane after treatment. These data suggest that miglustat therapy may be beneficial in SLOS.

163

Glucose kinetics in subjects with MELAS syndrome: interim results. *L. Emrick[1], A. El-Hattab[2], J. Hsu[3], F. Jarhoor[3], F. Scaglia[1], W. Craigen[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Division of Medical Genetics, Department of Child Health, University of Missouri Health Care, Columbia, MO; 3) US Department of Agriculture/ Agricultural Research Service Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX.

Background: The mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) syndrome is a maternally inherited mitochondrial disorder in which diabetes mellitus (DM) occurs in one third of affected individuals. The pathogenesis of DM in MELAS syndrome remains unclear. We hypothesize that DM develops in individuals with MELAS syndrome due to multiple defects in glucose metabolism, including decreased glucose utilization, increased glucose production, decreased insulin secretion, and increased insulin resistance. Individuals with MELAS syndrome who do not have DM may have altered glucose metabolism. **Methods:** We aimed to measure the rates of endogenous glucose production, gluconeogenesis, glucose oxidation, and glucose clearance via stable isotope infusion technique in subjects with MELAS syndrome with and without DM and in healthy control subjects. We measured the concentrations of fasting blood glucose, insulin, HbA1c; and assess insulin resistance using the Homeostatic Model Assessment (HOMA). After a 12 hour-fast subjects undergo an isotope infusion with a priming dose of $NaH^{13}CO_3$ and U-$^{13}C_6$ glucose followed by continuous infusion of U-$^{13}C_6$ glucose for 6 hours. Blood and breath samples were collected and analyzed for isotopic enrichments. **Results:** To date, 6 control subjects, 5 subjects with MELAS and DM, and 6 subjects with MELAS without DM have completed the study. Both groups of subjects with MELAS show increased glucose production and gluconeogenesis rates when compared to the control subjects. Subjects with MELAS and DM exhibit higher insulin resistance as calculated by HOMA, whereas subjects with MELAS without DM show a higher rate of glucose clearance. **Conclusions:** This interim analysis reveals that subjects with MELAS syndrome have abnormalities in glucose metabolism. Subjects with MELAS who do not have DM have higher rates of glucose production and gluconeogenesis that can predispose them to develop diabetes. Subjects with MELAS and diabetes showed both increased glucose production and higher insulin resistance, confirming that DM develops due to multiple defects in glucose metabolism in MELAS. The results of this study provide a better understanding of the pathophysiological mechanisms of DM in subjects with MELAS syndrome, which can influence the management and prognosis of the disorder and may provide further insights into the pathogenesis of DM in mitochondrial diseases in general.

164

Phenylbutyrate therapy for pyruvate dehydrogenase complex deficiency. *R. Ferriero[1], E. Lamantea[2], P.W. Stacpoole[3], L. Bonafe[4], B. Lee[5], M. Zeviani[2], N. Brunetti-Pierri[1,6].* 1) Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Unit of Molecular Neurogenetics Besta Institute, Milan, Italy; 3) Departments of Medicine (Division of Endocrinology and Metabolism) and Biochemistry and Molecular Biology, College of Medicine, University of Florida, Gainesville, Florida; 4) Division of Molecular Pediatrics, CHUV, Lausanne, Switzerland; 5) Department of Molecular and Human Genetics Baylor College of Medicine, Houston, Texas; 6) Department of Pediatrics, Federico II University of Naples, Naples, Italy.

Deficiency of pyruvate dehydrogenase complex (PDHC) is the most common disorder leading to lactic acidemia. Phosphorylation of specific serine residues of the E1-alpha subunit of the PDHC by pyruvate dehydrogenase kinase (PDK) inactivates the enzyme, whereas dephosphorylation restores PDHC activity. Phenylbutyrate prevents phosphorylation of the E1-alpha subunit of the branched-chain ketoacid dehydrogenase complex (BCKDC) and reduces plasma concentrations of neurotoxic branched chain amino acids in patients with maple syrup urine disease (MSUD), due to the deficiency of BCKDC. We hypothesized that, similarly to BCKDC, phenylbutyrate enhances PDHC enzymatic activity by increasing the portion of unphosphorylated enzyme. We found that wild-type human fibroblasts treated with phenylbutyrate have reduced levels of phosphorylated E1-alpha compared to untreated cells. To investigate the effect of phenylbutyrate in vivo, we administered phenylbutyrate to C57B6/L wild-type mice and we detected a significant increase in Pdhc enzyme activity and a reduction of phosphorylated E1-alpha subunit in brains, muscles, and livers compared to saline-treated mice. Next, we showed that phenylbutyrate increases PDHC activity in fibroblasts from PDHC-deficient patients. Being a drug already approved for human use, phenylbutyrate has potential for increasing the residual enzymatic activity of PDHC and improving the phenotype of PDHC deficiency.

165

Etiologies for Neurocognitive delays in Argininosuccinic Aciduria. *A. Erez[1], S. Nagamani[1], P. Campeau[1], O. Shchelochkov[1,2], J. Kho[1], K. Bissig[2], Q. Sun[1], N. Bryan[3], S. Cederbaum[4], B. Lee[1,5].* 1) Baylor College of Medicine, Houston, TX; 2) University of Iowa; 3) UT Houston, TX; 4) UCLA; 5) Howard Hughes Medical Institute.

Argininosuccinic aciduria (ASA) is caused by the deficiency of argininosuccinate lyase (ASL), the only enzyme able to generate endogenous arginine. All subjects with ASA are supplemented with arginine, while some need additional therapy with nitrogen scavengers. In spite of early diagnosis and initiation of therapy, patients with ASL deficiency have a higher incidence of neurocognitive abnormalities as compared to subjects with other urea cycle disorders, for which the cause is unclear. Recently, we reported that ASL is essential outside its catalytic role, for the assembly of a novel complex necessary for utilization of arginine for nitric oxide (NO) synthesis. Thus, subjects with ASA can have NO deficiency in spite of arginine supplementation. Since NO has an established role as a modulator of neuronal function, we hypothesized that NO deficiency contributes at least in part to the cognitive delays in ASA. In addition, in ASA, the elevation of argininosuccinic acid and the potential depletion of arginine metabolites like guanidinoacetate and creatine, could also contribute to neurotoxicity. As of yet, no study has been able to dissect the contribution of these altered metabolite(s) to the neurocognitive phenotype. To systematically analyze the levels of arginine metabolites in the neuronal tissue, we studied gene-therapy treated ASA mice, and generated induced pluripotent stem cells from ASA patients and control subjects. Our preliminary results support the notion that both increase in ASA and decreased arginine levels persist in the brain after correction of hepatic ureagenesis defect supporting the tissue specific requirement of ASL. Furthermore, we were able to show that the neuronal-NOS (NOS1) forms a complex with Asl and that this complex assembly is less efficient with Asl deficiency. In addition, we were able to show that adding NO donors to the conventional treatment of ASA, improved nitrosylation in the brain of ASA mice. Importantly, in a proof of concept study, treatment with NO donor in an ASA subject was associated with improvement in neuropsychological parameters. In summary, our preliminary results support a multifactorial etiology for the increased neurocognitive delay observed in ASA, involving NO deficiency and increased argininosuccinic acid levels with possible depletion of arginine metabolites. Our study is significant as it has clinical implications relevant to optimization of the current treatment modalities for ASA.

166

Enzyme substitution therapy for phenylketonuria delivered orally using a genetically modified probiotic: proof of principle. *J. Christodoulou[1,2], N. Al-Hafid[2], X.-Z. Tong[2], K. Carpenter[1,3], V. Wiley[1,4], S. Cunniingham[5], I.E. Alexander[1,5].* 1) Western Sydney Genetics Program, Children's Hospital at Westmead, Westmead, NSW, Australia; 2) Genetic Metabolic Disorders Research Unit, Children's Hospital at Westmead, Westmead, NSW, Australia; 3) NSW Biochemical Genetics Service, Children's Hospital at Westmead, Westmead, NSW, Australia; 4) NSW Newborn Screening Programme, Children's Hospital at Westmead, Westmead, NSW, Australia; 5) Gene Therapy Research Unit, Children's Hospital at Westmead, Westmead, NSW, Australia.

Phenylketonuria (PKU), caused in most cases by functional defects of the liver-specific enzyme, phenylalanine hydroxylase (PAH), is one of the commonest treatable genetic metabolic disorders worldwide. Dietary phenylalanine (phe) restriction remains the mainstay of therapy for most, although up to a third of individuals may be at least partially responsive to pharmacological doses of the PAH cofactor tetrahydrobiopterin (BH4). Unpalatability of the diet, the lack of uniform responsiveness to BH4 and its cost, justify the development of novel approaches to therapy. Phenylalaine ammonialyase (PAL) is a simple enzyme, requiring no cofactors, which can effectively catabolise phe to nontoxic transcinnamate. Currently clinical trials of an injectable pegylated form of PAL are under way. We are developing an alternate approach to the delivery of functional PAL, administered orally using a genetically modified probiotic. We have developed an expression vector, inducible with nisin, which expresses codon optimised yeast PAL. On transduction into a strain of Lactococcus lactis we demonstrate in vitro PAL enzyme activity reaching up to 20% that of commercially available purified PAL. In further in vitro experiments, we have demonstrated that this genetically modified (GM) organism is capable of effectively catabolising phe that has been delivered as part of intact polypeptides, and survives the gastric and small intestinal milieu. Moreover, using stable-isotopically labelled phe administered orally along with our GM probiotic to wildtype and Pahenu2 mice, the latter being a well-established model of PKU, we have demonstrated that the GM probiotic completely prevents the uptake of labelled phe from the gut into the blood stream in wildtype mice, and significantly reduces uptake in the mutant mice. These experiments provide proof of principle that orally administered PAL delivered in a GM probiotic may be a viable alternative approach to therapy for PKU. In ongoing work, we are refining the system to provide improved functional stability of the PAL enzyme and improved viability of the GM probiotic in the gut. We also plan to explore whether chronic administration of the GM probiotic leads to a sustained reduction of blood phe in the PKU mouse model.

167

A new inborn error of manganese metabolism caused by mutations in *SLC30A10*, a newly identified human manganese transporter. *K. Tuschl[1], P.T. Clayton[1], S.M.Jr. Gospe[2], S. Gulab[3], S. Ibrahim[3], P. Singhi[4], R.T. Ribeiro[5], M.S. Zaki[6], M. Luz del Rosario[7], S. Dyack[8], V. Price[8], R.A. Wevers[9], P.B. Mills[1]*. 1) Clinical and Molecular Genetics Unit, ICH Institute of Child Health, London, United Kingdom; 2) Departments of Neurology and Pediatrics, University of Washington and Seattle Children's Hospital, Seattle, United States; 3) Department of Pediatric Neurology, Aga Khan University Hospital, Karachi, Pakistan; 4) Department of Pediatrics, Postgraduate Institute of Medical Education and Research, Chandigarh, India; 5) Department of Neurology, Federal University of Sao Paulo, Sao Paulo, Brazil; 6) Clinical Genetics Department, National Research Center, Cairo, Egypt; 7) Department of Pediatrics, St. Lukes Medical Center, Quezon City, Philippines; 8) Department of Pediatrics, IWK Health Centre, Halifax, Canada; 9) Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Background We have identified an autosomal recessively inherited disorder of manganese metabolism that leads to manganese accumulation in liver and brain with characteristic MRI appearances of hyperintense basal ganglia on T1-weighted sequences. Most affected individuals present in childhood with difficulties walking and fine motor impairment due to dystonia or spastic paraparesis. Movement disorder is accompanied by liver cirrhosis and some patients have died at young age following cirrhosis complications. An adult-onset form of Parkinsonism associated with hepatomegaly and hypermanganesemia has also been described. Further hallmark features include polycythemia and features of iron depletion such as low ferritin levels and increased total iron binding capacity. Chelation therapy with disodium calcium edetate and iron supplementation lead to significant improvement of clinical symptoms and stabilization of blood manganese levels. **Methods** In order to identify the affected gene, we performed homozygosity mapping on two consanguineous families using an Illumina CytoSNP-12. The candidate gene was sequenced on an ABI sequencer in nine affected families. The function of the wild-type protein and the effect of sequence changes were studied in the manganese-sensitive yeast strain Δpmr1 using Gateway technology (Invitrogen). **Results** Homozygosity mapping identified *SLC30A10* as the affected gene, and homozygous sequence changes were found in all affected individuals. Previously, human *SLC30A10* was thought to belong to a class of zinc transporters. However, expression of human wild-type *SLC30A10* in the Δpmr1 yeast strain rescued growth in high manganese conditions confirming its role in manganese transport. The presence of missense and nonsense mutations in *SLC30A10* failed to restore manganese resistance. Evidently, evolutionary changes in the amino acid sequence of the protein have altered the substrate specificity of the transporter from zinc in yeast to manganese in mammalian cells. **Conclusion** SLC30A10 is the first recognized human manganese transporter that, when defective, causes a syndrome of hepatic cirrhosis, dystonia, polycythemia and hypermanganesemia. Chelation therapy with disodium calcium edetate and iron supplementation provide effective treatment. Hence, whole blood manganese determination should be part of the diagnostic work-up of dystonia and Parkinsonism, particularly when accompanied by hepatomegaly and polycythemia.

168

Combined methylmalonic acidemia and homocystinuria, cblC type: a prospective clinical protocol focusing on neurologic and neurodevelopmental parameters in a cohort of pre-school children diagnosed on expanded newborn screening. *J.D. Weisfeld-Adams[1,2], H.A. Bender[3], A.M. Akerstedt[3], E. Miles-Mason[3], T.P. Naidich[4], S. Lipson[5], T. Bottiglieri[6], S.P. Young[7], G.A. Diaz[1,2]*. 1) Department of Genetics & Genomic Sci, Mount Sinai School of Medicine, New York, NY; 2) Department of Pediatrics, Mount Sinai School of Medicine, New York, NY; 3) Department of Psychiatry, Mount Sinai School of Medicine, New York, NY; 4) Department of Radiology, Mount Sinai School of Medicine, New York, NY; 5) Department of Anesthesiology, Mount Sinai School of Medicine, New York, NY; 6) Neuropharmacology Laboratory, Baylor Research Institute, Dallas, TX; 7) Biochemical Genetics Laboratory, Duke University Medical Center, Durham, NC.

Combined methylmalonic acidemia and homocystinuria, cblC type, results from mutations in MMACHC, a gene critical for production of methylcobalamin and adenosylcobalamin, essential co-factors for methionine synthase and methylmalonyl-CoA mutase, respectively. Early-onset cblC disease is associated with disordered white matter development on MRI, and developmental delays and visual deficits are common. Proposed pathophysiologic mechanisms contributing to the phenotype include homocysteine-mediated vascular endothelial injury or neurotoxicity, bioenergetic strokes from methylmalonic acid (MMA)-mediated mitochondrial toxicity, oxidative stress from reactive oxygen species (ROS), depletion of brain methionine/S-adenosylmethionine (SAM)/creatine, and toxic accumulation of guanidinoacetate (GAA). We developed a comprehensive clinical protocol to prospectively examine various factors related to neurodisability in a cohort of preschool children (n=12) with molecularly-confirmed cblC, all of whom were diagnosed on newborn screening and treated from early infancy. Most participants completed all parts of the protocol, which, in addition to routine care, included neuropsychologic evaluation, brain MRI and MR spectroscopy with diffusion-weighted and diffusion-tensor imaging, evaluation of homocysteine remethylation metabolites in plasma, and health-related quality of life assessment. Patients exhibited a spectrum of genotypes and phenotypes. Notable radiologic findings include a high prevalence of developmental white matter abnormalities, especially callosal hypoplasia, while fewer patients than anticipated had basal ganglia abnormalities. Neuropsychologic evaluation demonstrated delays in a range of spheres, and subclinical abnormalities were demonstrated in some high-functioning children. Mean ± SD values for plasma metabolites were homocysteine 43.7 ± 12.7 μmol/L (nl <15), methionine 26.1 ± 10.8 μmol/L (nl 14–46), S-adenosylhomocysteine 40.5 ± 21.4 nmol/L (nl 13.2–18.2), SAM 149.4 ± 71.4 nmol/L (nl 33.0–94.8), MMA 2.6 ± 1.4 μmol/L (nl <0.6), GAA 0.9 ± 0.2 μmol/L (nl 0.1–1.7), creatine 51.8 ± 17.9 μmol/L (nl 3–114). GAA has been previously reported as markedly abnormal in cblC, but normal levels on LC-MS/MS testing in our cohort suggest that mechanisms other than GAA are responsible for cblC-related neurotoxicity. These findings expand understanding of cblC-related neurologic phenotypes, and provide further insights into their possible underlying mechanisms.

169

Whole-exome sequencing of 10,000 type 2 diabetes cases and controls from five major ancestry groups. *T.M. Teslovich[1], A.P. Morris[2], P. Fontanillas[3], M.A. Rivas[2], X. Sim[1], J. Flannick[3], N. Burtt[3], H. Chen[4], A.G. Day-Williams[5,6], A. Mahajan[2], G. Atzmon[7], P. Cingolani[8], L. Moutsianas[2], H.M. Highland[9], T2D-GENES Consortium.* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 4) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 6) Biogen Idec, Cambridge, MA, USA; 7) Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY, USA; 8) Department of Computer Science, McGill University and Genome Quebec, Montreal, QC, Canada; 9) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

As part of the Type 2 Diabetes (T2D) Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) study, we are performing whole-exome sequencing (~18k genes) of 10,000 individuals from five major ancestry groups: African American (AA), East Asian (EA), European, Hispanic, and South Asian (SA). The unique study design will yield a catalog of common, low-frequency, and rare coding variation across diverse populations. Using single variant and burden tests, we aim to: 1) identify novel T2D associations; 2) assess the evidence for heterogeneity in genetic effects between ancestry groups; and 3) localize underlying causal variants for T2D. In addition to exome-wide analysis, we are focusing fine mapping efforts in regions of known association from genome-wide association studies (GWAS) of T2D and related traits (e.g. fasting glucose and insulin). We hypothesize that genes underlying the common signals identified in GWAS also harbor low-frequency and rare variants that contribute to T2D risk. We seek to identify causal genes at GWAS loci by analyzing such variants, and subsequently to determine whether low-frequency and/or rare variation explain the common variant association signal. Sequencing, variant calling, and variant quality control are underway at the Broad Institute using standard pipelines. Association analyses have been implemented by the T2D-GENES analysis committee. A first data release of ~3500 samples from AA (N= 1024), EA (N=1604), and SA (N=852) cohorts has identified ~1.6M single nucleotide variants (SNVs) in the exome, 1.1M (71.5%) of which are novel (absent from the 1000 Genomes Phase I release). Within the dataset, ~555k (35.4%), 481k (30.6%), and 339k (21.6%) variants are unique to AA, EA, and SA samples, respectively, and only ~89k (5.6%) variants are present in all three groups. In 3500 samples, we have 80% power to detect variants with MAF=1% and OR=3, and 99.9% power to detect variants with OR=4 (α=5×10^{-8}). To date, in the exome, we have identified no novel common variants or low-frequency, highly penetrant single nucleotide variants associated with T2D. Results presented at the meeting will be based on analysis of SNVs, indels, and CNVs typed in >5k samples representing each of the five ancestry groups. Sequencing will be complete in late 2012. We anticipate that analysis of the full dataset will lead to identification of causal genes and variants and give insight into the genetic architecture of T2D.

170

Rare and low frequency coding variants are associated with LDL cholesterol levels: findings from the NHLBI Exome Sequencing Project. *L.A. Lange[1], Y. Hu[2], C. Xue[3], Z. Tang[4], C. Bizon[5], E.M. Lange[1, 4], J.D. Smith[6], E.H. Turner[6], G. Jun[2], H.-M. Kang[2], K.P. Li[4], G.M. Peloso[7, 14], C.L. Wassel[8], A.P. Reiner[9], E. Boerwinkle[10], B.M. Psaty[11], C.J. O'Donnell[12, 13], S. Kathiresan[13, 14, 15], K.E. North[16], D. Lin[4], G.P. Jarvik[17], L.A. Cupples[12, 18], C. Kooperberg[19], J.G. Wilson[20], D.A. Nickerson[6], G.R. Abecasis[2], S.S. Rich[21], R.P. Tracy[22], C.J. Willer[3] on behalf of the NHLBI Exome Sequencing Project.* 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 3) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI; 4) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 5) Renaissance Computing Institute, Chapel Hill, NC; 6) Department of Genome Sciences, University of Washington, Seattle, WA; 7) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 8) Division of Preventive Medicine, Department of Family and Preventive Medicine, University of California San Diego, La Jolla, CA; 9) Department of Epidemiology, University of Washington, Seattle, WA; 10) Human Genetics Center, Health Science Center, University of Texas, Houston, TX; 11) Department of Medicine, Cardiovascular Health Research Unit, University of Washington, Seattle, WA; 12) National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA; 13) Harvard Medical School, Boston, MA; 14) Broad Institute, Cambridge, MA; 15) Massachusetts General Hospital, Boston, MA; 16) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 17) Department of Medicine, University of Washington Medical Center, Seattle WA; 18) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 19) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 20) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS; 21) University of Virginia School of Medicine, Charlottesville, VA; 22) Departments of Pathology and Biochemistry, University of Vermont, Colchester, VT.

Elevated low-density lipoprotein cholesterol (LDL-C) is a treatable, heritable risk factor for cardiovascular disease, the leading cause of death in industrialized countries. To determine whether rare and/or low-frequency coding variants are associated with blood LDL-C levels, we performed exome sequencing and analysis of 2005 European-American (n = 1151) and African-American (n = 854) samples from seven population based cohort studies (Atherosclerosis Risk in Communities; Coronary Artery Risk Development in Young Adults; Cardiovascular Health Study; Framingham Heart Study; Jackson Heart Study; Multi-Ethnic Study of Atherosclerosis; Women's Health Initiative). Among sequenced individuals, 412 were selected because of extreme LDL-C levels (< 2nd percentile and > 98th percentile), and 1593 individuals were selected for studies of other phenotypes or as part of a deeply phenotyped random sample. We performed race-stratified gene burden tests for association between LDL-C and rare variants using the Combined Multivariate Collapsing (CMC) test (Li & Leal 2008) with minor allele frequency cut-offs of 5%, 1% and 0.1%, and meta-analysis of the race-specific results. Analyses were conducted separately (1) for all nonsynonymous variants and (2) for loss of function (defined as nonsense, splice and read-through) variants. Analysis limited to the 412 extreme LDL-C samples did not identify gene burden associations meeting exome-wide significance (p < 1×10-6). However, in the total sample of 2005 samples, we observed exome-wide significant evidence for association between LDL-C and nonsynonymous rare variants in *LDLR* (MAF < 0.1%; p = 3×10-9), low frequency nonsynonymous variants in *PCSK9* (MAF < 5%; p = 2×10-8), and singleton loss of function variants in *APOB* (p = 8×10-8). Among five other genes implicated in Mendelian LDL-C related phenotypes, two genes (*NPC1L1* and *ABCG5*) had suggestive evidence for association (p < 4×10-4). Our results suggest that no single rare variant association test is ideal for detecting all relevant genes, as the genetic architecture underlying association at each gene differed based on variant frequency and effect on protein function. This study provides important insight regarding study design and analytic approach for detecting association with rare variants for complex traits.

171

Exome sequencing of extreme phenotypes identifies DCTN4 and CAV2 as modifiers of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *M.J. Emond[1], T. Louie[1], J. Emerson[2, 3], S. McNamara[2], W. Zhao[1], R.A. Mathias[4], M.R. Knowles[5], F.A. Wright[6], M.J. Reider[7], H.K. Tabor[2, 8], D.A. Nickerson[7], K.C. Barnes[4], R.L. Gibson[2, 9], M.J. Bamshad[2, 7, 10]*. 1) Biostatistics, University of Washington, Seattle, WA; 2) Department of Pediatrics,University of Washington, Seattle, Washington, USA; 3) Center for Clinical and Translational Medicine, Seattle Children's Research Institute Seattle, Washington, USA; 4) Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA; 5) Cystic Fibrosis/ Pulmonary Research and Treatment Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; 6) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; 7) Department of Genome Sciences, University of Washington, Seattle, Washington, USA; 8) Trueman-Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, Washington, USA; 9) Division of Pulmonary Medicine, Seattle Children's Hospital, Seattle, Washington, USA; 10) Division of Genetic Medicine, Seattle Children's Hospital, Seattle, Washington, USA.

Evidence indicates that genetic factors beyond CFTR genotype have a strong influence on age at onset of chronic Pseudomonas aeruginosa (Pa) infection among individuals with CF. As part of the NHLBI Exome Sequencing Project (ESP), we sought to use whole exome sequencing to identify genetic variants associated with differential risk for chronic Pa infection among individuals from the EPIC Observational Study. Individuals in the early and late extremes for age of onset of chronic Pa infection were selected for exome sequencing (n=48 each). For each gene, we tested for a significant difference between phenotypic extremes in the proportion of individuals with rare, nonsynonymous variants while adjusting for confounders. In secondary analyses taking advantage of additional ESP exomes, we used small-sample-adjusted SKAT-O to compare the small number of exomes from children with early age of onset of chronic Pa to 2457 exomes from non-CF European American controls. In the extreme phenotype study, one gene, DCTN4 (dynactin 4), was found to be significantly associated with phenotype, with $p<0.05$ after correction for multiple testing (naïve $p=2.2\times10\text{-}6$). Twelve individuals carried rare variants at either of two sites in the early chronic extreme (rs11954652 or rs35772018) vs none in the late extreme. In the secondary analysis, CAV2 (caveolin-2) was the most significant gene ($p=8.7\times10\text{-}6$) after CFTR. We then sequenced DCTN4 and performed exome chip genotyping in 667 remaining EPIC OBS individuals and used Cox regression to validate associations between age at onset of chronic Pa infection and variants in DCTN4 and CAV2. Presence of either DCTN4 variant was associated with a significant increase in the hazard for chronic Pa: HR = 1.9, 95% CI=(1.2, 2.7), p=0.001. Individuals who were homozygous for rs11954652 (n=9) or heterozygous for rs35772018 (n=13) were at even higher risk: HR=3.7, (1.3, 10.2), p=0.02; and HR = 8.6, (2.4, 30), p=0.0009, respectively. On the other hand, variants at rs8940 in CAV2 had a protective effect with HR = 0.58 (95% CI=[0.35, 0.97], p=0.038), a finding consistent with experimental results for showing decreased Pa invasion in CAV2-deficient murine lung epithelial cells (Zaas, et al, 2006). These results provide strong evidence that DCTN4 and CAV2 are genetic modifiers of Pa susceptibility in CF and illustrate the utility of exome sequencing as a tool for identifying novel genetic modifiers of this complex phenotype.

172

A high resolution study of Type 2 Diabetes genetic architecture through whole-genome sequencing of 2850 European individuals: the Go-T2D Study. *J. Flannick[1], C. Fuchsberger[2], K.J. Gaulton[3], N.P. Burtt[1], H.M. Kang[2], C. Hartl[1], R.D. Pearson[3], the Go-T2D Consortium.* 1) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

Over the past decade, improvements in genotyping technology have enabled studies to explore the genetic architecture of Type 2 Diabetes (T2D) with increasing resolution. Recently, Genome Wide Association Studies have identified ~70 regions associated with T2D risk and produced valuable insights into disease etiology, but these studies focus on common (MAF>5%) SNPs and explain a limited proportion of the genetic variance. The Genetics of Type 2 Diabetes (Go-T2D) study aims to characterize T2D genetic architecture with unprecedented resolution through low-coverage (4x) whole-genome sequencing, deep (~70x) exome sequencing, and 2.5M SNP genotyping of 1,425 cases and 1,425 controls from Europe, with propagation of variants (via imputation) into 30k additional samples and testing of coding SNPs (via the Exome chip) in ~50k additional samples.

Of the targeted 2,850 samples, we successfully sequenced 2,772 whole-genome, sequenced 2,760 whole-exome, and genotyped 2,805 at 2.5M SNPs. Using a dataset of the first 908 samples, we developed methods to integrate the technologies into a single data set and test rare variants for association with T2D and related traits. Applying these methods to the 2,760 sample whole-exome dataset, we (a) observe most variants to be rare (e.g. 363,804 singletons out of 723,057 SNPs) but most variants within an individual to be common (e.g. 135 singletons out of 34,240 SNPs); (b) find no MAF <5% single variants with genome-wide significant ($p<5e\text{-}8$) T2D associations; and (c) find no fully-penetrant (e.g. case-unique) variant to have case frequency >0.6%. We also (d) identify several genes with an excess of loss of function variants in cases; (e) observe genes that cause Mendelian diabetic phenotypes to harbor an excess of coding MAF<1% variants in cases (p=.006), including 8 case-unique variants in PPARG (p=.005); and (f) find associations (p<.05) between rare variation in T2D-related transcription factor targets and T2D-related traits (e.g. PDX1 targets with fasting glucose, PPARG targets with BMI and fasting insulin).

Our dataset will complete by September, with a near-comprehensive catalog of variation in our samples and 80% power to detect a MAF .5% coding variant of relative risk 2 (at $\alpha=5e\text{-}8$). This work enables the highest resolution study of T2D genetic architecture to date as well as a comparison of methods and technologies to find risk variants for complex traits in general.

173

Mapping quantitative traits with integrated whole exome/genome/array panel in individuals of European descent. *X. Sim[1], M.A. Rivas[2], A.K. Manning[3,4,5], A.E. Locke[1], C.M. Lindgren[2], GoT2D Consortium.* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Michigan, MI, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, MA, USA; 4) Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, MA, USA; 5) Department of Genetics, Harvard Medical School, Boston, Massachusetts, MA, USA.

There has been tremendous progress in identifying common genetic variants contributing to human diseases and traits using genome-wide association studies. However, the majority of the heritability remains unexplained, likely due in part to inadequate understanding of the casual variants at these established loci and also to undetected less common variants. The GoT2D (Genetics of Type 2 Diabetes) Project aims to address these issues through a combination of sequencing (low-pass whole genome and deep whole exome) and genotyping (Omni2.5M) technologies in 2,850 European type 2 diabetes (T2D) cases and controls, and exome-chip genotyping across ~50,000 samples. This strategy allows us to discover novel loci, fine-map coding regions and established loci, and obtain high quality genotypes across a broad allele frequency spectrum genome-wide. Here, we focus on T2D related quantitative traits: glycemic, anthropometric, and lipids to gain new insights into the pathophysiology of metabolic disease. For each quantitative trait, we test the following hypotheses using a combination of single variant and burden tests: (i) single, low-frequency variants are associated with quantitative traits at novel or GWAS-implicated loci; (ii) novel genes haboring rare variants contribute to phenotypic variance. Analysis of exome-sequence data for lipid traits (LDL, HDL, total cholesterol, and triglycerides) identifies single variant associations at known GWAS loci. These include common missense variants at *APOE* (*P*=2.9e-12) with LDL, *GCKR* with TG (*P*=1.3e-08), and a low-frequency missense variant in *PCSK9* (*P*=2.7e-06) with LDL. In addition, analysis of protein altering variants using SKAT reveals association of *APOE* (LDL-*P*=2.4e-10,TC-*P*=6.6e-06) and *KLF4* (LDL-*P*= 4.3e-05, TC-*P*=9.0e-06) with both LDL and total cholesterol; as well as *ADIPOQ* (*P*=7.0e-05) with BMI. Preliminary analysis of severe loss-of-function mutations shows burden-test-based association with fasting insulin at *VWF* (*P*=2.9e-05) and LDL at *TRMT1* (*P*=8.4e-05). The completion of the integrated panel of SNPs, indels, and structural variants will provide a comprehensive survey of the genome for T2D-related quantitative phenotypes with follow-up in exome-chip and allow high resolution imputation in large numbers of European samples with GWAS-level data.

174

Whole genome sequence analyses describe the genetic architecture of complex traits: the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium. *A.C. Morrison[1], A. Voorman[2], A.D. Johnson[3], X. Liu[1], J. Yu[4], A. Li[1], D. Muzny[4], F. Yu[4], K. Rice[2], G. Zhu[5], G. Heiss[6], C.J. O'Donnell[3], B. Psaty[7], L.A. Cupples[3,8], R. Gibbs[4], E. Boerwinkle[1,4].* 1) Human Genetics Center, University of Texas at Houston, Houston, TX; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) NHLBI Framingham Heart Study, Framingham, MA; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 5) Department of Agronomy, Kansas State University, Manhattan, KS; 6) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 7) Cardiovascular Health Research Unit, University of Washington, Seattle, WA; 8) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Whole genome sequencing (WGS) in deeply-phenotyped population-based samples, the next frontier in human genetics, allows investigation of rare and common variation in both protein-coding and regulatory regions. WGS was completed on 962 individuals from the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium and high density lipoprotein cholesterol (HDL-C) was analyzed as a model quantitative trait. Forty-nine percent of the observed variants were novel in comparison to the 1,000 Genomes Project. The concordance rate between our WGS and GWAS genotype data was 99%. The TiTv ratios for intergenic, intronic and exonic genome regions were 2.07, 2.17 and 2.59, respectively. The largest proportion of variants was intergenic and 7% were in annotated functional domains, with non-coding RNAs housing the largest fraction. In aggregate, common variation (MAF >1%) contributed seven times more to the heritability of HDL-C than the fraction attributable to rare variation (MAF<1%). Screening this sample for known mutations causing Mendelian disorders related to dyslipidemia (e.g. Tangier's disease) identified individuals presenting HDL-C values at the extremes of the population distribution. WGS data allow for a global survey of the entire genomic landscape, complemented by annotation-based assessments. Three types of analyses and results were completed across the genome: single common SNP analyses, a sliding window aggregating the contribution of rare variants, and burden tests for a priori annotated regions. Results from two regions of the genome will be shown in detail as examples. In the first example, all three WGS analyses revealed the full array of common variation contributing to HDL-C in the 5' region of CETP, and also highlighted a CEBPβ transcription factor binding site at position 56995236, coinciding with rs1800775, the -629 variant previously suggested as functional via allelic reporter gene and gel shift assays. The second example highlights the strength of the agnostic sliding window approach, which identified a region on chromosome 4 near PARM1. The association signal is in a transcript of uncertain coding potential and a marker of histone and transcription factor activity. These results document the tractable nature of whole genome sequence analysis for revealing the genetic architecture of common complex phenotypes in humans.

175

Genome sequencing and analysis in Autism Spectrum Disorder. *S. Walker[1], A. Prasad[1], L. Lau[1], B. Thiruvahindrapuram[1], B. Fernandez[2], R. Yuen[1], C.R. Marshall[1], E. Fombonne[3], W. Roberts[4], L. Zwaigenbaum[5], P. Szatmari[6], S.W. Scherer[1]*. 1) Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Memorial University of Newfoundland, Disciplines of Genetics and Medicine, St. John's, A1B 3V6, Canada; 3) 3Montreal Children's Hospital and McGill University, Department of Psychiatry, Montreal, H3Z 1P2, Canada; 4) The Hospital for Sick Children, Autism Research Unit, Toronto, M5G 1L7, Canada; 5) University of Alberta, Department of Pediatrics, Alberta, T5G 0B7, Canada; 6) McMaster University, Department of Psychiatry and Behavioural Neurosciences, Hamilton, L8S 4K1, Canada.

With the aim of discovering rare genetic variants associated with Autism Spectrum Disorder (ASD), we are carrying out detailed genomic analysis combining high-resolution microarrays, exome and whole-genome sequence data in a cohort of 1000 Canadian families. Currently we are focusing on 1) examining individuals known to carry potentially pathogenic CNVs with the aim of identifying additional etiologic DNA sequence alterations and 2) studying individuals with complex ASD and extensive dysmorphology or medical complications under the hypothesis that there may be a greater incidence of highly damaging de novo mutations. From the first 300 individuals analysed with exome sequencing, we typically detect some 24,000 single nucleotide variants per individual, of which approximately 1,000 are novel and in coding regions. We have identified numerous distinct, potentially pathogenic sequence changes in genes previously associated with ASD susceptibility for example *NRXN3*, *CDH8* and *ASTN2* that appear to segregate with the phenotype and uncovered *de novo* variants implicating new genes *LYPD6B*, *ARRDC4* and *ASPM*. Rare inherited mutations resulting in likely haploinsufficiency have been discovered in known ASD genes *NRXN1* and *NLGN4X* with direct diagnostic applications for carrier individuals. Furthermore, incidental findings from our study have also precipitated clinical follow-up such as early screening in families with cancer syndromes. Further characterisation of ASD candidate risk genes will be achieved by assessment of allele frequencies in independent case and control populations. Moreover, novel `mutations' are being modeled for their functional characteristics using induced pluripotent stem cell lines. Our combined microarray and sequencing data support a multigenic model for Autism susceptibility and demonstrate that extensive analysis of both genotypes and phenotypes is necessary to further our understanding of complex disorders.

176

Deep whole genome sequencing in pedigrees illuminates the contribution of low frequency and private mutations to the genetic architecture of metabolic quantitative traits. *A.K. Manning[1,2,3], A.R. Wood[4], P. Fontanillas[1], G. Jun[5], P. Cingolani[6], M. Almeida[7], C. Fuchsberger[5], T.D. Dyer[7], M. Rivas[8], K. Gaulton[8], J. Maller[8], J. Curran[7], J. Grunstad[9], T.W. Blackwell[5], T.M Teslovich[5], D.M. Lehman[10], R. Grossman[9], J.M. Laramie[11], S.E. Lincoln[11], M. Boehnke[5], M.I. McCarthy[8,12], T.M. Frayling[4], R. Sladek[6], R. Duggirala[7], J. Blangero[7], G. Abecasis[5], D. Altshuler[1,2,3], T2D-GENES*. 1) Medical and Population Genetics Program, Broad Institute, Cambridge, MA; 2) Harvard Medical School, Boston, MA, USA; 3) Massachusetts General Hospital, Boston, MA, USA; 4) University of Exeter, Exeter, UK; 5) University of Michigan, Ann Arbor, MI, USA; 6) McGill University, Montreal, QC, Canada; 7) Texas Biomedical Research Institute, San Antonio, TX, USA; 8) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford UK; 9) University of Chicago, Chicago, IL, USA; 10) University of Texas Health Science Center at San Antonio, San Antonio, TX, United States; 11) Complete Genomics Inc, Mountain View, CA, USA; 12) Churchill Hospital, Oxford, UK.

Next-generation sequencing association studies of low frequency and private genetic variation may illuminate the biology of complex metabolic quantitative traits (QTs) such as fasting glucose, BMI, triglycerides (TG), LDL, and HDL. In unrelated individuals, each private mutation is observed only once, making it impossible to perform statistical tests of association. By sequencing many individuals from extended pedigrees, the T2D-GENES study provides multiple observations of many private variant and allows us to (1) identify variants influencing metabolic QTs from across the full spectrum of allele frequencies and (2) evaluate the role of private functional mutations on these traits. **Results:** In a preliminary freeze of the T2D-GENES data, we obtained whole genome sequences (60x) for 466 individuals from 20 Mexican American diabetes-enriched pedigrees and imputed into 959 additional family members. We successfully imputed 12M SNPs, including 62K non-synonymous polymorphisms (nsSNPs) We examined nsSNPs in genes that had previously been associated with monogenic lipid traits (18 genes), fasting glucose (16 genes) or BMI (50 genes). Using an approach that accounts for family relatedness, we confirmed associations of common SNPs with the known strongest effects (SNP/gene/trait/P: rs964184/APOA1/TG/5e-9, rs3764261/CETP/HDL/1e-6, rs10830963/MTN1RB/glucose/3e-4), but we did not observe excess association over null expectations for all SNPs or all nsSNPs. For glucose, BMI and lipid genes, we identified 44, 89, and 95 nsSNPs, of which 35, 65 and 60 were low frequency (<5%). We observed significant and suggestive enrichment of associations for the lipid and BMI genes, respectively: for HDL, TG and BMI we observed 9, 10 and 7 associations at $P<0.05$ (4 expected). We further evaluated the role of private loss of function (LOF) mutations by comparing the distribution of association signals for private LOF SNPs (N= 648) to other classes of private genetic variation: non-synonymous (N= 2444) and synonymous (N=1522). We did not see an enrichment of association signal in the class of private LOF SNPs. **Conclusion:** Testing of a large collection of private LOF variants in deeply sequenced large pedigrees failed to show enriched of association with QTs compared to other classes of coding variation. The analyses of the full spectrum of allele frequencies using a deep whole genome sequencing approach did not reveal additional enrichment of nsSNPs associated with metabolic QTs.

177

Whole Genome Sequencing of 2100 Individuals in the founder Sardinian Population. *C. Sidore[1,2,3], S. Sanna[3], A. Kwong[2], H.M. Kang[2], R. Cusano[3], M. Pitzalis[3], M. Zoledziewska[3], A. Maschio[2,3], F. Busonero[3], M. Lobina[3], A. Angius[3], R. Lyons[4], B. Terrier[4], C. Brennan[4], R. Atzeni[5], A. Mulas[3], M. Dei[3], M. Piras[3], S. Lai[5], F. Reinier[5], R. Berutti[5], C. Jones[5], M. Marcelli[5], M. Urru[5], M. Oppo[5], D. Schlessinger[6], G. Abecasis[2], F. Cucca[1,3].*
1) Università degli Studi di Sassari, Sassari, Italy; 2) Center for Statistical Genetics, Ann Arbor, University of Michigan, MI, USA; 3) Istituto di Ricerca Genetica e Biomedica, CNR, Monserrato, Cagliari, Italy; 4) University of Michigan, DNA Sequencing Core, Ann Arbor, MI, USA; 5) Center for Advanced Studies, Research, and Development in Sardinia (CRS4), AGCT Program, Parco Scientifico e tecnologico della Sardegna, Pula, Italy; 6) NIA, Laboratory of Genetics, Baltimore, MD, USA.

Genome-wide association studies have increasingly furthered our understanding of the molecular basis of many complex traits by finding, through genotyping and imputation, loci associated with many different traits. However, studies based on variants present in common genotyping arrays and imputation panels may not capture the fraction of human genome variation that is rare or geographically restricted and unique to specific populations. To advance our understanding of the genetics of a variety of traits in the Sardinian population, we are studying a sample of >6,000 individuals recruited from the population of a cluster of 4 small towns in Sardinia. Using whole genome sequencing, we sequenced DNAs from 2,120 Sardinian individuals enrolled either in this project or in a parallel project on autoimmune diseases, at an average depth of coverage of ~4X. We successfully identified and genotyped >17M single nucleotide polymorphisms (30.6% of them novel, not in dbSNP v135) with an error rate of 0.2% that is expected to decrease further by increasing the sample size (the estimated error rate was 0.5% and 0.3% in previous analyses of 505 and 1146, respectively). To increase the power to detect association, we are using the haplotypes generated by sequencing of these individuals to impute missing genotypes in the remaining ~6000 already genotyped with Immunochip and Metabochip. Strikingly, imputation using our Sardinian reference panel shows greatly increased accuracy when compared to an equal size reference panel of European haplotypes generated by the 1000 Genomes Project (average imputation accuracy, rsqr=0.90 compared to 0.75 for alleles with frequency 1–3%). With a larger reference panel, imputation accuracy of variants with frequency 1–3% reaches 0.94, giving us the possibility of analyzing the rare frequency domain in the Sardinian population. As an example of the advantages of analyzing population specific rare variation, we will discuss the Q39X mutation in the HBB gene, which is common in Sardinia (MAF ~5%) but very rare elsewhere. The variant is associated with a variety of blood phenotypes. For LDL cholesterol, the variance explained by this variant in Sardinia is higher than the variance explained by any of the variants previously found with standard GWAS analysis. Our approach thus increases the power of detecting population specific association.

178

Differential relatedness of African Americans to populations within West Africa. *K. Bryc[1,2,10], A. Williams[1,2,10], N. Patterson[2], S. Musani[3], M. Sale[4,9], W. Chen[5,9], J. Divers[6], M. Ng[7], D.W. Bowden[7], J.G. Wilson[8], D. Reich[1].* 1) Genetics, Harvard Medical School, Boston, MA; 2) Broad Institute, 7 Cambridge Center, Cambridge, MA; 3) Department of Medicine, University of Mississippi Medical Center, Jackson, MS; 4) Dept Medicine and Dept Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA; 5) Department of Public Health Sciences, Division of Biostatistics and Epidemiology, University of Virginia, Charlottesville, VA; 6) Department of Biostatistical Sciences, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC; 7) Center for Diabetes Research, Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC; 8) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS; 9) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 10) These authors contributed equally to this work.

African American genomes have been deeply studied with respect to the variability in their proportions of African and European ancestry, but there has been little success at using genetic data to determine the African source populations. To address these questions, we analyzed genome-wide SNP genotype data from 9,859 African Americans from six cohorts and 4,019 Africans from Sierra Leone, Gambia and Nigeria, and applied two independent methods for inferring ancestry.

First, we modeled each African American as a linear combination of European Americans and several sub-Saharan African groups, using a method that builds a linear ancestry model from population allele frequencies, and can obtain accurate estimates of ancestry proportion assuming that one has access to accurate surrogates for the true ancestral populations. However, a limitation is that it is not possible to disentangle which African populations are directly ancestral to African Americans, and which are only genetically similar to such ancestral populations.Second we identified long segments of the genome (at least 3 centiMorgans) that are shared identical-by-descent (IBD) between African American and sub-Saharan Africans, showing that the individuals who share the segments are likely to share ancestry at that segment of the genome within the last couple of thousand years. Assuming relatively little gene flow between African tribal groups, this allows us to directly connect African American individuals to specific African tribes.

Both methods detect significant but subtle ancestry differences in sharing to Africa between African American cohorts. To our knowledge, this is the first work that has documented African ancestry differences in African Americans using genome-wide data, offering opportunities for studying history beyond what is possible with uniparentally inherited mitochondrial DNA and Y chromosome data.

179

Fine scale population genetic structure of African Americans. *E.Y. Durand, M. Macpherson, B.T. Naughton, J. Mountain, C.B. Do.* 23andMe, Inc., Mountain View, CA 94043.

Admixture deconvolution refers to the inference of the geographic origin of chromosomal segments. It has a wide range of applications, from disease mapping to learning about history. In particular, the fine-scale study of recently admixed populations such as African Americans illuminates recent human migrations and is critical for better understanding medical genomics in less homogenous populations. However, current methods for admixture deconvolution generally yield only coarse information because they are limited in the number of ancestral populations they use (typically 2 or 3). Therefore the ancestry of chromosomal segments is identified only at the level of continental origin. It has been estimated that African Americans typically have 65–85% of their ancestry tracing to Africa and 15–25% tracing to Europe, but little is known about their subcontinental distribution of admixture. Little is known, also, about the contribution of Native Americans to the African American gene pool. Leveraging a panel of more than 7,500 individuals with known ancestry derived from a combination of several publicly available datasets and over 5,000 23andMe, Inc., customers reporting four grandparents with the same country-of-origin, we developed a novel method that permits the accurate assignment of chromosomal segments to more than 20 geographic regions. We applied this method to data for 8,500 unrelated African American 23andMe customers who had been genotyped at more than one million sites. We traced back the ancestry of chromosomal segments to subregions of Europe and Africa, thus obtaining an unprecedented fine-grained picture of the ancestry of African Americans. We also found that many African Americans have a significant amount (e.g. more than 5%) of Native American ancestry, consistent with admixture between the two groups as recent as four generations ago. Using a subset of 3,000 individuals who self reported their place of birth, we compared the distribution of African, European and Native American ancestry across different regions in the United States.

180

A model-based approach for analysis of spatial structure in genetic data. *W. Yang[1,4], J. Novembre[3,4], E. Eskin[1,2,4], E. Halperin[5,6,7].* 1) Department of Computer Science, UCLA, Los Angeles, CA; 2) Department of Human Genetics, UCLA, Los Angeles, CA; 3) Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, CA; 4) Bioinformatics IDP, UCLA, Los Angeles, CA; 5) International Computer Science Institute, Berkeley, California, USA; 6) Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv, Israel; 7) School of Computer Science, Tel Aviv University, Tel Aviv, Israel.

Characterizing genetic diversity within and between populations has broad applications in studies of human disease and evolution. Two key step towards this objective are spatially global ancestry inference, which aims at predicting geographical locations for the ancestries of individual, and spatially local ancestry inference, which aims at predicting the geographical locations for chromosome segments, or ancestry blocks. We propose a new approach, SPALL (SPatial Ancestry analysis LocaL), for solving the two inference problems in a unified probabilistic model. This model takes linkage disequilibrium into account and can be solved efficiently by Expectation Maximization (EM) algorithm in conjunction with forward-backward algorithm. This new method allows us to assign geographical locations for parents, grandparents, and ancestries from more generations ago of an given individual. It also allows us to assign geographical locations for each locus-specific variant. We analyzed a European and a worldwide dataset, and showed that the SPALL can actually predict locations with a high accuracy. The proposed model is build as a generalization of our recently published work called Spatial Ancestry Analysis (SPA), which explicitly models the spatial distribution of each SNP by assigning an allele frequency as a continuous function in geographic space. The method allows us to assign an individual, or an admixed individual to geographical locations instead of predefined categories of population. A software including all the proposed methods is freely available in our website http://genetics.cs.ucla.edu/spa.

181

People of the British Isles: An analysis of the genetic contributions of European populations to a UK control population. *S. Leslie[1], B. Winney[2], G. Hellenthal[3], S. Myers[4], P. Donnelly[3], W. Bodmer[2].* 1) Statistical Genetics, Murdoch Childrens Research Institute, Melbourne, Australia; 2) Department of Oncology, University of Oxford, UK; 3) The Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 4) Department of Statistics, University of Oxford, UK.

There is much interest in fine scale population structure in the UK, as a signature of historical migration events and because of the effect population structure may have on disease association studies. Population structure appears to have a minor impact on the current generation of genome-wide association studies, but will probably be important for the next generation of studies seeking associations to rare variants. Furthermore there is great interest in understanding where the British people came from. Thus far genetic studies have been limited to a small number of markers or to samples not collected to specifically address these questions. A natural method for understanding population structure is to control and document carefully the provenance of samples. We describe the collection of a cohort of rural UK samples (The People of the British Isles), aimed at providing a well-characterised UK control population. This will be a resource for research community as well as providing fine-scale genetic information on the history of the British. Using a novel clustering algorithm, approximately 2000 samples were clustered purely as a function of genetic similarity, without reference to their known sampling locations. When each individual is plotted on a UK map, there is a striking association between inferred clusters and geography, reflecting to a major extent the known history of the British peoples. A similar analysis is performed on samples from different parts of Europe. Using the European samples as `source populations' we apply a novel algorithm to determine the proportion of the genomes within each of the derived British clusters that are most closely related to each of the source populations. Thus we can observe the relative contribution (under our model) of each of these European populations to the genomes of samples in different regions of Britain. Our results strikingly reflect much of the known historical and archaeological record while raising some important questions and perhaps answering others. We believe this is the first detailed analysis of very fine-scale genetic structure and its origin in a population of very similar humans. This has been achieved through both a careful sampling strategy and an approach to analysis that accounts for linkage disequilibrium.

182

The applicability of the Balding-Nichols model to a dataset of over 100,000 Brazilian individuals. *R.V. Rohlfs[1], A. Bhaskar[1], V.R.C. Aguiar[2], K. Lohmueller[1], A.M. Castro[3], A.C.S. Ferreira[3], F.S.V. Malta[3], Y. Song[1], I.D. Louro[2], R. Nielsen[1]*. 1) Integrative Biology, University of California, Berkeley, Berkeley, CA; 2) Universidade Federal do Espírito Santo; Vitória, ES, Brazil; 3) Laboratório Hermes Pardini; Belo Horizonte, MG, Brazil.

The Balding-Nichols model of population differentiation has been broadly used in population genetic applications to understand patterns of genetic variation and population structure. In particular, the Balding-Nichols model has been used to model allele sharing between pairs of individuals in large datasets, which is relevant to forensic and relative identification studies. Yet, it is not clear that the model applies directly to this kind of pairwise analysis. Using a remarkable dataset of over 100,000 individuals from Brazil, we consider the applicability of the Balding-Nichols model to calculate expected rates of pairwise allelic matching. Surprisingly, we observe an excess of individuals sharing few alleles. This pattern cannot be explained by a low (or zero) coancestry coefficient value. However, this observation may be explained when considering the nature of pairwise individual comparisons in the context of a structured population where some groups of individuals are more recently descended from common ancestor population than others. While the Balding Nichols model accounts the increase in allele sharing when comparing individuals from the same subpopulation, it does not account for the expected decrease in allele sharing when comparing individuals from different subpopulations. These results may help inform interpretations of emergent pairwise relatedness analyses made possible by the increasing availability of large population genetic datasets. More generally, this work illustrates some of the challenges that arise when applying existing population genetic models to datasets of many thousands of individuals.

183

Rare genetic variants in deep sequencing of neutral regions from a homogeneous population refine models of recent explosive human population growth. *A. Keinan[1], E. Gazave[1], A. Coventry[1], S. Gottipati[1], D. Chang[1], L. Ma[1], D. Muzny[2], E. Boerwinkle[2,3], C. Sing[4], R. Gibbs[2], A. Clark[1]*. 1) Cornell University, Ithaca, NY; 2) Baylor College of Medicine, Houston, TX; 3) UT Houston Health Science Center, Houston, TX; 4) University of Michigan, Ann Arbor, MI.

The human population has grown by over three orders of magnitude within the last 400 generations. Genetic signatures of this growth have recently become evident through sequencing of a large number of individuals (Coventry et al. 2010; Tennessen et al. 2012; Nelson et al. 2012). We have recently illustrated how large sample sizes reveal rarer variants that are more informative for studying this recent epoch of growth (Keinan & Clark 2012). These recent studies modeling growth differ in their estimates of e.g. the extant effective population size of Europeans, which vary by an order of magnitude across studies. Because these studies focused on protein-coding genes, both natural selection and demography have impacted the site frequency spectrum (Kiezun et al. 2012). Here, to disentangle the signatures of demographic history from those of natural selection, we sequenced regions that are as neutral as possible by targeting loci very far from known or predicted genes and other functional elements, and that lack repeats and selection signals, for a total of 216 kb spanning 15 loci. Another crucial consideration in studying recent history from allele frequencies is that the sequenced population exhibit as little substructure as possible; otherwise the frequency of rare variants will be skewed by mutations that postdate the split of subpopulations. We analyzed the population structure of 9716 European Americans from the ARIC study using EIGENSOFT and selected 500 individuals that constitute a homogeneous cluster. For singletons to be reliably called, we sequenced the neutral regions in these individuals to a very high median coverage of 282X using Illumina HiSeq (paired 100 bp reads), and used GATK to call >2000 high-confidence SNPs. We found that 37% of the variants called were singletons. This is the largest dataset to date of neutral, non-genic regions in a relatively large sample from a homogeneous population. We utilized these data to refine models of recent history by fitting the observed site frequency spectrum to an array of models of recent population growth. These models include changes in the rate of exponential growth over time (supported by the historical record of Europe) and the possibility of a very recent epoch of faster-than-exponential growth. The resulting models are of unprecedented resolution and provide a foundation to modeling the genetic architecture of complex disease, with implications to methods for gene-disease association studies.

184

Estimating human population sizes using the coalescent with recombination. *S. Sheehan[1], K. Harris[2], Y.S. Song[1,3]*. 1) Department of Electrical Engineering and Computer Sciences, UC Berkeley, Berkeley, CA; 2) Department of Mathematics, UC Berkeley, Berkeley, CA; 3) Department of Statistics, UC Berkeley, Berkeley, CA.

Throughout history, the population size of modern humans has varied considerably due to changes in environment, culture, and technology. It has been further shaped by recent super-exponential growth [1] and still contains signatures of large-scale migration events such as the out of Africa bottleneck. More accurate estimates of population size changes and when they occurred could provide a clearer picture of human colonization history and shed light on the concept of effective population size. Li and Durbin [2] recently developed a coalescent-based hidden Markov model (HMM) for a pair of chromosomes to estimate past population sizes. However, this approach is restricted by the fact that, because of the small sample size, few coalescence events occur in the very recent or very ancient past, thus impeding inference during those epochs. Here, we present a general HMM that allows one to efficiently handle more samples, thus significantly improving the estimation of past population size changes.

Our work generalizes the framework of the sequentially Markov conditional sampling distribution (CSD) recently proposed by Paul et al. [3] The CSD describes the probability of observing a newly sampled haplotype given a set of previously sampled haplotypes, and it allows one to approximate the joint probability of multiple haplotypes as a product of approximate conditionals [4]. Because the CSD proposed by Paul et al. was derived from the diffusion process dual to the coalescent and the construction admits a natural genealogical interpretation, it can be modified in a principled way to incorporate past population size changes, which can be inferred within an expectation-maximization framework. Simulation results demonstrate that we can accurately reconstruct the true population size changes, with especially good power in the recent past. We apply our method to the genomes of multiple human individuals to obtain a detailed size change history during such recent times.

[1] Coventry et al. Nature Communications, 2010. [2] Li and Durbin. Nature, 2011. [3] Paul, Steinrucken, and Song. Genetics 2011. [4] Li and Stephens. Genetics, 2003.

185

Reconstructing historical contributions to modern gene pools using the Sequentially Markovian Coalescent Conditional Sampling Distribution. *A. Platt, J. Novembre*. UCLA, Los Angeles, CA.

The Sequentially Markovian Coalescent Conditional Sampling Distribution describes the composition of a single chromosome in relation to the remainder of a population sample. In it, the observed haplotype of the focal chromosome is represented as a mosaic of imperfect copies of other observed haplotypes. This is modeled as a Hidden Markov Model where the hidden state space consists of the identity of the haplotype from which a locus was copied as well as the amount of time since the copying took place. We use the formulation of Paul, Steinrücken, & Song to calculate the probability that a locus selected from a chromosome taken from one sub-population shares a most recent common ancestor with a haplotype from any other subpopulation, conditional on that common ancestor having existed during a particular period in history. This gives us a measure of the relative contributions other sub-populations have made to the current gene pool of our focal population and how those contributions have changed through time. This characterization lends itself to identifying past and present demographic scenarios. We present signature patterns of phenomena such as population structure, admixture, and colonization and range expansion in simulated data, and contrast these with patterns observed in the 1000 Genomes project.

186

On the Sardinian ancestry of the Tyrolean Iceman. *M. Sikora[1], M. Carpenter[1], A. Moreno-Estrada[1], B.M. Henn[1], P.A. Underhill[1], I. Zara[2], M. Pitzalis[3], C. Sidore[3,4,5], F. Reinier[2], M. Marcelli[2], A. Angius[3,4], C. Jones[2], T.T. Harkins[6], A. Keller[7,8], A. Zink[9], G. Abecasis[4], S. Sanna[3], F. Cucca[3], C.D. Bustamante[1]*. 1) Department of Genetics, Stanford University, Stanford, CA, USA; 2) CRS4, Center for Advanced Studies, Research and Development in Sardinia, Parco Scientifico e Tecnologico della Sardegna, Pula, Italy; 3) Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Monserrato, Italy; 4) Center for Statistical Genetics, Ann Arbor, University of Michigan, MI, USA; 5) Università degli Studi di Sassari, Dip. Scienze Biomediche, Sassari, Italy; 6) Genome Sequencing Collaborations Group, Life Technologies, Beverly, MA, USA; 7) Department of Human Genetics, Saarland University, Homburg, Germany; 8) Siemens Healthcare, Erlangen, Germany; 9) Institute for Mummies and the Iceman, EURAC research, Bolzano, Italy.

The complete genome of the 5,300 year old mummy of the Tyrolean Iceman, found in 1991 on a glacier near the border of Italy and Austria, has recently been published and yielded new insights into his origin and relationship to modern European populations. A key finding of this study has been an apparent recent common ancestry with individuals from Southern Europe, in particular Sardinians. This finding was interpreted as a genetic signature of the demic diffusion model of the expansion of Neolithic people into Europe during the spread of agriculture, although a possible recent migration of the ancestors of the Iceman to Central Europe could not be ruled out. Furthermore, the possibility that modern-day Sardinians present a remnant population of those early farmers mostly unaffected by subsequent migrations in the European mainland was not explored. In order to address these questions we analyzed the genome of the Iceman together with a large set of publicly available as well as newly generated genomic data, from both modern and ancient European individuals. We used unpublished data from whole genome sequencing of 452 Sardinian individuals, together with publicly available data from Complete Genomics and the 1000 Genomes project, to confirm that the Iceman is most closely related to contemporary Sardinians. An analysis of these data together with ancient DNA data from a recently published study on Neolithic farmers and hunter-gatherers from Sweden shows the Iceman most closely related to the farmer individual, but not the hunter-gatherers, with the Sardinians again being the contemporary Europeans with the highest affinity. Strikingly, an analysis including novel ancient DNA data from an early Iron Age individual from Bulgaria also shows the strongest affinity of this individual with modern-day Sardinians. Our results show that the Tyrolean Iceman was not a recent migrant from Sardinia, but rather that among contemporary Europeans, Sardinians represent the population most closely related to populations present in the Southern Alpine region around 5000 years ago. The genetic affinity of ancient DNA samples from distant parts of Europe with Sardinians also suggests that this genetic signature was much more widespread across Europe during the Bronze Age.

187

Gene-based epistasis analysis in genome-wide association studies. *L. Ma[1], A. Brautbar[2], E. Boerwinkle[3], C.F. Sing[4], A.G. Clark[1,5], A. Keinan[1]*. 1) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Section of Cardiovascular Research, Department of Medicine, Baylor College of Medicine, Houston, Texas; 3) Human Genetics Center, Health Science Center, University of Texas, Houston, Texas; 4) Department of Human Genetics, School of Medicine, University of Michigan, Ann Arbor, Michigan; 5) Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York.

Epistasis is likely to play an important role in complex diseases and traits. Here, we first present an analysis of a genome-wide association study (GWAS) that led to the discovery of a novel gene-gene interaction: We tested for gene-gene interactions affecting the level of four lipids, restricting the space of tests by using prior knowledge of established GWAS hits, protein-protein interactions (PPI), and pathway information. Using genotype data from 9713 European Americans (EA) from the ARIC study, we identified an interaction between *HMGCR* and a locus near *LIPC* in their effect on HDL-C levels (Bonferroni corrected $P_c = 0.002$). Using a locus-based validation procedure, we successfully validated this gene-gene interaction in the EA cohorts from the Framingham Heart Study ($P_c = 0.002$) and the Multi-Ethnic Study of Atherosclerosis (MESA; $P_c = 0.006$). The interaction between these two loci is also significant in the African American sample from ARIC ($P_c = 0.004$) and in the Hispanic American sample from MESA ($P_c = 0.04$). In order to reduce the number of tests and thereby gain statistical power, we developed novel gene-based interaction tests by combining marker-based interaction tests of pairs of markers in each of two genes. We derived analytical formulas for the correlation between marker-based tests due to linkage disequilibrium using either genotype data or an external reference panel, and incorporated it into the tests. The tests extend four p-value combining methods, a minimum p-value method, a gene-based association test using an extended Simes procedure, a truncated tail strength method, and a truncated p-value product method. Extensive simulations confirm accurate type-I error rates of all tests and show that these methods are more powerful than an interaction test based on principal components. The two truncated tests are more powerful than the former two in cases where the causal interaction is not directly observed or of multiple causal interactions. Along with a curated PPI network, we applied our methods to ARIC EA samples and identified five gene-based interactions affecting HDL-C levels which were missed by marker-based interaction analyses. One gene-gene interaction, between *SMAD3* and *NEDD9*, is successfully replicated in an EA sample from MESA. Our gene-based interaction tests have the power and potential to identify gene-gene interactions in current and future association studies.

188

Building Human Phenotype Networks from Shared Genetic Risk Variants. *C. Darabos[1], K. Desai[2], M. Giacobini[3], M. Lupien[2], J.H. Moore[1]*. 1) Department of Genetics, Institute for Quantitative Biomedical Sciences, Geisel School of Medicine, Dartmouth College, Lebanon, NH; 2) University of Toronto, Canada; 3) University of Torino, Italy.

Elucidating relationships between human traits or diseases is becoming increasing important in the study of complex genetic disorders. Ultimately, a thorough understanding of these connections may provide clinical tools to design common drug targets. These links can be mapped out as human phenotype networks (HPN). HPNs have traditionally focused on gene centric analysis, drawing links between genes, genes and diseases, or gene products. In the present study, we propose a non-gene centric method relying on genetic risk factors, such as single nucleotide polymorphisms (SNPs) associated with human diseases and phenotypes that map to any given genomic region. We visualize the results in a phenotypic network (PN). For each trait in our network, we extract the set of risk-associated (ra)SNPs, called a risk-associated variome (RAV), from the catalog of published genome-wide association studies (GWAS). To address the low genomic coverage provided by GWAS, we associate each raSNP with all SNPs found in linkage disequilibrium (ldSNPs) using the HapMap project data. This imputed variome (iRAV) allows us to establish connections between disease/traits that share haplotype blocks, i.e. that share their raSNPs or ldSNPs. The weight of the interaction is computed as the normalized number of SNPs clusters shared between the iRAVs from two phenotypes. Several links within our PN confirm prior knowledge, such as the clustering of immune-related disorders. The inclusion of behaviors and traits, as opposed to diseases only, also proves very informative. The PN shows a relation between C reactive protein (CRP) and Alzheimer's disease via shared SNPs. This particular link was only recently clinically identified. Furthermore the PN reveals unlikely connections between diseases in terms of shared drug targets, i.e. lung cancer and systemic lupus erthymatosus. Cimetidine, used in treating lung cancer has also been administered orally to cases of lupus nephritis patients to improve renal function. Our results demonstrate how seemingly incidental disease connectivity via shared variants can be used to extract clinically relevant information about diseases and potential shared treatment. We anticipate our network to be a starting point for more focused biological studies of diseases based on risk associated variants.

189

Incorporating network dynamics to prioritize genes through genome wide association studies. *L. Hou[1], M. Chen[2], C.K. Zhang[3], J. Cho[4], H. Zhao[1]*. 1) Division of Biostatistics, Yale School of Public Health, New Haven, CT; 2) Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX; 3) Keck Biotechnology Laboratory Biostatistics Resource, Yale University School of Medicine, New Haven, CT; 4) Internal Medicine, Yale University, New Haven, CT.

Although genome-wide association studies (GWAS) have successfully identified many susceptibility loci for common diseases, these loci only explain a small portion of heritability. It is believed that many loci remain to be discovered, because association signals are often weak and inseparable from background noise. In order to increase power, one potentially fruitful direction is to incorporate functional genomics information, especially gene networks, to prioritize GWAS signals. Most current methods that utilize network information to predict or prioritize disease genes are based on the &ldquoθilt by association" principle. In these methods, networks are treated as static, and genes associated with the same disease are assumed to locate closer with each other than random pairs in the network. In contrast, we propose a novel &ldquoθilt by rewiring" principle. Studying the dynamics of gene networks between patient and control groups, this principle assumes that a number of disease genes undergo extensive rewiring to carry out disease-specific functions, while most interactions in the network are stable, regardless of disease condition, to maintain basic functions of the cell. In order to study the rewiring diagram of networks, we compared gene co-expression networks in patients with Crohn's disease (CD) and controls. The results demonstrate dramatic rewiring of gene expression; specifically, a significant proportion of co-expression interactions in CD are absent in controls, and these interactions are enriched in disease association signals. Furthermore, in immune system related genes (Reactome pathway), the density of rewiring is 5 times of that at whole genome. In order to integrate this network rewiring feature and GWAS signal, we utilized the Markov random field modeling framework to prioritize genes associated with CD. We show that the prioritized genes through incorporating network dynamics are enriched (hypergeometric test, pvalue less than 0.05, fold change larger than 1.3) in immune system related pathways (Reactome), such as cytokine signaling, signaling by interleukins, compared to those identified without taking network rewiring into account.

190

Large-scale multi-phenotype meta-analysis evaluates pleiotropic effects at *FADS1* and *GIPR* loci. *V. Lagou[1,2], R. Mägi[3], K. Fischer[3], M. Akerlund[4], I. Surakka[5,6], M. Kaakinen[7], J.S. Ried[8], A. Mahajan[1], M. Horikoshi[1,2], L. Marullo[1,9], K. Strauch[8,10], C. Gieger[8], S. Ripatti[5,6,11], A.P. Morris[1,2], V. Lyssenko[4,12], I. Prokopenko[1,2] for the ENGAGE (European Network for Genetic and Genomic Epidemiology) consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; 3) Estonian Genome Center, University of Tartu, Tartu, Estonia; 4) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden; 5) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland; 6) National Institute for Health and Welfare, Helsinki, Finland; 7) Institute of Health Sciences and Biocenter Oulu, University of Oulu, Oulu, Finland; 8) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 9) Department of Evolutionary Biology, Genetic Section, University of Ferrara, Ferrara, 44121, Italy; 10) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; 11) Wellcome Trust Sanger Institute, United Kingdom; 12) Lund University Diabetes Centre, CRC at Skåne University Hospital, Malmö, Sweden.

Cardiometabolic human phenotypes are correlated epidemiologically and may share common genetic factors influencing their variability. So far, the majority of genetic association studies have tested univariate associations with individual traits. However, analysis of the effects for multiple traits simultaneously is known to provide increased power and may detect possible pleiotropic effects. We aimed to explore pleiotropic effects at two loci, *FADS1* and *GIPR*, with prior evidence of associations with multiple metabolic traits. Genome-wide associations are known for *FADS1* variants with lipids, fasting glucose (FG) and resting heart rate, among others. At *GIPR*, genome-wide associations are described with body mass index (BMI), type 2 diabetes and 2 hours post load glucose (2hGlu). We implemented the method described by Clarke and colleagues that utilizes multiple logistic regression with genetic marker as the outcome and phenotypes as explanatory variables. Four multi-phenotype sets were generated considering the current knowledge on *FADS1* and *GIPR* variants, as well as differential availability of phenotypes in the participating cohorts. All possible combinations of phenotypes within each set were assessed in each cohort and derived likelihoods from each model were meta-analyzed across studies. The Bayesian Information Criterion (BIC) statistic was used to prioritise models subsequently combined through meta-analysis. Analysis of pleiotropic effects was performed in up to 34,048 European individuals from six studies within the ENGAGE consortium. The multi-trait meta-analysis led to selection of models that gave the best fit. Simulations confirmed that the meta-analysis BIC-statistic was neither biased towards too complex nor too simple models and it selected the most plausible model within a range of realistic scenarios. At *FADS1*, the best models within multi-phenotype sets underscored independent effects for BMI and pulse pressure (BIC=55975, BIC_{null}=55985, $P_{LRT}=4.4\times10^{-7}$) and for weight and stroke (BIC=4447, BIC_{null}=4469, $P_{LRT}=9.35\times10^{-9}$). For *GIPR*, BMI (BIC=79110, BIC_{null}=79181, $P_{LRT}=1.3\times10^{-19}$) and a combination of weight, FG, 2hGu, total cholesterol (BIC=24421, BIC_{null}=24452, $P_{LRT}=6.8\times10^{-14}$) were prioritized. Results from this study are being tested in up to 200,000 individuals availble within ENGAGE. The proposed design and methodology for assessing pleiotropic effects at genetic loci proves feasible and can be applied in large-scale studies.

191

Building and assessing protein-protein interaction networks from genome wide association results in cancer. *L.T. Hiraki[1,2], A.D. Joshi[1,2], S. Lindstrom[1,2], A.T. Chan[2,3], S. Chanock[4], P. Kraft[1,2].* 1) Program of Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Epidemiology, Harvard School of Public Health, Boston MA; 3) Gastrointestinal Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 4) National Cancer Institute, Bethesda, MD.

Genome wide association studies (GWAS) have identified hundreds of independent single nucleotide polymorphisms (SNPs) associated with cancer risk. However, most of these SNPs are in linkage disequilibrium with several proximal genes, and identifying which of these genes are functionally involved in carcinogenesis remains a challenge. Mendelian disease studies have demonstrated that protein products of causal genes tend to physically interact, so we hypothesized that genes that show evidence for protein-protein interactions (PPIs) with genes at other GWAS-identified and Mendelian cancer loci are likely involved in carcinogenesis. We used the "InWeb" database, which summarizes multiple PPI experiments, and the analytic framework implemented in the Disease Association Protein-Protein Link Evaluator (DAPPLE) software to explore PPI networks for genes within loci identified by GWAS to be associated with breast (BRCA), prostate (PCA) and colorectal (CRC) cancer. A total of 26 BRCA, 25 PCA and 16 CRC SNPs were identified from literature review of GWAS (as of May 2012). We found significantly more direct connections among GWAS-identified CRC loci than expected by chance [6 genes: CDH1, LIMA1, SCG5,SMARCD1, EIF3H, RPS21]. There was no significant enrichment for direct connections for BRCA or PCA (p=0.33, 0.77 respectively). At particular loci, DAPPLE identified several genes more connected to other risk loci than expected ($p<0.05$, adjusted for multiple testing): MAP3K1 for BRCA and C11orf53, EIF3H, and GREM1 for CRC. Several genes outside of the GWAS-identified loci also demonstrated higher interaction with seed proteins than expected by chance ($p < 0.1$) : TRAF2, DLK1, FKBP5,MRPL13, PAK1, SMARCD1, MRPL11, TNFRSF10B for BRCA, IGFBP3, IGFBP2, ZFPM1, DDX3X, HSPD1, LMNA, AR, KPNA2 for PCA and GABPA, CTNNA1, ERBB2IP, BMPR1B for CRC. This approach leverages GWAS and PPI information to highlight potential mechanistic pathways and therapeutic targets for cancer. We plan to expand our investigation to include other cancers and to assess whether candidate genes nominated by DAPPLE are enriched for novel mutations using the large, agnostic GWAS conducted by the Collaborative Oncology Gene-environment Study (COGS).

192

A smoothed functional principal component analysis for pathway analysis with next-generation sequencing data. *J. Zhao[1], Y. Zhu[1], E. Boerwinkle[2], M. Xiong[2].* 1) Department of Epidemiology, Tulane University, New Orleans, LA; 2) Human Genetics Center, School of Public Health, University of Texas HSC-Houston.

Pathway analysis typically tests the association of a predefined set of genes with disease based on known biology of a disease. Although pathway analysis has been developed and successfully applied to association studies of common variants, the methods for testing the association of genetic pathways with rare variants have not been well studied. Existing statistics suffer from inflated rate of false positives and low power in pathway-based association tests for rare variants. The high false-positive and low true-positive rates of the current methods are mainly due to the lack of ability to account for gametic phase disequilibrium and population stratification in association analysis. Here we developed a novel statistic that takes into account information for both gametic phase disequilibrium and population stratification to detect disease-associated pathways for next-generation sequencing data, which usually delivers a large amount of rare variants. This method, based on the smoothed functional component analysis (SFCA), captures position-level variant information and accounts for the gametic phase disequilibrium among SNPs. Simulation studies under different scenarios indicated that the SFCA-based statistic has correct type 1 error rates in testing pathway associations with disease and can be used for common or rare variants or the combination of these two types of polymorphisms. The statistical power of the SFCA-based statistic was also evaluated and compared with 20 existing statistics. We found that this novel statistic has much higher power than other existing statistics under all the scenarios considered. To further assess its performance, we also applied this method to exome sequencing data to identify genetic pathways that might be associated with early onset myocardial infarction (MI). Three pathways were found to be significantly associated with MI after Bonferroni correction. In addition, the proposed SFCA-based statistic exhibited much smaller p-values than existing methods in detecting pathways of rare or common variants or both with MI.

193

Variants in exons and in transcription factors affect gene expression in trans. *A. Kreimer, I. Pe'er.* Computer Science, Columbia University, New York, NY.

In recent years many genetic variants (eSNPs) have been found to be associated with gene expression. However, the causal variants and the regulatory mechanisms by which they act remain mostly unknown. Here we present a comprehensive analysis of trans-eSNPs, integrating SNPs that are fully ascertained from genomic sequencing data with transcriptional profiling (RNA-seq) in the same cohort. When considering interpretable genomic regions containing candidate eSNPs, we observe enrichment of such variants in exons. We thus focus on exonic eSNPs, and consider eSNPs within the span of Transcription Factors (TFs) for comparison. In both cases, these variants define the spanning source gene, along with the respective gene target of association. We map the source and target genes onto a Protein-Protein Interaction (PPI) network and study their topological properties. When considering pairs of eSNP exon source with its corresponding target, the stronger their association, the closer they are within the PPI network (permutation $p<9.9\bullet10\text{-}4$) and the higher the degree of the target (permutation $p<0.002$). Expression analysis demonstrates that these source-target pairs are more likely to be co-expressed ($p<5.4\bullet10\text{-}5$) and the eSNP tends to have a cis effect, modulating the expression of the source gene ($p<2.3\bullet10\text{-}13$). In contrast, source-target pairs with a TF eSNP are not observed to have such properties. We do observe these latter pairs to reside within the same PPI cluster more than expected by chance (permutation $p<0.0043$), and to assemble functionally enriched units of a TF source along with its gene targets. Our results suggest two modes of trans regulation: TF variation frequently acts via a modular regulation mechanism, with multiple targets that share a function with the TF source. Notwithstanding, exon variation often acts by a local cis effect, and propagates through shorter paths of interacting proteins across functional clusters of the PPI network.

194

"The continuation of theory by other means": ForSim as a forward simulator for improved understanding of the genetic architecture of complex traits and its evolution. *K.M. Weiss[1], J.H. Lee[2], J.D. Terwilliger[3,4,5], B.W. Lambert[1].* 1) Dept Anthropology, Pennsylvania State Univ, University Park, PA; 2) Department of Epidemiology, Sergievsky Center, Columbia University, NY; 3) Department of Genetics and Development, Department of Psychiatry, Columbia University, NY; 4) Division of Medical Genetics, NY State Psychiatric Institute, NY; 5) Department of Chronic Disease Prevention, Public Health Genomics Unit, National Institute of Health and Welfare, Helsinki.

von Clausewitz famously said that war is the continuation of politics by other means. Like war, biological traits are often so complex as to be difficult to untangle. They are generally produced by many different functional genomic elements and their regulation and non-additive interactions. Each is subject to variation within and among populations. Since each gene is different and each trait controlled by different genes (as well as environmental effects), no single theory captures the genetic architecture in the unitary way of theories in physics and chemistry. Empirical studies find statistical pattern in data but in few instances is the actual causal architecture known. Weak inference and the variability among samples undermine prediction of phenotypes from genotypes. Computer simulation can at least generate data less formally constrained by pre-specified equations, and can attempt to mimic empirical data with an eye towards focusing follow-up studies to improve our understanding. ForSim is a very flexible C++ cross-platform freely available simulation program that can accommodate many different parameters. Traits evolve in a forward evolutionary (population history) way, including random mutation, allelic effects, recombination, mating, natural selection (or drift), and other parameters. Users specify flexible phenogenetic relationships. Multiple populations with flexible selection and mating (mate-choice and gene flow) patterns can be included conditions can be changed during the run. ForSim is by design a basically brute-force approach, enabled by modern computing speed to make minimal formal presumptions. The program produces diverse, user-specified output, including complete pedigrees for the simulated population(s) of user-specified depth or complete histories of simulated SNPs. Thus various evolutionary scenarios can be simulated, and results can be used to construct and test diverse sampling strategies, such as GWAS, linkage analysis, or biobank data. Improvements are being made to increase the scale and scope of what can be practicably simulated.

195

Incorporating phylogenetic conservation and pedigree information in tests of rare-variant association. *H. Hu[1], C.D Huff[2], H. Coon[3], S. Guthery[4], S. Tavtigian[5], J.C. Roach[6], Z. Kronenberg[1], J. Xing[7], A.F.A Smit[6], G. Glusman[6], A.K. Holloway[8], V. Garg[9], B. Moore[1], R. Hubley[6], W.S. Watkins[1], H. Li[6], S.Z. Montsaroff[6], D.E. Abbott[6], L.E. Hood[6], K.S. Pollard[8], D.J. Galas[10], D. Srivastava[8], M.G. Reese[11], L.B. Jorde[1], M. Yandell[1].* 1) Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; 2) Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX; 3) Psychiatry Department, University of Utah School of Medicine, Salt Lake City, UT; 4) Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT; 5) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT; 6) Institute for Systems Biology, Seattle, WA; 7) Department of Genetics, Rutgers University, Piscataway, NJ; 8) Gladstone Institute of Cardiovascular Disease and University of California, San Francisco, San Francisco, CA; 9) Department of Pediatrics, The Ohio State University and Center for Cardiovascular and Pulmonary Research, Research Institute at Nationwide Children's Hospital, Columbus, OH; 10) Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 7, avenue des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg; 11) Omicia, Inc., Emeryville, CA.

Personal genome sequencing presents new challenges and opportunities for clinical interpretation of novel variants. GWAS approaches compare the allele frequency of each variant between cases vs. controls. In contrast, rare-variant association tests aggregate variants into discrete features to obtain greater statistical power. The Variant Annotation, Analysis, and Search Tool (VAAST) employs a variant association test that combines amino acid substitution and allele frequency information using a composite likelihood ratio test (CLRT). Here, we report two new algorithms that expand VAAST to incorporate phylogenetic sequence conservation and pedigree information, improving accuracies on both Mendelian and complex genetic diseases. We tested our phylogenetic conservation algorithm, implemented in the package VAAST 2.0, on variants from the Human Gene Mutation Database and rare variants from the BRCA1/2 genes. VAAST 2.0 prioritizes variants more accurately than other variant function prediction methods, such as SIFT, PolyPhen-2 or Mutation-taster. Moreover, VAAST 2.0 also outperforms other popular rare-variant association tests, such as KBAC, WSS, SKAT and Variable Threshold using published datasets for Crohn disease (NOD2), hypertriglyceridemia (LPL), and breast cancer (ATM, CHEK2, and XRCC2). We also demonstrate that VAAST 2.0 improves search accuracy on simulated datasets across a wide range of allele frequencies, population attributable disease risks, and allelic heterogeneity, factors that compromise the accuracy of many published rare-variant association tests. Our algorithm for pedigree analysis, implemented in the pVAAST package, incorporates the observed inheritance pattern at each locus from one or more families directly into the CLRT to increase the accuracy and greatly decreases bioinformatic complexity. The inheritance pattern of each variant is verified automatically using identity-by-descent mapping, and statistical significance is calculated in a manner that accounts for pedigree relationships and the observed pattern of variation in cases and controls. pVAAST supports many types of sequenced-based familial disease-gene searches, with inheritance models for dominant, simple recessive, compound heterozygote, and de novo mutations. We demonstrate pVAAST's utility by analyzing familial sequence data for recessive (Miller syndrome), dominant (cardiac septal defects), and complex genetic diseases (autism, breast cancer, and Crohn disease).

196

***TMTC4*: A novel candidate gene for callosal development.** *L. Fernandez[1], J. Li[1], M. Wakahiro[1], E. Rider[1], T. Bartman[2], E. Sherr[1]*. 1) Neurology, Univ California, San Francisco. San Francisco, CA; 2) Neonatology, Nationwide Children's Hospital. Columbus, OH.

Background: Agenesis of the corpus callosum (AgCC) is a common brain malformation with a birth incidence greater than 1:4,000 that can cause autism, intellectual disability, and seizures. Although AgCC can be associated with genetic syndromes, the underlying genetic causes in most cases remain elusive. We previously mapped a candidate AgCC gene, *TMTC4* (transmembrane tetratricorepeat 4), at the breakpoint of a balanced translocation in an AgCC patient and found additional potentially pathogenic ultrarare mutations in our AgCC cohort. **Approach:** We use zebrafish (*Danio rerio*) as a model to test functionally and genetically whether *TMTC4* plays a critical role in the development of the CC and whether ultrarare mutations identified in AgCC patients may be pathogenic. **Methods & Results:** We utilized morpholinos (MOs) targeted against the splice donor and acceptor sites of the zebrafish paralogue of *TMTC4* to downregulate expression of its transcript. After injecting MOs into the embryo, the phenotype was evaluated at 24 and at 48 hours. The injected embryos showed a severe phenotype: very small head and small or no eyes, or a mild severity of these defects (intermediate phenotype); compared to non-injected embryos, and to embryos injected with a 5-mispaired specific control. On average, 30% of the injected embryos had a severe phenotype. We then showed that coinjection of wild type (WT) human mRNA for *TMTC4* along with the MO pair rescues the phenotype, with only 16% of embryos showing severe phenotype. We also functionally tested the pathogenicity of deleterious point mutations identified in our AgCC patient cohort by analyzing whether injection of the human *TMTC4* mRNA engineered to contain these point mutations do not rescue the phenotype, while the WT mRNA can rescue. Our preliminary data have shown that injection of the human *TMTC4* mRNA engineered to contain these point mutations do not rescue the phenotype, while the wild-type mRNA can rescue. The failure of the mutated human mRNA to rescue the pathogenic phenotype suggests that the human mutation results in loss of the normal gene function. **Conclusions:** We believe *TMTC4* is a strong candidate for callosal development. The identification of novel genes and novel pathways associated with AgCC will improve our knowledge of this complex malformation and related disorders.

197

CLK2 missense mutation in a family with pontocerebellar hypoplasia type 7. *V.R.C. Eggens[1], Y. Namavar[1,3], M.A. Haagmans[2], K. Fluiter[1], E.J. Bradley[1], P.G. Barth[3], B.T. Poll-The[3], F. Baas[1]*. 1) Genome Analysis, Academic Medical Centre, Amsterdam, Amsterdam, Netherlands; 2) Department of Clinical Genetics, Academic Medical Centre, Amsterdam, The Netherlands; 3) Division of Pediatric Neurology, Emma's Childrens Hospital, Academic Medical Centre, Amsterdam, The Netherlands.

Pontocerebellar hypoplasias (PCH1-7) represent a group of autosomal recessive neurodegenerative disorders with prenatal onset. Common characteristics of all subtypes include hypoplasia of pons and cerebellum, and severe mental and motor impairments. PCH7 is characterized by genital abnormalities in males in addition to pontocerebellar hypoplasia. Mutations in genes encoding for the tRNA splicing endonuclease (TSEN) complex, mitochondrial arginyl-tRNA synthetase (RARS2), exosome component 3 (EXOSC3) and vaccinia related kinase 1 (VRK1) have been associated with PCH type 1,2,4,5 and 6. Up to now, no locus for PCH7 has identified. We performed exome sequencing on two siblings with PCH7 and their healthy parents. A missense mutation (p.A390S) in cdc2-like kinase 2 (CLK2) was identified homozygous in both patients and heterozygous in the parents. To verify this candidate gene, we performed knockdown experiments in zebrafish. Knockdown of clk2 by morpholino (MO) injections in zebrafish embryos resulted in a smaller head and brain region at 24 hours post fertilization when compared to fish injected with a control MO. In addition, LNA in situ hybridizations on human embryonic brain tissue showed CLK2 mRNA expression in the cerebellum at 8 weeks gestational age. CLK2 encodes for a SR kinase involved in mRNA splicing. The role of CLK2 in mRNA splicing is in line with TSEN, RARS2 and EXOSC3 mutations causing PCH, as these genes are all involved in RNA processing. In addition, CLK2 is associated with tra2-beta1, which plays a role in sex determination in drosophila. These findings fit with the genital abnormalities found in PCH type 7.

198

Missense mutations in ITPR1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia. *J. Warman Chardon[1], L. Huang[2], M. Carter[3], K. Friend[4], T. Dudding[5,6], J. Schwartzentruber[7], R. Zou[8], P. Schofield[9], S. Douglas[1], D. Bulman[8,10], K. Boycott[2,1]*. 1) Genetics, Chidren's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada; 3) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Genetic Medicine, Women's and Children's Hospital, SA Pathology, North Adelaide, Australia; 5) Hunter Genetics, Warratah, NSW, Australia; 6) University of Newcastle, Newcastle, NSW, Australia; 7) McGill University and Genome Quebec Innovation Centre, Montréal, QC, Canada; 8) Ottawa Hospital Research Institute, University of Ottawa, Ottawa, ON, Canada; 9) Centre for Translational Neuroscience and Mental Health, University of Newcastle, NSW, Australia; 10) Division of Neurology, Ottawa Hospital and University of Ottawa, Ottawa, ON, Canada.

Congenital nonprogressive spinocerebellar ataxia is characterized by early gross motor delay, hypotonia, gait ataxia, mild dysarthria and dysmetria. The clinical presentation remains fairly stable and may be associated with cerebellar atrophy. To date, only a few families with autosomal dominant congenital nonprogressive spinocerebellar ataxia have been reported. Linkage to 3pter was demonstrated in one large Australian family and this locus was designated spinocerebellar ataxia type 29. We studied the Australian family and an unreported Canadian family diagnosed with autosomal dominant congenital nonprogressive spinocerebellar ataxia and identified missense mutations in the ITPR1 gene in both families. ITPR1 encodes inositol 1,4,5-trisphosphate receptor, type 1, a ligand-gated ion channel that mediates calcium release from the endoplasmic reticulum. Deletions of ITPR1 are known to cause spinocerebellar ataxia 15, a very slowly progressive form of cerebellar ataxia with onset in adulthood. Our study demonstrates for the first time that, in addition to spinocerebellar ataxia 15, alteration of ITPR1 function can cause a distinct congenital nonprogressive ataxia; highlighting important clinical heterogeneity associated with the ITPR1 gene and a significant role of the ITPR1-related pathway in the development and/or maintenance of the normal functions of the cerebellum.

199

Vps37A causes a novel form of complex Hereditary Spastic Paraparesis. *T. Falik-Zaccai[1,2,5], Y. Zivony-Elboum[1,5], W. Westbroek[3], D. Savitzki[4], Y. Shoval[1], Y. Anikster[6], A. Waters[7], R. Kleta[7,8]*. 1) Institute of Human Genetics, Western Galilee Hosp, Nahariya, Israel; 2) The Galilee Faculty of Medicine - Bar Ilan, Tzfat, Israel; 3) Section on Human Biochemical Genetics, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda MD, USA; 4) Department of Child Development, Western Galilee Hospital-Nahariya, Israel; 5) Rappaport Faculty of Medicine, Technion, Haifa, Israel; 6) Metabolic Unit, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 7) Nephro-Urology Unit, Great Ormond Street Hospital, London WC1N 3JH, UK; 8) Division of Medicine, University College London, London NW3 2PF, UK.

Hereditary Spastic Paraplegia (HSP) comprises a heterogeneous group of neurodegenerative disorders characterized by progressive lower limb spasticity, retrograde degeneration of the crossed cortico-spinal tracts, and thinning of the posterior columns in the spinal cord. Complicated forms (CHSP) are characterized by the addition of such neurological features as spastic quadriparesis, seizures, dementia, amyotrophy, extrapyramidal disturbance, cerebral or cerebellar atrophy, optic atrophy, and peripheral neuropathy, as well as by extra neurologic manifestations such as dysmorphism, albinism, retinitis pigmentosa. CHSP forms are generally inherited as autosomal recessive (AR) traits. Currently, more than 40 HSP loci and 21 causative genes for pure and complicated HSP forms have been identified. We report members of two unrelated kindred of Arab Moslem origin who present with infantile spastic paraparesis of upper and lower limbs, mild intellectual disability, kyphosis, pectus carinatum, and hypertrichosis. We performed neurological and developmental examinations on the affected individuals. We conducted whole genome linkage and haplotype analyses, followed by sequencing of candidate genes; RNA and protein expression studies, and investigations on knockdown morpholino oligonucleotide injected zebrafish. We characterize a novel form of AR CHSP. MRI studies of brain and spinal cord were normal. Within a single significantly linked locus we identified a homozygous missense mutation c.1146A>T (p.K382N) in the Vacuolar Protein Sorting 37A (Vps37A) gene, fully penetrant and segregating with the disease in both families. Mobility was significantly reduced in Vps37A knockdown morpholino oligonucleotide injected zebrafish, supporting the causal relationship between mutations in this gene and the phenotype described in our patients. We provide evidence for involvement of Vps37A, a member of the endosomal sorting complex required for transport (ESCRT) system, in upper motor neuron disease. The ESCRT system has been shown to play a central role in intracellular trafficking, in the maturation of multivesicular bodies and the sorting of ubiquitinated membrane proteins into internal luminal vesicles. Further investigation of mechanisms by which dysfunction of this gene cause CHSP will contribute to the understanding of intracellular trafficking of vesicles by the ESCRT machinery and its relevance to CHSP.

200

Genome-wide association study identifies two novel susceptibility loci for musician's dystonia. *K. Lohmann[1], A. Schmidt[1], A. Schillert[2], S. Winkler[1], K. Siegesmund[1], H.-C. Jabusch[3], M. Kasten[1,4], J.L. Groen[5], C. Hemmelmann[2], J. Hagenah[1], J. Graf[1], N. Brüggemann[1], A. Grünewald[1], F. Baas[5], A. Münchau[6], K.E. Zeuner[7], S. Schreiber[8], G. Deuschl[7], M.A.J. de Koning-Tijssen[5], E. Altenmüller[9], A. Ziegler[2], C. Klein[1]*. 1) Section Neurogenetics at the Department of Neurology, University Lübeck, Lübeck, Germany; 2) Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany; 3) Institute for Musician's Medicine, University of Music, Dresden, Germany; 4) Department for Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany; 5) Department of Neurology, Academic Medical Center, Amsterdam, The Netherlands; 6) Department of Neurology, University Medical Centre Eppendorf, Hamburg, Germany; 7) Department of Neurology, Christian Albrechts University, Kiel, Germany; 8) Institute of Clinical Molecular Biology, Christian Albrechts University, Kiel, Germany; 9) Institute of Music Physiology and Musicians' Medicine, Hanover University of Music, Drama and Media, Hanover, Germany.

Dystonia is a movement disorder characterized by involuntary twisting, repetitive movements and abnormal postures. Focal dystonia is the most common form of dystonia and about 25% of patients have a positive family history. Little is known about genetic risk factors and no genome-wide association study (GWAS) has been reported for dystonia to date, possibly due to its marked phenotypic and genotypic heterogeneity which decreases the power to find novel loci. To identify genetic risk factors for dystonia, we performed a GWAS restricted to a homogeneous subtype of dystonia, i.e. musician's dystonia (MD), along with writer's dystonia (WD). While MD occurs when a (professional) musician is playing his instrument, WD is present during writing, a much more common task. Interestingly, WD often occurs in relatives of patients with MD or even in patients with MD after termination of their career. Both tasks, playing an instrument and writing, are accomplished through highly trained fine finger and hand movements. We genotyped 193 German patients (141 MD, 52 WD) and 1190 population-based controls from Germany (PopGen) on the Affymetrix® Genome-Wide Human SNP Array 6.0. Genotypes at 558245 SNPs in 177 patients (127 MD, 50 WC) and 984 controls passed stringent quality control. Analysis for association of single SNPs with MD/WD susceptibility was performed with additive SNP coding using linear logistic regression. A total of 18 SNPs from 11 different genetic regions revealed $P<10^{-5}$, and one SNP per region was put forward to the first replication phase using an independent sample of 116 German MD patients and 125 healthy musicians. The three SNPs with $P<0.05$ in this replication phase were selected for a second replication in 155 patients with WD and 278 newly collected population-based controls. Genome-wide significance ($P<5*10^{-8}$) was found for two intronic variants in genes on chromosome 17 (SNP1: $P=1.45*10^{-9}$; OR=3.12; 95% CI=2.16-4.51) and chromosome 9 (SNP2: $P=2.39*10^{-8}$; OR=1.66; CI=1.39–1.99), respectively. Both genes are neuronally expressed and their products function as enzymes in cell signaling and activation. This homogeneous patient group enabled identification of two novel risk factors for MD with large effect size that will influence both clinical practice and future research. Our data implicate novel pathways in the pathophysiology of dystonia.

201

Autosomal Recessive Axonal Neuropathy with Neuro-Myotonia: a novel disease entity caused by mutations in *HINT1*. *J. Baets[1,2,3], M. Zimon[2,4], L. Almeida-Souza[2,5], J. Nikodinovic[6], Y. Parman[7], E. Battaloglu[8], V. Guergueltcheva[9], I. Tournev[9], M. Auer-Grumbach[10], T. Müller[11], P. Van Damme[12], W.N. Löscher[13], N. Barisic[14], Z. Mitrovic[15], S.C. Previtali[16], H. Topaloglu[17], G. Bernert[18], A. Beleza-Meireles[19], S. Todorovic[6], B. Ishpekova[9], K. Peeters[2,4], A.F. Hahn[20], S. Züchner[21], V. Timmerman[2,5], V. Milic Rasic[6], A.R. Janecke[11,22], A. Jordanova[2,5,23], P. De Jonghe[1,2,3].* 1) Neurogenetics Group, VIB Department of Molecular Genetics, University of Antwerp, Belgium; 2) Neurogenetics laboratory, Institute Born-Bunge, University of Antwerp, Antwerpen, Belgium; 3) Department of Neurology, Antwerp University Hospital, Antwerp, Belgium; 4) Molecular Neurogenomics Group, VIB Department of Molecular Genetics, University of Antwerp, Belgium; 5) Peripheral Neuropathies Group, VIB Department of Molecular Genetics, University of Antwerp, Belgium; 6) Clinic for Neurology and Psychiatry for Children and Youth, University of Belgrade, Belgrade, Serbia; 7) Department of Neurology, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey; 8) Department of Molecular Biology and Genetics, Bogazici University, Istanbul, Turkey; 9) Department of Neurology, Medical University-Sofia, Sofia, Bulgaria; 10) Department of Internal Medicine, Division of Endocrinology and Metabolism, Medical University of Graz, Graz, Austria; 11) Department of Pediatrics II, Innsbruck Medical University, Innsbruck, Austria; 12) Department of Neurology, University Hospital Leuven, University of Leuven, Leuven, Belgium; 13) Department of Neurology, Innsbruck Medical University, Innsbruck, Austria; 14) Department of Paediatrics, University of Zagreb, Medical School, University Hospital Centre Zagreb, Zagreb,Croatia; 15) National Center for Neuromuscular Diseases, Department of Neurology, University Hospital Center Zagreb, Zagreb, Croatia; 16) Department of Neurology, San Raffaele Scientific Institute, Milano, Italy; 17) Department of Paediatric Neurology, Faculty of Medicine, Hacettepe University, Ankara, Turkey; 18) Department of Paediatrics, Gottfried von Preyer'sches Kinderspital, Vienna, Austria; 19) Department of Genetics, Coimbra Paediatric Hospital, CHUC-EPE, Portugal; 20) Department of Clinical Neurological Sciences, London Health Sciences Centre, University of Western Ontario, London, Ontario, Canada; 21) Hussman Institute for Human Genetics, University of Miami Miller School of Medicine, Miami, Florida, USA; 22) Division of Human Genetics, Innsbruck Medical University, Innsbruck, Austria; 23) Department of Medical Chemistry and Biochemistry, Molecular Medicine Center, Medical University-Sofia, Sofia, Bulgaria.

Inherited peripheral neuropathies are common neuromuscular disorders known for their clinical and genetic heterogeneity. Although striking clinical features may be of use to distinguish genetic subtypes, many families remain without molecular diagnosis to date. The association between hereditary peripheral neuropathy and neuromyotonia was previously recognized. Neuromyotonia is a syndrome characterized by delayed muscle relaxation due to hyper-excitability of the peripheral nerve axons. In the current study we identified mutations in *HINT1* (*histidine triad nucleotide binding protein 1*), a novel gene for autosomal recessive (AR) hereditary neuropathy. We report detailed clinical data based on standard methods including nerve conduction studies, concentric needle EMG, and muscle/nerve biopsies in a cohort of 50 patients from 33 unrelated nuclear families harboring *HINT1* mutations. A number of striking features was noted that together constitute a disease entity specific to *HINT1*-neuropathies. Disease onset was typically in the first decade of life, patients presented with progressive gait difficulties in combination with variable degrees of muscle cramps and weakness in hands. A distal pattern of weakness was present and 33/50 had distal sensory abnormalities. Strikingly, delayed muscle relaxation of the hands was noted in 36/50 patients. Nerve conduction studies were compatible with motor and sensory axonal neuropathy in the majority of patients with 13/50 having a pure motor axonal neuropathy. Concentric needle EMG showed high-frequency spontaneous motor unit action potentials (neuromyotonic discharges) in 39/50 patients. All 50 patients had clinical or electrophysiological signs of neuromyotonia or a combination of both. In a screening cohort of 262 unrelated index patients with AR hereditary neuropathy we found *HINT1* mutations in 11%. In a more specific subset of patients presenting with AR neuropathy and clinical or electrophysiological signs of neuromyotonia, the *HINT1* mutation frequency was 76%. Autosomal recessive axonal neuropathy with neuromyotonia (ARAN-NM) due to *HINT1* mutations is a new disease entity within the field of inherited peripheral neuropathies. This novel myotonic syndrome has to be distinguished from related disorders such as myotonic dystrophy and the non-dystrophic myotonias. Our findings have major implications for future molecular genetic testing in patients and families.

202

De novo gain of function KCNT1 channel mutations cause seizures and developmental delay in malignant migrating partial seizures of infancy. *G. Barcia[1,2], M.R. Fleming[3], A. Deligniere[1], V. Gazula[3], M.R. Brown[3], M. Langouet[4], H. Chen[5], J. Kronengold[3], A. Abhyankar[6], R. Cilio[7], P. Nitschke[4], A. Kaminska[1], N. Boddaert[1], J.L. Casanova[6], I. Desguerre[1], A. Munnich[4], O. Dulac[1,2], L.K. Kaczmarek[3], L. Colleaux[4], R. Nabbout[1,2].* 1) Department of Pediatric Neurology, Centre de Reference Epilepsies Rares, Department of Paediatric Radiology, Clinical Electrophysiology Unit, Hôpital Necker Enfants Malades, APHP, Paris, France; 2) Inserm U663, Université Paris Descartes, PRES Sorbonne Paris Cité, Hôpital Necker Enfants Malades, Paris, France; 3) Departments of Pharmacology and Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut, 06520, USA; 4) INSERM U781, Université Paris Descartes, Sorbonne Paris Cité, Institut IMAGINE, Hôpital Necker-Enfants Malades, Paris, France; 5) Department of Biological Sciences, State University of New York, Albany, NY 12222, USA; 6) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, New York, USA; 7) Division of Neurology, Bambino Gesu' Children's Hospital, IRCCS, Rome, Italy.

Malignant migrating partial seizures of infancy (MMPSI) is a rare epileptic encephalopathy of infancy characterized by the occurrence of polymorphous focal seizures and arrest of psychomotor development in the first 6 months of life. To identify the disease-causing gene, we collected DNA samples from 12 individuals fulfilling the criteria for MMPSI and performed exome sequencing in 3 probands and their unaffected parents. A single gene, KCNT1, was affected by distinct heterozygous missence variants (p.Arg428Gln and p.Ala934Thr) in 2 unrelated probands. Both mutations were confirmed by Sanger sequencing and occured de novo. Direct sequencing of KCNT1 in the 9 remaining patients identified the same de novo p.Arg428Gln missense mutations in two further patients and 2 distinct de novo mutations p.Arg474His and p.Ile760Met in two other patients KCNT1 encodes a sodium-activated potassium (KNa) channel widely expressed in the nervous system. Its activity contributes to the slow hyperpolarization that follows repetitive firing. It regulates the rates of bursting and enhances the accuracy with which action potentials lock to incoming stimuli. Electrophysiological studies demonstrated that MMPSI results from de novo gain-of-function mutations that led to constitutive activation of the channel, mimicking and occluding the effects of phosphorylation of the C-terminal domain by protein kinase C. Our findings further confirm the genetic etiology of MMPSI and ascribe MMPSI to the large family of channelopathies. In addition to regulating ion flux, the KCNT1 channel interacts with the mRNA binding protein, FMRP, whose defect results in the most common form of inherited intellectual disability, namely fragile X syndrome. We thus propose that mutations in KCNT1 provide the first pathophysiological basis of a channelopathy linking dysfunction of firing, thus epilepsy, to impaired function of a protein involved in cognitive development, FMRP, causing arrest of psychomotor development. Finally, this finding paves the road for a new pharmacological approach of MMPSI based on inhibition of this constitutively activated channel.

203

Investigating the genetic etiology of familial epilepsies using next-generation sequencing. *E.K. Ruzzo[1,3], E.L. Heinzen[1,3], R. Wedel[2], K.V. Shianna[1], I.E. Scheffer[4,5], S.F. Berkovic[4], R. Ottman[2,3], D.B. Goldstein[1,3]*. 1) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708, USA; 2) G. H. Sergievsky Center, Columbia University, New York, NY, USA; 3) Epigen Consortium; 4) Epilepsy Research Centre and Department of Medicine, The University of Melbourne, Austin Health, Heidelberg, Victoria, Australia; 5) Florey Neuroscience Institutes and Department of Pediatrics, The University of Melbourne, Royal Children's Hospital, Melbourne, Australia.

The genetic etiology of the vast majority of epilepsy cases is unknown. Epilepsy is highly heritable but also shows extreme locus heterogeneity, which confounds gene discovery in case-control cohorts. To reduce heterogeneity, we exploited multiplex epilepsy families (average of 6 affected per family). Our underlying hypothesis is that in some of these families, there is a single variant, rather than multiple segregating variants, that is responsible for all instances of epilepsy within the family. Furthermore, a gene harboring potentially pathogenic variants in multiple families is likely to influence epilepsy susceptibility. We performed whole-genome or exome sequencing (>30x) of 81 samples from 39 families, aligned the short-sequence reads, called SNVs and indels, and identified high quality variants that were likely to impact the encoded protein product and that were exceedingly rare in or completely absent from control genomes. In families with ≥2 sequenced cases, variants not seen in all cases were removed from the candidate list. Variants exclusively found in families or enriched in familial cases were evaluated for familial cosegregation. A few families had no cosegregating variants, while many families show multiple variants cosegregating with epilepsy. Not surprisingly, no single variant was found in a large number of families. However, a number of genes harbor extremely rare functional variants in ≥2 families. These genes, together with several known epilepsy genes and other interesting candidate genes (60 genes in all), are currently being sequenced using HaloPlex target enrichment in two cases from each of 240 additional multiplex families. This experiment will be critical since discovery of multiple families harboring pathogenic variants in any of these candidate genes will yield much greater statistical power than any individual family. Genes of high interest include: seizure threshold 2 (*SZT2*); CCR4-NOT transcription complex, subunit 3 (*CNOT3*); CNOT, subunit 1 (*CNOT1*); and glutamate receptor, ionotropic, N-methyl D-aspartate 2A (*GRIN2A*). These genes harbor rare likely damaging variants in four, three, one, and two families, respectively. The candidate genetic variants identified in this study likely include true risk factors for epilepsy and we are actively collecting the evidence needed to prove these associations by studying additional multiplex epilepsy families.

204

Autoregulation of the DYT6-gene THAP1. *A. Erogullari[1], P. Seibler[2], D. Braunholz[1], A. Grünewald[2], R. Depping[3], J. Eckhold[1], A. Rakovic[2], T. Lohnau[2], G. Gillessen-Kaesbach[1], C. Klein[2], K. Lohmann[2], F.J. Kaiser[1]*. 1) Human Genetics, University of Luebeck, Luebeck, Germany; 2) Section of Clinical and Molecular Neurogenetics at the Department of Neurology, University of Luebeck, Luebeck, Germany; 3) Institute of Physiology, Center for structural and cell biology in Medicine, University of Luebeck, Luebeck, Germany.

Mutations in the THAP1 gene cause the primary torsion dystonia DYT 6. This type of dystonia belongs to a heterogeneous group of movement disorders characterized by sustained involuntary muscle contractions, repetitive movements and abnormal postures of the affected body parts. THAP1 encodes a transcription factor of 213 amino acids with a characteristic THAP zinc finger domain at its N-terminus which mediates specific binding of THAP1 to bipartite THAP-binding sequences (THABS). Previously, we and others demonstrated that THAP1 specifically binds to the TorsinA (TOR1A) promoter to regulate TOR1A gene expression, the gene mutated in patients with DYT1 dystonia. Searching for other target genes of THAP1 we gained evidence for regulation of its own promoter in an autoregulative manner. As a preliminary step we characterized the THAP1 promoter region by using in-silico prediction and luciferase reporter gene assays. Interestingly, five THABSs are localized in the THAP1 core promoter. The specific binding of THAP1 to its core promoter region was demonstrated by chromatin immunoprecipitation (ChIP) analysis in human neuroblastoma cells (SH-SY5Y) and electromobility shift assays (EMSA). Luciferase reporter gene assays revealed strong repression of the THAP1 core promoter activity by wildtype THAP1 whereas DYT6-causing mutations in THAP1 significantly decreased THAP1-mediated repression. To test for in-vivo changes on endogenous expression levels, we re-programmed fibroblast cells from a THAP1 mutation carrier (Leu159fs180X) and controls to induced pluripotent stem (iPS) cells that were subsequently differentiated into neurons. Quantitative PCR in these cells revealed a significant increase of THAP1 expression in mutant THAP1 cells as compared to the wildtype cells suggesting an autoregulation of THAP1 in vivo. We identified a biological feedback-loop in the regulation of THAP1 expression. Interestingly, this may compensate for expressional changes of dystonia-related THAP1 target genes caused by THAP1 mutations, at least for several years until a certain biological threshold is exceeded. Notably, the mean onset of DYT6 dystonia is at 16 years. This hypothesis needs to be further explored to better understand the biological function of the THAP1 protein.

205

Identification of 23 novel prostate cancer susceptibility loci using a custom array (the iCOGS) in an international consortium, PRACTICAL. *R. Eeles[1], A. Amin Al Olama[2], S. Benlloch[2], E. Saunders[1], D. Leongamornlert[1], M. Tymrakiewicz[1], M. Ghoussaini[2], C. Luccarini[2], J. Dennis[2], S. Jugurnauth-Little[1], T. Dadaev[1], The. PROTECT GROUP[3], G. Giles[4], G. Severi[4], F. Wiklund[5], H. Gronberg[5], C. Haiman[6], F. Schumacher[6], B. Henderson[6], L. Le Marchand[7], S. Lindstrom[8], P. Kraft[8], D. Hunter[8], S. Gapstur[9], S. Chanock[10], S. Berndt[10], The. PRACTICAL CONSORTIUM[1], Z. Kote-Jarai[1], D. Easton[2].* 1) Cancer Gen Unit, Inst Cancer Res, Sutton, Surrey, United Kingdom; 2) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, UK; 3) Surgical Oncology (Uro-Oncology: S4), University of Cambridge, Box 279, Addenbrooke's Hospital, Hills Road, Cambridge, UK and Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK; 4) Cancer Epidemiology Centre, The Cancer Council Victoria, 1 Rathdowne street, Carlton Victoria, Australia; 5) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, US; 7) University of Hawaii Cancer Center, Honolulu, HI, US; 8) Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Pubic Health, Boston, MA, US; 9) Epidemiology Research Program, American Cancer Society, Atlanta, GA, US; 10) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, 20892, US.

Prostate cancer (PrCa) is the most frequently diagnosed male cancer in developed countries. To identify common PrCa susceptibility alleles, we genotyped 211,155 SNPs on a custom chip (iCOGS) in blood DNA from 25,074 PrCa cases and 24,272 controls from 32 studies in the international PRACTICAL consortium. The chip was composed of GWAS hits from a meta analysis of 4 GWAS [11,085 cases and 11,463 controls from UK,Australia, Cancer Genetic Markers of Susceptibility (CGEMS); Cancer of the Prostate in Sweden (CAPS) and the Breast and Prostate Cancer Cohort Consortium (BPC3)], fine mapping of known regions and genes in candidate pathways. Results from the GWAS component are reported here. Twenty-three new PrCa susceptibility loci were identified at genome-wide significance (P<5×10-8). Neighbouring genes include MDM4, NOTCH4, MMP7, RAD51L1, and HOXB13. There was no strong evidence for heterogeneity in the per-allele OR among studies. All alleles are common (MAF 8–49%) with per allele ORs from 1.06 to 1.15. All but two exhibit a pattern of association consistent with a log-additive model, as observed for most common cancer susceptibility alleles. Aggressive disease was defined as Gleason score > 8, PSA >100, disease stage of distant (i.e. outside the pelvis) or death from PrCa. The majority of SNPs showed clear association when analysis was restricted to aggressive disease (13 SNPs significant at P<.01) and for 22 of the 23 SNPs the estimated OR was in the same direction for aggressive and non-aggressive disease. Two SNPs were associated with PSA level in controls. Six SNPs showed a trend in OR with age at diagnosis, with a higher OR at younger ages. GeneGo pathway enrichment analysis showed the most strongly associated pathways were cell adhesion and extracellular matrix (ECM) remodelling (P=2.6×10-6), epidermal growth factor receptor (EGFR) signaling and immune-response. Comparison of the direction of effects in the GWAS and COGS replication stages indicate that more than 1500 loci on the array are associated with this disease. More than 70 PrCa susceptibility loci, explaining 30% of the familial risk in this disease, have now been identified. Based on the combined risks conferred by these loci, the top 1% of the risk distribution has a 4.7fold increased risk compared with the average of the population. These results will facilitate population risk stratification for clinical studies.

206

Large-scale genotyping identifies more than 40 novel breast cancer susceptibility loci. *K. Michailidou[1], P. Hall[2], A. Gonzalez-Neira[3], M. Ghoussaini[4], J. Dennis[1], R.L. Milne[5], M.K. Schmidt[6], J. Chang-Claude[7], S.E. Bojesen[8,9], M.K. Humphreys[1], Q. Wang[1], M. Garcia-Closas[10], P.D.P. Pharoah[1,4], G. Chenevix-Trench[11], A.M. Dunning[4], J. Benitez[12,13], D.F. Easton[1,4], Breast Cancer Association Consortium.* 1) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 2) Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden; 3) Human Genotyping Unit, Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; 4) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 5) Genetic & Molecular Epidemiology Group, Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; 6) Division of Molecular Pathology, Netherlands Cancer Institute, Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands; 7) Division of Cancer Epidemiology, Deutsches Krebsforschungszentrum, Heidelberg, Germany; 8) Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark; 9) Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark; 10) Division of Genetics and Epidemiology and Breakthrough Breast Cancer Centre, The Institute of Cancer Research, Sutton, Surrey, UK; 11) Queensland Institute of Medical Research, Brisbane, Australia; 12) Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; 13) Centro de Investigación en Red de Enfermedades Raras (CIBERER), Spain.

Breast cancer is the most common cancer among women. To date, 26 breast cancer susceptibility loci have been identified, accounting for ~ 9% of the familial risk of the disease. To identify novel breast cancer susceptibility loci, we conducted a meta-analysis of nine genome-wide association studies that included 10,052 breast cancer cases and 12,575 controls of European origin, from which we selected 29,675 SNPs for replication, based on evidence for association using 1df trend test, a test placing increased weight for cases with positive family history or a 2df genotype test. These SNPs were genotyped in 45,290 cases and 41,880 controls of European origin in the Breast Cancer Association Consortium (BCAC), as part of a collaborative genotyping experiment involving four consortia (COGS). We identified SNPs at 41 novel loci at genome-wide significance ($P<5\times10^{-8}$). Pathway analyses indicates that these loci are overrepresented in genes involved in DNA stability or DNA repair, and genes involved in tumor aggressiveness, tumorigenesis or metastasis in model systems. Thirteen of the loci exhibited stronger associations for estrogen receptor (ER)-positive than for ER-negative tumors. Based on the per-allele OR estimates from the BCAC replication data, the new 41 loci explain approximately an additional 5% of the familial risk of breast cancer. Comparison of the direction of the effects in the GWAS and replication stages suggest that more than 1,000 further loci are involved in breast cancer susceptibility.

207

Genome-wide association study in *BRCA1* mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *A.C. Antoniou[1], X. Wang[2], L. McGuffog[1], A. Lee[1], M.M. Guedet[3], K.B. Kuchenbaecker[1], P. Soucy[4], J. Simard[4], K. Offit[5], D.F. Easton[1], G. Chenevix-Trench[6], F.J. Couch[7], Consortium of Investigators of Modifiers of BRCA1/2.* 1) Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic,Rochester, MN, USA; 3) Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA; 4) Cancer Genomics Laboratory, Centre Hospitalier, Universitaire de Québec and Laval University, Quebec City, Canada; 5) Clinical Genetics Service, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 6) Genetics Department, Queensland Institute of Medical Research, Brisbane, Australia; 7) Department of Laboratory Medicine and Pathology, and Health Sciences Research, Mayo Clinic, Rochester, MN, USA.

BRCA1 associated breast and ovarian cancer risks can be modified by common genetic variants. Through a 2-stage GWAS of 11,705 *BRCA1* mutation carriers from the Consortium of Investigators of Modifiers of *BRCA1/2* (5,920 with breast, 1,839 with ovarian cancer), and replication in 2,646 *BRCA1* carriers, we identified two novel ovarian cancer risk modifier loci: 17q21.31 ($P=1.4\times10^{-8}$, per-allele HR=1.27,95%CI:1.17–1.38) and 4q32.3 ($P=3.4\times10^{-8}$, per-allele HR=1.20, 95%CI:1.17–1.38). The 4q32.3 locus was not associated with ovarian cancer risk in the general population (Odds Ratio=1.00, 95%CI:0.97–1.04, P=0.76) or *BRCA2* carriers (HR=1.08, 95%CI: 0.96–1.21, P=0.22), suggesting a *BRCA1*-specific association. The 17q21.31 locus was also associated with ovarian cancer risk in 8,211 *BRCA2* carriers ($P=2\times10^{-4}$), with several SNPs providing $P=2\times10^{-8}$-8×10^{-11} for association in *BRCA1* and *BRCA2* carriers combined. In addition, we identified a novel breast cancer risk modifier locus at 1q32 for *BRCA1* carriers ($P=2.7\times10^{-8}$, per-allele HR=1.14, 95%CI:1.09–1.20). Based on the joint distribution of the 7 known ovarian cancer risk modifying loci, we estimated that the 5% of *BRCA1* mutation carriers at lowest risk will have a lifetime risk of developing ovarian cancer of 28% or lower whereas the 5% at highest risk will have a lifetime risk of 63% or higher. Similarly, based on the 10 known breast cancer risk modifying loci, the breast cancer lifetime risks for the 5% of *BRCA1* carriers at lowest risk are predicted to be 28–50% compared to 81–100% for the 5% at highest risk. Such differences in risk may have important practical implications for risk prediction and clinical management for *BRCA1* carriers.

208

Identification of the first locus to modify breast cancer risk specifically in BRCA2 mutation carriers. *K. Kuchenbaecker[1], M. Gaudet[2], J. Vijai[3], R. Klein[4], T. Kirchhoff[5], L. McGuffog[1], D. Barrowdale[1], A. Dunning[6], A. Lee[1], P. Hall[7], F. Couch[8,9], J. Simard[10], D. Altshuler[11,12,13], D. Easton[1,6], G. Chenevix-Trench[14], A. Antoniou[1], K. Offit[3], Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) collaborators.* 1) Centre for Cancer Genetic Epidemiology, Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 2) Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA; 3) Clinical Genetics Service, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 4) Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 5) Division of Epidemiology, Department of Environmental Medicine, New York University School of Medicine, New York, NY, USA; 6) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 7) Division of Breast Cancer Research, Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, UK; 8) Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; 9) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; 10) Cancer Genomics Laboratory, Centre Hospitalier Universitaire de Québec and Laval University, Quebec City, Canada; 11) Department of Molecular Biology and Medicine, Massachusetts General Hospital, Boston, MA, USA; 12) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 13) Departments of Genetics and Medicine, Harvard Medical School, Boston, MA, USA; 14) Genetics and Population Health Division, Queensland Institute of Medical Research, Brisbane, Australia.

Common genetic variants contribute to the observed variation in breast cancer risk for BRCA2 mutation carriers. To date, all identified common genetic modifiers have been found through population-based genome-wide association studies (GWAS). To seek additional breast cancer risk modifying loci, we carried out a two-stage GWAS using samples from BRCA2 mutation carriers participating in the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). A total of 19,029 SNPs identified through stage 1 of the GWAS were genotyped in 10,048 samples from female BRCA2 mutation carriers. Associations with breast cancer risk were evaluated using a score test statistic based on modelling the retrospective likelihood of the observed genotypes given the disease phenotypes. Analyses were stratified by country of residence and Ashkenazi Jewish ancestry. After quality control filtering, data from 18,086 SNPs were available for analysis in 3,881 BRCA2 mutation carriers diagnosed with breast cancer and 4,330 unaffected BRCA2 mutation carriers. For six regions previously shown to be associated with breast cancer risk for BRCA2 mutation carriers, including FGFR2 and PTHLH, we found SNPs with p-values smaller than those previously reported. We identified a novel susceptibility locus at 6p24 (per allele HR=0.85, 95% CI 0.80–0.90, P=3.9×10-8). The locus was not associated with breast cancer risk either in the general population, based on 42,599 breast cancer cases and 46,451 controls from the Breast Cancer Association Consortium (OR= 1.00, 95%CI: 0.98–1.02, P=0.74), or breast cancer risk for BRCA1 mutation carriers (HR=0.99, 95%CI: 0.94–1.04, P=0.75), based on 5,920 BRCA1 carriers with breast cancer and 5,783 without from CIMBA. The 6p24 locus lies within a region containing TFAP2A, which encodes a transcriptional activation protein that interacts with tumor suppressor genes. Our results suggest that 6p24 is the first modifier locus that is associated with breast cancer risk specifically in BRCA2 mutation carriers. This locus might provide further insight into the biology of breast cancer development for these women.

209

Fine-scale mapping and functional analysis of the breast cancer 11q13 (CCND1) locus. *M. Ghoussaini[1,7], K.B. Meyer[2,7], S. Edwards[3,7], J.D. French[3,7], K. Michailidou[4], S. Ahmed[1], S. Khan[5], M.J. Maranian[1], C.S. Healey[1], P.D.P. Pharoah[4], H. Nevanlinna[5], M.A. Brown[3], G. Chenevix-Trench[6], D.F. Easton[4], A.M. Dunning[1], BCAC.* 1) Centre for Cancer Genetic Epidemiology, Department of Oncology, Univ Cambridge, Cambridge, United Kingdom; 2) Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK; 3) School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia; 4) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 5) Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 6) Department of Genetics, Queensland Institute of Medical Research, Brisbane, Australia; 7) These authors contributed equally to the work.

We previously identified a variant on 11q13 associated with breast cancer risk through a genome-wide association study (rs614367; OR = 1.21; 95% CI 1.17–1.24; p=10-39). In an attempt to determine the causal variant(s) underlying this association, we analysed 4,405 genotyped and imputed variants in 41 case-control studies within the Breast Cancer Association Consortium (BCAC) which includes 89,050 European subjects. We identified three independent association signals; in each case the SNPs were associated with Estrogen receptor (ER)-positive but not ER-negative breast cancer. The strongest signal maps to a transcriptional enhancer element. SNP rs554219, the best candidate causal variant [Odds Ratio (OR) per allele = 1.33; 95% Confidence Interval (95%CI) 1.28–1.37; p-value 10–66], displays differential binding of transcription factor ELK4, reduces the enhancer activity in luciferase assays and is associated with reduced cyclin D1 (CCND1) levels in tumours. A second, independent SNP, rs78540526, lies in the same enhancer element and also appears to be functional. A third SNP, rs75915166 [OR per minor allele = 1.38; 95% CI 1.32–1.44; p-value10-46], creates a GATA3 binding site within a silencer element. The effect sizes of the three SNPs are the largest of any GWAS-discovered breast cancer locus identified to date and account for ~2% of the familial risk of breast cancer in Europeans. Rs75915166 appeared monomorphic in 12,893 subjects of Asian ancestry from 9 case-control studies, but the risk alleles of rs554219 (OR=1.64; p-value=1.3×10-4 and rs75915166 (OR=1.42, p-value=3.6×10-2), while rare in Asians (<1%), showed similar effect sizes to Europeans, providing further support that these SNPs may have directly causative effects. Finally, we identified, using chromatin conformation studies, a long-range physical interaction restricted to ER+ breast cancer cells between the enhancer and silencer elements and with CCND1, an oncogene with key role in cell cycle regulation. We conclude that these three variants are likely to be causally related to breast cancer risk and act by controlling CCND1 expression.

210

Three independent loci within the TERT-CLPTM1L locus associated with telomere length and risk of breast and ovarian cancer. *G. Chenevix-Trench[1], S.E. Bojesen[2, 3], K.A. Pooley[4], S. Johnatty[1], J. Beesley[1], K. Michailidou[4], J. Tyrer[4], S.L. Edwards[5], H.C. Shen[6], K. Lawrenson[6], H. Pickett[7,8], M. Stutz[7], C. Smart[9], J. French[5], P.L. Mai[10], M.H. Greene[10], S. Gayther[6], R. Reddel[7,8], P.D.P. Pharoah[4], E.L. Goode[11], A. Berchuk[12], D.F. Easton[4], A.C. Antoniou[4], A.M. Dunning[4] on behalf of CIMBA, OCAC and BCAC.* 1) Department of Genetics & Computational Biology, Queensland Institute of Medical Research, Brisbane, QLD, Australia; 2) Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark; 3) Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark; 4) Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK; 5) School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA; 7) Cancer Research Unit, Children's Medical Research Institute, Westmead, NSW, Australia; 8) Sydney Medical School, University of Sydney, Sydney, NSW, Australia; 9) The UQ Centre for Clinical Research, The University of Queensland, Royal Brisbane and Women's Hospital, Herston, QLD, Australia; 10) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA; 11) Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA; 12) Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA.

Germline mutations in TERT cause dyskeratosis congenita, a cancer susceptibility disorder in which affected persons have exceedingly short telomeres. Common variants around TERT have been associated with multiple cancer types; these associations might be mediated through changes of telomere length (TL). To identify variants within this region associated with TL and breast or ovarian cancer risk, 110 SNPs were genotyped in 103,991 breast cancer cases and controls, 44,308 ovarian cancer cases and controls, and 11,705 BRCA1 mutation carriers. TL was determined in whole blood DNA from 53,645 individuals. We found three independent peaks of association with the various phenotypes. Peak 1, in the TERT promoter, contained SNPs associated with TL (P=5.8×10-7), overall breast cancer risk (P=4.7×10-10), estrogen-negative breast cancer risk (P=1.2×10-9) and cancer risk for BRCA1 mutation carriers (P=1.1×10-5). Luciferase assays showed that a construct carrying the risk alleles of three highly correlated, candidate functional variants in the TERT promoter abrogated reporter activity, suggesting the risk alleles act by reducing TERT expression. A SNP in Peak 2, which spans TERT introns 2–4, showed the strongest association with TL (P=2.3×10–14) and risk of serous low malignant potential (LMP) ovarian cancer (P=1.3×10–15). SNPs in Peak 3, also spanning TERT introns 2–4, displayed the strongest associations with estrogen receptor negative breast cancer (P=1.2×10–12), breast cancer risk for BRCA1 mutation carriers (P=1.6×10–14) and serous invasive ovarian cancer (P=1.3×10–11) but were not associated with TL. Luciferase assays showed that the regions around Peaks 2 and 3 can act as silencers and that one of the risk-associated SNPs in Peak 3 decreases relative luciferase signal by ~30%. Analysis of transcripts produced from a minigene construct showed that one of the risk-associated SNPs in Peak 3 causes expression of a previously unreported TERT mRNA splice variant, and analysis of chromatin structure revealed this SNP occupies a novel site of potential regulatory activity in stromal and myoepithelial cells from reduction mammoplasty samples. Our results demonstrate three independent peaks within TERT that are associated with TL, breast cancer, serous LMP and/or serous invasive ovarian cancer risk. However, functional studies indicate that only one of the three association peaks supports the hypothesis that increased cancer risk is mediated through shorter telomeres.

211

Statistical fine mapping of regions containing melanoma susceptibility genes identified through genome-wide association studies. *J.H. Barrett[1], J.C. Taylor[1], M. Brossard[2], A.M. Goldstein[3], P.A. Kanetsky[4], E.M. Gillanders[5], J.A. Newton Bishop[1], D.T. Bishop[1], F. Demenais[2], M.M. Iles[1], GenoMEL consortium.* 1) Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, St James's University Hospital, University of Leeds, Leeds, United Kingdom; 2) INSERM, U946, Fondation Jean-Dausset-CEPH, Paris, France; 3) Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute. National Institutes of Health, Bethesda, Maryland, USA; 4) Centre for Clinical Epidemiology & Biostatistics, University of Pennsylvania, Philadelphia, Pennyslvania, USA; 5) Inherited Disease Research Branch, National Human Genome Research Institute, US National Institutes of Health, Baltimore, Maryland, USA.

Genome-wide association (GWA) studies of melanoma have resulted in the identification and confirmation of approximately 15 loci associated with disease risk. In most of these regions the causal variant(s), and sometimes even which genes are implicated, are still not known. We have applied statistical methods to data from the GenoMEL GWA study (Barrett et al, Nature Genetics, 2011; 43:1108-13) to refine each locus of interest and identify the most parsimonious model(s) explaining the association. Genotypes were imputed in at least 2Mb around each locus using IMPUTEv2 (Howie et al, PLoS Genetics, 2009) with 1000 Genomes (March 2012 release) as reference panel after stringent quality control. All genotyped single nucleotide polymorphisms (SNPs) and imputed SNPs with an INFO score >0.8 were analysed using a gene dosage approach in standard single locus logistic regression analysis adjusting for population structure. This analysis was repeated conditioning on the top genotyped and then the top imputed SNP. Hyperlasso (Hoggart et al, PLoS Genetics, 2008), which implements a form of penalized logistic regression, was applied to all SNPs in the region to select a set of associated SNPs, and the model including these selected SNPs was further characterised using multiple logistic regression. In the 16q24.3 region close to *CDK10* and *MC1R* this approach allowed us to show (and confirm by sequencing) that the signal is explained by the known functional variants in *MC1R*, despite the fact that the original signal is closer to other genes. The results obtained varied markedly across the other loci. At one extreme, for the region on 11q14-q21 around the tyrosinase gene, the association was completely explained by one SNP (the most strongly associated imputed SNP, which is in strong linkage disequilibrium (r^2=0.91) with the missense variant rs1126809). For over half the loci a single SNP explaining the association could not be identified. For example, for the 5p15.33 region around *TERT* and *CLPTM1L*, despite the fact that the initial signal was confined to a narrow region, the simplest model included a two-SNP haplotype and possibly one other SNP, suggesting either a more complex mechanism or the involvement of SNP(s) neither genotyped nor imputed. Statistical fine mapping is a useful first step in the attempt to identify the causal mechanisms giving rise to association signals. Next steps will include sequencing of more narrowly defined regions and functional experiments.

212

Combining expression phenotypes with high density imputation to identify melanoma risk genes. *M. H. Law[1], G. W. Montgomery[1], K. M. Brown[2], A. E. Cust[3], N. G. Martin[1], G. J. Mann[4], N. K. Hayward[1], S. MacGregor[1], Q-MEGA and AMFS Investigators.* 1) Queensland Institute of Medical Research, Brisbane, QLD, 4029, Australia; 2) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; 3) Cancer Epidemiology and Services Research, Sydney School of Public Health, The University of Sydney, NSW, 2006, Austraia; 4) Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, NSW 2145 Australia.

Statement of purpose GWAS have successfully identified a large number of genetic loci associated with melanoma, and while for many the underlying gene has been identified others require further dissection. We recently reported on a melanoma GWAS using samples from the Queensland study of Melanoma: Environment and Genetic Associations (Q-MEGA) and the Australian Melanoma Family study (AMFS). One novel loci we identified, 1q21.3, contains a large number of plausible melanoma candidate genes including *ARNT* [MIM: 126110] and *SETDB1* [MIM: 604369]. Re-imputation of our GWAS data to the latest 1000 genomes reference will allow inclusion of newly collected samples and a greater coverage depth allowing improved fine mapping of 1q21.3 and previously identified loci with complex LD structures (e.g. 20q11.2). Even with assignment of the signal to a single gene, functional confirmation of genetic associations is essential. Common genetic variants can influence gene expression, and combining GWAS data with genome wide expression data from melanoma cell lines can both assist with fine-mapping and strengthen initial genetic associations.

Methods AMFS and Q-MEGA sample sets were sub-grouped by genotype array to minimise stratification and individual batch effects. Sub-groups were imputed to the April 2012 update of 1000 genomes reference panel using IMPUTE v2.2.0. Individual imputed datasets were analysed using SNPTEST, with imputed dosage scores and genotyped variants treated as having additive effects. Stratification was corrected by principal components analysis. The resulting high density datasets were combined by meta-analysis. Expression data for melanoma cell lines were normalised and underwent quality control using the R BioConductor package. Melanoma associated genetic variants were analysed for association with the expression of cis (within 1Mb) genes.

Results and Conclusion In the multi-gene 1q21 loci independent animal data and gene based association test of the Q-MEGA/AMFS GWAS data points towards a role for *SETDB1*, whereas preliminary expression analysis suggest the adjacent *ARNT* gene. This analysis will be expanded to the other loci associated with melanoma.

213

Meta-analysis identifies four new loci for testicular germ cell tumor.

C.C. Chung[1], Z. Wang[1,2], P.A. Kanetsky[3], C. Turnbull[5], K. McGlynn[1], R.L. Erickson[16], M.H. Greene[1], M.A.T. Hildebrandt[6], R.I. Skotheim[7,8], C. Kratz[1], M.B. Cook[1], F. Schumacher[9], R. Koster[10], M. Yeager[1,2], K.B. Jacobs[1,2], S.M. Schwartz[11,12], D.T. Bishop[13], H.K. Gjessing[14,15], V. Cortessis[9], N. Rahman[5], X. Wu[6], S.J. Chanock[1], K.L. Nathanson[4]. 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA; 2) Core Genotyping Facility, Division of Cancer Epidemiology and Genetics, Advanced Technology Program, SAIC-Frederick Inc., NCI-Frederick, Frederick, Maryland, USA; 3) Department of Biostatistics and Epidemiology, Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA; 4) Department of Medicine, Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA; 5) Division of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, UK; 6) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA; 7) Department of Cancer Prevention, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway 8Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway; 8) Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway; 9) Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA; 10) Department of Medical Oncology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 11) Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA; 12) Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 13) Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James' University Hospital, Leeds, UK; 14) Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; 15) Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway; 16) Global Emerging Infections System, Walter Reed Army Institute of Research, Silver Spring, Maryland, USA.

We conducted a two-stage meta-analysis to identify new loci for testicular germ cell tumor (TGCT). In the discovery phase, meta-analysis of 931 affected individuals and 1,975 controls from three genome-wide association studies (GWAS), two from the National Cancer Institute and one previously reported from the University of Pennsylvania was performed. After excluding known susceptibility loci, 17 tagging SNPs from nine regions ($P < 10^{-4}$) were selected for follow-up in six independent sample sets totaling 3,215 affected individuals and 7,562 controls. In combined analysis, TGCT status was significantly associated with four new loci: 4q22.2 (rs17021463; $P = 1.11 \times 10^{-8}$, per allele OR 1.19, 95%CI 1.12–1.26) in an intron of hematopoietic prostaglandin D synthase gene, *HPGDS*; 7p22.3 (rs12699477; $P = 5.59 \times 10^{-9}$, per allele OR 1.21, 95%CI 1.14–1.29) in an intron of mitotic arrest deficient-like 1, *MAD1L1*; 16q22.3 (rs4888262; $P = 7.72 \times 10^{-9}$, per allele OR 1.29, 95%CI 1.18–1.40) in exon 8 of the ring finger WD domain 3, *RFWD3*; and 17q22 (rs9905704; $P = 4.32 \times 10–13$, per allele OR 0.79, 95%CI 0.75–0.85 and rs7221274; $P = 4.04 \times 10^{-9}$, per allele OR 0.83, 95%CI 0.78–0.89) in introns of the testis expressed 14 gene, *TEX14*, and the protein phosphatase Mg^{2+}/Mn^{2+} dependent 1E gene, *PPM1E*, respectively, between the two of which is *RAD51C*, a DNA repair gene in which truncating mutations have been demonstrated to confer susceptibility to ovarian cancer. We have identified four new TGCT susceptibility loci and plan to initiate fine mapping to nominate the most promising variants for functional studies.

214

Next generation sequencing detects mutations in *ISPD* as a common cause of Walker-Warburg syndrome with defective glycosylation of α-dystroglycan. *T. Roscioli[1,2], E-J. Kamsteeg[1], K. Buysse[1], I. Maystadt[3], J. van Reeuwijk[1], C. van den Elzen[1], E. van Beusekom[1], M. Riemersma[1,4], R. Pfundt[1], L.E.L.M. Vissers[1], M. Schraders[1,5], M.F. Buckley[1,2], H.G. Brunner[1], H. Zhou[1], J.A. Veltman[1], C. Gilissen[1], G.M.S. Mancini[6], M.A. Willemsen[4], D. Petković Ramadža[7], D. Chitayat[8,9], C. Bennett[10], E. Sheridan[10], E.A.J. Peeters[11], G.M.B. Tan-Sindhunata[12], H. Kayserili[13], O. Abd El-Fattah El-Hashash[14,15], D.L. Stemple[16], D.J. Lefeber[4, 17], Y-Y. Lin[16], H. van Bokhoven[1]*. 1) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 2) School of Women's and Children's Health, Sydney Children's hospital and the University of New South Wales, Sydney, Australia; 3) Centre de Génétique Humaine, Institut de Pathologie et de Génétique, Gosselies, Belgium; 4) Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Department of Otorhinolaryngology, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen; 6) Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 7) Department of Pediatrics, University Hospital Centre, Zagreb, Croatia; 8) Mount Sinai Hospital, The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, University ofToronto, Toronto, Canada; 9) The Hospital for Sick Children, Division of Clinical and Metabolic Genetics, Toronto, Canada; 10) Department of Clinical Genetics, St James's University Hospital, Leeds, United Kingdom; 11) Medisch Centrum Haaglanden, Den Haag, The Netherlands; 12) Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands; 13) Medical Genetics Department, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey; 14) Pediatric department, Farwaniya Hospital, KuwaitCity, Kuwait; 15) Ministry of Health, Cairo, Egypt; 16) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB101SA, United Kingdom; 17) Laboratory for Genetic, Endocrine and Metabolic Disease, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

We present the results of exome sequencing identifying mutations in a novel gene, *ISPD*, as a common cause of Walker-Warburg syndrome (WWS). Bacterial IspD (encoding isoprenoid synthase domain containing) is a nucleotidyl transferase belonging to a large glycosyltransferase family, but the role of the orthologous protein in chordates is obscure to date, as this phylum does not have the corresponding non-mevalonate isoprenoid biosynthesis pathway. WWS is an autosomal recessive multisystem disorder characterized by eye and brain abnormalities with congenital muscular dystrophy and aberrant α-dystroglycan (αDG) glycosylation. Genotyping thirty WWS patients using dense human SNP and CGH arrays identified three affected individuals with homozygous deletions affecting *ISPD*. Concurrently, exome sequencing and filtering based on an autosomal recessive model identified a homozygous *ISPD* mutation in a region of shared homozygosity in three affected members of a WWS family. This c.647C>A transversion predicting a p.Ala216Asp substitution, showed complete segregation in the family. Additional sequencing identified mutated alleles in five additional families. Knockdown of the zebrafish orthologue recapitulates the major aspects of the human WWS phenotype with hypoglycosylation of αDG. Moreover, co-injection experiments support a co-operative interaction between *ispd*, *fukutin* and *fkrp* in αDG glycosylation in zebrafish. These results implicate a role for *ISPD* in αDG glycosylation in maintaining sarcolemma integrity in vertebrates.

215

The identification of a novel gene identified by exome sequencing reveals the upstream components of the RAS/MAPK signaling pathway involved in Noonan syndrome. *H. Yntema[1], W. Nillesen[1], J. Paardekooper Overman[2], M. Bonetti[2], J. de Ligt[1], H. Venselaar[3], M. Tartaglia[4], S.G.M. Frints[5], L.E.L.M. Vissers[1], J. den Hertog[2,6], I. van der Burgt[1]*. 1) Human Genetics, Radboud University Nijmegen Medical Cente, Nijmegen, Netherlands; 2) Hubrecht Institute-KNAW and University Medical Center, Utrecht, The Netherlands; 3) Nijmegen Centre for Molecular Life Sciences (CMBI), Nijmegen, The Netherlands; 4) Dipartimento di Biologia Cellulare e Neuroscienze, Istituto Superiore di Sanità, Rome, Italy; 5) Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, The Netherlands; 6) Institute of Biology, Leiden, The Netherlands.

Background Noonan syndrome (NS) is a relatively common, clinically variable developmental disorder. Cardinal features include postnatally reduced growth, distinctive facial dysmorphism, congenital heart defects and hypertrophic cardiomyopathy, variable cognitive deficit and skeletal, ectodermal and hematologic anomalies. Noonan syndrome is transmitted as an autosomal dominant trait, and is genetically heterogeneous. So far, heterozygous mutations in nine genes (*PTPN11*, *SOS1*, *KRAS*, *NRAS*, *RAF1*, *BRAF*, *SHOC2*, *MAP2K1* and *CBL*) have been documented to underlie this disorder or clinically related phenotypes. Although molecular testing of these genes now confirms a diagnosis in approximately 75% of affected individuals, still a lot of patients with an obvious Noonan syndrome phenotype are unexplained. An exome sequencing approach was used to identify new candidate genes for NS. **Results** Exome sequencing was performed in seven individuals with NS without a mutation in the known Noonan syndrome genes. In one patient, we identified and confirmed a unique non-synonymous de novo mutation in a gene encoding an extracellular protein that can bind to the low density lipoprotein receptor-related protein 1 (LRP1). The fact that LRP1 is, upon ligand binding, capable of activating the Mitogen-activated protein kinase (MAPK)/Extracellular signal-regulated kinase (ERK) cascade, provides an essential link between our novel gene and the pathway known to be involved in NS. Mutation analysis of this gene in a cohort of 35 Noonan syndrome patients revealed three additional pathogenic mutations. Expression of these mutations in zebrafish embryos resulted in developmental defects that are comparable to defects induced by mutations in other known Noonan syndrome genes. **Conclusions** Our findings provide strong experimental support for a novel gene involved in Noonan syndrome and provide insight in the mechanism upstream of the RAS-MAPK pathway. *Note: The name of the gene will be disclosed at the meeting.*.

216

DYNC2H1 Mutations are commonly found in Jeune Asphyxating Thoracic Dysplasia (JATD) without extraskeletal features while IFT140 mutations cause JATD with renal involvement. *M. Schmidts[1], H.H. Arts[2], Z. Yap[1], E.M.H.F Boengers[2], D. Anthony[1], M.M. Oud[2], S. alTurki[3], L. Duijkers[2], J. Stalker[3], J.B. Yntema[4], A. Hoischen[2], R. Bogdanovic[5], A. Peco-Antic[6], C. Gillisen[2], H. Kayserili[7], I. Veltman[2], A. Kutkowska[8], E.J. Kamsteeg[2], R.C.M. Hennekam[9], P. Scambler[1], P.L. Beales[1], C. UK10K Consortium[3], N.V.A.M. Knoers[10], R. Roepman[2], H.M. Mitchison[1].* 1) Molecular Medicine Unit, Institute of Child Health, University College London, London, United Kingdom; 2) Department of Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 3) UK10K, Wellcome Trust Genome Center, Cambridge, UK; 4) Department of Paediatrics, Radboud Univerisity Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Department of Nephrology, Institute of Mother and Child Health Care, Belgrade, Serbia; 6) Department of Nephrology, University Children's Hospital, Belgrade, Serbia; 7) Department of Medical Genetics, Istanbul University, Istanbul, Turkey; 8) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland; 9) Department of Pediatrics, Academic Medical Center University of Amsterdam, Amsterdam, The Netherlands; 10) Department of Medical Genetics, University Medical Centre Utrecht, The Netherlands.

Background: JATD is a rare autosomal-recessively inherited ciliary chondrodysplasia exhibiting high perinatal lethality. Mutations in several genes encoding mostly intraflagellar transport proteins such as *IFT80*, *DYNC2H1*, *IFT144*, *WDR35*, *NEK1*, *TTC21B* and *IFT140* have been shown to be causative but no phenotype-genotype correlations have been made to date except that the few reported *IFT80* patients presented with mild skeletal disease. Furthermore, genetic diagnosis in JATD has been hampered in the past by the underlying genetic heterogeneity of the disease and the fact that *DYNC2H1* is a very large gene. **Study design:** To define the molecular basis of JATD we have performed whole exome sequencing (WES) of 63 probands in combination with Sanger sequencing. *IFT80* mutations were excluded prior to WES. **Results:** WES showed biallelic mutations in *DYNC2H1* in 25 (39%) of all patients examined, while 3 patients from 2 families were found to be heterozygous for a single allele only. Strikingly, none of the patients was found to carry two nonsense alleles suggesting the human phenotype is at least partly hypomorphic. Biallelic mutations in *IFT140* were detected in 1 proband while 5 patients were found to be heterozygous. No mutations were found in *WDR19*, *WDR35*, *NEK1* or *TTC21B*. Subsequent Sanger sequencing of *IFT140* in JATD patients with severe renal involvement revealed 2 more cases of compound heterozygous mutations. Interestingly, no significant extraskeletal involvement was reported for our *DYNC2H1* patients while all biallelic *IFT140* patients suffered from end stage renal disease in childhood. Furthermore, bioinformatic analysis for heterozygous unique missense variants in *IFT140* in 46 JATD vs 134 control samples revealed a significant enrichment of heterozygous variants with predicted deleterious effect (Fisher exact test, odds ratio 12.5, $p = 0.0156$) indicating a disease modifying role for IFT140 mutations. **Summary:** *DYNC2H1* mutations account for approximately a third of all JATD cases and mutations in other known JATD genes are rare. While extraskeletal disease was rare in our *DYNC2H1* cases, severe renal disease was observed in all our patients with biallelic mutations in *IFT140*. We therefore propose a clinically and diagnostically relevant genotype-phenotype correlation and report an increased burden of *IFT140* deleterious variants not being disease causative in JATD, indicating that certain alleles may confer modifier effects on disease expression.

217

Dominant missense mutations in ABCC9 cause Cantú syndrome. *G. van Haaften[1], M. Harakalova[1], J.J.T. van Harssel[1], P. Terhal[1], S. van Lieshout[1], K. Duran[1], I. Renkens[1], D.J. Amor[2,3], L.C. Wilson[4], E.P. Kirk[5], C.L.S. Turner[6], D. Shears[7], S. Garcia-Minaur[8], M.M. Lees[4], A. Ross[9], H. Venselaar[10,11], G. Vriend[10,11], H. Takanari[12], M.B. Rook[12], M.A.G. van der Heyden[12], M.E. Swinkels[1], I.J. Scurr[13], S.F. Smithson[13], N.V. Knoers[1], J.J. van der Smagt[1], I.J. Nijman[1], W.P. Kloosterman[1], M.M. van Haelst[1,14], E. Cuppen[1,15].* 1) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands; 2) Murdoch Children's Research Institute, Royal Children's Hospital, Victoria, Australia; 3) Department of Paediatrics, University of Melbourne, Victoria, Australia; 4) Department of Clinical Genetics, Great Ormond Street Hospital, London, UK; 5) Department of Medical Genetics, Sydney Children's Hospital, Sydney, Australia; 6) Department of Clinical Genetics, Royal Devon and Exeter Hospital, Exeter, UK; 7) Clinical Genetics Department, Churchill Hospital, Oxford, UK; 8) Institute for Medical and Molecular Genetics, La Paz University Hospital, Madrid, Spain; 9) Department Clinical Genetics, Foresterhill, Aberdeen, UK; 10) Center for Molecular and Biomolecular Informatics (CMBI), Nijmegen, The Netherlands; 11) Nijmegen Center for Molecular Life Sciences (NCMLS), Radboud University Nijmegen Medical Center, The Netherlands; 12) Department of Medical Physiology, University Medical Center Utrecht, Utrecht, The Netherlands; 13) Department of Clinical Genetics, St. Michael's Hospital, Bristol, UK; 14) Section of Genomic Medicine, Imperial College London, London, UK; 15) Hubrecht Institute, The Royal Dutch Academy of Arts and Sciences, University Medical Center Utrecht, Utrecht, The Netherlands.

Cantú syndrome is characterized by congenital hypertrichosis, characteristic facial features, osteochondrodyplasia and cardiac defects. We used family-based exome sequencing and identified a de novo mutation in ABCC9. Subsequently we discovered novel dominant missense mutations in ABCC9 in a total of 14 out of 16 Cantú syndrome cases. ABCC9 is part of an ATP dependent potassium (K_{ATP}) channel, which couples the metabolic state of a cell with its electrical activity. All mutations alter amino acids in or close to the transmembrane domains of ABCC9. Using electrophysiological measurements we show that mutations in ABCC9 reduce the ATP-mediated inhibition, which results in opening of the channel. Moreover, similarities between the phenotype of Cantú patients and side effects of the K_{ATP} channel agonist minoxidil indicate that the mutations in ABCC9 result in channel opening. Thanks to the availability of ABCC9 antagonists our findings may have direct implications for the treatment of Cantú patients.

218

Reduced dosage of ERF causes complex craniosynostosis in humans and mice, and links ERK1/2 signalling to regulation of osteogenesis.

S.R.F. Twigg[1], I. Paraki[2], S.J. McGowan[1], M. Allegra[2], A.L. Fenwick[1], V.P. Sharma[1], E. Vorgia[2], A. Zaragkoulias[2], E. Sadighi Akha[3], S.J. Knight[3], H. Lord[4], T. Lester[4], L. Izatt[5], A.K. Lampe[6], S.N. Mohammed[5], F.J. Stewart[7], A. Verloes[8], L.C. Wilson[9], D. Johnson[10], S.A. Wall[10], P. Hammond[11], J. Hughes[1], S. Taylor[1], G. Mavrothalassitis[2], A.O.M. Wilkie[1,10]. 1) Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK; 2) School of Medicine, University of Crete and Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, Heraklion, Crete, 71003, Greece; 3) The Oxford Partnership Comprehensive Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Molecular Genetics Laboratory, Oxford University Hospitals NHS Trust, Oxford, UK; 5) Clinical Genetics, Guy's and St Thomas' NHS Foundation Trust, London, UK; 6) South East of Scotland Clinical Genetics Service, NHS Lothian, Edinburgh, UK; 7) Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast, UK; 8) Département de Génétique, Hôpital Robert Debre, Paris, France; 9) North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Trust, London, UK; 10) Craniofacial Unit, Oxford University Hospitals NHS Trust, Oxford, UK; 11) Molecular Medicine Unit, UCL Institute of Child Health, London, UK.

We used exome sequencing to analyze the DNA from a 7-year old boy with multisuture craniosynostosis. After excluding previously described variants and genomic regions that did not segregate with the phenotype, 135 nonsynonymous changes remained. One of these (R183X) was present in *ERF*, encoding an inhibitory ETS-family transcription factor. ERF is a phosphorylation target of ERK1/2, key effectors of the RAS-MAPK signal cascade. We considered *ERF* as a candidate because activation of ERK1/2 was previously demonstrated in craniosynostosis. To analyze the role of *ERF* mutation in the phenotype, we sequenced the gene in 411 samples from unrelated subjects with craniosynostosis. Heterozygous mutations suggestive of loss of function, including 4 *de novo* changes, were present in a further 11 cases but not in 288 normal controls. 7 of 12 probands had syndromic multisuture synostosis, representing a 13-fold enrichment compared to other diagnostic groups. Notably, despite the multisuture involvement many affected individuals presented later in childhood than usually occurs in craniosynostosis, and primary surgery was frequently delayed. In the mouse, heterozygous loss-of-function of *Erf* is not associated with any abnormal phenotype, whereas homozygous loss causes placental defects resulting in death by E10.5. To explore the function of Erf during development, we engineered mice harbouring a conditional allele (Erf^{loxP}). Both heterozygous ($Erf^{loxP/+}$) and homozygous ($Erf^{loxP/loxP}$) conditional mice were phenotypically normal, but compound conditional/null heterozygotes ($Erf^{loxP/-}$) had domed heads that became apparent from P21. MicroCT scanning demonstrated craniosynostosis, but no other specific skeletal abnormalities were evident. Analysis of *Erf* cDNA showed reduced transcription of the conditional allele (~48% vs wild-type). Thus, as in humans, the cranial sutures appear particularly sensitive to reduced Erf dosage, but the threshold level required for phenotypic manifestation is lower in mice. In mouse embryonic calvariae we found increased expression of late markers of osteogenesis, indicating enhanced differentiation of membraneous bones, but the osteogenic regulator *Runx2* was reduced. Moreover, ChIPseq analysis showed that ERF binds preferentially to enhancers containing RUNX motifs. Based on these data we propose that Erf normally antagonises positively acting Ets factors acting in concert with Runx2 to activate osteogenesis during membraneous ossification.

219

Mutations in the multidomain protein MEGF8 identify a new subtype of Carpenter syndrome associated with defective lateralization.

D.L. Lloyd[1], S.R. Twigg[1], N. Elcioglu[2], D. Jenkins[3], C.D.O. Cooper[4], N. Akarsu[5], E. Taskiran[5], N. Al-Sannaa[6], A. Annagür[7], G. Gillessen-Kaesbach[8], I. Stefanova[8], S.J.L. Knight[9], J.A. Goodship[10], B. Keavney[10], P.L. Beales[3], O. Gileadi[4], S. McGowan[1], A.O.M. Wilkie[1]. 1) Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom; 2) Department of Pediatric Genetics, Marmara University Medical Faculty, Istanbul, Turkey; 3) Molecular Medicine Unit, UCL Institute of Child Health, 30 Guilford Street, London, UK; 4) SGC, University of Oxford, ORCRB, Roosevelt Drive, Oxford, UK; 5) Hacettepe University Medical Faculty, Department of Medical Genetics, Gene Mapping Laboratory, Sihhiye, 06100, Ankara, Turkey; 6) Pediatrics Services Division, Dhahran Health Center, Saudi Aramco Medical Services Organization, Saudi Arabia; 7) Division of Neonatology, Selçuk University, Meram Faculty of Medicine, Konya, Turkey; 8) Institut für Humangenetik Lübeck, Universitätsklinikum Schleswig-Holstein, Ratzeburger Allee 160, D-23538 Lübeck, Germany; 9) NIHR Biomedical Research Centre, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, UK; 10) Institute of Genetic Medicine, University of Newcastle, International Centre for Life, Central Parkway, Newcastle upon Tyne, UK.

Carpenter syndrome is a rare, autosomal recessive disorder characterised by craniosynostosis, typically with fusion of the midline sutures, and other malformations including brachy- and polysyndactyly. We previously showed that mutation of *RAB23*, a gene involved in negatively regulating SHH signalling, is the most common cause of Carpenter syndrome. Here, we describe loss-of-function mutations in a second gene, *MEGF8*, in 4 unrelated patients presenting with this diagnosis. We performed whole exome sequencing of a sample from a child with bicoronal synostosis, polysyndactyly and dextrocardia, clinically diagnosed with Carpenter syndrome but in whom *RAB23* mutations were excluded. As the parents were first cousins, we prioritised variants in regions of homozygosity. Within the largest such region, on chromosome 19, we predicted a homozygous mutation (c.4496G>A; p.R1499H) within a Kelch domain of a highly conserved 2,778 amino acid protein, MEGF8. A missense mutation in *Megf8* was previously shown to cause polydactyly and heterotaxy in mice, closely phenocopying the patient. We next performed DNA sequencing on samples from a panel of 22 *RAB23*-negative children (or carrier parents) with possible Carpenter syndrome and 15 isolated heterotaxy patients, and identified likely pathogenic mutations of *MEGF8* in 3 subjects from the former cohort. The mutations were p.[R448X];[S2367G], p.[G199R];[G199R], and c.3349+4_3349+5insAA;c.7069-2A>G. As well as craniosynostosis and polysyndactyly, in all subjects there was a distinctive dysmorphic facies, and 3 had abnormal cardiac lateralisation (dextrocardia in 2 cases and transposition of great arteries in the third). Morpholino knockdown of *Megf8* expression in zebrafish yielded 83% of embryos with an epiboly defect. This was rescued with human wildtype *MEGF8* RNA (12% epiboly defect), but with mutant *MEGF8* RNA encoding any of the 3 missense mutations, rescue was substantially diminished (60–73% epiboly defect). Furthermore, we obtained folded proteins for an EGF-laminin domain containing the p.S2367G substitution and demonstrated a perturbation of secondary structure using circular dichroism spectroscopy. We conclude that recessive mutations of *MEGF8* associated with residual biological activity cause a phenotype similar to Carpenter syndrome, but with the additional feature of heterotaxy. Our findings are consistent with *MEGF8* performing essential roles in normal left-right patterning, limb and cranial suture development.

220

Increased frequency of *FBN1* variants in adolescent idiopathic scoliosis. *J.G. Buchan[1], D.A. Alvarado[4], M.C. Willing[2], M.B. Dobbs[4,5], C.A. Gurnett[2,3,4]*. 1) Genetics, Washington University School of Medicine, Saint Louis, MO; 2) Pediatrics, Washington University School of Medicine, Saint Louis, MO; 3) Neurology, Washington University School of Medicine, Saint Louis, MO; 4) Orthopedics, Washington University School of Medicine, Saint Louis, MO; 5) Shriners Hospitals for Children, Saint Louis, MO.

Adolescent idiopathic scoliosis (AIS) is a common childhood spine deformity that affects up to 3% of the population. The etiology of AIS is poorly understood although there is a strong genetic component in some patients. Connective tissue disorders such as Marfan syndrome (MFS) are associated with a high incidence of scoliosis, with approximately 60% of patients with MFS developing spinal curvature. MFS is caused by mutations in *fibrillin 1* (*FBN1*), an extracellular matrix glycoprotein with over 600 mutations described across 65 exons. Due to the large size of *FBN1*, genetic testing is costly and many diagnoses are made only by clinical examination using established criteria. However, recognition of MFS can be challenging due to a broad phenotypic expression and nonspecific features, including scoliosis. Therefore, we used exome sequencing to evaluate 60 unrelated AIS patients to determine the frequency of rare *FBN1* variants. Clinical genetics evaluations, including the revised Ghent systemic features score, were also obtained when possible. Rare (MAF<1%) or novel *FBN1* variants were identified in 7/60 (11.7%) patients compared to 4/108 (3.7%) unrelated controls from an in-house database (p=0.046). The frequency of rare *FBN1* variants was also increased compared to 1000 Genomes (63/1092; 5.8%) (p=0.063). *FBN1* mutations that had previously been associated with MFS or fibrillinopathy were identified in three AIS patients, and five novel variants of unknown significance were identified in four additional cases. Clinical genetics evaluations revealed that some AIS patients with *FBN1* variants met the revised Ghent nosology for MFS, while the majority did not. Targeted sequencing of *FBN1* in 200 additional AIS patients is being performed. Overall, our data suggests that rare *FBN1* variants occur at a higher frequency in AIS compared to controls. Further studies are needed to determine whether these *FBN1* variants cause MFS or a related fibrillinopathy, isolated AIS, or represent benign polymorphisms.

221

Exome sequencing in idiopathic scoliosis reveals rare variants in *VANGL1*, a planar cell polarity gene involved in axial development. *S. Sharma[1], J.A. Herring[2,3], X. Gao[1], D. Zhang[1], C. Wise[1,4]*. 1) Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA; 2) Department of Orthopedics, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA; 3) Department of Orthopedics, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA; 4) McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA.

Idiopathic scoliosis (IS) is a common spinal disorder that is generally considered multi-genic, yet rare families with Mendelian inheritance are reported. Prior population-based genetic studies of IS have yielded associations with common variants, but causal mutations, rare or common, are lacking. In the present study, we performed family-based exome sequencing to identify a highly penetrant disease mutation in a family with dominant inheritance of IS. The majority of affected cases in this family required treatment, and four cases warranted surgical correction. We filtered for novel, co-segregating mutations. No single novel variant segregated perfectly with disease. One variant, a c.676C>T (L226F) mutation in the *VANGL1* gene, segregated with IS in all but one family member who had very mild IS and no affected offspring (five adult daughters). *VANGL1* encodes van Gogh-like 1, an integral membrane protein that participates along with its homolog *VANGL2* in a planar cell polarity (PCP) pathway involved in early axial development. *VANGL1* or *VANGL2* mutations have been reported in human patients with neural tube defects (NTDs). We further sequenced the *VANGL1* coding region in non-Hispanic white IS cases (N= 182) and controls (N=380). Results suggested enrichment of novel or rare (<1%) *VANGL1* mutations (burden test P = 0.0029). We noted that one novel *VANGL1* mutation, R263C, was analogous to the *VANGL2* mutant responsible for the recessive mouse Looptail (Lp^{m2Jus}) phenotype characterized by a curled tail. We explored the functional consequences of this and the novel L226F mutation in 293F cell lines stably expressing wild type or mutant GFP-tagged VANGL1 proteins. Distinct from the wild type protein, VANGL1 mutants were found largely in the intracellular compartment, whereas wild type protein localized to the plasma membrane as expected. These results suggest that IS-associated variants affect VANGL1 localization. This was apparently not mediated by dishevelled (DVL1), a known VANGL1 effector protein, as co-IP experiments did not find differences in DVL1 binding affinity to mutant or wild type VANGL1. This study illustrates the challenges of rare variant discovery for common diseases. While these data support rare-variant *VANGL1* associations in IS, implicating a pathway of early axial development in later-onset disease, further *VANGL1* validation studies in larger IS populations are warranted.

222

Recessive mutations in *FKBP10*, a PPIase known to cause type XI OI, extend the phenotype to a congenital contracture syndrome (Kuskokwim disease), and cause diminished collagen cross-linking in matrix. *A.M. Barnes[1], M. Weis[2], W.A. Cabral[1], E. Makareeva[3], E.L. Mertz[3], W. Paton[4], G. Duncan[5], C. Trujillo[6], S. Leikin[3], D.R. Eyre[2], S.J. Bale[7], J.C. Marini[1]*. 1) BEMB, NICHD/NIH, Bethesda, MD; 2) Orthopaedic Research Laboratories, University of Washington, Seattle, WA; 3) SPB, NICHD/NIH Bethesda, MD; 4) Alaska Native Medical Center, Anchorage, AK; 5) Christchurch Hospital, Christchurch, New Zealand; 6) Genetics Unit, Dr. Erfan & Bagedo General Hospital, Jeddah, Saudi Arabia; 7) Gene Dx, Gaithersburg, MD.

Recessive osteogenesis imperfecta (OI) is caused by defects in genes whose products interact with type I collagen for modification and/or folding. Recently, mutations in *FKBP10*, encoding the ER chaperone and isomerase FKBP65, have been shown to cause both recessive OI and Bruck Syndrome (OI with contractures). We identified an *FKBP10*-null mutation in a 5-generation Palestinian pedigree, associated with moderately severe recessive OI, as well as an in-frame deletion in *FKBP10*, in an Alaskan pedigree with Kuskokwim Disease. Kuskokwim Disease is characterized by congenital contractures and osteopenia; this is the first description of a predominantly contracture disorder caused by *FKBP10* mutations. The child with moderate type XI OI has a homozygous *FKBP10* frameshift mutation (c.1271_1272delCCinsA). *FKBP10* transcripts in proband fibroblasts are 4% of control; with absent FKBP65 protein. The Alaskan pedigree has an in-frame *FKBP10* deletion of a single residue (c.875_877del, p.Tyr293del) in the 3rd PPIase domain of FKBP65 which leads to normal to increased levels of *FKBP10* transcripts and residual protein. Both pedigrees show minimal changes in type I collagen migration on gel electrophoresis with a slight increase in hydroxylysine levels. On matrix deposition of fibroblasts, the mature cross-linked collagen fraction was decreased by >95% in the null mutation and by 45–85% in the Kuskokwim pedigree. Mass spectrometry revealed that hydroxylation of the telopeptide lysine involved in collagen cross-linking was <1% in the *FKBP10*-null mutation and 2–10% in the Kuskokwim, compared to ~60% hydroxylation of collagen secreted from normal fibroblasts. The lack of cross-linking affected the organization of matrix. The *FKBP10*-null cells deposited a very sparse and disorganized matrix. The Kuskokwim pedigree cells deposited a normal amount of matrix, yet the matrix was still disorganized, compared to the orderly fibrils seen in control. The decrease in the amount of matrix was confirmed in both pedigrees by Raman microspectroscopy: the *FKBP10*-null cells had ~30% of control matrix, while 60–95% of control matrix was detected from the Kuskokwim cells, with some correlation to patient severity. Taken together, these two types of *FKBP10* mutations reveal a similar mechanism, in that loss of FKBP65 leads to reduced collagen telopeptide hydroxylation, decreased collagen cross-linking, and decreased quality of collagen matrix, weakening bone tissue.

223

PRIMUS: Pedigree Reconstruction and Identification of the Maximum Unrelated Set. *J. Staples, D. Nickerson, J. Below.* Department of Genome Sciences, University of Washington, Seattle, WA.

Recently, researchers have successfully leveraged familial relationships to attain the necessary power in analyses to identify rare causes of disease (e.g. Kabuki and Freeman-Sheldon syndromes), leading to a renewed interest in family-based analysis of genetic diseases, in which obtaining accurate pedigree information is crucial. Reconstruction of pedigrees is useful both to verify these clinically ascertained pedigrees as well as to reconstruct cryptic pedigrees. For example, pedigree reconstruction from genetic data has been used in forensic analysis of mass graves, genetic studies of large cohorts with cryptic relatedness, and genealogical research; in such cases genealogical data is unavailable or incomplete and the pedigree is difficult, if not impossible, to obtain. Previous methods for reconstructing pedigrees are incapable of handling missing samples (gaps) in the family, large and/or multigenerational pedigrees, or non-monogamous relationships. We have developed a program that uses genome-wide IBD estimates to quickly reconstruct large, arbitrary, human and non-human pedigrees that may include gaps as distant as first cousins. This program is part of our software package called PRIMUS. PRIMUS uses genome-wide estimates of IBD to identify family networks and predict the type of familial relationship between each pair of individuals in the dataset, reconstructs all possible pedigrees that fit these relationships, and identifies the most likely pedigree(s) given the data. The reconstructed pedigree(s) is output as a six-column PED file as well as a post-script image file. PRIMUS correctly reconstructed 288 out of 293 known multi-generational prostate cancer pedigrees of varying sizes (Stanford et al., 2009), and the number will increase as PRIMUS is able to better handle complex relationships such as double first cousins and consanguinity. PRIMUS reconstructs highly informative pedigrees in seconds, but slows as information content drops. Building on previous work on the cryptic relatedness in the HapMap3 data, we have generated pedigrees for the extended/cryptic family structures in the ASW, CEU, LWK, MKK, MXL, and YRI populations. Finally, PRIMUS can use affection status and reconstructed pedigrees to select the optimal samples for exome or whole-genome sequencing. PRIMUS will aid researchers in verifying family data and generating previously unknown pedigrees from DNA, allowing them to utilize these familial relationships in further analyses.

224

Pharmacogenoinformatics: Novel approach of Insilico drug designing based on genetic variation of MDR1 gene involved in statin resistance. *A. Munshi, M. Sai Babu, A. Venkateswara Rao, A. Jyothy.* Insistute of Genetics and hospital for genetic diseases, Hyderabad, Begumpet, Hyderabad., India.

Statins are the most prescribed drugs, highly effective in reducing the risk of cardiovascular and cerebrovascular events, primarily by lowering low density lipoprotein (LDL) cholesterol. Although large clinical trials found a 27% average relative risk reduction of major coronary events, there is large variability in benefits from statin therapy. Researchers have found three SNPs (C3435T, G2677T/A, C1236T) of MDR1gene, which codes for P-Glycoprotein (P-gp) (a drug efflux transporter), responsible for the reduced bioavailability of statins. We aimed to design a new drug molecule based on synonymous and nonsynonymous SNPs of MDR1 gene, which is not a substrate to P-gp and acts directly on β-hydroxy methylglutaryl coenzyme A reductase (HMG-CoA), a target site for statins, using Insilico tools. Structural changes in mRNA due to synonymous and nonsynonymous SNPs were evaluated by SNPfold. The 3D structures of normal and mutant proteins of P-gp and HMG-CoA reductase were modeled by Molecular Operating Environment (MOE). A new lead molecule was designed from native structure by VegaZZ and parameters of drug were validated with Hyperchem, and Pharmacophore mapping was done using LigandScout. We docked the lead molecule with normal and mutant P-gp and found no interactions with P-gp showing that it is not a substrate for P-gp. However, it forms clear hydrogen bond interactions with HMG-CoA reductase. This is a novel approach in the field of bioinformatics and pharmacogenomics (pharmacogenoinformatics) for the development of new drug molecules based on the SNPs of genes involved in drug metabolism in a particular population.

225

A general, integrated variant prioritization method for rapid determination of disease causing mutations from next generation sequencing data. *B.D. O'Fallon, W. Wooderchak-Donahue, P. Bayrak-Toydemir.* Clinical and experimental pathology, ARUP Labs, Salt Lake City, UT., USA.

Accurate identification of genetic variants associated with human disease is a fundamental problem facing medical geneticists. Modern, high throughput sequencing techniques may generate thousands to millions of genetic variants for each sample, thus individual consideration of each candidate is often impractical. While many computational techniques exist to rank or classify such variants, few take into account both the severity of an individual mutation as well as its relevance to a particular disorder. Here, we propose a variant ranking algorithm that includes both the likelihood that a given variant has a deleterious effect as well as the likelihood that the variant is associated with a phenotype in question. The algorithm combines several measures of evolutionary conservation and biochemical similarity to determine the severity of a given variant, and searches gene interaction databases, Gene Ontology (GO) terms, Pubmed abstracts, and NCBI gene summaries for information relevant to the phenotype. We show that by combining multiple orthogonal measures of effect prediction our procedure can meaningfully prioritize large lists of variants and greatly reduce the number requiring individual consideration. We applied the ranking technique to a variety of exome samples including individuals with hereditary hemorrhagic telangiectasia (HHT), Noonan syndrome (NS), and common variable immune deficiency (CVID), recapitulating known causative mutations in the HHT and CVID samples, and uncovering multiple promising variants in previously uncharacterized samples. When sufficient clinical and molecular information is available to characterize a disorder, this integrated approach may facilitate rapid identification of disease causing mutations.

226

Visually Integrating and Exploring High Throughput Phenome-Wide Association (PheWAS) Results Using PheWAS-View and PhenoGram. *S.A. Pendergrass[1], S. Dudek[1], D.C. Crawford[2], M.D. Ritchie[1].* 1) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 2) Center for Human Genetics Research, Vanderbilt University.

In Phenome-Wide Association Studies (PheWAS), the association between single nucleotide polymorphisms (SNPs) and an extensive range of phenotypic measurements are calculated in a high throughput, unbiased manner. The PheWAS approach is complementary to Genome Wide Association studies (GWAS) that calculate the association between hundreds of thousands of SNPs and one or a limited range of phenotypes. Extensively exploring the association between phenotypic structure and genotypic variation through PheWAS produces a set of complex and comprehensive results. Visualization of these data is integral to fully inspecting, analyzing, interpreting, and sharing the results. Thus we have developed PheWAS-View, software for creating a variety of visual summaries of the SNP, gene, phenotype, and association information resulting from PheWAS investigations. PheWAS-View can be used to plot results on a larger summary level as well as the individual result level. While this tool was developed specifically for PheWAS, PheWAS-View can be applied to any high throughput bioinformatics data where thousands of association results are being explored. PhenoGram is additional software developed for visualization of multiple phenotype association results. PhenoGram presents results for multiple phenotypes across an ideogram of the 22 human autosomes, as well as the sex chromosomes, at the chromosomal locations of the SNPs associated with the phenotypes. The format is similar to that of the NHGRI GWA plot of published genome-wide associations (Hindorf LA, et al., Proc Natl Acad Sci USA. [May 27, 2009] and www.genome.gov/gwastudies). Both software packages allow users to investigate the complex high-throughput results of PheWAS studies, or any other study investigating a complex set of association results across phenotypes. PheWAS-View and PhenoGram are freely available for non-commercial research institutions, via a user-friendly web interface for using the software, as well being available in a command-line version. For full details see: http://ritchielab.psu.edu/ritchielab/software.

227

PhenoDB: a new web-based tool for the collection, storage and analysis of phenotypic features. *A. Hamosh[1], J. Hoover-Fong[1], V.R. Sutton[2], N. Sobreira[1], C. Boehm[1], F. Schiettecatte[3], D. Valle[1]*. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Univ Sch Med, Baltimore, MD; 2) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 3) FS Consulting, Salem, MA.

Historically, DNA diagnostic labs reasonably assumed that an individual undergoing testing for a single gene or a panel of genes had a phenotype consistent with mutations in the tested gene(s). With the advent of whole exome and genome sequencing, comprehensive phenotypic information plus knowledge of pedigree structure and previous clinical testing are essential for interpretation of the sequence data. Image data (photos, videos, radiographs, CTs, and MRIs) are also valuable. To meet these needs and as part of the Centers for Mendelian Genomics, we developed PhenoDB, a secure, web-based portal for entry and analysis of phenotypic and related information. Phenotypic features are organized hierarchically according to the major headings and subheadings of the OMIM clinical synopses with sub-divisions according to structure and function descending to the most granular features. Each string ends with an Other Feature textbox for optional free-text entries that are reviewed regularly so that other terms may be added. All features use the preferred term from Elements of Morphology with numerous synonyms. There are ~2900 features that are fully searchable and mapped to the UMLS and HPO. A familiar user can enter a family with a complex phenotype in <2 minutes, including unaffected individuals (optional) and sample availability. PhenoDB allows for ascertainment of all relevant information in a family and is searchable by family, OMIM number, phenotypic feature, mode of inheritance, genes screened, etc. Multiple usertypes are defined with user-specific permissions for viewing and editing information: Submitters can see their own families and track their progress; phenotype review committee members see all families in a standardized format but are blind to identifiers; ELSI committee members can see and comment on consent forms but cannot see family information; analysis and interpretation committee members can see phenotypic information to guide their analysis. Our goal is to utilize this database to format phenotypic data for submission to dbGAP for appropriately consented individuals. This application is built using Django, a Python based open source web development tool, and uses MySQL as the underlying database. As it is likely to be useful to others embarking upon clinically relevant activities, we will make it freely available for non-profit organizations through the Johns Hopkins Institute of Genetic Medicine website (http://igm.jhmi.edu).

228

A novel metabolomics analysis workflow provides new biological insights into the genetic basis of human metabolic variation. *H. Dharuri[1], P. Henneman[1], D.O. Mook-Kanamori[3], K. Suhre[3], K. Willems van Dijk[1,2], P.A.C. 't Hoen[1]*. 1) Department of Human Genetics, Leiden University Medical Center, Leiden, RC, Netherlands; 2) Department. of Internal Medicine, Leiden University Medical Center, Leiden, Netherlands; 3) Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar, Education City, Qatar Foundation, PO Box 24144, Doha, State of Qatar.

We are interested in the association of genetic polymorphisms with metabolite levels. These intermediate phenotypes generally demonstrate larger effect sizes and potentially point at pathways relevant to disease. However, the traditional genome-wide association of SNPs with all measured metabolites comes with considerable multiple testing problems. Hereby, we present a method that aims to circumvent this problem by selectively testing genes that operate in the vicinity of the metabolite potentially affecting its levels. As a proof of principle we have evaluated the sensitivity and the ability of our method to make novel discoveries using published studies of GWAS with intermediate phenotypes. We have developed a software tool that automates the association of metabolites with SNPs near genes in the pathway relevant to the synthesis and degradation of the metabolite. The software makes use of a work flow management technology called Taverna. This allows us to conduct our analysis across diverse datasets and applications like Biomart, and metabolite and pathway databases like BioCyc, and KEGG. The software has been used to generate relevant gene/SNP sets for metabolites of interest with the ultimate goal being able to use this set for association studies. Evaluation of the sensitivity of our method in terms of identifying true positives in previous GWAS studies has been very encouraging. Given the increased power of our method, we have been able to identify novel loci from previous studies that had failed to reach the stringent genome-wide threshold. Our method was able to correctly identify previously known associations like fatty acid delta-5 desaturase (FADS1) and fatty acid elongase (ELOVL2) genes with phosphatidylcholines, acyl-coA dehydrogenases: ACADS, ACADM and ACADL genes with acyl carnitines among other important loci. A novel locus that we have discovered is the aldehyde dehydrogenase (ALDH1L1) gene that is found associated with glycine and glycine/serine ratios. This gene, also known as folate dehydrogenase, plays an important role in folate metabolism and acts immediately upstream of the serine hydroxymethyl transferase (SHMT) gene which is involved in the inter-conversion of glycine and serine. The increased power of our method to identify mechanistically relevant SNP-metabolite pairs will be demonstrated.

229

Integration of large-scale gene annotation, electronic medical records, and incidence data to produce phenotype-specific posterior probabilities to aid interpretation of genome-wide variant data. *I.M. Campbell[1], S.W. Cheung[1], A. Patel[1], S.R. Lalani[1,2], P. Stankiewicz[1], M.B. Ramocki[2,3], J.R. Lupski[1,2,3], C.A. Shaw[1]*. 1) Department of Molecular and Human Genetics, Baylor College Medicine, Houston, Texas; 2) Texas Children's Hospital, Houston, Texas; 3) Department of Pediatrics, Baylor College of Medicine, Houston, Texas.

Curation and interpretation of variants identified by genome-wide testing is made challenging both by the large number of events harbored by individual personal genomes and the extent of rare variation. This challenge is further exacerbated by the increasing genomic resolution and variant calls identified during the conceptual switch from locus specific to genome-wide genetic testing. Conventional determination of phenotypic relevance relies on patterns of higher incidence in affected individuals versus controls; however, as an increasing amount of ascertained variation proves to be rare or private to family groups and clans, incidence data alone have less utility to resolve pathogenic from benign. Our group and others have turned to computational approaches, including machine learning, to enhance interpretation of copy number variation (CNV). To develop these techniques, we leveraged large-scale knowledge sources, including human tissue gene expression, gene ontology, protein-protein interaction data and other annotation systems together with broad sets of known phenotype-associated training genes. As one example of this approach, we constructed a pathogenicity score specific for epilepsy, generated for every annotated human gene. To fine-tune our method, we have applied the epilepsy pathogenicity score to a cohort of 295 individuals with abnormal chromosomal microarray results and confirmed epilepsy gathered from affiliated hospitals by electronic medical record review. Seventy-four patients harbored CNVs that were diagnostic based on previous reports, but 221 (75%) had CNVs that were rare, novel, and non-recurrent. We used our “epileptogenicity score” to prioritize the CNVs most likely to be pathogenic. Moreover, we discovered CNVs harbored by our patients with epilepsy had significantly higher pathogenicity scores compared to those of patients referred for non-neurologic indications. We also applied our pathogenicity score to a replication cohort of patients referred from outside centers with less well-defined clinical characteristics. Finally, we explore combining gene pathogenicity scores with global and phenotype-specific incidence data to calculate posterior pathogenicity probabilities for each gene and CNV. Our results demonstrate the utility of integrative data approaches in medical genomics that incorporate patient medical records and bioinformatic techniques to enhance the interpretation of whole genome variant data from personal genomes.

230

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: CREX, computerized methodology to identify health conditions using the EMR for GWAS. *S. Sciortino[1], L. Walter[1], D. Ranatunga[1], L. Shen[1], D. Ludwig[1], J. Kay[1], L. Sakoda[1], N. Risch[2], C. Schaefer[1]*. 1) Division of Research, Kaiser Permanente, Oakland, CA; 2) Institute for Human Genetics, UCSF, San Francisco, CA 94143.

The Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort has genotype data on over 100,000 participants. In order to characterize many health conditions of interest found among respondents for sample-size estimation and GWAS, we sought to extend methodologies that utilized the Electronic Medical Record (EMR) in an automated way to characterize multiple disease entities without the expense and limitation of numerous, complex algorithms specific to individual diseases and conditions. We tested a probabilistic approach that scored clinical decisions recorded in the EMR to capture specific diagnostic and treatment domains. We first considered the domain encompassed by physician diagnosis, then added pharmacy prescriptions. We assessed sensitivity and specificity against internal registries at the Kaiser Permanente Division of Research. We also used the methodology to characterize a Type 2 diabetes phenotype for a GWAS. We tested a single diagnostic domain in a logistic model based on an ICD-9 taxonomy and found high sensitivity and specificity when compared with internal registries for breast, lung, colon and prostate cancers; Barrett's esophagus; Hepatitis B and C; HIV; Crohn's disease; ulcerative colitis, and diabetes. We then assessed an additive logistic regression model to distinguish among members with Types 1 and 2 diabetes in the GERA cohort, utilizing the ICD-9 taxonomy, the earliest age at diagnosis available in the EMR or by self-report, and anti-diabetic drug utilization. The model exhibited a sensitivity of 96.3% and specificity of 99.6% for Type 1 diabetes and a sensitivity of 94.0% and specificity of 98.4% for Type 2 diabetes when compared with the gold standard internal diabetes registry. We were able to identify an additional 60 cases of Type 1 diabetes and to remove 900 cases of undetermined diabetes type from cases and controls for a GWAS on Type 2 diabetes. Utilizing diagnostic information in the EMR as independent domains in probabilistic models to accomplish phenotype creation appears to be a reliable approach to facilitate rapid and large scale analysis of numerous disease phenotypes. The method is agnostic to input taxonomies as long as the EMR record contains sufficient and reliable atomic detail. It can also be adapted for machine learning given expert user feedback when gold standard databases are not available.

231

An Informatics Approach to Analyzing the Incidentalome. *M.C. Adams[1], J.S. Berg[1], N. Nassar[2], C. Bizon[2], K. Lee[1], C.P. Schmitt[2], K.C. Wilhelmsen[1,2], J.P. Evans[1].* 1) Genetics, UNC-Chapel Hill, Chapel Hill, NC; 2) Renaissance Computing Institute, Chapel Hill, NC.

Next generation sequencing (NGS) has transformed medical genetics research and appears poised to revolutionize clinical diagnosis of genetic diseases. However, the vast amounts of data and inevitable discovery of clinically relevant incidental findings pose challenges to the adoption of these techniques in the clinic, necessitating novel analytic approaches. We recently described a conceptual strategy for classifying genes into three broad "bins" to facilitate informed consent, analysis, and return of incidental findings in a clinical setting. Bin 1 contains genes in which a mutation would trigger specific medical action. Bin 2 contains genes known to be associated with human diseases, for which evidence does not support any specific action; bin 2 is further stratified based on the potential for harm. 2016 genes linked with Mendelian diseases were categorized into Bin 1, Bin 2b, and Bin 2c, and we implemented this method using a python script to query a database containing variant data from 80 whole genome sequences. We utilized allele frequencies from the 1000 Genomes project and disease mutations ("DM" variants) from the Human Gene Mutation Database (HGMD) to assist in the selection of variants for manual review. The algorithm effectively identified incidental variants of clinical relevance. Restricting analysis to rare (<5% allele frequency) truncating variants drastically reduced the number of variants in each bin. An additional HGMD query improved the yield for missense mutations, with an average of 74 "DM" variants per person. However, 251 of the 871 unique "DM" variants (29%) had allele frequencies >5%, resulting in a disproportionate number of "DM" variants per genome (78%) having >5% allele frequency and suggesting that errors exist in the HGMD. The final algorithm, which combined strict filtering of protein truncating variants with the rare "DM" variants from HGMD facilitated the discovery of known disease-causing mutations and novel, predicted deleterious mutations, with an average of 17.4 variants per person that required manual analysis. This method is an efficient and practical way identify clinically relevant incidental findings, readily adaptable to other types of clinical genomic analyses, scalable to the demands of a clinical laboratory workflow, and flexible with respect to advances in medical genetics and genomics.

232

Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *M. Taniguchi[1,2], K. Kobayashi[2], M. Kanagawa[2], CC. Yu[2], T. Oda[2], A. Kuga[2], H. Kurahashi[3], H.O. Akmen[4], S. DiMauro[4], T. Yokota[5], S. Takeda[6], T. Toda[2].* 1) Departrment of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan; 2) Department of Neurology / Molecular Brain Science, Kobe University Graduate School of Medicine, Kobe, Japan; 3) Division of Molecular Genetics, Institute for Comprehensive Science, Fujita Health University, Aichi, Japan; 4) Department of Nerurology, Columbia University Medical Center, New York, USA; 5) Department of Medical Genetics, Fuculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada; 6) Department of Molecular Therapy, National Institue of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan.

Fukuyama muscular dystrophy (FCMD; MIM253800), one of the most common autosomal recessive disorders in Japan, was the first human disease found to result from ancestral insertion of a SINE-VNTR-*Alu* (SVA) retrotransposon into a causative gene. In FCMD, the SVA insertion occurs in the 3'-untranslated region (UTR) of the *fukutin* gene. The pathogenic mechanism for FCMD is unknown, and no effective clinical treatments exist. Here we show that aberrant mRNA splicing, induced by SVA exon-trapping, underlies the molecular pathogenesis of FCMD. Quantitative mRNA analysis pinpointed a region that was missing from transcripts in FCMD patients. This region spans part of the 3' end of the fukutin coding region, proximal part of the 3' UTR, and the SVA insertion. Correspondingly, *fukutin* mRNA transcripts in FCMD patients and SVA knock-in (KI) model mice were shorter than the expected length. Sequence analysis revealed an abnormal splicing event, provoked by a strong acceptor site in SVA and a rare alternative donor site in fukutin exon 10. The resulting product truncates the fukutin C-terminus and adds 129 amino acids encoded by the SVA. Introduction of antisense oligonucleotides (AONs) targeting the splice acceptor, the predicted exonic splicing enhancer, and the intronic splicing enhancer prevented pathogenic exon-trapping by SVA in FCMD patient cells and model mice, rescuing normal fukutin mRNA expression and protein production. AON treatment also restored fukutin functions, including *O*-glycosylation of α-dystroglycan (α-DG) and laminin binding by α-DG. Moreover, we observe exon-trapping in other SVA insertions associated with disease (hypercholesterolemia, neutral lipid storage disease) and human-specific SVA insertion in a novel gene. Thus, we have discovered in human disease a role for SVA-mediated exon-trapping and demonstrated the promise of splicing modulation therapy as the first radical clinical treatment for FCMD and other SVA-mediated diseases.

233

Hematopoietic stem cell transplantation for adolescent and adult onset cerebral X-linked adrenoleukodystrophy. *T. Matsukawa[1], T. Yamamoto[1], S. Seo[2], K. Kumano[2], M. Ichikawa[2], Y. Takahashi[1], H. Ishiura[1], J. Mitsui[1], M. Tanaka[1], J. Goto[1], M. Kurokawa[2], S. Tsuji[1].* 1) Dept Neurology, Univ Tokyo, Tokyo, Japan; 2) Dept Hematology, Univ Tokyo, Tokyo, Japan.

Background: There have been accumulating evidences supporting the efficacy of the hematopoietic stem cell transplantation (HSCT) for childhood-onset cerebral X-linked adrenoleukodystrophy (ALD) when performed at an early stage of the disease. To date, there have been only two reported cases of adult-onset cerebral ALD (ACALD) treated with HSCT and the clinical efficacy remains to be established. Objective: To evaluate the clinical efficacy of HSCT for patients with adolescent/adult-onset cerebral ALD. Methods: To determine the optimum timing for HSCT in the early stage of adolescent/adult-onset-cerebral ALD, we have been following 15 ALD patients (adolescent cerebral ALD (AdolCALD); 1, AMN with later development of cerebral ALD (AMN-Cer); 2, cerebello-brainstem ALD; 2, AMN; 9, and Addison only; 1). The average observation period was 4.5 years. The patient has been observed at the interval of 3–6 months to detect appearance of the early cerebral symptoms. The indications for HSCT include an early stage of the disease and the presence of Gadolinium (Gd)-enhancing white matter lesions on MRI. To evaluate the clinical outcome after HSCT, we designed a protocol of HSCT for adolescent/adult-onset ALD, including neurological exams, cognitive function tests, ADL rating scales, laboratory tests, imaging and electrophysiological tests. Conditioning regime included intravenous busulphan and cyclophosphamide. Subject 1 (AdolCALD): Loss of motivation appeared at age 17. Dysarthria and ataxia appeared at age 18. These symptoms became progressively worse. The patient received HSCT from a 6/6 HLA-matched unrelated donor 7 months after presentation of dysarthria and ataxia. Clinical status has been stable for 4 years after HSCT and dysarthria has slightly improved. Enhancement on brain MRI disappeared 49 days after HSCT, although mild progression of atrophy of brainstem was observed. Loes score has increased by only 1 point (from 10 to 11) 4 years after HSCT compared with that before HSCT. Subject 2 (AMN-Cer): HSCT was conducted very recently, and the patient has been followed for 3 months with stable clinical course. Conclusion: The present study suggests the efficacy of HSCT for adolescent/adult-onset cerebral ALD. It will be important to determine the optimum timing of HSCT for adolescent/adult-onset cerebral ALD to accomplish a good outcome from HSCT.

234

Treating Pelizaeus-Merzbacher disease with clinically applicable compounds, curumin and chloroquine: preclinical studies. *K. Inoue[1], T. Morimura[1], Y. Numata[1], L-H. Yu[1], L. Gotoh[1], R. Yamamoto[1], N. Inoue[1], B. Antalfy[2], K. Deguchi[1], H. Osaka[3], Y. Goto[1].* 1) Dept. Mental Retardation and Birth Defect Research, National Institute of Neuroscience, NCNP, Kodaira, Tokyo, Japan; 2) Pathology, Texas Children's Hospital, Houston, US; 3) Neurology, Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan.

PLP1 amino acid substitutions cause accumulation of misfolded protein and induce endoplasmic reticulum (ER) stress, causing Pelizaeus-Merzbacher disease (PMD), a cureless hypomyelinating disorder of the central nerve system. Because PMD is an orphan disease of children, medicines with low entry barrier are preferred to be actually used to the patients. By targeting ER stress, we looked for safe and clinically applicable compounds as potential medication for PMD. Here we identified that curcumin and chloroquine can independently mitigate the cellular phenotype of PMD *in vitro* and/or *in vivo*. We treated mice carrying A242V PLP1 mutation (*Msd* mice) or HeLa cells transiently expressing $PLP1^{A242V}$ with either curcumin or chloroquine. Curcumin, a food compound from turmeric, was given to *Msd* mice orally at $180mgkg^{-1}day^{-1}$ from the postnatal day 3. *Msd* mice lived 25% longer than the wild-type mice. TUNEL assay and caspase 3 immunostaining revealed that oligodendrocytes undergoing apoptosis were reduced in number in the brain. In contrast, no apparent improvement in motor function, neurological phenotype, myelin formation and ER stress maker expression was observed. Meanwhile, chloroquine, a classic antimalarial medicine, decreased the amount of the mutant PLP1 by preventing translation through enhancing phosphorylation of eIF2α, and ameliorates the ER stress induced by the mutant protein in HeLa cells. These effects were also observed in the spinal cords of the mutant male mice treated with chloroquine injected intraperitoneally (0.5 mg/ml PBS, 10 μl/g body weight). Furthermore, chloroquine up-regulated the *Mbp* and *Plp1* transcripts in the mutant mice, suggesting an enhanced survival and/or proceeded differentiation of premature oligodendrocytes as a result of attenuation of ER stress. However, chloroquine treatment did not extend the lifespan of *Msd* mice. In conclusion, curcumin and chloroquine, a food compound and an FDA-approved well-tolerated medicine, may serve as safe and clinically applicable medications for the treatment of PMD.

235

Systemic L-threo-dihydroxyphenylserine corrects neurochemical abnormalities in a mouse model of Menkes disease. *S. Kaler[1], A. Donsante[1], P. Sullivan[2], D. Goldstein[2], C. Holmes[2].* 1) Intramural Res Prog, NICHD/NIH, Bethesda, MD; 2) Intramural Res Prog., NINDS/NIH, Bethesda, MD.

Menkes disease is a lethal infantile neurodegenerative disorder caused by mutations in the copper-transporting ATPase, ATP7A. Treatment with early copper replacement can prevent death and neurodegeneration, although survivors often manifest symptoms of dysautonomia during childhood and teenage due to partial deficiency of dopamine-β-hydroxylase (DBH), a copper-dependent enzyme that converts dopamine to norepinephrine. Like individuals with congenital absence of DBH, Menkes disease survivors exhibit persistent serum and cerebrospinal fluid neurochemical abnormalities, as well as orthostatic hypotension, syncope, and chronic diarrhea, reflecting impaired sympathetic noradrenergic function. The drug L-threo-dihydroxyphenylserine (L-DOPS), which is metabolized directly to norepinephrine, bypasses the block in DBH-mediated norepinephrine synthesis. In a mouse model of Menkes disease, mottled-brindled (mo-br), we tested whether systemic administration of L-DOPS would correct the central and peripheral neurochemical abnormalities. All animal experimental procedures were approved by the NICHD Animal Studies Committee. C57BL/6-Atp7a mo-br breeding pairs were obtained from Jackson Laboratories (Bar Harbor, ME). For experiments, litters were culled to 4–5 pups. Shortly after birth, toe biopsies were performed for genotyping. At 8, 10, and 12 days of age, wild type and mutant mice received either 200 μg of L-DOPS per gram body weight or an appropriate volume of a mock carrier via intraperitoneal (i.p.) injection. Mo-br mice receiving i.p. L-DOPS showed significant increases in brain levels of norepinephrine ($P<0.001$)and its deaminated metabolite, dihydroxyphenylglycol (DHPG, $P<0.05$), and significantly improved ratios of proximal:distal metabolites in the catecholamine biosynthetic pathway ($P<0.01$). The overall catecholamine pathway effects of L-DOPS treatment were less pronounced in mo-br serum, although serum norepinephrine levels still increased significantly ($P<0.001$). These data provide compelling evidence that L-DOPS bypasses the defect in DBH deficiency in this murine model and suggest that oral L-DOPS treatment in Menkes disease survivors may ameliorate symptoms of dysautonomia.

236

Response to VPA therapy in SMA patients is concordant from blood to neurons and influenced by CD36. *B. Wirth[1], L. Heesen[1,2], I. Hölker[1], T. Bauer[3], J. Schreml[1], K. Zimmermann[1], M. Thoenes[1], M. Walter[4], J. Dimos[5], M. Peitz[2], O. Brüstle[2], R. Heller[1], L. Garbes[1]*. 1) Institute of Human Genetics, Institute of Genetics and Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany; 2) Institute of Reconstructive Neurobiology, LIFE & BRAIN Center, University of Bonn and Hertie Foundation, Bonn, Germany; 3) Institute of Pharmacology, University of Cologne, Cologne, Germany; 4) The Microarray Facility, University of Tübingen, Tübingen, Germany; 5) iPierian Inc., 951 Gateway Blvd, South San Francisco, CA 94080 USA.

Spinal muscular atrophy (SMA) is the most common genetic condition causing infant letality and no cure is available. SMA is caused by functional loss of *SMN1* leading to progressive degeneration of spinal α-motor neurons. Current therapeutic approaches focus on *SMN2*, a copy gene of *SMN1*. Indeed, *SMN2* produces only small amounts of correctly spliced full-length transcripts however, these are sufficient to influence SMA severity. Consequently, compounds activating *SMN2*, like histone deacetylase inhibitors(HDACi), ameliorate the SMA phenotype in mice and humans. The antiepileptic drug valproic acid (VPA), a short chain fatty acid HDACi, increases SMN levels *in vitro* as well as in animal models and humans. Treatment of 16 SMA patients with VPA confirmed our previous results that 1/3 of SMA patients are positive responders showing increased full-length SMN2 levels in white blood cells whereas the remaining are non- or negative- responders to VPA. Since blood is not the actual target tissue for SMA, we asked whether intra-individual responses are mirrored in a second cell system and analyzed fibroblasts from 15 VPA-treated SMA patients. In ~60% a concordant response was detected between circulating SMN2 transcripts and *in vitro* SMN levels in fibroblasts upon VPA treatment. Moreover, we generated GABAergic neurons from iPSCs of a positive and a non-responder and found a similar response to VPA in CNS neurons. This proves that VPA triggers similar effects in blood, fibroblast and CNS neurons derived from the same individual. To identify the pivotal factor suppressing VPA response, we compared the transcriptomes of positive and non-responder fibroblasts under mock- and VPA treatment. Strikingly, in non-responders no single transcript was significantly differentially expressed upon VPA treatment. Indeed, VPA did not trigger SMN2 promotor hyperacetylation in non-responders. We identified increased levels of *CD36*, an ubiquitously expressed fatty acid translocase, as the most likely underlying cause of VPA non-responsiveness. Importantly, in HEK293 cells, stably transfected with *CD36*, response to VPA was abolished. Our data provide evidence that monitoring response to VPA in fibroblasts allows inferring a response in CNS neurons and may be used to identify positive responders prior to patient treatment. This finding might have significant implications not only for SMA but also for far more common diseases typically treated with VPA such as epilepsy or migraine.

237

Melatonin: a new biomarker reflecting brain serotonin metabolism in individuals with phenylketonuria: evaluation of large neutral amino acid therapy by a randomized, double-blind crossover study. *S. Yano[1], K. Moseley[1], C. Azen[2]*. 1) Pediatrics/Gen Div, 1G24, LAC+USC Med Ctr, Keck School of Medicine, USC, Los Angeles, CA; 2) Clinical and Translational Science Institute, Keck School of Medicine, USC, Los Angeles, CA.

Background: Phenylketonuria (PKU) is one of the most common genetic metabolic defects. Early diagnosis by newborn screening and phenylalanine (Phe) restriction has been successful in preventing intellectual disability in the majority of patients. However, current studies indicate that PKU individuals have lower executive function and a higher prevalence of depressive and anxiety disorders even with blood Phe levels in the recommended range. This is believed to be due to deficiencies of neurotransmitters including serotonin and dopamine. Nevertheless, the management of PKU has been solely based on blood Phe levels with little attention to neurotransmitter precursor amino acids. Up until now there was no noninvasive method to optimize dietary treatment based on neurotransmitter metabolism. Methods: We studied blood and urine melatonin, which is a serotonin metabolite in the pinealocytes, in individuals with PKU in a randomized double-blind crossover study consisting of three 3-week phases in ten adult PKU subjects: Phase 1 (washout), Phase 2 (supplementation of large neutral amino acid (LNAA) tablets or placebo), and Phase 3 (alternate supplementation). An overnight protocol to measure blood melatonin and urine 6-sulfatoxymelatonin and dopamine in first void urine specimens was conducted after each phase for PKU subjects and once in ten control subjects. Findings: This study showed significantly decreased nocturnal blood melatonin production and low urine 6-sulfatoxymelatonin levels in PKU subjects after Phase 1 compared to controls, and significant increases with LNAA supplementation compared to the washout and placebo phases. Urine dopamine levels showed the identical changes with 6-sulfatoxymelatonin. Blood Phe did not show significant changes after LNAA supplementation. Interpretation: Urine 6-sulfatoxymelatonin and dopamine may serve as independent biomarkers, different from blood Phe, reflecting serotonin and dopamine metabolism in the brain. Monitoring these biomarkers would optimize dietary therapy with medical food products which may prevent long term neuropsychological disturbances. Funding: Funded by Applied Nutrition.

238

Beyond cholesterol: Antioxidant treatment for patients with Smith-Lemli-Opitz syndrome. *E. Elias[1], R. Braverman[2], S. Tong[3]*. 1) Dept Genetics & Pediatrics, Children's Hosp CO, Aurora, CO; 2) Dept of Ophthalmology, Children's Hosp CO; 3) Statistician, CTRC, Children's Hosp CO.

Introduction: Smith-Lemli-Opitz Syndrome (SLOS) is a devastating disorder in the final step of cholesterol biosynthesis, affecting the conversion of 7-dehydrocholesterol (7-DHC) into cholesterol. Patients with SLOS present with a complex phenotype including cognitive disabilities, autism, severe photosensitivity, deafness, and progressive retinal dystrophy. These complex issues can in part be explained by cholesterol deficiency and 7-DHC accumulation, but cholesterol supplementation alone does not correct the biochemical or clinical abnormalities. Since 2008, an antioxidant preparation in addition to cholesterol supplementation has been used to treat patients with SLOS, based on data showing improved retinal function in animals given this combination. **Methods:** 12 patients with biochemically confirmed SLOS were treated on an IRB approved protocol, with the preparation AquADEKS, in addition to a concentrated cholesterol suspension. The patients underwent electroretinograms, ophthalmologic exams under anesthesia and auditory brainstem responses at baseline and on an annual basis. **Results:** ERG data, reports of retinal evaluations and ABR's were available from patients on cholesterol supplementation alone from 2001–2008, and were compared with results following the start of antioxidant treatment. Previously reported ERG abnormalities in SLOS patients include diminished response to light (decreased amplitude), as well as a slower response (prolonged implicit time). On cholesterol alone, these abnormalities continued to worsen over time. However, when antioxidants were added, statistically significant improvement in ERG function (increased amplitude and diminished implicit time) were seen. **Conclusions:** Antioxidant medication plus cholesterol supplementation leads to improvement in retinal function in patients with SLOS. It has recently been shown that 7-DHC is oxidized to toxic substances called oxysterols, 200x more readily than is cholesterol, and that neurotoxic oxysterols accumulate in the brain and retinae of SLOS animals. Treatment with antioxidants is thus hypothesized to reduce oxysterol accumulation, and thus protect the retina from further damage. Preliminary analysis also shows similar improvement in prolonged implicit times on the ABR test. Further studies are currently underway to analyze oxysterol compounds in human patients with SLOS, and to evaluate which particular antioxidants might be most beneficial.

239

Positive effects of short course androgen therapy on the neurodevelopmental outcome in boys with 47, XXY syndrome at 36 and 72 months of age. *C. Samango-Sprouse[1,3], E. Stapleton[3], T. Sadeghin[3], F. Mitchell[3], T. Dixon[3], M. Kingery[3], A. Gropman[1,2].* 1) George Washington University of the Health Sciences, Washington, DC; 2) Department of Neurology, Children's National Medical Center, Washington, DC; 3) NDC for Young Children, Davidsonville, MD.

Background: The effects of early androgen treatment on neurodevelopmental performance in prepubertal males with 47, XXY have not been well investigated. Androgens have a profound effect on modulating neurodevelopment, brain function and behavioral outcomes from as early as 16 weeks gestation throughout adulthood. Androgen insufficiency has been described during puberty in several studies in XXY and suggests that hormone replacement therapy may have a positive outcome on brain function. Males with XXY have CNS abnormalities including endocrine and neurocognitive deficits with language based learning disabilities, dyspraxia, and musculoskeletal anomalies.

Objective: To determine if an early course of androgen treatment (3 injections of testosterone enanthate, 25mg, each) could have a positive impact on neurodevelopmental function in XXY boys immediately and later.

Methods: 101 prenatally diagnosed males with 47, XXY participated in comprehensive neurodevelopmental assessments. One group (n=34) received androgen treatment in infancy and the second was untreated (n=67). Statistical analysis was completed to determine if there was a treatment effect at 36 and 72 months on multiple domains of development.

Results: At 36 months, there was a significant positive treatment effect in multiple neurodevelopmental domains on the WISC-IV of Vocabulary p=.0007, Comprehension p=.0099, VIQ p=.0225 and FSIQ p=.0203. A positive treatment effect was observed in Vocabulary p=.0052, Comprehension p=.0165, intellectual abilities of VIQ p=.0081 and FSIQ p=.0270 and neuromotor skills p=.0197 at 72 months.

Conclusion: Improved function was observed in neurodevelopmental performance in XXY males at 36 and 72 months when treated with a short course of androgen in infancy and revealed significant improvement in areas of the brain with known androgen receptors, which have been described as deficient in XXY males. Demonstrating that an early course of hormonal replacement may have an extended positive effect and giving support to the link between neurobiological treatment and neurodevelopmental outcome and the possibility for newborn screening resulting in early treatment for XXY associated developmental disabilities.

240

A mechanism and treatment strategy for pregnancy-associated aortic dissection in Marfan syndrome. *J.P. Habashi[1], N. Huso[2], D. Bedja[2], G. Rykiel[2], J.J. Doyle[2], H.C. Dietz[2,3].* 1) Dept Pediatric Cardiology, Johns Hopkins Univ School of Medicine, Baltimore, MD; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Univ School of Medicine, Baltimore, MD; 3) Howard Hughes Medical Institute, Baltimore, MD.

Aortic root dilation and tear is the leading cause of mortality in Marfan syndrome (MFS). Recent work in mouse models has revealed a prominent role for increased TGFβ signaling, more specifically ERK activation, in the pathogenesis of vascular disease. The angiotensin II type 1 receptor blocker losartan or the MEK inhibitor RDEA-119, both of which abrogate pathologic ERK activation, normalize aortic growth and pathology in MFS mice. Women with MFS tolerate a ≈15% risk of death due to aortic dissection with each pregnancy. While this risk has been attributed to hemodynamic stress, the majority of dissections occur after delivery, and this risk is not attenuated by Cesarean section. We have now observed a startling rate (95%) of aortic dissection associated with pregnancy in our mgR/mgR mouse model of MFS (homozygous for a hypomorphic Fbn1 allele), compared to age-matched nulliparous mgR/mgR littermates($p<0.0001$), with all deaths occurring in the 4 week postpartum period. In consideration of the timing of pathogenic events in pregnancy, oxytocin release emerged as a promising candidate. Oxytocin is needed to initiate uterine contraction and milk letdown, peaks in late gestation, is sustained during breastfeeding, and mediates its effects on peripheral tissues through ERK activation. Moreover, oxytocin receptor expression is upregulated in the aorta in response to estrogen and pregnancy. These data lead to the hypothesis that oxytocin contributes to the predisposition for aortic growth and tear in MFS. In support of this, the removal of mgR/mgR females from their pups on the day of delivery, thereby stopping lactation-induced oxytocin release, confers a significant survival benefit compared to mgR/mgR female that lactate(74% survival vs. 5%, $p<0.0001$). Additionally, we show an increased rate of ascending aortic growth during the 7 week period of pregnancy and postpartum lactation(1.26±0.74) in comparison to age-matched nonpregnant mgR/mgR females(.33±0.39, $p<0.05$); removal of the pups fully normalizes aortic growth in mgR/mgR mice(0.13±0.69). Administration of exogenous oxytocin to age-matched nonpregnant mgR/mgR mice increases the rate of aortic growth (0.72±0.73) and dissection(40%). Ongoing work is examining the protective effects of knocking out the oxytocin gene or treatment with the oxytocin antagonist Atosiban in MFS mouse models. This study reveals a potential strategy to modify the vascular risks of pregnancy in MFS.

241

Pharmacogenomics, ancestry and clinical decision making for global populations. *E. Ramos[1], A. Doumatey[1], H. Huang[1], D. Shriner[1], G. Chen[1], S. Callier[2], J. Zhou[1], A. Adeyemo[1], H. Mcleod[3], C. Rotimi[1]*. 1) Center for Research on Genomics and Global Health, NIH/NHGRI, Bethesda, MD; 2) Department of Clinical Research and Leadership, School of Medicine and Health Sciences, George Washington University, Washington, DC; 3) Institute for Pharmacogenomics and Individualized Therapies, University of North Carolina, Chapel Hill, NC.

A significant component of individualizing patient care will be largely attributable to our understanding of the influence the human genome has on adverse drug reactions, dosing and other treatment modalities (i.e., pharmacogenomics). However, best practices for laying the groundwork towards treating a single individual include the consideration of genetic variation from multiple populations. Therefore, we examined 19 global populations sampled from 5 continents allowing for analysis at various levels, including inter- and intracontinental comparisons of variants relevant to drug metabolism. Specifically, we genotyped several African-ancestry populations using a platform that contains nearly 2000 variants selected from roughly 230 genes known to be involved in the absorption, distribution, metabolism, or excretion of drugs. For the remaining populations, we extracted these markers from the publically available genome sequences of the 1000 Genomes project. Minor allele frequencies (MAFs) were calculated and compared across all populations in addition to measurements of population differentiation. The set of markers were analyzed as a whole; however, a subset of markers (42 total) identified to be clinically useful was highlighted. We identified several clinically actionable single nucleotide polymorphisms (SNPs) that vary among populations in this study. For example, rs9923231, associated with warfarin metabolism (an anti-platelet drug) showed MAFs of nearly the entire range (from 0.02 in Yorubans to 0.95 in Han Chinese from Beijing). However, variation was not limited to just global comparisons. We also observed markers that varied in frequency within a given country illustrated by rs776746 (cyclosporine metabolism) in the Kenyan samples (MAF of 0.88 and 0.54 for Luhya and Masaai, respectively). In addition, the MAF for rs1801280 that predicted the acetylator phenotype of NAT2 was 0.71 in Spanish Iberian samples compared to other western European counterparts or Latin American samples, which ranged from 0.34 to 0.46. We highlighted in this study clinically actionable pharmacogenomic markers where group labels such as “black” or “Hispanic” could be a barrier to safe and effective drug selection. The data highlights the importance of casting a wide net when trying to assess the profile of clinically relevant genetic variation. Our data also speak to other ethical and social issues such as access as well as relevant public policy implications.

242

Cell line profiling in Oncology (CELLO) as a discovery platform for systematic identification of genetic and genomic biomarkers of drug sensitivity. *J. Zhong, H. Niu, J. Cai, S. Middleton, H. Bian, J. Hakenberg, C. Saisanit, F. Birzele, W. Berkofsky-Fessler, J. Rosinski, N. Sanapareddy, Z. Albertyn, B. Chen, S. Bader, G. Chen, M. Xia, L. Vassilev, A. Belousov, L. Essioux.* Roche, Nutley, NJ.

Profiling genetic and genomic alterations with drug sensitivity is a way to develop a tailored approach to treating patients with cancer. To systematically identify genetic and genomic biomarkers of sensitivity and resistance to cancer therapeutics under clinical and preclinical investigation, we profiled around three hundred cell lines in Oncology (CELLO), representing a comprehensive coverage of the cancer tissues and genetic diversity of human cancers. All cell lines were characterized by genome-wide sequencing of coding exons (Exome-seq), genome-wide detection of copy number variations (aCGH), and gene expressions (RNA-seq); together with pharmacological profiling of several Roche compounds. In this presentation, we will present the biostatistical and bioinformatics analysis on various levels of CELLO data. First, we found that exome-seq achieved reliable detections of rare variations in cancer cell lines, characterized by 0.98 positive predictive value and 0.92 negative predictive value comparing with Sanger resequencing validations. Second, the mutated cancer genes most associated with drug sensitivity were often direct targets of the compound. For example, mutation of TP53 is most significantly associated with resistance to MDM2 inhibitor that negatively regulates p53 protein levels (P<10-10). Third, wide ranges of responses remained unexplained by the mutation status of the target genes for most of the studied compounds. We discovered that a systematic search through genetic variations and gene expression biomarkers can reveal the genetic mechanism involving multiple genes and/or transcripts, and help identify pre-treatment signatures predicting drug responses to treatments. For example, we identified a classifier, containing P53 exome-seq mutations, chromosome 17 deletions and MDM2 copy number gains, that could distinguish sensitive cell lines from resistant cell lines to MDM2 inhibitor with 0.901 area under the curve (95% confidence interval, 86.3 to 91.0%). In summary, our results indicate that comprehensive cell-line collections, together with well-planned analysis workflow tailored for the complexity of cancer genome, provides a powerful biomarker discovery platform to gain insights into mechanisms of cancer therapeutics and to generate and verify hypotheses for clinical testing of the personalized health care.

243

Screening of the *TPMT* gene before thiopurine treatment results in a lower leucopenia occurrence in patient with inflammatory bowel disease. *M.J.H. Coenen[1], C.J. van Marrewijk[1], L.J.J. Derijks[2], S.H. Vermeulen[1,3], O.H. Klungel[4], A.L.M. Verbeek[3], H. Scheffer[1], B. Franke[1], H.J. Guchelaar[5], D.J. de Jong[6], TOPIC study.* 1) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 2) Department of Clinical Pharmacy, Máxima Medical Centre, Veldhoven, The Netherlands; 3) Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 4) Department of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute of Pharmaceutical Sciences, Utrecht University, The Netherlands; 5) Departments of Clinical Pharmacy and Toxicology, University Medical Center, Leiden, The Netherlands; 6) Department of Gastroenterology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Thiopurines play an important role in the treatment of inflammatory bowel diseases (IBD). Nowadays, thiopurines are becoming the first treatment of choice to prevent a complicated disease course. Unfortunately, more than 20% of the patients discontinue therapy due to severe adverse drug reactions among which leucopenia is one of the most serious side effects. Thiopurine S-methyltransferase (*TPMT*) pharmacogenetics has been proven effective for optimizing safety and efficacy of thiopurine treatment. Nonetheless, in clinical practice it is still only used on a limited scale. We aimed to investigate the added value of pre-treatment *TPMT* genotyping on the occurrence of leucopenia. We performed a prospective randomized clinical trial including 850 IBD patients starting on thiopurine treatment as part of the Dutch multicentre Thiopurine response Optimisation by Pharmacogenetic testing in IBD Clinics (TOPIC) study. Patients were randomly assigned to undergo pre-treatment screening for three common variants in *TPMT* (*TPMT**2, *3A and *3C) or to undergo standard treatment based on the Dutch treatment guidelines. Patients heterozygous for a *TPMT* variant received 50% of the standard thiopurine dose and patients homozygous for the tested variants received 0–10%. To assess the effect of pre-treatment genotyping we compared patients that were genotyped before treatment with patients that received standard treatment for the occurrence of leucopenia (white blood cell count $<3.0*10^9$ /l) in the first 5 months after treatment initiation. Of the patients 62% was diagnosed with Crohn's disease and 38% with ulcerative colitis. 64% of the patients were treated with azathioprine and 36% with 6-mercaptopurine. In total 65 patients developed leucopenia. In the pre-treatment genotyped group (n=428) as well as the group receiving standard treatment (n=422) 42 patients (10%) carried at least one genetic variant in the *TPMT* gene. Analysis of the cases carrying a genetic variant showed a statistically significant reduction of the number of leucopenia cases in the group that underwent pre-treatment genotyping: 2.4% (n=1) versus 21.4% (n=9), p-value 0.003. Using the largest prospective cohort of IBD patients treated with thiopurine we showed that pre-treatment genotyping followed by dose adjustment results in a lower prevalence of leucopenia. This study strongly implies that pharmacogenetic testing for *TPMT* should be used as standard care to individualize treatment of IBD patients.

244

PGRNseq: a new sequencing-based platform for high-throughput pharmacogenomic implementation and discovery. *A.S. Gordon[1], J.D. Smith[1], Q. Xiang[2], M.L. Metzker[2], R.A. Gibbs[2], E.R. Mardis[3], D.A. Nickerson[1], R.S. Fulton[3], S.E. Scherer[2].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Washington University Genome Center, St. Louis, MO.

Understanding the genetic basis of an individual's response to therapeutic drugs (pharmacogenetics) is a unique area of research with significant translational impact for medicine. Although pharmacogenetics has a rich history, we lack a complete picture of the common and rare genetic variation that influences an individual's response to medication. Known genetic variants with effects on important clinical phenotypes, including clopidogrel efficacy and warfarin maintenance dose, highlight the potential translational utility of pharmacogenetic analysis. The emergence of next-generation sequencing offers a promising new tool to explore the links between drug response and genetic variation, both common and rare. To characterize the spectrum of variation in human populations and to evaluate how these differences are linked to drug responses, the National Institutes of Health's Pharmacogenomics Research Network (PGRN) has developed a new platform, PGRNseq. PGRNseq is a low-cost, high-throughput next-generation sequencing platform centered around the custom capture of 84 genes with strong drug phenotype associations. Sequence captured from these genes includes coding regions and 2kb upstream to assess variation within potential regulatory regions. PGRNseq's design includes known variants present on other commercially available pharmacogenetic platforms for backwards compatibility with existing datasets. To test the performance and accuracy of this new tool, we sequenced 32 diverse trios from HapMap and 1000Genomes using the PGRNseq platform. Analysis of Mendelian inconsistencies across test trios identified paralogous regions in which better read mapping and variant calling are needed. In uniquely mapping regions, we found 99.9% genotype concordance at all overlapping sites with orthogonal datasets from HapMap and 1000Genomes. PGRNseq is able to assess known variants of clinical utility such as CYP2C9*3; we found 99.9% genotype concordance at all such sites. Copy number variation is also related to drug response in several genes (e.g. CYP2D6); we are developing methods to discover and type such events using data from PGRNseq. Low-cost, high-throughput platforms such as PGRNseq are needed in order to move genomics forward from the bench to the bedside. Our initial data suggests that PGRNseq could be successfully deployed in a clinical setting to inform patient care and generate large-scale, high-quality data for future genotype-phenotype association studies.

245

Genetic Variation in the GRK4 gene associates with susceptibility to hypertension and response to angiotensin receptor blockade. *M. White[1], Z. Wang[2], H. Sanada[3,4], M. Yoneda[3,4], S.M. Williams[1], J. Bartlett[1], L. Gordon[1], S. Chen[5], L. Asico[2], C. Escano[2], V. Villar[2], C. Zeng[6], L. Wong[7], J. Jones[2], R. Felder[8], G. Eisner[9], P. Jose[2].* 1) Human Genetics, Vanderbilt Univ, Nashville, TN; 2) 1Center for Molecular Physiology Research, Children's National Medical Center, and Department of Pediatrics, George Washington University School of Medicine and Health Sciences, Washington, DC 20010, USA; 3) 2Division of Health Science Research, Fukushima Welfare Federation of Agricultural Cooperatives, Fukushima, Japan; 4) 3Department of Nephrology, Hypertension, Diabetology, Endocrinology and Metabolism, Fukushima Medical University, School of Medicine, Fukushima, Japan; 5) Department of Physiology and Pharmacology, The University of Georgia College of Veterinary Medicine, Athens, GA 30602, USA; 6) Department of Cardiology, Daping Hospital, Third Military Medical University, Chongqing, PRC; 7) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; 8) 8Department of Pathology, University of Virginia Health Sciences Center, Charlottesville, VA 22908, USA; 9) Department of Medicine, Georgetown University Medical Center, Washington, DC 20007, USA.

Human essential hypertension (HT) is a chronic condition characterized by elevated arterial blood pressure. HT is a complex disorder, with an undefined genetic architecture likely to be influenced by multiple genetic and environmental factors. We hypothesized that variants in the renin-angiotensin system (RAS) and G-protein coupled receptor kinase (GRK) 4 gene associated with susceptibility to HT, and that GRK4 variants play a role in response to angiotensin-1 receptor blocker (ARBs) treatment. In a Japanese cohort of HT cases (n=588) and controls (n=486), we tested if any of these gene variants associated with HT and/or ARB response. Single locus Chi-square square analysis revealed that all three variants in the GRK4 gene were significantly associated with HT in both genotypic (6.65×10-7<pvalues<8.24×10–12) and allelic (5.54×10-8<pvalues<1.34×10–11) tests. A genetic variant located in the DRD1 gene was also associated with HT at the allelic level (pvalue=0.02). Logistic regression analyses, adjusted for age, BMI, and family history of hypertension, showed evidence of significant independent effects for all three GRK4 variants (2.69×10-6<pvalues<0.0006; 2.33<OR<1.42) as well as the DRD1 variant (pvalue= 0.0387, OR=0.729). Haplotype analysis compared the wild-type GRK4 three locus haplotype (G-C-C) with the variant haplotype (T-T-T) and reported an odds ratio of 3.47 (pvalue=1.25×10–11). Multi-locus analysis, using MDR, determined that the most predictive model included the GRK4 142V and GRK4 486V variants (test accuracy=0.62, pvalue<0.001). A separate cohort of Japanese patients with newly diagnosed HT (n= 881) was used to assess response to ARB as a function of GRK4 genotypes. Logistic regression analysis was used to determine the OR of response (ORR) for patients with GRK4 variants versus those with no GRK4 variants. A significant ORR was uncovered for the individuals with one copy (ORR 0.31, p<0.0001) and two copies (ORR 0.38, p<0.001) of variant alleles at GRK4 486V. Carriers of variant alleles at GRK4 142V displayed a significantly larger drop in systolic BP than non-carriers (19.36mm Hg vs. 14.58mm Hg, p=0.0012). Carriers of variant alleles at all three GRK4 polymorphisms were also significantly less likely to respond to ARBs (ORR 0.15, p<0.001). Our combined results support the conclusion that variants in the GRK4 gene are associated with HT, as well as response to ARB treatment in HT patients.

246

Genome-wide discovery of drug-dependent human liver enhancers. *R.P. Smith[1,2], K.M. Morrissey[1], X. Sun[1,2], T.J. Hoffman[2,3], K.M. Giacomini[1], N. Ahituv[1,2]*. 1) Department of Bioengineering and Therapeutic Sciences UCSF, San Francisco, CA; 2) Institute for Human Genetics, UCSF, San Francisco, CA; 3) Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA.

Interindividual variation in gene regulatory elements plays a causative role in adverse and ineffective drug reactions. However, our knowledge of the location and function of drug-dependent elements remains poor. This is particularly true in the case of enhancers, which can be over 1 Mb from their target gene. To uncover drug-associated enhancers in a genome-wide manner, we performed ChIP-Seq on primary human hepatocytes treated with the antibiotic rifampin or vehicle control. We identified 1,297 regions bearing a drug-dependent regulatory signature, exhibiting a conditional enrichment of pregnane X receptor (PXR/NR1I2) occupancy as well as three known enhancers marks (p300, H3K4me1, H3K27ac). These regions are highly enriched near genes involved in liver metabolism and response to xenobiotics, particularly those of the cytochrome p450 (CYP) family of enzymes. A parallel RNA-Seq analysis revealed 363 rifampin-induced genes whose loci largely overlap with the drug-dependent enhancer regions. Forty-two sequences bearing the rifampin-dependent enhancer signature were selected for in vitro enhancer assays in hPXR-transfected HepG2 cells. Nineteen of these drove basal and/or rifampin-dependent expression >2 fold over controls. Five of these enhancers were exclusively active upon rifampin treatment, and thus would not have been identified by genome-wide approaches using physiologically normal tissues. To probe the role of drug-dependent enhancers on human drug response, we identified nucleotide variants that overlap rifampin-induced PXR/P300 binding sites and are in linkage disequilibrium with pharmacogenomic-related genome-wide association studies (GWAS). Fifteen of these sites harbor common variants linked to GWASs for warfarin dosing, response to various drug treatments (antineoplastic agents, statins, antidepressants, anti-seizure medications), physiological readouts such as bilirubin and liver enzyme levels, and dialysis-related mortality. Functional testing of GWAS-linked sites in hPXR-transfected HepG2 cells revealed that several drive robust expression of a reporter gene when induced by rifampin. Although this study focuses on PXR activation by rifampin, our methodology is universally applicable to any drug/assay combination. The elements identified by our approach are functional, and are likely to contain many undiscovered causative variants of adverse drug reactions.

247

Genome-wide association study of vancomycin pharmacokinetics using a de-identified biorepository. *S.L. Van Driest[1], T.L. McGregor[1,2], Z. Lu[3], S. Vear[1], C.B. Creech[1], P.J. Kannankeril[1], K.B. Brothers[1,4], A. Potts[5], E. Bowton[6], J.T. Delaney[7], Y. Bradford[2], S. Wilson[2], L. Olson[2], D.C. Crawford[2], B. Saville[3], D.M. Roden[7,8], J.C. Denny[7,9]*. 1) Pediatrics, Vanderbilt University, Nashville, TN; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Biostatistics, Vanderbilt University, Nashville, TN; 4) Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, TN; 5) Pharmacy, Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, TN; 6) Office of Research, Vanderbilt University, Nashville, TN; 7) Medicine, Vanderbilt University, Nashville, TN; 8) Pharmacology, Vanderbilt University, Nashville, TN; 9) Biomedical Informatics, Vanderbilt University, Nashville, TN.

Serum levels of vancomycin, a widely-used antibiotic, directly relate to efficacy and renal toxicity. Although excreted renally as unchanged drug, vancomycin levels and the renal elimination rate constant (Ke) vary among individuals and are incompletely predicted by clinical factors. We sought to identify loci associated with vancomycin pharmacokinetics through genome-wide association (GWA). Using BioVU, the Vanderbilt biorepository linking de-identified electronic health record (EHR) data to DNA samples, we identified individuals with intravenous vancomycin therapy with documented dose and schedule, a vancomycin trough obtained 3 or more dosing intervals after the first dose, and European ancestry. Those < 18 years of age, undergoing dialysis before or during drug course, or receiving a heart transplant during drug course were excluded. Ke for each individual was estimated using trough, dose, dosing interval, and actual body weight. Age, sex, height, serum creatinine and concomitant diuretic and/or nephrotoxic drugs were extracted from the EHR. Multiple imputation was used for missing covariate information. DNA was genotyped using Illumina Omni1-Quad BeadChips. Association of genotypes to the log-transformed primary outcome (vancomycin trough) and secondary outcome (Ke) was determined using linear regression using PLINK, assuming an additive model with covariate adjustment. GWA identified one locus associated with vancomycin trough (n=745 subjects) at genome-wide significance; analysis of Ke (n=733) also identified this locus and suggested a second. rs3002142 at chr1q41 was associated with a 25% increase in the geometric mean of the vancomycin trough (beta 0.22, P=9.5×10-9) and a 14% decrease in the geometric mean of the Ke (beta -0.16, P=4.9×10-9). The nearest gene, *MIA3*, is 3382bp downstream. *HHIPL, TAF1A, AIDA*, and *BROX* are within 100kb. At 5q14.3, rs10085144 was associated with similar increase in trough (+15%, beta 0.14, P=6.3×10-6) and decrease in Ke (-11%, beta -0.11, P=4.5×10-8); *EDIL3*, the nearest gene, is 560kb upstream. For both candidate SNPs, there are no SNPs in known genes in linkage disequilibrium (r2>0.8) within 500kb based on 1000 Genomes CEU data. This first report of GWA of vancomycin pharmacokinetics finding novel associations of loci at 1q41 and 5q14.3 identifies potential mechanisms underlying vancomycin pharmacokinetic variability and may lead to improved vancomycin outcomes through pharmacogenomics.

248

Integrating multiple levels of phenotypic information to map genetic determinants of glucocorticoid sensitivity. *J. Maranville, S. Baxter, D. Witonsky, A. Di Rienzo.* University of Chicago, Human Genetics, Chicago, IL, 60637.

Genetic variation influences organism-level phenotypes (e.g. disease status) through effects on underlying molecular processes. This generates correlations across individuals between genotype and phenotype, as well as correlations between intermediate and higher level phenotypes. By combining multiple levels of phenotypic information into a single mapping experiment, we can extend on traditional association mapping (i.e. genotype-phenotype correlations) to interrogate also the relationship among phenotypes measured at different levels. This improves our ability to map phenotypic variability (on all levels) and provides insights into the mechanisms that connect genotype to organism-level phenotypes by identifying intermediate molecular processes. To this end, we measured *in vitro* cellular sensitivity to glucocorticoids (GCs), steroid hormones that are also widely used as pharmaceuticals, and transcriptome-wide response to GCs in peripheral blood mononuclear cells in a panel of 88 African-American healthy donors. Consistent with previous work, we found that transcriptional response ($\log_2$ fold change in expression) at multiple genes was correlated with *in vitro* cellular sensitivity. These included *NFKB1*, which has been previously correlated with GC response, and 27 additional genes (FDR<0.1). In a genome-wide association study, we found a major effect QTL for *in vitro* cellular sensitivity ($p=4.5\times10^{-8}$, r2=0.26). We found that this QTL is also associated with transcriptional response at 161 genes (FDR<0.1). Furthermore, we found that the genes most strongly associated with *in vitro* sensitivity are significantly more likely to be influenced by this QTL than other expressed genes, with 14 of the 27 genes associated with GC sensitivity also marginally associated with genotype (OR=13.5, $p=1.2\times10^{-9}$). These results suggest that this locus is a trans-eQTL that interacts with GC treatment to influence overall GC sensitivity. This work integrates genetic data and multi-level phenotypic information to shed light on the genetic architecture and molecular processes that underlie variation in sensitivity to a common pharmaceutical and key physiological regulator. Additional levels of phenotypic information could also be incorporated, ranging from molecular processes that mediate genotypic effects on expression to clinical outcomes that depend on *in vitro* sensitivity, eventually uncovering all the links in the chain of mechanisms connecting genotype to phenotype.

249

Common and rare genetic variation in Maturity-onset Diabetes of the Young genes influence response to interventions for diabetes prevention. *L. Billings[1,2,3], I. Tipton[4], A. Warner[1], J. McAteer[1], A. Shuldiner[5], D. Ehrmann[6], A. Manning[7,8,9], D. Dabelea[10], P. Franks[11], S. Kahn[12], T. Pollin[13], W. Knowler[14], D. Altshuler[7,8,9], K. Jablonski[4], J. Florez[1,2,3,7].* 1) Center Human Genetic Res, Massachusetts Gen Hosp, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, Massachusetts; 3) Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts; 4) The Biostatistics Center, George Washington University, Rockville, Maryland; 5) Department of Medicine, Division of Endocrinology, Diabetes & Nutrition, University of Maryland School of Medicine, Baltimore, Maryland; 6) Department of Medicine, Pritzker School of Medicine, University of Chicago, Chicago, Illinois; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts; 8) Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA; 9) Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; 10) Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Colorado; 11) Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University Diabetes Center, Skåne University Hospital Malmö, SE-205 02, Malmö, Sweden; 12) Division of Metabolism, Endocrinology and Nutrition, VA Puget Sound Health Care System and University of Washington, Seattle, Washington; 13) Departments of Medicine (Division of Endocrinology, Diabetes & Nutrition) and Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, MD; 14) Diabetes Epidemiology and Clinical Research Section, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, Arizona.

Mutations in HNF4A, HNF1A, IPF1, HNF1B, GCK, and NEUROD1 genes cause beta-cell dysfunction leading to Maturity-onset Diabetes of the Young (MODY). Genome-wide association studies have identified common genetic variants associated with type 2 diabetes and glycemic traits. It is not yet known whether common and rare genetic variation in MODY genes influences the response to insulin-sensitizing interventions. We sequenced the exons, intron-exon junctions and promoter regions of the MODY genes in 190 multiethnic participants in the Diabetes Prevention Program, followed by genotyping of the discovered variants in 3,566 participants. Subjects were randomized to placebo, metformin, or intensive lifestyle for diabetes prevention. We examined 224 common variants and an aggregate of 19 uncommon missense variants for association with diabetes incidence, baseline and change in insulinogenic index (IGR) and oral disposition index (DIo) after 1 year of the intervention. Assuming an additive model, a Cox proportional hazards model tested the association with diabetes incidence and multivariate linear regression tested for associations with glycemic traits. Analyses were stratified by treatment group for SNP*treatment interaction (Pint<0.05). The increase in diabetes risk for each additional A allele of rs11868513 at HNF1B in the placebo group (HR 1.69 [95% CI 1.36–2.10], P=2.4×10-6) was negated by the metformin (HR 0.87 [0.65–1.16], P=0.33) and lifestyle (HR 1.08 [0.79–1.47], P=0.64) interventions (Pint=0.0007). Among the metformin-treated participants, each additional G allele of rs11086926 at HNF4A conferred resistance to the preventive effects of metformin (HR 1.81 [1.35–2.43], P=8.3×10-5) compared to the placebo (HR 0.81 [0.61–1.11], P=0.2) and lifestyle (HR 1.29 [0.91–1.83], P=0.15) interventions (Pint=0.002). Variants in NEUROD1, IPF1 and HNF1B were associated with baseline log IGR. The T allele of rs832646 (NEUROD1) was associated with a deterioration in log IGR among placebo participants (β=-0.74 [0.15], P=1×10-4) and the C allele of rs3212185 (HNF4A) showed an enhancement in log DIo in response to the lifestyle intervention (β=0.74 [0.15], P=2×10-4) over one year. Neither of these associations was seen among the other treatment groups. An aggregate of uncommon missense variants was not associated with any outcome. These results provide evidence that genetic variation among MODY genes influences response to insulin-sensitizing interventions.

250

Germline Mosaicism Does Not Explain the Maternal Age Effect on Trisomy. *R. Rowsey[1], B. Murdoch[1], P. Hunt[1], C. Dickerson[2], T. Woodruff[2], T. Hassold[1]*. 1) School of Molecular Biosciences and Center for Reproductive Biology, Washington State University, Pullman, WA 99164; 2) Department of Obstetrics and Gynecology and Reproductive Sciences, University of California, San Francisco, CA 94143.

It has been over 50 years since the presence of an additional chromosome was first linked to a clinical disorder; i.e., in 1959 trisomy 21 was found to be the cause of Down syndrome. From subsequent studies, it has become clear that the vast majority of trisomic conditions are of maternal origin and that the incidence of trisomy increases with advancing maternal age. A variety of hypotheses have been proposed to explain the association between incidence and age, but no one model satisfactorily explains all aspects of the age effect. However, Hulten and colleagues (Reproduction; 139: 1–9, 2010) recently proposed a provocative model - the oocyte mosaicism selection model - that has generated considerable attention. In this model, age dependent aneuploidy is linked to events occurring in the fetal ovary, specifically to pre-meiotic nondisjunctional events in a proportion of germ cells as they mitotically proliferate; consequently, small populations of germ cells will carry an extra chromosome. It is then assumed that the presence of an additional chromosome delays meiotic progression, thereby causing these cells to be ovulated later in reproductive life, resulting in an age-dependent increase in aneuploid eggs. This model carries enormous clinical implications; accordingly, we initiated studies to confirm or refute it. By combining immunofluorescence and chromosome-specific FISH, we examined the chromosome content of human oocytes from 7 female fetuses with gestational ages between 16 and 23 weeks. We focused our analysis on leptotene oocytes, since at this early stage of prophase the individual chromosomes are not yet synapsed with their homologous partners. We analyzed over 3,000 leptotene cells, scoring the number of FISH signals for three chromosomes commonly associated with human trisomies (i.e. 13, 16, and 21). In contrast to the prediction of the oocyte mosaicism selection model, we found no evidence of trisomy mosaicism for any of these chromosomes. Accordingly, we conclude that errors in pre-meiotic germ cells are unimportant in the genesis of human aneuploidy, and cannot explain the association between maternal age and trisomy.

251

Female meiosis II errors prevalence and their impact on human embryo viability. *A. Kuliev, Z. Zlatopolsky, I. Kirillova, J. Cieslak-Janzen.* Reprod Gen Inst, Chicago, IL.

The fact that human aneuploidies originate mainly from female meiosis I (MI) is based on studies of spontaneous abortions and live-born children, representing only those aneuploidies that are compatible with implantation and recognized pregnancies. However, the lack of meiosis II (MII) errors in this material may simply be due to their failure to survive implantation, so we performed the direct testing of MI and MII outcomes in 20,946 oocytes from IVF patients of over 35 years of age, to investigate the actual prevalence of both of these errors at the zygote stage. For this purpose the first and second polar bodies (PB1 and PB2) were removed and tested by FISH analysis, using five-color probe specific for chromosomes 13, 16, 18, 21 and 22, or array-CGH. 9812 (46.8%) of tested oocytes were detected to be aneuploid, from 20% in patients of 35, to over 40% in patients of 40 years of age. In contrast to the above mentioned concept of female meiosis I origin of chromosomal aneuploidies, no difference was observed in prevalence of aneuploidies originating from MI (31.1%) and MII (33.7%). However, only 30.4% of MI, and 39.8% of MII errors were represented by isolated defects, with 29.8% of the chromosomally abnormal oocytes representing the outcomes of sequential MI and MII errors, indicating that only one third of MII errors may be due to the preceding MI errors. However, as half of MII errors are still observed independent from MI errors, the genotype of the resulting zygote cannot be evaluated without testing for MII errors. Significant differences were also observed in chromosome specific origin of MII errors, chromosomes 16 and 22 errors originating predominantly from Meiosis II. The results provide strong evidence for MII errors prevalence comparable to MI, the lack of which in recognized pregnancies is due to their significant impact on embryo viability incompatible with implantation and post-implantation development.

252

A population isolate reveals enriched recessive deleterious variants underlying neurodevelopmental traits. *O. Pietilainen[1,2,3], J. Suvisaari[5], W. Hennah[2], V. Leppa[2], T. Paunio[2,3,4], M. Torniainen[5], S. Ripatti[1,2], S. Ala-Mello[6], K. Rehnstrom[1], A. Tuulio-Henriksson[5], T. Varilo[2], J. Tallila[1], K. Kristiansson[2], M. Isohanni[7], J. Kaprio[2], J. Eriksson[8], M. Jarvelin[9], R. Durbin[1], J. Lonnqvist[4,5], M. Hurles[2], H. Stefansson[10], N. Freimer[11], M. Daly[12], A. Palotie[1,2,12]*. 1) The Wellcome Trust Sanger Institute, Cambridge, Cambridge, United Kingdom; 2) Institute for Molecular Medicine Finland FIMM, Helsinki, Finland; 3) National Institute for Health and Welfare, Public Health Genomics Unit, Helsinki, Finland; 4) University of Helsinki and Helsinki University Central Hospital, Department of Psychiatry, Helsinki, Finland; 5) National Institute for Health and Welfare, Department of Mental Health and Substance Abuse Services, Helsinki, Finland; 6) Helsinki University Central Hospital, Department of Clinical Genetics, Helsinki, Finland; 7) Department of Psychiatry, Institute of Clinical Medicine, University of Oulu, Finland; 8) National Institute for Health and Welfare, Chronic Disease Epidemiology and Prevention, Helsinki, 90014, Finland; 9) Institute of Health Sciences, University of Oulu, Oulu, Finland; 10) deCODE genetics, 101 Reykjavik, Iceland; 11) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, UCLA, Los Angeles, California, USA; 12) The Broad Institute of MIT and Harvard University, Cambridge, Ma, USA.

Low frequency variants (MAF <5%) likely contribute to susceptibility for complex traits, but their study is challenging in admix populations. We hypothesize that population isolates that have experienced bottlenecks would have an enrichment of specific low frequency variants some of which could be predisposing to complex traits. This enrichment could benefit especially identification of variants with recessive effects. To test this hypothesis, we studied homozygous deletions in a prospective birth cohort from an isolated Northern Finnish population (N=4,931). The role of rare deletions being clearly establish in abnormal neuronal development led us to constrain our initial analysis to seven supposedly relevant phenotypes including diagnosis of schizophrenia, intellectual deficit, learning difficulties, epilepsy, neonatal convulsion, impaired hearing and cerebral palsy/perinatal brain damage. The analysis included 32,487 homozygous deletions in 205 loci of which 11% included exons of one or more genes. Among the seven traits studied, the strongest association was found with impaired hearing and a deletion on 15q15.3, overlapping STRC, previously associated with deafness (p = 10-4). The largest identified homozygous deletion was 240 kb on 22q11.22 and was associated with intellectual deficit (p<0.02). The deletion showed significant regional enrichment in an internal north-eastern isolate with 3-fold risk of schizophrenia compared to elsewhere in the country. Follow up of the deletion in 265 schizophrenia patients and 5140 controls revealed an allelic association with schizophrenia (p= 0.02, OR = 1.9) and was further replicated in 9,539 cases and 15,677 controls of European origin (p = 0.03, OR = 2.1). After screening over 13,106 Finns, we identified four individuals being homozygous for the deletion, all diagnosed with schizophrenia and/or intellectual disability. The deletion overlaps a gene encoding for TOP3B and was found to down regulate its expression to half among heterozygous carriers and zero in homozygous carriers (p < 10-10). Our results demonstrate the effect of multiple consecutive population bottlenecks in the enrichment of sizable deletions contributing to abnormal neuronal development. In addition the findings highlight the usefulness of population isolates in studying rare and low frequency variants in complex traits.

253

The role of trans-acting factors on recombination in oocytes with non-disjoined chromosomes 21. *C.D. Middlebrooks[1], N. Mukhopadhyay[2], S.W. Tinker[1], E.G. Allen[1], L.J.H. Bean[1], F. Begum[2], R. Chowdhury[3], V. Cheung[3,4,5], E. Feingold[2,6], S.L. L. Sherman[1]*. 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America; 3) Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America; 4) Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America; 5) Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America; 6) Department of Human Genetics, Graduate School of Public Health University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America.

Oocytes with nondisjoined chromosomes 21 (chr21) that result in trisomy 21 (T21) after fertilization have been found to display aberrant recombination patterns. In oocytes with meiosis I (MI) nondisjunction (NDJ) errors, there is increased frequency of bivalents with no recombination or a single recombinant in the telomeric region of chr21. In those with meiosis II (MII) errors, there is elevated recombination at the centromeric region of chr21. We sought to determine whether trans-acting factors play a role in the altered recombination patterns found on nondisjoined chr21. To answer this question, we compared genome-wide recombination (GWR) patterns from families with T21 probands (N=114 families) due to a maternal MI or MII NDJ error with those from families with normal meiotic outcomes. To determine recombination patterns for T21 families, we genotyped 5,656 SNPs across the genome of probands, parents and maternal grandparents. For control families, we obtained GWR data on normal meiotic outcomes from publically available GWAS data. We used linear regression models to determine whether the frequency or location of recombination on chr21 predicted GWR patterns, stratified by meiotic outcome group (MI error, MII error and normal meiotic outcome). Among normal events we found a positive association between the number of chr21 recombinants and total number of GWR. No association was found among MI or MII events. To assess location of recombination, we conducted several analyses based on the definition of the location of GWR. In one assessment, each chromosome arm was divided into three regions: the 20% most centromeric, the 20% most telomeric and the remaining medial region. In our linear regression analysis, the outcome measure was the proportion of centromeric (or telomeric) GWR and our predictor was the presence of recombination in the centromeric (or telomeric) region of chr21. In normal outcomes, we found an association between a telomeric recombination on chr21 and the proportion of genome wide telomeric events. This association was not observed in MI or MII outcomes. There was no association with respect to centromeric events for any of the outcome groups. Analyses defining location of events in other ways (e.g., absolute number of basepairs instead of proportion) are underway. To date, our data indicate that trans-acting factors play a role in defining number of GWR events in normal meiosis, but may be perturbed in oocytes with NDJ events.

254

Large-scale Function-based Enhancer Discovery. *D.E. Dickel[1], Y. Zhu[1], A.S. Nord[1], J.A. Akiyama[1], A. Visel[1,2], L.A. Pennacchio[1,2]*. 1) Genomics Division, Lawrence Berkeley National Lab, Berkeley, CA; 2) United States Department of Energy Joint Genome Institute, Walnut Creek, CA.

Distant-acting transcriptional enhancers play a critical role in human biology and disease, but their identification and functional validation at a genomic scale remains a major challenge. Current large-scale mapping methods rely on enhancer-associated epigenomic marks, which are incompletely understood and thus miss many enhancers. To enable unbiased, high-throughput enhancer identification, we coupled pluripotent cell reporter assays and flow cytometry with DNA sequencing to screen thousands of candidate sequences in parallel for enhancer activity. Libraries of sequences, coupled to fluorescent reporter proteins, are introduced into mouse embryonic stem cells (ESCs) where they are targeted in single copy to a reproducible locus to minimize ectopic and positional effects. Using fluorescence-activated cell sorting followed by sequencing, enhancers driving robust and reproducible reporter expression can be identified. To demonstrate the efficiency of the approach in screening large genomic intervals of interest, we examined hundreds of kilobases of mouse and human sequence at 1kb resolution, including key ESC pluripotency loci. Our analysis identified multiple enhancers at these loci. Their activities were validated through complementary assays, demonstrating the utility of our approach for functional enhancer screens across extended genomic loci. We also used this approach to test a genome-wide library of non-coding ultraconserved sequences for enhancer function in ESCs. Our massively-parallel testing in conjunction with validation assays indicates a remarkable contrast between the function of these sequences in ESCs compared to later embryonic stages. While half of the tested sequences were previously shown to be in vivo enhancers during mouse embryogenesis, nearly all are inactive in ESCs, demonstrating major changes in genome-wide enhancer architecture throughout early embryonic development and supporting an evolutionary model in which enhancers active at the phylotypic stage of embryonic development are most conserved among vertebrates. The unbiased sequencing-driven approach that enabled these studies can be adapted to numerous other cell types through the use of cellular differentiation, making it a powerful method for identifying enhancers active in a variety of disease states and developmental processes.

255

A single enhancer on human chromosome 11 directly controls >1,000 promoters and distal regulatory elements genome-wide. *J.A. Stamatoyannopoulos[1], H. Wang[1], G.J. Cost[2], H. Quh[1], Y. Santago[2], J. Belton[3], R. McCord[3], S. Orlando[2], S. Neph[1], L. Zhang[2], T. Canfield[1], E. Giste[1], R. Sandstrom[1], R.S. Hansen[1], R.E. Thurman[1], P.D. Gregory[2], J. Dekker[3], F.D. Urnov[2]*. 1) Dept Genome Sci, Univ Washington, Seattle, WA; 2) Sangamo Biosciences, Point Richmond, CA; 3) University of Massachusetts, Worcester, MA.

Enhancers and locus control regions (LCRs) are believed to function chiefly in cis via direct interactions with promoters and other regulatory regions, with declining potency as a function of genomic distance. Here we show that removal of the beta-globin LCR by genome editing in erythroid cells results in precise, dose-dependent (heterozygous vs. homozygous deletion) repression or activation of >1,000 tissue-specific and ubiquitous enhancers and promoters genome-wide. These effects are equally potent at regulatory elements located in cis on chromosome 11 and those on other chromosomes. The affected elements share a unique pattern of transcription factor binding that mirrors that at the LCR. Genome-scale analysis by chromatin conformation capture (HiC) reveals that the affected elements preferentially interact both with the LCR and with one another, forming a larger co-regulatory structure. The LCR appears to program chromatin state through a `hit-and-run' mechanism in which transient interactions with a distant site are sufficient to specify its activity state with high fidelity. Our results reveal an extensive and previously-hidden layer of epigenetic control circuitry that pervades the human genome. The results also suggest that naturally-occurring non-coding deletions that remove regulatory DNA may have pleiotropic genome-wide effects.

256

Identification of trait- and disease-relevant genetic polymorphisms in microRNA target sites. *S. Busche[1], B. Ge[2], T. Kwan[2], K. Wong[3], S.-H. Chen[1], M. Georges[4], D. Ginzinger[3], T. Pastinen[1,2]*. 1) Human Genetics, McGill University, Montreal, QC, Canada; 2) McGill University and Genome Quebec Innovation Center, Montreal, QC, Canada; 3) WaferGen Biosystems, Inc., Fremont, CA, USA; 4) Faculty of Veterinary Medicine, University of Liege, Liege, Belgium.

MicroRNAs (miRNAs) are approximately 22 nucleotide long noncoding RNAs that post-transcriptionally control the function of eukaryotic genomes by acting as specific adaptors recognizing and regulating targeted mRNAs. Candidate studies suggest genetic polymorphisms in miRNA target sites (poly-miRTS) to be associated with various disorders including Tourette's syndrome, Parkinson's disease, hypertension, and cancer. We aim to create a genome-wide catalogue of poly-miRTS to provide the fundamental basis for the characterization of miRNA-related genetic alterations with an impact on human traits and disease. To identify poly-miRTS in a genome-wide scale, we monitored RNA levels and differences in allelic expression with and without active miRNA gene regulation in human fibroblasts. To shutdown miRNA-mediated gene regulation we interfered with the miRNA processing machinery by silencing Drosha, Exportin-5, and Dicer. Importantly, interference with any of these genes does not prevent siRNA processing. Upon tightly coordinated triple-silencing comprehensive miRNA expression profiling by quantitative real-time PCR carried out on WaferGen SmartChip technology identified an approximate 80% overall reduction of mature miRNA levels. We then assessed the impact of overall miRNA inhibition on mRNA expression levels by carrying out RNA-sequencing and SNP genotyping on Illumina 5M arrays. Looking at overall transcript expression levels, we unexpectedly observed similar amounts of up- and downregulated mRNAs upon silencing. By further monitoring genome-wide differences in allelic rather than total gene expression we anticipated to efficiently eliminate non-specific effects of silencing miRNA expression, which are expected to target both mRNA alleles equally. We identified hundreds of mRNAs displaying altered allelic expression, potentially due to poly-miRTS. Careful follow-up studies will reveal true polymorphic miRNA target sites impacting human traits and disease.

257

Mapping functional p53 response elements and their variants in human genome. *X. Wang[1], M.R. Campbell[1], V.G. Cheung[2], D.A. Bell[1]*. 1) Laboratory of Molecular Genetics, National Institute of Environmental Health Science, Research Triangle Park, NC 27709, USA; 2) Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

The p53 tumor suppressor acts as a master transcriptional regulator, controlling the expression of a variety of genes. In response to stress, p53 is activated and binds to response elements (REs) in the genome to regulate genes involved in major cellular pathways, such as cell-cycle arrest, DNA repair, apoptosis, and microRNA expression. Sequence-specific DNA binding is critical to p53 function. Missense mutations in p53 protein can disrupt DNA binding and are common in human cancers. On the other hand, polymorphisms in the DNA sequences bound by p53 could change the binding affinity to p53 and affect target gene expression. To date, p53 is known to mediate approximately 150 responsive genes in human. To get an extensive survey of p53REs and their variants, we have used chromatin immunoprecipitation with parallel sequencing (ChIP-Seq) technology to identify genome-wide p53 binding in human lymphoblastoid cell lines in response to the DNA-damaging chemotherapeutic reagent doxorubicin. Using ~15 million uniquely mapped 36-base sequence reads of ChIP DNA from a pool of 2 HapMap CEU cell lines, we found 2984 p53-bound genomic regions with high confidence and 73 of them contain known p53REs. The de novo motif discovery confirmed that the top enriched motif matched the p53 binding motif, and 2751 regions contain one or more putative p53REs. We observed more than 60% p53 bound regions were within the upstream, 5'UTR, and 1st introns, and were adjacent to genes with significant changes in expression under doxorubicin treatment. We identified SNPs within putative p53REs in ChIP-seq peaks and explored the impact of SNPs on p53 binding and transcription by analyzing the association between genotype and gene expression association in 362 HapMap CEU individuals. Examining genotypes imputed using 1000 Genomes as reference and baseline expression profiles, we found highly significant associations preferentially occurred within 200 kb of TSS, including 66 associations with p-value < 10-6. Among them, 18 associations conferred by SNPs that their major alleles were the predicted strong binding alleles and corresponded to higher expression levels. We also found significant enrichment of p53 ChIP-seq SNPs among SNPs identified from genome-wide association studies. This study greatly expands the p53 transcriptional landscape and reveals new insights into how variation in the p53 network may affect stress response, environmental susceptibility, and disease etiology.

258

A SNP associated with skin cancer and pigmentation disrupts a melanocyte enhancer in an intron of *IRF4*. *D.U. Gorkin[1], S.K. Loftus[2], D. Lee[3], M.A. Beer[1,3], W.J. Pavan[2], A.S. McCallion[1]*. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, Baltimore, MD 21205, USA; 2) Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; 3) Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21201, USA.

Genome-Wide Association Studies (GWAS) have identified Single Nucleotide Polymorphisms (SNPs) associated with hundreds of human phenotypes. However, efforts to identify the causative variants underlying these associations have been hindered by limited knowledge of the location and sequence composition of functional non-coding sequences. Focusing on phenotypes that involve pigment cells (melanocytes), we recently developed two tools that facilitate the identification of variants impacting functional non-coding sequences: 1) a genome-wide catalog of 2,489 melanocyte enhancers identified by ChIP-seq (EP300/H3K4me1) with a high validation rate in functional assays of 86% *in vitro* (43/50) and 70% *in vivo* (7/10), and 2) a vocabulary of sequence motifs derived by machine learning (Lee D. et al. 2011) that are predictive of melanocyte enhancer function with power to indentify additional enhancers genome-wide in both the mouse and human genomes. We used these tools to identify a melanocyte enhancer in an intron on *IRF4* that contains a SNP (rs12203592) associated with skin cancer (melanoma, basal cell carcinoma, and squamous cell carcinoma), nevus count, freckling, hair color, and eye color (Han J. et al. 2011; Duffy D.L. et al. 2010; Han J. et al. 2008; Gathany A.H. et al. 2009; Eriksson N. et al. 2010). The sequence containing rs12203592 drives strong reporter gene expression in cultured melanocytes, and shows additional hallmarks of melanocyte enhancer activity including EP300 binding and H3K4me1 enrichment in mouse melanocytes, and DNase I hypersensitivity in human primary melanocytes and melanoma cell lines. Based on our enhancer sequence vocabulary, the risk allele of rs12203592 (T) is predicted to have a strongly negative impact on enhancer function. We demonstrate that this allele significantly diminishes the ability of the enhancer to drive reporter expression in melanocytes (P= 2.7e-5). We will discuss our progress in uncovering the molecular basis of the association between rs12203592 and melanocyte-related phenotypes, as well as how knowledge of functional non-coding sequences can be systematically applied to identify variants that disrupt the function of these sequences.

259

Genetic and functional abnormalities of the melatonin biosynthesis pathway in patients with bipolar disorder. *S. Jamain[1,10], B. Etain[1,3,10], A. Dumaine[1,10], F. Bellivier[1,2,3,10], C. Pagan[7,8,9,10], L. Francelle[1,10], H. Goubran-Botros[7,8,10], S. Moreno[7,8,10], J. Deshommes[1,3,4,10], K. Moustafa[1,10], K. Le Dudal[5,10], F. Mathieu[1,10], C. Henry[1,2,3,10], J.P. Kahn[6,10], J.M Launay[9,10], T.W. Mühleisen[11,12], S. Cichon[11,12,13], T. Bourgeron[7,8,10], M. Leboyer[1,2,3,10]*. 1) Psychiatrie Génétique, INSERM U955, EQ 15, Créteil, France; 2) Université Paris Est, Faculté de Médecine, Créteil, 94000, France; 3) AP-HP, Hôpital H. Mondor - A. Chenevier, Département de Psychiatrie, Créteil, 94000, France; 4) AP-HP, Hôpital H. Mondor - A. Chenevier, Plateforme de Ressources Biologiques, Créteil, 94000, France; 5) INSERM, Centre d'Investigation Clinique 006, Hôpital H. Mondor - A. Chenevier, Pôle Recherche Clinique Santé Publique, Créteil, 94000, France; 6) Département de Psychiatrie et de Psychologie clinique, CHU de Nancy, Hôpital Jeanne-d'Arc, Toul, 54200, France; 7) Génétique Humaine et Fonctions Cognitives, Institut Pasteur, Paris, 75015, France; 8) CNRS URA 2182 “Genes, synapses et cognition”, Institut Pasteur, Paris, France; 9) Service de Biochimie, Hôpital Lariboisière, AP-HP, Faculté de Pharmacie, Paris, France; 10) Fondation Fondamental, Créteil, 94000, France; 11) Department of Genomics, Life & Brain Center, University of Bonn, D-53127 Bonn, Germany; 12) Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany; 13) Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, D-52425, Juelich, Germany.

Patients affected by bipolar disorder (BD) frequently report abnormalities in sleep/wake cycles. In addition, they showed abnormal oscillating melatonin secretion, a key regulator of circadian rhythms and sleep patterns. The acetylserotonin O-methyltransferase (ASMT) is a key enzyme of the melatonin biosynthesis and has recently been associated with psychiatric disorders such as autism spectrum disorders and depression. In this study, we analysed rare and common variants of ASMT in patients with BD and unaffected control subjects and performed functional analysis of these variants by assaying the ASMT activity in their B-lymphoblastoid cell lines. We sequenced the coding and the regulatory regions of the gene in a discovery sample of 345 patients with BD and 220 controls. We performed an association study on this discovery sample using common variants located in the promoter region and showed that rs4446909 was significantly associated with BD ($p=0.01$) and associated with a lower mRNA level ($p<0.0001$) and a lower enzymatic activity ($p<0.05$) of ASMT. A replication study and a meta-analysis using 480 independent patients with BD and 672 controls confirmed the significant association between rs4446909 and BD ($p=0.002$). These results correlate with the general lower ASMT enzymatic activity observed in patients with BD ($p=0.001$) compared with controls. Finally, several deleterious ASMT mutations identified in patients were associated with low ASMT activity ($p=0.01$). Thus, we determined how rare and common variations in ASMT might play a role in BD vulnerability and suggest a general role of melatonin as susceptibility factor for BD.

260

Massively-parallel sequencing of the brain transcriptome reveals differential expression of novel genes in bipolar disorder. *N. Akula[1], J. Barb[2], X. Jiang[1], J. Wendland[1], K. Choi[3], S. Sen[4], B.K. Lipska[5], J.E. Kleinman[5], H.C. Bravo[6], D.T. Chen[1], P.J. Munson[2], F.J. McMahon[1]*. 1) Human Genetics Branch, National Institute of Mental Health Intramural Research Program, National Institutes of Health, US Dept of Health and Human Services, Bethesda, MD, USA; 2) Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, US Dept of Health and Human Services, Bethesda, MD, USA; 3) Department of Psychiatry, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; 4) National Human Genome Research Institute, National Institutes of Health, US Dept of Health and Human Services, Bethesda, MD, USA; 5) Section on Neuropathology, Clinical Brain Disorders Branch, National Institutes of Health, US Dept of Health and Human Services, Bethesda, MD, USA; 6) Department of Computer Science, Institute for Advanced Computer Studies and Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD USA.

Massively-parallel sequencing of mRNA (RNA-seq) is a novel approach to gene expression studies that provides a direct estimate of transcript abundance and can detect alternative splicing, low abundance transcripts, novel transcripts, allele-specific expression, and post-transcriptional modifications, such as RNA-editing. Here we used RNA-seq to characterize differential expression of brain transcripts in bipolar disorder. We performed deep-sequencing (~110M-300M paired-end reads) of high quality total RNA extracted from dorsolateral prefrontal cortex (DLPFC) obtained post-mortem from 10 cases diagnosed with bipolar I disorder and 11 age- and sex-matched, psychiatrically-healthy controls. Library preparation, fragmentation and PCR enrichment of target RNA was followed by paired-end sequencing on the Illumina GA-IIx (9 samples) or HiSeq platform (12 samples). After quality control, reads were mapped and aligned to the reference genome. Principal Component Analysis was performed with JMP, differential expression was analyzed with DESeq, and results were combined across sequencing platforms by meta-analysis. Gene-set enrichment analysis (GSEA) was performed with DAVID. Enrichment for genome-wide association (GWAS) signals was tested by permutation of results from a published GWAS meta-analysis (Chen et al 2011). The first 3 principal components explained > 65% of the variance and significantly separated cases and controls. Meta-analysis identified 3 differentially expressed genes at a false discovery rate (FDR) of < 5%: PROM1, LINC00173, and CD34. A set of 1,309 unique genes differentially expressed at a nominal p-value < 5% were subjected to GSEA, demonstrating functional enrichment of 10 GO categories at FDR < 5%. Genes in 4 of these categories (“homophilic cell adhesion,” “ion homeostasis,” “passive membrane transporter activity,” and “channel activity”) showed significant enrichment of p-values in the GWAS data. A total of 62 differentially expressed genes replicated by microarray in an independent sample of 30 cases and 30 controls (hypergeometric $p=1.37 \times 10\text{-}7$). These results demonstrate that RNA-seq may reveal differential expression of genes in biologically-relevant pathways that were not detected consistently in previous microarray-based gene expression studies.

261

Rare and common gain-of-function alleles of the serotonin transporter gene, *SLC6A4*, associated with Tourette disorder. *P.R. Moya[1], J.R. Wendland[2], A.M. Andrews[3], L.M. Rubenstein[1], K.R. Timpano[4], G.A. Heiman[5], J.A. Tischfield[5], R.A. King[6], S. Rammamoorthy[7], F.J. McMahon[8], D.L. Murphy[1]*. 1) Laboratory of Clinical Science, National Institute of Mental Health, Bethesda, MD; 2) F. Hoffman-La Roche AG, pRED, Pharma Research and Early Development, DTA CNS, Grenzacherstrasse 124, Basel, Switzerland, CH-4070; 3) Semel Institute for Neuroscience and Human Behavior and California Nano Systems Institute, University of California, Los Angeles, CA; 4) Department of Psychology, University of Miami, Coral Gables, FL; 5) Human Genetics Institute of New Jersey and Department of Genetics, Rutgers University, Piscataway, NJ; 6) Yale Child Study Center, New Haven, CT; 7) Department of Neurosciences, Medical University of South Carolina, Charleston, SC; 8) Human Genetics Branch, NIMH-IRP, Bethesda, MD.

To evaluate possible genetic contributions to Tourette disorder (TD), functional alleles in the serotonin transporter gene *SLC6A4* were evaluated in TD probands, their relatives and controls (total Caucasian N = 1284). Among common alleles, the higher-expressing triallelic 5-HTTLPR/rs25531 L_A variant was more prevalent in TD probands than controls ($\chi^2 = 5.75$, p = 0.017, OR = 1.35). The L_A variant was significantly more frequent in probands with TD alone than in those with TD plus OCD (Fisher's exact test, p = 0.004, OR = 3.46). Further, the rare gain-of-function SERT I425V coding variant was found in three male siblings with TD and/or OCD and their father. Two of these siblings and their father had a congenital renal/vesicuolureteral disorder. The cumulative count of SERT I425V thus becomes 1.42% in OCD/TD spectrum conditions vs. 0.20% in controls, with a recalculated, family-adjusted significance of $\chi^2 = 9.28$, p = 0.002, OR = 5.72 (total N = 2552 genotyped). This first-reported large case-control study of *SLC6A4* variants in TD adds weight to the concept of greater SERT expression and function as a potential contributor to serotonergic system abnormalities in TD and other related spectrum disorders. This report provides an unusual combination of both common and rare variants in one gene, all found associated with potentially causative *SLC6A4* over-expression and regulatory consequences from SERT I425V.

262

***GLRB* is the third major gene-of-effect in hyperekplexia or startle disease.** *S.K. Chung[1], A. Bode[2], C.A. Hunt[1], A. Derrick[1], T.D. Cushion[1], S. Wood[1], C. Drew[1], O.W. Howells[1], R.H. Thomas[1], J.G. Mullins[1], J. Lynch[2], M.I. Rees[1]*. 1) Neurology Research and Molecular Neuroscience, College of Medicine, Swansea, United Kingdom; 2) Queensland Brain Institute, University of Queensland, Brisbane, Australia.

Glycinergic neurotransmission is a major inhibitory influence in the CNS and dysregulation is associated with a paediatric startle disorder, hyperekplexia. Glycinergic neurons are enriched in the brainstem and spinal cord as well as other highly-specialized areas of the human brain. The α1-subunit of the inhibitory glycine receptor (*GLRA1*) and its cognate postsynaptic glycine transporter (*SLC6A5*) are well-established genes-of-effect in hyperekplexia, nevertheless, 60% of cases remain gene-negative. The β-subunit of the glycine receptor (*GLRB*) is the heteropentameric partner of GLRA1 and was screened through 92 gene-negative hyperekplexia patients using a Sanger sequencing approach. All variants were analysed for in vitro functional consequences. This study identified 10 recessive and dominant *GLRB* variants in 12 independent hyperekplexia index-cases. This included 4 nonsense, 2 indel, and 1 recurrent splicing variants where the biological consequence is unambiguous. A further 2 missense mutations were also discovered and investigated for electrophysiological properties. They showed a marked reduction in maximal current and / or decreased glycine sensitivity supporting molecular modelling predictions for damaging effects. This comprehensive study describes the definitive assignment of *GLRB* as the 3rd major gene for hyperekplexia and impacts on the genetic diagnosis and biological causation of this neonatal/paediatric disorder.

263

Functional analysis of rare chimeric genes in schizophrenia. *C. Rippey[1], C. Remmers[3], M. Cahill[3], A. Nord[1], T. Walsh[2], M. Lee[2], M. Gasperini[2], P. Penzes[3], J. McClellan[4], M.-C. King[1,2]*. 1) Dept Genome Sciences, Univ Washington, Seattle, WA; 2) Department of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Physiology, Northwestern University, Chicago; 4) Department of Psychiatry, University of Washington, Seattle, WA.

Structural rearrangements are continuously arising in the human genome and are known to contribute to neurodevelopmental illnesses such as schizophrenia. Chimeric genes result when such rearrangements fuse together portions of two different genes, creating a novel gene that may differ from its parent genes in localization, regulation, or function. Because these changes are so dramatic, gene fusions are uniquely suited to play crucial roles in both evolution and disease. We are focusing on newly arisen chimeric genes in patients with schizophrenia. We hypothesize that brain-expressed chimeric genes contribute to schizophrenia pathogenesis, and that we can gain insight into the neuronal pathways involved by studying the effects of these chimeras on neurons. To test these hypotheses, we screened DNA from 124 individuals with schizophrenia and 240 controls using arrayCGH and scanned genome-wide for copy number variants (CNVs) longer than 30kb not present in the Database of Genomic Variants or in other samples run on the same platform in our lab. From these events, we selected those predicted to delete or duplicate the 5' end of one gene and the 3' end of another, as these CNVs should produce chimeras. We detected four such events in cases, three resulting from tandem duplications and one from a deletion. We confirmed the breakpoints of each of these events by PCR and sequencing. For all events, at least one of the two parent genes was shown to be expressed in human brain RNA by gene-specific RT-PCR. In order to evaluate the effects of these events on protein stability and localization, as well as on neuronal morphology and connectivity, we obtained lymphoblast RNA from each individual and targeted primers to predicted chimeric mRNA. We then cloned full-length chimeric transcripts and full-length parent genes into V5 epitope-tagged expression vectors. We transfected these constructs into cultured cortical neurons. Each fusion gene tested produced a stable protein detectable both by western blot and immunostaining. For the chimeric genes, we observed changes both in protein localization and in dendritic architecture. This latter observation recapitulates changes seen in animal models of schizophrenia as well as post-mortem studies of schizophrenia brains. De novo formation of chimeric genes represents a novel mechanism for major mental illness, as well as implicating new genes and pathways in schizophrenia.

264

Excess homozygosity in the major histocompatibility complex(MHC) in schizophrenia. *S. Mukherjee[1,2], S. Guha[1,2], M. Ikeda[5], N. Iwata[5], A.K. Malhotra[1,2], I. Pe'er[3], A. Darvasi[4], T. Lencz[1,2]*. 1) Psychiatry Research, North Shore LIJ health System, The Zucker Hillside Hospital, Glen Oaks, New York, NY; 2) Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Manhasset, NY; 3). Computer Science, Columbia University, New York, NY; 4) Genetics, Hebrew University, Jerusalem, Israel; 5) Department of Psychiatry, School of Medicine, Fujita Health University, Toyoake, Aichi, Japan.

The major histocompatibility complex (MHC) region on chromosome 6 has been associated with schizophrenia in genome wide association studies (GWAS). Notably, GWAS to date have generally focused on additive allelic effects. However, the ecological literature suggests that homozygosity at the MHC locus may be associated with vulnerability to disease. We applied a novel approach to study homozygosity in the MHC region in an ethnically homogenous schizophrenia case-control cohort to understand the etiology of schizophrenia. We genotyped 904 schizophrenia cases and 1640 healthy controls drawn from the Ashkenazi Jewish (AJ) population using the Illumina HumanOmni1-Quad array. Phased (using Beagle 3.0) genotype data was processed with GERMLINE and DASH algorithms (under default parameters) to perform pairwise comparisons across all chromosomes, thereby identifying chromosomal segments (haplotypes) shared identical-by-descent (IBD). We extracted all such segments shared IBD across at least 3 chromosomes in our dataset, and compared homozygosity at such segments in cases and controls. We focused on specifically on data for the extended MHC locus (chr6: 25Mb–35Mb). For the replication study, we analyzed the genotype data for 548 schizophrenia cases from Japan and 542 matched healthy controls(Japan cohort). We found a significant excess of homozygosity in schizophrenia cases compared to controls in the MHC (p-value =2.2e-16, OR=1.3 in AJ cohort and p-value = 0.002, OR = 1.42 in the Japan cohort). By contrast, amount of homozygosity-by-descent was relatively equal across the rest of the genome (OR=0.98 in AJ cohort and OR = 1.008 for Japan dataset). Within the extended MHC locus, we observed that most of the homozygous regions which were over-represented in cases relative to controls were located within the classical MHC class I regions. Homozygosity in the classical MHC region appears to convey significant risk for schizophrenia.

265

Significant risk of new mutations for Huntington disease: CAG-size specific risk estimates of intermediate allele repeat instability. *A. Semaka, C. Kay, C. Doty, J.A. Collins, M.R. Hayden.* Medical Genetics, University of British Columbia, Centre for Molecular Medicine & Therapeutics, Vancouver, BC, Canada.

Intermediate alleles (IAs) for Huntington disease (HD) have between 27–35 CAG repeats. While they do not confer the HD phenotype, they are prone to paternal germline CAG repeat instability. Consequently, IAs may expand into the HD range (≥36 CAG) upon transmission to the next generation producing a new mutation. Quantified risk estimates for IA repeat instability are extremely limited but urgently needed to inform genetic counselling. Small-pool PCR analysis was used to quantitatively assess the frequency and magnitude of repeat instability in sperm of males with an IA. Approximately 18,763 sperm cells from 31 different IAs were examined. A significant non-linear relationship was observed between CAG size and the frequency of instability ($r=0.794$, $p<0.001$). Control alleles (≤26 CAG, n=35) were highly stable, with only 2.2%; of sperm (n=490/22,446) demonstrating instability, which was biased towards contraction (1.6%) compared to expansion (0.6%). There was a 6-fold increase in instability over the intermediate CAG size range, with 27 CAG alleles (n=5) demonstrating 5.5% (n=161/2,907 sperm) instability and 35 CAG alleles (n=4) having 33.0% (n=756/2,290) instability. IAs exhibited a higher frequency of contractions until 33 CAG when a switch toward a preponderance of expansions occurred. As a group, 3.4% (n=610/18,763 sperm) of IAs expanded into the HD range producing a new mutation. The new mutation rate ranged from 0.1% to 21.0% for 27 and 35 CAG alleles, respectively, with a 9-fold increase occurring between 34 and 35 CAG. The majority of new mutations were within the reduced penetrance HD range (36–39 CAG, 3.1%, n=566/18,763 sperm), compared to the full penetrance range (≥40 CAG, 0.3%, n=45/18,763). The magnitude of instability was quantified by the repeat length variation (i.e. +1 CAG, -5 CAG). The overall magnitude of instability increased with increasing CAG size. For control alleles, the repeat length variation of contractions was greater than expansions (range: -10 CAG to +3 CAG). Conversely, for IAs the magnitude of expansions was greater than contractions (range: -12 CAG to +20 CAG). This is the largest study to provide quantified CAG size-specific risk estimates of IA repeat instability. These figures are vital to providing persons with IAs accurate genetic counselling and will assist them in their reproductive decision-making. This data also increases our knowledge on the dynamics of CAG repeat instability in HD.

266

Mutations in AKT3 Lead to Hemimegalencephaly. *A. Poduri[1,2], G.D. Evrony[2,3,4], X. Cai[2,3,4], P.C. Elhosary[1,4], R. Beroukhim[2,5,6,7], M.K. Lehtinen[2,4,8], L.B. Hills[4], E.L. Heinzen[9], A. Hill[4], R.S. Hill[4,7], B.J. Barry[4], B.F.D. Bourgeois[1,2], J.J. Riviello[1,2,10], A.J. Barkovich[11], P.M. Black[12], J. Madsen[2,13], K.L. Ligon[2,8,14,15], C.A. Walsh[2,4,7].* 1) Neurology, Boston Children's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; 4) Division of Genetics, Manton Center for Orphan Disease Research, and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA; 5) Department of Cancer Biology and Center for Cancer Genome Discovery, Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA; 6) Department of Medicine, Brigham and Women's Hospital, Boston, MA; 7) Broad Institute, Cambridge, MA; 8) Department of Pathology, Boston Children's Hospital, Boston, MA; 9) Center for Human Genome Variation, Duke University, Durham, NC; 10) Comprehensive Epilepsy Center, New York University Langone Medical Center, New York, NY; 11) Department of Radiology, University of California San Francisco, CA; 12) World Federation of Neurosurgical Societies, Nyon, Vaud, Switzerland; 13) Departments of Neurosurgery, Boston Children's Hospital and Brigham and Women's Hospital, Boston, MA; 14) Department of Medical Oncology and Department of Pathology and Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA; 15) Department of Pathology, Brigham and Women's Hospital, Boston, MA.

Statement of purpose: Hemimegalencephaly (HMG) is a developmental brain disorder characterized by an enlarged, malformed cerebral hemisphere. While HMG has been associated with syndromes such as Proteus syndrome, genetic causes of isolated HMG have not been described. Methods: We identified cases of HMG who underwent epilepsy surgery. All research was conducted in accordance with the IRB of Boston Children's Hospital. We studied DNA from HMG tissue and leukocytes from the same individuals when possible. We used copy number analysis of SNP data, quantitative PCR confirmation, and sequencing, to test whether HMG reflects somatic mutations of genes critical to brain development. We studied one case using TOPO cloning and single cell sequencing to quantify mosaicism. Results: (1)We initially identified 8 cases of HMG and recently identified an additional 21 with DNA from brain, leukocytes, and/or buccal cells available. We found that 2/8 initial HMG brain samples showed trisomy of chromosome 1q, which includes *AKT3*, a serine threonine kinase known to regulate brain size. SNP data and qPCR-calculated copy number of 2.4–2.8 indicated mosaic trisomy. We also showed that the mutation was present in brain but not blood in one case. (2)A third brain sample from the original 8 cases revealed a mutation in *AKT3* (c.49G→A, creating p.E17K, previously reported as an activating mutation) present in ~35 percent (by TOPO cloning) of brain cells and absent from the patient's leukocytes. Single cell analysis confirmed the mutation is present in both neurons and glia. (3)Preliminary study of the new cases suggests a novel *AKT3* mutation, W330X, in a buccal cell sample from an individual with HMG. Conclusions: We provide a genetic explanation for three cases of HMG involving somatic mosaic activation of *AKT3* in brain, either by duplication or activating point mutation leading to hemispheric overgrowth. The *AKT3* E17K mutation is paralogous to E17K mutations in *AKT1* and *AKT2*, recently described in other overgrowth syndromes. We conclude that somatic mutations arising de novo in the developing brain can be responsible for HMG, suggesting that somatic mosaic mutations may represent an important cause of complex neurogenetic disease. A fourth case shows a novel truncating mutation in *AKT3*, the pathogenicity of which is under study. Given the brain-predominant role of *AKT3*, it is also possible that constitutional *AKT3* mutations can also lead to HMG.

267

De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *J. Lee[1,2], M. Huynh[3], G. Mathern[3], J. Gleeson[2]*. 1) Translational Neurogenetics Laboratory, Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305–701, Republic of Korea; 2) Institute for Genomic Medicine, Rady Children's Hospital, Howard Hughes Medical Institute, University of California, San Diego, CA 92093; 3) Departments of Neurosurgery and Psychiatry & BioBehavioral Medicine, Mattel Children's Hospital, David Geffen School of Medicine, University of California, Los Angeles, CA 90095.

De novo somatic mutations in focal areas are well documented in diseases like neoplasia but are rarely reported in malformation of developing brain. Hemimegalencephaly (HME) is characterized by overgrowth of either one of the two cerebral hemispheres. The molecular etiology of HME remains a mystery. The associated intractable epilepsy can be relieved by the surgical treatment of hemispherectomy, allowing sampling of diseased patient tissue. Exome sequencing and mass spectrometry analysis in paired brain/blood samples from HME patients (n=20) identified de novo somatic mutations in 30% of patients in the PIK3CA, AKT3, and MTOR genes. A recurrent PIK3CA c.1633G>A mutation was encountered in four separate patients. Identified mutations were present in ~8–40% of sequenced alleles in various brain regions and were associated with increased neuronal S6 protein phosphorylation in patient brains, indicating aberrant activation of mTOR signaling. Thus HME is likely a genetically mosaic disease due to gain-of-function in PI3K-AKT3-mTOR signaling.

268

Minimal differences in single nucleotide variation calls between blood and cell-line derived DNA from the same individuals. *C.M. Schafer[1], N.G. Campbell[2], G. Cai[3], J.S. Sutcliffe[2], J.D. Buxbaum[3], K. Roeder[1], ARRA Autism Sequencing Consortium.* 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA., USA; 2) Vanderbilt Brain Institute, Departments of Molecular Physiology & Biophysics and Psychiatry, Vanderbilt University, Nashville, TN, USA; 3) Seaver Autism Center for Research and Treatment, Mount Sinai School of Medicine, New York, NY, USA.

Whole-exome (WES) and genome sequencing are now common tools to study genetic disorders. Scientists have assembled large collections of whole blood-derived DNA samples for use in genetic studies, but replenishable DNA resources for long-term research require substantial investment to create immortalized lymphoblastoid cell lines (LCLs). Concerns now exist about use of LCL DNA for sequencing, since transformation and serial passaging of cells may result in a large number of DNA sequence changes. We compared results from WES of DNA from both whole blood and LCLs for 16 subjects, with LCLs of low passage number (<5) predominantly established at the NIMH repository at Rutgers (http://www.rucdr.org/). We sought to empirically investigate the implications of using LCL instead of blood DNA and to explore the ability of standard analysis pipelines to filter low-quality genotype calls. We initially found a large number of discordant genotype calls between blood and LCL samples (n=15,099; ~1000/subject), despite good depth and strong confidence in the individual calls. Further filtering, however, reduced candidate disagreements to 864. For these remaining candidates, read-level analyses of sequence data involved evaluation of base-call, mapping quality and read direction. This permitted segregation of candidate blood/LCL differences into categories of “confidence”. Sanger sequencing showed several instances in which discordant genotypes appeared as low to very low level mosasicism in the LCL based on the presence of a small amplitude peak (SAP) corresponding to the called non-reference allele. We estimate overall that the vast majority of the 864 candidate differences are false positives and that our most confident predictions were effective in detecting valid sequence differences, including LCLs with putative mosaicism for the non-reference allele. Ultimately, there were no unambiguous, high-quality cases of mutation taking place from blood to cell line. Mutation did occur in a fraction of the loci, but these appear to be mosaics in cell line, and hence could potentially be identified and filtered using improved diagnostic measures. Hence, this work confirms the validity of the use of NIMH cell lines - and other cell lines of similar low passage number - for sequencing studies. Moreover, while it supports the effectiveness of standard approaches for discarding low-quality variant calls, directions for improvement are evident and simple to implement.

269

The impact of genetic variation on diabetes-related quantitative traits from whole exome sequences: The T2D-GENES Consortium. *H.M. Highland[1], X. Sim[2], A. Manning[3,4,5], M. Rivas[6], G. Atzmon[7], S. Choi[8], B.K. Cornes[9,10], J. Dupuis[11,12], J.C. Florez[3,10,13], P. Fontanillas[3], T. Frayling[14], E.R. Gamazon[15], I.-S. Huh[16], H.K. Im[17], J. Kim[18], Y.J. Kim[19], C.M. Lindgren[6], A. Locke[2], J.B. Meigs[9,13], A.P. Morris[6], N. Palmer[20], I. Prokopenko[6,21], T.M. Teslovich[2], T2D-GENES Consortium.* 1) Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48105, USA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA; 5) Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; 6) University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; 7) Departments of Medicine and Genetics, Albert Einstein College of Medicine; 8) Interdisciplinary program in Bioinformatics, Seoul National University, Gwanak-gu, Seoul 151-742, Korea; 9) General Medicine Division, Massachusetts General Hospital; 10) Department of Medicine, Harvard Medical School; 11) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; 12) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA; 13) Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital; 14) Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK; 15) Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL 60637, USA; 16) Seoul National University, Gwanak-gu, Seoul 151-742, Korea; 17) Department of Health Studies, The University of Chicago, Chicago, IL 60637, USA; 18) Department of Statistics, Seoul National University, Kwan-ak St. 599, Kwan-ak Gu, Seoul, South Korea 151-741, Republic of Korea; 19) Center for Genome Science National Institute of Health, KCDC, Korea; 20) Department of Biochemistry and Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston Salem, NC 27157, USA; 21) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK.

Type 2 diabetes, which is increasing at epidemic rates, is associated with high blood pressure, dyslipidemia, and obesity. While many genetic loci have been associated with various diabetes-related traits, most studies have focused on populations of European ancestry. The T2D-GENES consortium was formed to investigate the genetics of type 2 diabetes in ethnically-diverse populations. To identify ubiquitous and population-specific variants associated with diabetes and diabetes-related traits, 10,000 individuals from five ethnic backgrounds are undergoing whole exome sequencing using Agilent TruSeq exome capture at the Broad Sequencing Center. Approximately half in each group have been diagnosed with type 2 diabetes, the remainder are controls. The ethnic groups are African-Americans, Mexican-Americans, Europeans (Finnish and Ashkenazim), South Asians (Indian Asians from London and Singapore), and East Asians (Chinese in Singapore and Koreans). Here we report preliminary results from an initial data freeze of deep exome sequencing of 3,581 individuals, consisting of African-American, and East and South Asian samples for anthropometric measures, lipids, and blood pressure and, for the non-diabetic subjects, glycemic traits. We identified 1,651,864 exome variants consisting of 599,841 missense variants; 12,073 nonsense variants; 375,110 synonymous variants; and 664,604 noncoding variants. We have detected genome-wide significant associations in covariate adjusted, single variant analyses at known GWAS loci: *CETP* ($p = 3.9\times10^{-12}$) and *LPL* ($p = 1.7\times10^{-8}$) for HDL, *TOMM40* ($p = 1.7\times10^{-10}$) for LDL, and *APOA5* ($p = 1.2\times10^{-13}$) for triglycerides. Other suggestive associations include *RNF26* ($p = 6.3\times10^{-7}$) with fasting insulin, *ZBTB41* ($p = 5.9\times10^{-7}$) with diastolic blood pressure, and *PDE3A* ($p = 1.8\times10^{-6}$) for BMI which was previously shown to be associated with HDL and adiponectin. Rare variant analysis with SKAT across all individuals identified suggestive associations with the genes *GCNT2* ($p = 5.4\times10^{-5}$) for fasting glucose, *TTN* ($p = 6.9\times10^{-5}$) for HDL, *THYN1* ($p = 8.2\times10^{-5}$)for cholesterol, *INPP5E* ($p = 6.3\times10^{-5}$) for systolic blood pressure, and *AGFG2* ($p = 2.7\times10^{-5}$) for BMI. Investigation of the entire set of 10,000 individuals will permit an unprecedented evaluation of the contribution of rare and common genetic variation to diabetes-related traits in ancestrally diverse populations.

270

Whole-Exome Sequencing in Multiplex Families Identifies Novel Rare Variants in Multiple Sclerosis. *A.H. Beecham[1], J.L. McCauley[1], A. Hadjixenofontos[1], P.L. Whitehead[1], I. Konidari[1], A. Aviram[1], Y. Pasco[1], S.L. Hauser[2], J.R. Oksenberg[2], D.J. Hedges[1], J.M. Vance[1], J.L. Haines[3], M.A. Pericak-Vance[1]*. 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Department of Neurology, School of Medicine, University of California, San Francisco, CA, USA; 3) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN, USA.

Multiple sclerosis (MS) is a common neurodegenerative disease, affecting more than 1.3 million individuals worldwide. Given the number of multiplex families found in MS, it is plausible that rare variants may contribute to MS in these families. We have used next-generation sequencing to scan multiplex families for rare variants. DNA was captured using the Agilent 50Mb kit and sequenced using Illumina HiSeq2000. Alignment and genotype calling were performed with Burrows-Wheeler Aligner and Genome Analysis ToolKit. Filtering 127 individuals from 26 multiplex families, assuming a dominant model with incomplete penetrance, we identified high quality coding (nonsense, missense, or splice), conserved, potentially damaging, and rare (MAF ≤ 0.01) variants in OCLN, C6orf170, SAMD3, TLN2, TBC1D21, and FANCI, which segregated completely in at least 2 families. We also identified variants in confirmed MS genes, EOMES and CLEC16A, segregating in 1 family each. These 8 variants were genotyped in 8,090 individuals (4,009 controls, 1,762 affected, and 2,319 unaffected from MS families). As expected for a rare allele, the strongest association, seen with C6orf170 (p=0.066), was not significant. An additional 570 variants met all filtering criteria. Using a gene based approach, we sought to determine if any of the 578 variants segregating in at least 1 family were in the same gene and how many families had distinct variants segregating in the same gene. We found 18 genes with at least 2 distinct variants segregating in 2 different families. Of those, 4 genes (DNAH1, DCHS2, ADAMTSL1, and SAMD3) had at least 3 distinct variants segregating in 3 families. Assuming a recessive model, we identified high quality coding, conserved, potentially damaging, and rare (homozygote frequency ≤ 0.05) variants segregating completely in at least 1 family. We identified 2 new genes having a distinct variant segregating in multiple families (C1orf103 and MCM10) as well as 3 genes (C14orf21, TRD5, and COL11A2) that were also identified assuming dominance. In total, we identified 27 novel genes having rare variants which segregate completely in at least 2 families. In particular, COL11A2 had 2 distinct variants segregating in 2 families assuming dominance, for a total of 3 distinct variants segregating in 3 different families, one being recessive. COL11A2 is in the HLA complex and has shown association with outcome prediction of MS, thus being an excellent candidate locus.

271

A Mendelian randomisation study on vitamin D status and blood pressure: a meta-analysis in up to 89,042 individuals. *KS. Vimaleswaran[1], DJ. Berry[1], A. Cavadino[1], MR. Järvelin[2,3,4], E. Hyppönen[1], D-CarDia collaboration.* 1) MRC Centre for the Epidemiology of Child Health, UCL Institute of Child Health, London, United Kingdom; 2) Institute of Health Sciences and Biocenter Oulu, University of Oulu, Oulu, Finland; 3) Department of Children, Young People and Families, National Institute for Health and Welfare, Oulu, Finland; 4) Department of Biostatistics and Epidemiology, School of Public Health, MRC-HPA Centre for Environment and Health, Imperial College, Faculty of Medicine, London, United Kingdom.

Vitamin D deficiency is highly prevalent and associated with blood pressure (BP) and hypertension in observational studies. Although vitamin D-mediated suppression of renin-angiotensin system could provide one mechanistic explanation, the causality of association has not been established. We used a Mendelian Randomization (MR) approach, that helps to overcome confounding and reverse causality affecting other observational studies, to investigate the causal relationship between vitamin D status [measured by 25-hydroxyvitamin D, 25(OH)D] and BP and hypertension. We examined associations of four vitamin D-related SNPs (chosen from a recent genome-wide association study) with 25(OH)D (for validation) and with BP/hypertension (for causal association) individually and in combination using separate allele scores for SNPs involved in synthesis (comprising *DHCR7* and *CYP2R1*) and metabolism (comprising *GC* and *CYP24A1*) of 25(OH)D. Meta-analysis included 89,042 individuals from 30 studies, with all models adjusted for age, age^2, gender, BMI, and region. 25(OH)D levels were associated with systolic (SBP) [-0.14 (95% CI: -0.20,-0.08), $p=2.4\times10^{-6}$] and diastolic BP (DBP) [-0.05 (95% CI: -0.09,-0.01), p=0.02] and hypertension (OR: 0.98 (95% CI: 0.97, 0.99), $p=2\times10^{-4}$). All four SNPs were strongly associated with 25(OH)D (*DHCR7*, $p=4.43\times10^{-29}$; *CYP2R1*, $p=8.28\times10^{-22}$; *GC*, $p=4.11\times10^{-148}$; *CYP24A1*, $p=3.77\times10^{-10}$). Of the four SNPs, *CYP2R1* (rs12794714; p.Ser59Ser) was associated with higher DBP (increase per allele, 11% (95% CI: 1%, 21%), p=0.03), while there was no association with SBP (p=0.31) or hypertension (p=0.16). When meta-analysing the data together with results from the International Consortium for Blood Pressure (ICBP) to increase statistical power (N=133,857, overlapping studies excluded) both *CYP2R1* [increase per allele, 12% (95% CI: 3%, 20%), p=0.006] and the synthesis score [increase per allele, 7% (95% CI: 1%, 14%), p=0.02] were significantly associated with DBP. Instrumental variable analyses using *CYP2R1* SNP as the instrument suggested a modest causal association between vitamin D status and DBP, with increasing 25(OH)D concentration from 25 to 75 nmol/l associated with 3.2% lower DBP (p=0.009). In conclusion, these data support a causal role of higher 25(OH)D in leading to reductions in blood pressure, providing further support for important non-skeletal effects for vitamin D.

272

APOE modulates the relationship among triglycerides, cholesterol, and CHD through pleiotropy and gene-gene interactions. *T.J. Maxwell[1], C.M. Ballantyne[2], J.M. Cheverud[3], C.S. Guild[4], C.E. Ndumele[5], E. Boerwinkle[2]*. 1) Human Gen Ctr, UT Houston Hlth Sci Ctr, Houston, TX; 2) Section of Atherosclerosis and Vascular Medicine, Baylor College of Medicine, and Center for Cardiovascular Disease Prevention, Methodist DeBakey Heart and Vascular Center, Houston, TX; 3) Department of Anatomy and Neurobiology, Washington University, St. Louis, MO; 4) Cameron Guild MD, Department of Medicine, University of Mississippi Medical Center,Jackson, MS; 5) Division of Cardiology, Johns Hopkins University School of Medicine, Baltimmore, MD.

We introduce the idea of loci that modify the relationship between traits mediated through GxG interactions. These relationship loci (rQTL) exist when the correlation between multiple traits varies by genotype. rQTL are often involved in gene-by-gene (GxG) or gene-by-environmental interactions, making them a powerful tool for detecting GxG. Empirically, we present an analysis of APOE with respect to lipid traits and incident CHD that additionally led to the discovery various loci that interact with APOE to affect these traits. Using ARIC GWAS data, we found that the relationship between Total Cholesterol (TC) and ln(Triglycerides) (TG) varies by APOE isoform genotype in African (AA) and European American (EA) populations (10-5 AA; p=10-7 EA) replicating earlier findings by Boerwinkle et al. (1987) in a small European sample. The overall correlation (0.24 AA; 0.33 EA) in both populations is reflective of the e3e3 genotype, which is by far the most common genotype. The e4 allele tends to reduce the correlation while the e2 allele tends to increase it. This led to a priori hypotheses that APOE genotypes may affect the relationship of TC and/or TG with incident CHD. We found that APOE*TC was significant (p=0.016) for African Americans (AA) but not European Americans (EA) while APOE*TG was significant for EA (p=0.027) but not AA. The e2e4 and e4e4 genotypes have values <1 in both cases suggesting that for individuals with either of these two genotypes, TC in EA and TG in AA have either no relationship with CHD or if anything a negative relationship. On the other hand, the e2e2 and e2e3 genotypes result in much stronger associations of TG (in AA) and TC (in EA) with CHD. In both overall populations, CHD & TC and CHD & TG are positively correlated and the marginal coefficients for both are similar to the coefficient for the within e3e3 class. The opposing effects and low frequencies of non-e3e3 genotypes result in having little impact on the overall correlations. This is an example of a relationship QTL (rQTL) its use to find a significant and interesting relationship between these important lipid traits and CHD. The scans for loci that significantly interact with APOE produced ten loci for African Americans (1 for CHD, 6 for TC, 1 for LDL, 2 for HDL). Six loci were found for European Americans, all for TC. rQTL are a powerful tool to identify loci that modify the relationship between risk factors and disease and point to GxG interactions.

273

Statistical inference of tissue-consistent and tissue-specific eQTLs. *T. Flutre[1,2], X. Wen[3], J. Pritchard[1], M. Stephens[1,4]*. 1) Human Genetics, University of Chicago, Chicago, IL; 2) Génétique et Amélioration des Plantes, INRA, Versailles, France; 3) Biostatistics, University of Michigan, Ann Harbor, MI; 4) Statistics, University of Chicago, Chicago, IL.

Understanding the genetic basis of variation in gene expression is now a well-established route to identifying regulatory genetic variants, and has the potential to yield important novel insights into gene regulation, and, ultimately, the biology of disease. Statistical methods for identification of eQTLs in a single tissue or cell type are now relatively mature, and several studies have shown the benefits in power obtained by the use of appropriate statistical methods, notably data normalization, robust testing procedures, and using dimension reduction techniques to control for unmeasured confounding factors. Here we consider statistical analysis methods for an important problem that until now has received less attention: combining information effectively across expression data from multiple tissues. The aims of such studies include both the identification of regulatory variants that are shared across tissues (“tissue-consistent”) and that are specific to one or a few tissues (“tissue-specific”). The simplest approach is to analyse data on each tissue separately, and then to examine overlap of results among tissues. Here we argue that substantial gains in both power and interpretability can be made by the use of more sophisticated analytical methods recently developed for meta-analysis of genetic association studies, that instead analyze all tissues simultaneously, taking account of the potential heterogeneity in effects among tissues. We illustrate these methods, and their potential benefits, on a dataset of Fibroblasts, LCLs and T-cells from the same set of 80 individuals (Dimas *et al.* 2009). We show a large gain in power when analyzing all tissues jointly: at an FDR of 0.05, the joint analysis identified 50% more eQTLs than separate analyses in each tissue. Compared with previous analyses of these data, we find a much higher rate of tissue-consistent eQTLs: our analyses estimated 69% of identified eQTLs to be shared across all 3 tissues (credible interval [61%,77%]). Moreover, among eQTLs shared between 2 of the 3 tissues (14%), the vast majority (12%) were shared between LCLs and T-cells, and absent in Fibroblasts, consistent with biological expectations of similarities among these cell types. We also consider the challenges in extending these methods to larger data sets, consisting of dozens of tissues, such as those currently being collected as part of the NIH GTEx project.

274

Estimates of penetrance for common pathogenic copy number variations. *J.A. Rosenfeld[1], B.P. Coe[2], E.E. Eichler[2,3], H. Cuckle[4], L.G. Shaffer[1]*. 1) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute, University of Washington, Seattle, WA; 4) Department of Obstetrics and Gynecology, Columbia University Medical Center, New York, NY.

Over the past several years, molecular cytogenetic techniques such as microarray-based comparative genomic hybridization (aCGH) have facilitated the discovery of pathogenic copy number variations (CNVs) that may be inherited from phenotypically normal parents. Although the parental carriers are healthy, their children manifest a variety of phenotypes including developmental delays, autistic features, and congenital anomalies. Population studies show that these CNVs are enriched among individuals with abnormal phenotypes, implicating them as &ldquo′edisposing” or “susceptibility” loci. As the use of microarrays in prenatal settings increases, more fetuses will be identified with these CNVs of reduced penetrance and variable expressivity. This can lead to counseling dilemmas and parental anxiety, especially in low-risk pregnancies, because the associated neurodevelopmental phenotypes cannot be ascertained prenatally and because it is difficult to quantify the risk to the fetus. We conducted a Bayesian analysis, based on the CNVs' frequencies in control populations (n=22,246) and in our database of >48,000 postnatal aCGH samples, to calculate empiric estimates of the penetrance for these CNVs. The background risk for congenital anomalies/developmental delay/intellectual disability was assumed to be ~5%. We calculated penetrance estimates for 1q21.1 proximal duplications, 1q21.1 distal deletions and duplications, 15q11.2 deletions, 16p13.11 deletions, 16p12.1 deletions, 16p11.2 proximal and distal deletions and duplications, 17q12 deletions and duplications, and 22q11.21 duplications. Estimates for the risk of an abnormal phenotype ranged from 10.4% for 15q11.2 deletions (the CNV with the highest frequency among controls) to 62.4% for distal 16p11.2 deletions (a CNV only seen once in our large control cohort). This model can be used to provide more precise estimates for the penetrance, and thus the chance of an abnormal phenotype, for many CNVs encountered in the prenatal setting. By providing the penetrance, additional, critical information is given to prospective parents in the genetic counseling session.

275

Combining Illumina gene expression microarrays from different tissues: methodological aspects. *K. Heim[1], C. Schurmann[2], A. Schillert[3], C. Müller[4], T. Zeller[4], C. Herder[5], J. Kruppa[3], T. Illig[6,7], G. Homuth[2], K. Strauch[8,9], A. Peters[10], H. Wallaschofski[11], M. Dörr[12], T. Meitinger[1,13], P.S. Wild[14,15], S. Blankenberg[4], U. Völker[2], M. Roden[5,16], A. Teumer[2], H. Prokisch[1,13], A. Ziegler[3] on behalf of the MetaXpress consortium.* 1) Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 2) Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt University Greifswald, Germany; 3) Institute for Medical Biometry and Statistics, University at Lübeck, Lübeck University Hospital Schleswig-Holstein, Campus Lübeck, Germany; 4) University Heart Center Hamburg, Clinic for General and Interventional Cardiology, Hamburg, Germany; 5) Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany; 6) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 7) Hannover Unified Biobank, Hannover Medical School, Hannover, Germany; 8) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; 9) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 10) Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 11) Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Germany; 12) Department of Internal Medicine B - Cardiology, Angiology and Pneumology, University Medicine, Greifswald, Germany; 13) Institute of Human Genetics, Technical University Munich, Munich, Germany; 14) Center for Thrombosis and Hemostasis, University Medical Center Mainz, Johannes Gutenberg University Mainz, Germany; 15) Department of Medicine 2, University Medical Center Mainz, Johannes Gutenberg University Mainz, Germany; 16) Department of Metabolic Diseases, University Clinics Düsseldorf, Heinrich-Heine University, Düsseldorf, Germany.

Microarray profiling of gene expression data is widely applied in molecular biology and functional genomics. Experimental and technical variations make meta-analysis of different studies challenging. In a total of 3,331 samples, all from German population-based cohorts and measured on the Illumina Human HT-12 v3 BeadChip array, we investigate the influence of preprocessing and technical effects in blood monocyte and whole blood gene expression data. First, a principal component analysis on the probe's intensities was performed. The EigenR2 algorithm was applied to identify the amount of variance explained by pre-defined variables. The possible impact of these variables was tested by associating the expression levels with BMI and a randomly generated normal distributed phenotype. Possible negative hybridization effects due to SNPs located within a probe's sequence were analyzed by testing the association of 2,148 probes and their enclosed SNPs. Both, in whole blood, and in monocyte samples, the first principal component (PC) explained more than 95%; of the variation. The technical factors that influenced the overall expression level most were the Illumina chip design, the plate layout after amplification, the RNA quality and the time between blood donation and RNA isolation (whole blood samples) or the time between RNA isolation and amplification (monocytes samples). These technical parameters were used as covariates to reduce the residual variance and for lowering the number of false negatives in association studies. Adjusting additionally for the first PC did hardly reduce the mean standard errors of the association results between the expression levels and the random phenotype. The strongest non-technical influences were observed for white blood cell composition parameters, sex effects and somatometric parameters including BMI. Despite many SNPs located within the probes' sequences were associated with a decrease of expression levels per mismatch allele, an increase was identified in almost 45%; of the tests. We conclude that SNPs within probes do not have an effect on hybridization efficiency. Adjustments for technical factors improve results and reliability of gene expression analysis and should be part of every gene expression analysis. Further analyses within this consortium included the association between gene expression levels and BMI, whereas different pathways showed up depending on the cell types the expression data was generated from.

276

A DNA variant caller adapted to assess mitochondrial DNA variation in lymphocytes from 1,000 Sardinians. *J. Ding[1], C. Sidore[2,3], O. Meirelles[1], M.K. Trost[2], F. Busonero[2,3], R. Nagaraja[1], F. Cucca[3], G.R. Abecasis[2], D. Schlessinger[1].* 1) Laboratory of Genetics, National Institute on Aging, Baltimore, MD; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy.

The degree to which mitochondrial DNA (mtDNA) varies heritably and somatically has been much discussed, but has not been systematically analyzed on a population basis. To approach such an analysis of mtDNA variation, genotype calling programs developed for nuclear DNA must be modified, because each cell has 100–10,000 mtDNA copies that can vary at any site (i.e., heteroplasmy). We have developed an algorithm that is adapted to identify variants in mtDNA; it incorporates in a likelihood calculation the sequencing error rate at each base in the sequence reads and is flexible to allow for different allele fractions at a variant site across all individuals. The program has been employed to assess homoplasmies and heteroplasmies in mtDNA sequences of lymphocytes from ~1,000 SardiNIA Project participants. The distribution of the number of homoplasmies per individual is bimodal. One group shows relatively fewer homoplasmies (mode of 11), whereas the other has a mode of 32. Compared to the current reference phylogenetic tree, both groups belong to European haplogroups - the first falls into the HV subgroup; the second is predominantly correlated with several other clades, including J, T and K subgroups. The number of heteroplasmies with a minor allele fraction threshold of 4% varies considerably among individuals, with most in a range of 0–3, but some showing considerably more. The overall heteroplasmy increases with age, but the slope is small, yielding an average increase of ~1 heteroplasmy between ages 20 and 80. As expected, mothers and their children share essentially all homoplasmies but share a lesser proportion of heteroplasmies. These results are consistent with the accumulation of heteroplasmies during the life span. Thus far, the determination of heteroplasmies is limited to those that reach a threshold level (i.e., 4%) of minor allele fraction in the cell samples. A complete census of mtDNA variation would require determinations of the rarer heteroplasmies; but the current approach is limited by levels of sequencing accuracy. The results to date provide information about mtDNA haplogroups and the inheritance of homo- and heteroplasmies in Sardinia; and about the extent of accumulation of heteroplasmies during aging. The algorithm can be further extended in several ways: for example, to study mtDNA from cloned normal cells in greater depth, and to investigate the nuclear DNA variability in cancer cells.

277

Direct Measure of Human Somatic Base-Substitution Mutation Rate in Monozygotic Twins. *J.B. Richards[1,2], R. Li[1], A. Montpetit[3], T.D. Spector[2], C. Polychronakos[4]*. 1) Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, Lady Davis Institute, Jewish Genral Hospital, McGill University, Montreal, Quebec, Canada; 2) Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 3) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 4) Departments of Pediatrics and Human Genetics, McGill University, Montreal, Quebec, Canada.

Mutation is the ultimate source of genetic variation and can be disease causing or improve reproductive fitness. The understanding of the nature of human mutation accumulation is essential to many aspects of medical genetics and human evolution. Germline mutation rates have been widely studied by utilizing the data at loci causing Mendelian disease, comparing putatively neutral evolving nucleotide sequences between humans and chimpanzees, and comparing the genome sequences of relatives. The estimate of human mutation rate per nucleotide site, per generation is consistent from these different approaches, which has been suggested to be ~10-8. However the somatic rate, which is distinct from the germline mutation rate, has not been thoroughly investigated in humans, but is central to the etiology of many diseases, such as cancer. In this study we genome-wide genotyped 92 pairs of monozygotic twins and directly measured the somatic mutation rate by single base substitution in 506,786 high quality genotype sites. We confirmed candidate somatic mutations by Sanger sequencing 33 pairs of monozygotic twins for whom DNA was available. Two somatic mutations were verified from two distinct pairs of twins. One was a G to A transition located in the intron of EDIL 3 gene. The other mutation was a G to T transversion on a CpG dinucleotide site in the 5'UTR of the TCF19 gene. We assumed that these two mutations occurred early in embryonic development since they did not occur in loci likely to confer survival advantage and Sanger sequence is only able to identify mutations that are present in the majority of a cell population. Given that the final number of active hematopoietic stem cells that a human individual needs is estimated as 385, we estimated the somatic mutation rate in the early development ranges between 3.13 × 10-10 and 1.20 × 10-7 per nucleotide per cell division. These data allow us to estimate that each individual carries, on the average, approximately 359 post-zygotic mutations that happened early in development or conferred enough of a selective advantage to be present in the majority of blood cells. These findings provide direct evidence that somatic mutations do occur and can lead to differences in genomes between otherwise identical twins, suggesting that mosaicism due to such mutations is reasonably common among the trillions of mitosis that occur over the human lifespan.

278

Estimating human mutation rate using autozygosity in a founder population. *C.D. Campbell[1], J.X. Chong[2], M. Malig[1], A. Ko[1], B.L. Dumont[1], L. Han[2], L. Vives[1], B.J. O'Roak[1], P.H. Sudmant[1], M. Abney[2], C. Ober[2], E.E. Eichler[1,3]*. 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Human Genetics, The University of Chicago; 3) Howard Hughes Medical Institute.

Understanding the rate and pattern of new mutation is critical to our understanding of human disease and evolution. Direct estimates have been complicated by the relatively modest number of events per generation, difficulties in distinguishing somatic (including cell line-derived) from germ-line mutations, and biases in targeted capture and resequencing. We took advantage of the extensive autozygosity (i.e. homozygosity by recent decent) in a genealogically well-defined population of Hutterites to estimate the mutation rate over multiple generations, since heterozygous mutations in these regions are necessarily recent in origin. We sequenced whole genomes from whole blood DNA of 15 individuals from five parent-offspring trios. Based on comparisons to genomes from other populations, we observed a 33-fold increase in autozygous basepairs in segments greater than 2 Mbp in Hutterite individuals. We focused on 44 segments of autozygosity greater than 5 Mbp in the Hutterites, since segments of this size were absent in the genomes of non-Hutterite individuals. Using genotyping data from the extended Hutterite pedigree, we computed the number of meioses separating each pair of autozygous alleles and determined the number of heterozygous single nucleotide variants (SNVs) in each segment. We validated 72 SNVs from 498 Mbp of autozygous DNA providing an SNV mutation rate of 1.21×10^{-8} (95% confidence interval $1\text{-}1.5\times10^{-8}$) mutations per basepair per generation. We observed a nine-fold increase in the mutation rate for bases within CpG dinucleotides (10.3×10-8 mutations per CpG basepair and per generation) and strong evidence ($p = 0.002$) for a paternal bias in the origin of new mutations. In addition, we found evidence of a long-range gene conversion event (at least 6 kbp) occurring within six meioses at the *C4*, *TNX* locus on chromosome 6. Interestingly, even after correcting for CpG bias, we observed a nonrandom distribution of heterozygous SNVs (both novel and known) in the autozygous segments ($p = 0.001$) suggesting mutational hotspots or sites of long-range gene conversion.

279

The Myth of Random Mating: Evidence of ancestry-related assortative mating across 3 generations in Framingham, MA. *R. Sebro[1,2], G. Peloso[3,4], J. Dupuis[5,6], N. Risch[1,7,8]*. 1) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 2) Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA; 3) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 5) Department of Biostatistics, Boston University School of Public Health, Boston MA; 6) The National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA; 7) Department of Biostatistics and Epidemiology, University of California, San Francisco, San Francisco, CA; 8) Division of Research, Kaiser Permanente, Oakland, CA.

The factors that influence spouse selection are important to geneticists because the mating pattern determines the genetic structure of a population. There has been evidence of positive assortative mating (PAM) related to several phenotypic traits like height. Ancestrally-related PAM is necessary for genetic population stratification, which means spouses are more likely to share genes of common ancestry. Prior studies have shown strong ancestry-related assortative mating among Latino populations. Here, Caucasian spouse pairs from the Framingham Heart Study (FHS) Original and Offspring cohorts (N=885) genotyped on Affymetrix 500K were analyzed using principal components (PC) analysis. Data from individuals genotyped in HapMap and the Human Genome Diversity Project (HGDP) were projected onto these PCs to facilitate interpretation. Based on these and other data, the first principal component delineates the prominent northwest-to-southeast European cline. In our data, there was clear clustering on this axis, probably separating individuals of English/Irish/German ancestry from those of Italian ancestry. The second principal component also reveals strong clustering, and likely reveals individuals of Ashkenazi Jewish ancestry. In the Original (older) cohort, there is a very strong correlation in PC1 between the spouses ($r=0.73$, $P=2e\text{-}22$) and also for PC2 ($r=0.80$, $P=4e\text{-}29$). In the Offspring cohort the spouse correlations were lower but still highly significant: $r=0.38$, $P=3e\text{-}28$ for PC1 and $r=0.45$, $P=9e\text{-}40$) for PC2. Examination of scatter plots for spouse pairs in the two generations reveals both a reduction in clustering and lower but still evident correlation in the Offspring cohort. Of genetic impact, we observed highly significant Hardy-Weinberg disequilibrium (homozygote excess) for SNPs loading heavily on PC1 and PC2 across 3 generations, and also highly significant linkage disequilibrium between the same set of SNPs located on different chromosomes. These results are consistent with demographic patterns of social homogamy which have existed in Framingham over several generations, and a general trend of reduced homogamy over time. While Framingham is not representative of the general US population, its historic mating patterns serve as a reminder that assumptions of Hardy Weinberg and Linkage Equilibrium need to be made with caution when applied to genetic loci that are related to ancestry in any population.

280

Combined analysis of loss-of-function variants in protein-coding genes from over 16,000 individuals. *D.G. MacArthur[1,2], M. Lek[1,2], K. Shakir[2], S. Balasubramanian[3], E. Lim[1,2], B.M. Neale[1,2], L. Habegger[3], S. Gabriel[2], P. Sullivan[4], S. Kathiresan[1,2], M.I. McCarthy[5], M. Boehnke[6], S. Purcell[1], S.A. McCarroll[2,7], M.B. Gerstein[3], D. Altshuler[1,2], M.A. Depristo[2], M.J. Daly[1,2].* 1) Massachusetts General Hospital, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) Yale University, New Haven, CT; 4) University of North Carolina School of Medicine, Chapel Hill, NC; 5) University of Oxford, Oxford, UK; 6) University of Michigan, Ann Arbor, MI; 7) Harvard Medical School, Boston, MA.

Recent large-scale genome sequencing studies have revealed that genetic variants predicted to cause complete loss-of-function (LoF) of protein-coding genes are found in all humans. While common LoF variants reveal genes that can be inactivated without major phenotypic effects, rare LoF variants are highly enriched for deleterious sequence changes likely to have an impact on human health. Exploring the phenotypic consequences of LoF variants will identify novel disease risk variants, provide insight into the function of previously uncharacterized human genes, and potentially identify new molecular targets for therapeutic intervention.

We have performed the largest ever systematic survey of LoF variants, using exome data from over 16,000 individuals sequenced as part of a variety of population genetic and disease-specific studies. To ensure consistently accurate variant-calling across the data-set, raw sequence data from these individuals have been processed, recalibrated and jointly analyzed for evidence of variation. All genetic variants identified in this cohort have been analyzed using a custom informatic pipeline for the identification and filtering of LoF variants, which removes most common classes of sequencing and annotation artifact, resulting in a high-confidence catalogue of human gene-inactivating variants.

Our results provide a view of the spectrum of human functional genetic variation extending down to extremely low population frequencies. We describe the global properties of human LoF variants, including their frequency spectrum and distribution across genes, and the implications of these properties for human evolutionary history, for large-scale studies exploring the genetic basis of complex and monogenic diseases, and for identification and validation of molecular targets for therapeutic development. We also discuss systematic analyses of the impact of predicted LoF variants on gene expression and mRNA splicing.

Finally, we outline the development of new genotyping arrays for rapid, cost-effective association studies with this systematic collection of LoF variants in large phenotyped human cohorts.

281

Abundant selection explains low diversity on human Y chromosomes. *M. Wilson Sayres[1,2], K. Lohmueller[1,2], R. Nielsen[1,2].* 1) Integrative Biology, University of California, Berkeley, Berkeley, CA; 2) Statistics, University of California, Berkeley, Berkeley, CA.

The human Y chromosome exhibits levels of diversity that are significantly lower than expected under neutral population genetic theory. Variance in male reproductive success (reducing the effective population size of males relative to females) has recently been proposed as an alternative neutral model to explain reduced diversity on the Y relative to mtDNA. Generally Y chromosomes are not included in whole genome analyses, so explicit tests of this hypothesis have yet to be conducted. Here we show that neutral models with unequal male and female effective population sizes are not consistent with observed genome-wide diversity on autosomes, X, Y and mtDNA across completely sequenced males. Instead, a model including selection is needed to explain the departure of observed Y diversity from expectations. We found that models with similar estimates of the strength of background selection can explain diversity for both the Y chromosome and mitochondrial genomes. Our results suggest that strong selection is necessary for explaining the evolutionary history of the human Y chromosome, and argue against the concept of the “junk” Y chromosome.

282

The genomic geography of close relatives across Europe. *P. Ralph, G. Coop.* Evolution and Ecology, UC Davis, Davis, CA.

Numbers of common ancestors shared at various points in time across populations can tell us about recent demography, migration, and population movements. These rates of shared ancestry from the past tens of generations can be inferred from genomic data, thereby dramatically increasing our ability to infer population history much more recent than was previously possible with population genetic techniques. We have analyzed patterns of IBD in a dataset of thousands of Europeans from across the continent, which provide a window into recent European geographic structure and migration over the past 3,000 years. These patterns are formed on many scales, from recombination modification near several inversions and centromeres, to the Slavic population expansion of around 600AD.

283

Evolutionary history and adaptation inferred from whole-genome sequences of diverse African hunter-gatherers. *J. Lachance[1], B. Vernot[2], C. Elbers[1], B. Ferwerda[1], A. Froment[3], J. Bodo[4], G. Lema[5], W. Fu[2], T. Nyambo[5], T. Rebbeck[6], K. Zhang[7], J. Akey[2], S. Tishkoff[1].* 1) Departments of Biology and Genetics, University of Pennsylvania, Philadelphia, PA 19104 USA; 2) Department of Genome Sciences, University of Washington, Seattle, WA 98185 USA; 3) IRD-MNHN, Musée de l'Homme, 75116 Paris, France; 4) Ministère de la Recherche Scientifique et de l'Innovation, BP1457, Yaoundé, Cameroon; 5) Department of Biochemistry, Muhimbili University College of Health Sciences, Dar es Salaam, Tanzania; 6) Perelman School of Medicine Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA 19104 USA; 7) Department of Bioengineering, Institute for Genomic Medicine and Institute of Engineering in Medicine, University of California at San Diego, San Diego, CA 92093 USA.

Hunter-gatherer populations have distinctive patterns of genetic variation. To reconstruct modern human evolutionary history and identify loci that have shaped hunter-gatherer adaptation, we sequenced the whole genomes of five individuals in each of three different hunter-gatherer populations at >60x coverage: Western African Pygmies from Cameroon and Khoesan-speaking Hadza and Sandawe from Tanzania. In these 15 genomes we identify 13.4 million variants, many of which are novel, substantially increasing the set of known human variation. We identify evidence of introgression of archaic lineages in all three hunter-gatherer populations and the distribution of time to most recent common ancestors from putatively introgressed regions is similar to that observed for introgressed regions in Europeans. All three hunter-gatherer populations maintain high levels of genetic diversity, although there is evidence of a population bottleneck and inbreeding among the Hadza. We find that natural selection continues to shape the genomes of hunter-gatherers, that deleterious genetic variation is found at similar levels across ethnically diverse African populations, and that non-African populations contain a larger proportion of deleterious variants than African populations. In addition, the genomes of each hunter-gatherer population contain unique signatures of local adaptation. These highly-divergent genomic regions include genes involved in immunity, metabolism, olfactory and taste perception, reproduction, and wound healing. Within the Pygmy population, we identify multiple highly differentiated loci that play a role in growth and anterior pituitary function and find statistically significant associations with height.

284

Mapping the human genome's missing pieces using population admixture. *G. Genovese[1,2,3], R.E. Handsaker[1,3], H. Li[1,3], N. Altemose[4], A.M. Lindgren[5], K. Chambert[1], B. Pasaniuc[6], A. Price[1,6], D. Reich[3], C.C. Morton[1,3,5], M.R. Pollak[1,2], J.G. Wilson[7], S.A. McCarroll[1,3].* 1) Broad Institute, Cambridge, MA; 2) Beth Israel Deaconess Medical Center, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) University of Oxford, Oxford, United Kingdom; 5) Brigham and Women's Hospital, Boston, MA; 6) Harvard School of Public Health, Boston, MA; 7) University of Mississippi Medical Center, Jackson, MS.

Almost 30 megabase pairs of euchromatic human genome sequence has no home on the reference human genome assembly; these sequences, including many protein-coding genes, are paralogous to known human genome sequences but have no known location in the human genome themselves. We describe an approach for localizing the human genome's missing pieces by utilizing the statistical patterns of genome sequence variation that have been created by recent admixture of human populations that were historically separated. We describe an approach for localizing the human genome's missing pieces by utilizing the patterns of genome sequence variation created by population admixture. We first leverage the data available from the 1000 Genomes Project to identify variation in assembled sequence currently not mapped to the canonial human genome reference. Then we leverage the long-range “admixture linkage disequilibrium” present in the genomes of African-American individuals to map novel sequence missing from the human genome reference and to map cryptic segmental duplications. We mapped several unlocalized genomic scaffolds spanning over five million base pairs of the human genome's unplaced euchromatic sequence, including about a dozen protein-coding genes, and identified eight large (>40kbp) and novel inter-chromosomal segmental duplications (the largest of which, ~240kbp, has never been described in the literature). We find that most of these sequences are hidden in the genome's heterochromatic regions, particularly its pericentromeric regions and the short arms of the acrocentric chromosomes. We evaluated a sample of these mappings by FISH analysis, with successful validation of the novel locations in each case. We also find that cryptic, centromeric genes are expressed at an RNA level, though their expression patterns often diverged from those of their known paralogs and that these missing pieces are more prone to CNVs than the rest of the euchromatic genome. Our approach, based on mapping clones through population admixture, is complementary to conventional clone tiling path approaches based on overlapping sequence at the end of clones, and might play an important role in completing physical maps of the euchromatic part of the human genome, particularly in cases where euchromatic sequence is buried inside extensive repeat-rich sequence.

285

When ancestry runs deep: Trans-species polymorphisms in apes. *L. Segurel[1,2], E. Leffler[1], Z. Gao[1], S. Pfeifer[3], A. Auton[4], O. Venn[4], L. Stevison[5], A. Venkat[1,2], J.L. Kelley[6], J. Kidd[6], C. Bustamante[6], R. Bontrop[7], M. Hammer[8], J. Wall[5], P. Donnelly[3,4], G. McVean[3,4], M. Przeworski[1,2,9].* 1) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 2) Howard Hughes Medical Institute, University of Chicago, Chicago, IL, USA; 3) Department of Statistics, University of Oxford, Oxford, UK; 4) Wellcome Trust Centre for Human Genetics, Oxford, UK; 5) UCSF, San Francisco, CA, USA; 6) Stanford University, Stanford, CA, USA; 7) Biomedical Primate Research Centre, Rijswijk, the Netherlands; 8) University of Arizona, Tucson, AZ, USA; 9) Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA.

Balancing selection refers to the maintenance of more than one allele in the population for a longer time than would be expected from genetic drift alone. In cases where balancing selection pressures have existed for millions of years, different species can share the same polymorphism identical by descent, resulting in a trans-species polymorphism. This mode of selection is thought to be extremely rare, with only MHC and ABO known in humans. However, modeling work suggests that it leaves a small footprint in genetic variation, so additional cases may have gone undetected. With these considerations in mind, we searched for trans-species polymorphisms between humans and chimpanzees using genome-wide resequencing data for 10 western chimpanzees from the PanMap project and 179 humans from the 1000 Genomes Pilot 1 data. We considered all orthologous sites polymorphic for the same alleles in both species, and found them to be in excess of what would be expected by chance after controlling for local variation in mutation rates. We then focused on shared haplotypes, i.e. cases of two or more close shared SNPs with the same linkage disequilibrium patterns in both species, in order to minimize the contribution of recurrent mutation to allele sharing between species. In addition to the MHC region, we identified over 100 cases, a set significantly enriched for transmembrane glycoproteins, which are often involved in interactions with pathogens. To further rule out the possibility of deep coalescent events by chance, we examined patterns of variation in seven samples of Gorilla gorilla. We discovered 25 cases shared among all three species, which we verified by Sanger sequencing. In a subset, within species diversity levels were unusually high and the tree of haplotypes clustered by allelic type rather than by species, providing definitive evidence for trans-species polymorphisms. Interestingly, when we considered non-synonymous polymorphisms shared between the three species, transmembrane glycoproteins again emerged as enriched. Thus, our scan revealed numerous targets of long-lived balancing selection in apes, narrowed down to only a couple of possible causal SNPs, and revealed a common theme to targets of balancing selection, likely related to immune response. Given our conservative criteria, our findings indicate that ancient balancing selection may be more common than previously believed.

286

Dark matter of the diseasome: Annotating personal genomes for gene regulatory disease risk alleles. *G. Bejerano.* Beckman B-300, 279 Campus Drive West, Stanford University, Stanford, CA 94305-5329 USA.

Two years ago our lab has developed http://GREAT.stanford.edu to derive insights into transcription factor (TF) function from ChIP-seq data. With the insights derived from this platform we have developed a powerful approach to annotate the human genome for transcription factor function, binding sites and target genes in a multitude of biological contexts. Building on over two million facts about protein function, multiple resources documenting experimental enhancer function, transcription factor binding sites and open chromatin, as well as over 1,000 non-redundant transcription factor binding site motifs we are able to annotate the human reference genome for gene regulation function with unprecedented detail and accuracy. Armed with this knowledge we turn our attention to personal genomes where the majority of current effort is devoted to coding sequence variation analysis. Using the published personal genome and medical history of an individual, where the original authors have focused mostly on coding mutations alone (Ashley et al., Lancet, 2010), we find many additional gene regulatory mutations. We rank these mutations by multiple factors including GWAS association to a relevant phenotype or disease and/or binding site disruption of regulating factors and target genes associated with relevant phenotypes or disease. Follow up functional assays implicate disease risk alleles with expression level modulation. Moreover, aggregate analysis of patient personal mutations predicted to alter transcription efficiency reveal set-level susceptibility for cholesterol and lipoprotein level abnormalities, heart disease risk etc. In summary, by building a unified experimental-predictive framework for annotating the human genome for gene regulatory function, we expose a rich layer of patient specific gene regulatory mutations very likely contributing significantly to the disease associations observed in the patient's medical record.

287

Causal mutation discovery using next generation sequencing data: Development and application of a pipeline to reduce false positive calls and to map regions of shared homozygosity and IBD. *S. Gulsuner[1], T. Walsh[1], A.C. Watts[1], M.K. Lee[1], T. Ozcelik[2], M.C. King[1].* 1) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA, USA; 2) Department of Molecular Biology and Genetics, Bilkent University, Ankara, TURKEY.

Next generation sequencing technologies have brought enormous successes for disease gene discovery but also challenges for data analysis, particularly in genomic regions with low or low quality sequence coverage. Errors in variant calling may lead to missing true variants or to calling many false positives. The false discovery rate can be reduced by optimizing variant calling thresholds such as quality of base pair identification, mapping, and alignment. However, such optimization strategies are often associated with the loss of true variants. We present and apply a pipeline for variant identification and verification using aligned sequences of related individuals. It is comprised of three modules: (1) an identification pipeline for de novo variants where data of parents and siblings are aligned in order to rule out false positive calls in children, false negative calls in parents, and indel artifacts; (2) a homozygosity mapping and IBD analysis module; and (3) a variant read depth module that reveals variants that may have been missed due to sequence coverage and quality issues. We applied module (1) to a large trio-based gene discovery project and reduced the number of variant calling errors by 74%, thereby significantly streamlining the experimental validation protocol for potential de novo variants. We also applied the pipeline to the discovery of the gene responsible for mega corpus callosum and microcephaly with developmental delay, and epilepsy in a brother and sister whose unaffected parents were first cousins. Our error correction pipeline significantly improved homozygosity mapping and IBD analysis and facilitated the rapid identification of the causal allele in this family.

288

A New Framework for Large-Scale Genomic Variant Discovery and Validation using Pooled Sequencing Data. *G. del Angel, M. Carneiro, E. Banks, R. Poplin, C. Hartl, M. dePristo, del Angel.* Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge,MA, MA.

As the costs of actual genome sequencing decrease, the bulk of a sequencing project cost is associated with sample preparation and library construction. Pooled sequencing has been an attractive alternative to limit this cost, but its drawbacks are well known, most importantly its lack of specificity for discriminating low-frequency variants from sequencing errors. We propose a new pooled sequencing method that is general enough that can be used for discovering candidate variant sites in pooled experiments, as well as for validating putative genomic variant genomic sites. Our method consists of two innovations: first, we include a barcoded reference sample with known genotypes in the pooled dilution, allowing us to build a statistical model of a genomic site's error properties. We also propose a new mathematical framework for estimating genotype likelihoods and allele frequencies in pooled data that is general enough to be used in any sequencing framework where allele frequencies must be computed from higher ploidy genomic data. To prove the effectiveness of this framework, we designed a first of its kind large-scale validation experiment. We produced a hybrid capture array for 55,000 genomic target SNPs and indels and validated variant pools made from 1100 samples using deep targeted resequencing at these sites. The accuracy of our method is consistent with the case where non-pooled barcoded samples are used. We additionally show that this pooled analytical framework can be used to call genomic mitochondrial variants. We can reliably detect heteroplasmic variants with 100% concordance in a maternal transmission test, and we can also find de-novo mitochondrial variants.

289

Discovery of genomic variants from RNA-sequencing data. *R. Piskol, G. Ramaswami, J.B. Li.* Genetics, Stanford University, Stanford, CA.

Background: The detection of novel genomic variation can yield important insights into human diseases including cancer. Recent efforts in form of the HapMap and 1000 Genomes Project have led to an immense increase in the knowledge of common human genetic variation. Linking this variation to human diseases has facilitated estimation of disease risk and discovery of causal variants as shown in the cases of Alzheimer's disease, Crohn's disease, and type 1 and 2 diabetes. To date, most disease related variants have been identified from whole genome sequencing (WGS) or exome sequencing experiments, despite much richer sources of RNA-seq data. Employing RNA-seq data to achieve the same feat, however, has remained more challenging due to the complexity of the transcriptome and the resulting technical difficulties during its computational analysis.

Methods: Here we present a simple yet highly accurate method to identify single nucleotide variants (SNVs) from RNA-sequencing experiments. Our approach consists of (1) a modified RNA-seq read mapping procedure that allows alignment of reads to the reference in a splice-aware manner, (2) variant calling using the Genome Analysis Toolkit and (3) vigorous filtering of false positive calls. The steps of our computational pipeline are inspired by common practice for mapping, variant calling and variant filtering in whole genome sequencing experiments. They were modified to account for the specific characteristics of RNA-sequencing experiments, including errors introduced during RNA-seq library preparation, sequencing, and difficulties in read mapping due to highly similar genomic regions.

Results: As a proof of concept, we have applied our method to matched RNA and genome sequencing data from a lymphoblastoid cell line and peripheral blood mononuclear cells. For the two data sets we are able to detect 181,951 and 309,144 RNA-seq variants, respectively. We find that our method achieves a precision of 97.5% and 95.5% for the two data sets, when comparing the identified RNA-seq variants to genomic variants and variants deposited in dbSNP. Furthermore, RNA-seq allows us to identify ~40% of all known WGS-variants in exonic regions, despite the limitation of our method to expressed genes only.

290

zCall: A Rare Variant Caller for Array-based Genotyping. *J.I. Goldstein[1,2], A. Crenshaw[3], J. Carey[3], G. Grant[3], J. Maguire[2], M. Fromer[1,2,5], C. O'Dushlaine[4], J.L. Moran[4], K. Chambert[4], C. Stevens[2], P. Sklar[5], C. Hultman[6], S. Purcell[1,4,5], S. McCarroll[2,4,7], P.F. Sullivan[8], M.J. Daly[1,2], B.M. Neale[1,2,4], Swedish Schizophrenia Consortium, ARRA Autism Sequencing Consortium.* 1) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Genetic Analysis Platform, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 5) Division of Psychiatric Genomics, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY; 6) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 7) Department of Genetics, Harvard Medical School, Boston, MA; 8) Department of Genetics, University of North Carolina, Chapel Hill School of Medicine, Chapel Hill, NC.

With recent developments in sequencing, the human genetics community has aggregated an extensive list of coding region variants for the purposes of designing array-based genotyping platforms that capture these variants (e.g., the Illumina Infinium HumanExome BeadChip and the Affymetrix Axiom Exome Array Plate). The overwhelming majority of variation assayed on the exome chip has a minor allele frequency < 1%. For rare variation, however, only the common allele homozygote cluster is well populated limiting the efficacy of existing genotype calling algorithms. To address this need, we developed a new method to accurately call rare variation for array-based genotyping called zCall. This caller is implemented as a post-processing step after a default calling algorithm has been applied. The algorithm uses the intensity profile of the common allele homozygote cluster to define the location of the other two genotype clusters. Using data from a previous Swedish schizophrenia association study and from the ARRA Autism Sequencing Consortium, we demonstrate improved detection of rare alleles when applying zCall to 1,316 samples that have both Illumina Infinium HumanExome BeadChip and exome sequencing data available. For example, we used zCall to increase the sensitivity of GenCall, Illumina's default genotype caller, by approximately 7% such that 99.12% of 10,075 singleton sites that were present on both the array and in the sequence data were called correctly.

291

Copy number detection and variant classification in the DDD project. *T.W. Fitzgerald[1], K.I. Morley[1], M. van Kogelenberg[1], E. Bragin[1], P. Vijayarangakannan[1], A. Tivey[1], S. Clayton[1], S. Gribble[1], C. Wright[1], D. FitzPatrick[2], H. Firth[1], J. Barrett[1], N. Carter[1], M. Hurles[1].* 1) Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Western General Hospital, Edinburgh.

Although the application of DNA microarrays for the detection of copy number variation (CNVs) is well established, individual change point detection algorithms often display variable performances. The definition of an optimal set of parameters for achieving a certain level of performance is rarely straightforward, especially where data qualities vary. We have developed a combined change point detection package, CNsolidate, which makes use of multiple weighted algorithms. Using this approach, we are able to rank detections based on differential weighting functions between component algorithms, which substantially improves the type1 and type2 error rates relative to other approaches. The Deciphering Developmental Disorders (DDD) project makes use of a number of advanced variant prediction approaches, including the accurate assignment of population based frequency estimates. CNV status assignment based on frequency information is important for studies involving rare disease. In addition to generating microarray data on ~1,000 control individuals, we use a consensus approach to generating common CNV reference sets, calculating frequency estimates across studies displaying differential sensitivities across the genomic range. These consensus data on normal population variation are available as a dedicated track in DECIPHER. Here we assess the benefits of using high performance detection and filtering methods on the number of pathogenic variant classifications for patients in the DDD project. More generally, we assess overall detection rates and describe a number of technical challenges when performing large-scale CNV data analysis.

292

Removal of mapping biases in sequence-based functional data improves regulatory element identification at heterozygous variants. *M. Buchkovich, K.L. Mohlke, T.S. Furey.* Department of Genetics, University of North Carolina, Chapel Hill, NC.

Sequence-based genome-wide data that detect epigenetic modifications, transcription factor binding and nucleosome-depletion identify regions of transcriptional regulation. The accuracy of these data at polymorphic sites is reduced by read alignments that are missing due to mismatch penalties incurred by non-reference allele-containing sequences. These missing reads may limit identification of regulatory elements at sites with the non-reference allele present. We quantified the ability of an allele-aware aligner, GSNAP, to recover missing reads at heterozygous sites compared to a commonly used aligner, BWA. First, we mapped DNase-seq sequences from GM12878, a human lymphoblastoid cell line, to a customized genome consisting of the hg19 sequence with 1,061,092 sites changed to match a homozygous non-reference allele in GM12878. The customized genome ensured that our comparison focused only on alignment differences caused by heterozygous sites. During mapping, BWA only used the reference allele, but GSNAP considered both alleles at user-provided GM12878 heterozygous sites. GSNAP mapped 2.8 million reads to heterozygous sites, which represents a 28.5% increase over the 2.2 million reads mapped to these sites by BWA. Next, we identified 100,000 regions, or peaks, with the strongest signal in each dataset. Heterozygous variants were present in 2.3% more peaks identified using GSNAP than using BWA. In addition, the signal strength of peaks containing heterozygous variants found by both GSNAP and BWA was significantly higher in GSNAP peaks (Mann-Whitney P= .0057). Finally, we identified heterozygous sites throughout the genome at which one allele was significantly overrepresented based on a binomial test. The reference allele was enriched at a majority (95%) of the sites found solely by BWA alignments, but was only enriched at 55% of sites found solely by GSNAP alignments, showing that GSNAP is better able to identify sites enriched for the non-reference allele. These data show that at heterozygous sites, GSNAP aligns more reads, increases the number and signal strength of peaks, and identifies a more even distribution of reference and non-reference allele enrichment. These improvements increase the utility of sequencing based data in identifying regulatory elements at heterozygous sites.

293

SNP Discovery in Diverse Human Populations by Rapid, Very-low-cost Next-generation Sequencing of Reduced Representation Libraries. *T.F. Cooke, M.C. Yee, M. Muzzio, R. Bell, O.E. Cornejo, C.D. Bustamante, E.E. Kenny.* Department of Genetics, School of Medicine, Stanford University, California 94305, USA.

As we sequence large numbers of ethnically diverse human genomes and exomes, a picture is emerging in which the majority of genetic variants are both rare and population-private. Next generation sequencing provides an unbiased means of discovering these low-frequency alleles, enabling the study of fine-scale population structure and recent demographic events, but its cost is still relatively high for population studies. We developed a genotyping-by-sequencing (GBS) method in which single nucleotide variants (SNVs) are discovered de novo in short segments of genomic DNA adjacent to restriction sites (a reduced representation library) by resequencing multiplexed samples on a single Illumina sequencing lane. In humans, our target region comprises roughly 45 Mb (or 1.5% of the genome), yielding approximately 45,000 SNVs per sample. Up to 20 individuals can be sequenced in batch for a lower cost per sample compared to commercial genotype arrays, and without ascertainment bias towards European populations. We describe several innovations in GBS methods: (i) Previous GBS libraries suffer from inefficient cluster recognition on Illumina flow cells because all inserts begin with identical restriction site overhangs. We solve this problem by using enzymes that cut 10–14 bp away from their recognition site, leaving random sequence at the insert ends. (ii) We ensure that a large proportion of sequenced regions are common to all samples by size-selecting our libraries on a Caliber Labchip, which enriches for a subset of the genomic DNA fragments and increases sequencing depth. In our first trial library of 6 samples, nearly 50% of reads mapped to regions that were covered to at least 8x across all samples. (iii) We are developing methods to reduce false homozygous calls at heterozygous sites due to restriction site polymorphism or PCR-inflated read counts by measuring the true number of genomic fragments represented by our data at each site. We designed sequencing adapters that harbor a random 8-mer barcode that will enable us to distinguish between PCR duplicate reads and non-PCR duplicates that are identical by chance. We are optimizing GBS on DNA samples from the 1000 Genomes Project, a worldwide catalog of human genetic variation, for which whole-genome sequencing data is publicly available. Given the low cost and rapid preparation of GBS libraries, we propose GBS as a key method for unbiased SNP discovery in large samples from diverse human populations.

294

HIBAG — HLA Genotype Imputation with Attribute Bagging. *X. Zheng[1], J. Shen[2], C. Cox[2], J. Wakefield[1], M. Ehm[2], M. Nelson[2], B. Weir[1]*. 1) Department of Biostatistics, University of Washington at Seattle, Seattle, WA; 2) Quantitative Sciences, GlaxoSmithKline, Research Triangle Park, NC 27709.

The major histocompatibility complex (MHC) is frequently found to be associated with disease and adverse drug reactions. While these associations can be identified using whole-genome SNP typing, associations found in this region, typically will require additional typing of classical human leukocyte antigen (HLA) alleles, which is often cost prohibitive for large populations. Here, we propose HIBAG, an attribute bagging approach, which takes advantage of the extended haplotype structure within the MHC to predict HLA alleles using dense SNP genotypes as available from genome-wide SNP panels. This method is implemented in the freely-available HIBAG R package that includes pre-fit classifiers based on our study data (n = 2,668 subjects of European ancestry with four-digit HLA alleles). Compared to current methods, like HLA*IMP and BEAGLE, HIBAG provides a readily available imputation approach without having to upload genotype information to a website or access large training datasets. We assessed and demonstrated our pre-fit classifiers using HLA data from the British 1958 birth cohort study of Wellcome Trust Case Control Consortium ($n \approx$ 1,000 subjects). Prediction accuracies for *HLA-A*, *B*, *C*, *DQB1* and *DRB1* range from 94.0% to 98.2% with call rates over 94% using a set of SNP markers common to the Illumina 1M Duo, OmniQuad, OmniExpress, 660K and 550K platforms. The HIBAG method, freely available software and accompanying pre-fit classifiers will enable any researcher to predict HLA types with high accuracy using SNP markers available on the most commonly used GWAS platforms.

295

Systematic identification of causal mutations in Mendelian disorders using exome sequence data. *M. Lek[1], N.F. Clarke[3,4], L.B. Waddell[3,4], B. Thomas[1], M.A. DePristo[2], M.J. Daly[1,2], K.N. North[3,4], D.G. MacArthur[1,2]*. 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) Institute for Neuroscience and Muscle Research, The Children's Hospital at Westmead, Sydney 2145, NSW, Australia; 4) Discipline of Paediatrics and Child Health, Faculty of Medicine, University of Sydney, Sydney, NSW 2006, Australia.

Exome sequencing has proven to be a powerful and cost-effective approach for the identification of causal mutations in many patients suffering from rare, severe Mendelian diseases. However, exome analysis unambiguously identifies a causal mutation in only 30–50% of sequenced families, indicating much work remains to be done to increase the yield of causal variants from sequencing-based approaches. Causal mutations can be missed by current exome sequencing approaches for a variety of reasons. Variants such as small insertions and deletions, multi-nucleotide polymorphisms, and larger structural variants are often under-called, as are those present in non-coding or poorly-captured regions, and the functional impact of detected mutation may be mis-assigned due to incomplete or erroneous functional annotation of the human genome. In other cases projects will fail due to mis-specification of inheritance modes, or the effects of unusual biological processes such as parental imprinting. We describe the development of an integrated pipeline for the identification of causal variants from exome data that reduces the impact of many of these challenges. Firstly, we take advantage of an extended approach to variant-calling that more accurately detects complex variant classes from large-scale exome data. Secondly, we have developed an improved approach to functional annotation spanning both coding and non-coding variation, including haplotype-based calling to appropriately aggregate the functional impact of multiple variants present on the same local haplotypic context. Finally, we present an online browser for the intuitive analysis of family-based exome data, permitting researchers and clinicians to rapidly explore the effects of altering inheritance modes and function/quality filters on the identification of potential causal mutations. We have applied this integrated approach to exome data from 80 individuals in 21 families affected by a range of neuromuscular diseases. We describe the detection of novel sequence variants with strong evidence for causality in these patients, and provide case studies indicating the value of improved variant-calling and functional annotation for the identification of disease-associated mutations.

296

Exome sequencing of a large cohort of patients with congenital digestive system disorders. *M. Yourshaw[1], S.F. Nelson[1,3], M.G. Martín[2]*. 1) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90025, USA; 2) Department of Pediatrics, Division of Gastroenterology and Nutrition, Mattel Children's Hospital and the David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA; 3) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at the University of California at Los Angeles, Los Angeles, CA 90095, USA.

Congenital digestive system disorders are uncommon, poorly understood, devastating conditions often caused by malabsorption of multiple nutrients. Affected patients have an adverse clinical course that includes life-long or prolonged total intravenous nutrition, and/or intestinal transplantation. Current therapeutic modalities are primitive, and are associated with significant morbidity, mortality, and daunting medical care costs. The diagnostic odyssey for these patients typically involves an extensive clinical search for additional phenotypic information, including imaging, biopsies, and functional assessments, sometimes followed by genetic testing to confirm a diagnosis. Genetic testing of specific candidate genes is often uninformative. We hypothesize that any of a large number of genes may be defective and will affect several developmental pathways. Whole exome sequencing of protein coding DNA can identify novel mutations in known genes as well as suggest novel candidates for further study. We extracted DNA from approximately 50 cases of congenital diarrheal disorders, including generalized malabsorptive diarrhea or secretory diarrhea. We enriched DNA samples for protein-coding regions, sequenced them on the Illumina massively parallel sequencing platform, aligned reads to the human reference genome, and assigned genotypes. Using custom software and databases, we identified variants that were likely to be damaging and had allele frequencies in the general population consistent with the known very rare disease incidence observed clinically. We performed dideoxy sequencing to confirm sequencing results and familial segregation of candidate disease-causing alleles. In approximately half of the cases we identified novel mutations in known genes that explained the patients' conditions, including *ADAM17* (inflammatory skin and malabsobtive diarrhea), *ENPP1* (rickets), *EPCAM* (Tufting enteropathy), *MYO5B* (diarrhea with microvillus atrophy), *NEUROG3* (generalized malabsorptive diarrhea), *PCSK1* (PC1/3 deficiency), SI (sucrase isomaltase deficiency), and *SLC5A1* (glucose/galactose malabsorption). Notably, in some instances knowledge of the genetic mutation led to actionable improvements in patient care. Many of the remainder are novel variants in newly implicated candidate genes that are the subject of ongoing functional studies. Conclusion. Exome sequencing can be an efficient tool for the diagnosis of difficult cases of congenital digestive system disorders.

297

Novel defect in kinetochore assembly causes short stature and microcephaly of postnatal onset. *C.Y. Hung[1], J.E. Dallman[2], O. Rittinger[3], J.W. Bauer[4], M. Tekin[1], O.A. Bodamer[1]*. 1) University of Miami, Department of Human Genetics, Hussman Institute of Human Genomics, Miami, FL; 2) University of Miami, Department of Biology, Miami, FL; 3) Medical University of Salzburg, Department of Pediatrics, Salzburg, AUT; 4) Medical University of Salzburg, Department of Dermatology, Salzburg, AUT.

Introduction: Microcephaly is a genetically heterogeneous disorder due to neurodegeneration or due to defects in either neurogenesis or mitosis of neurons. Although pathogenic mutations in several genes were identified to cause microcephaly, a large number of patients remain undiagnosed. We report two siblings with microcephaly, short stature, facial dysmorphism and intellectual impairment born to consanguineous Turkish parents with a pathogenic mutation in a gene of kinetochore assembly. Methods: Autozygosity mapping identified several candidate genes within the largest shared regions of homozygosity. Sanger sequencing of one of the genes detected a homozygous splice-site mutation in both affected siblings and heterozygosity in a non-affected brother and both parents. The mutation was absent in more than 600 Turkish control alleles as well as in the SNP database (dbSNP). Cloning and sequencing of the RT-PCR products showed three alternative isoforms in both patients. Two of the three isoforms cause out-of-frame deletions resulting in a predicted truncated protein affecting the C-terminal histone homology domain. The terminal histone homology domain is conserved throughout species. The third isoform results in an in-frame deletion without predicted changes to the histone domain. The absence of the wild-type isoform in both patients was confirmed using the DNA 7500 BioAnalyzer chip. A zebrafish model was established using in-vivo injections of morpholinos into zebrafish ova. Morpholinos were designed to modify splicing towards the 3' end prior to the conserved histone homology region in order to resemble the human molecular phenotype. Dose response experiments in the zebrafish showed a decrease in head and body size proportional to the morpholino dosage. The use of morpholinos targeting further 5' sequences caused early lethality dependent on the proximity of the morpholino targeting site to the 5' end of the gene. Conclusion: We report to our knowledge the first defect in a gene of the kinetochore assembly resulting in microcephaly and short stature. The presence of an in-frame isoform may help to explain the absence of early lethality and the relatively mild phenotype in both siblings. Additional studies are currently under way including functional studies and screening of patients diagnosed with microcephaly and short stature of unknown cause.

298

Mutations in PIGO, a member of the GPI anchor synthesis pathway, cause hyperphosphatasia with mental retardation syndrome. *P.M. Krawitz[1], Y. Murakami[2], J. Hecht[1], U. Krüger[1], S.E. Holder[3], G.R. Mortier[4], B. Chiaie[4], E. Baere[4], M.D. Thompson[5], T. Roscioli[6], S. Kielbasa[7], T. Kinoshita[2], S. Mundlos[1], P.N. Robinson[1], D. Horn[1]*. 1) Medical Genetics, Charité, Berlin, Berlin, Germany; 2) Department of Immunoregulation, Osaka University, Japan; 3) North West Thames Regional Genetics Service, Harrow, UK; 4) Department of Medical Genetics, Antwerp, Belgium; 5) Dpartment of Laboratory Medicine and Pathobiology, University of Toronto, Canada; 6) School of Women's and Children's Health, Sidney, Australia; 7) Center for Human and Clinical Genetics, Leiden, The Netherlands.

We have recently identified mutations in PIGV in individuals with hyperphosphatasia mental retardation (HPMR) syndrome, an autosomal recessive form of mental retardation with facial dysmorphism, seizures, brachytelephalangy, and persistent elevated serum alkaline phosphatase (hyperphosphatasia). However, not all patients with HPMR syndrome harbor mutations in PIGV. The purpose of the current study was to investigate the molecular etiology of HPMR syndrome in PIGV-negative patients and to establish a next- generation sequencing based screening approach for GPI pathway diseases. We herein present mutations found in PIGO that impair GPI-anchor synthesis. More than a hundred cell surface proteins are attached to the plasma membrane by covalent attachment to a glycophosphatidylinositol (GPI) anchor that is assembled in the endoplasmic reticulum (ER) and added to the C-terminus of the proteins. Biosynthesis of GPI anchors involves more than 30 different genes. Genetic defects in various components of the GPI pathway have been identified in a number of phenotypically diverse diseases. Somatic mutations in hematopoietic stem cells in PIGA cause paroxysmal nocturnal hemoglobinuria, which manifests as bone marrow failure, hemolytic anemia, smooth muscle dystonias, and thrombosis. Germline promoter mutations in PIGM resulting in severe GPI-AP deficiency were found in patients with portal and hepatic vein thrombosis and intractable absence seizures. An autosomal recessive syndrome caused by mutations in PIGN and characterized by dysmorphic features and multiple congenital anomalies, severe neurological impairment, chorea, and seizures leading to early death was described. CHIME Syndrome characterized by colobomas, heart defects, ichthyosiform dermatosis and mental retardation and ear anomalies is caused by mutations in PIGL. However in contrast to these mutations reported in PIGA, PIGM, PIGN and PIGL, our current results show that mutations PIGV and PIGO, which are involved in later stages of GPI anchor synthesis and are associated with accumulation of incomplete GPI bearing mannose.do result in hyperphosphatasia. Therefore, our results suggest a molecular mechanism of hyperphosphatasia in HPMR, and suggest that mutations in other genes encoding enzymes involved in later stages of GPI anchor synthesis remain to be found in further cases of HPMR.

299

The 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *S. Jacquemont[1], F. Zufferey[1], E.H. Sherr[2], N.D. Beckmann[1], E. Hanson[3], A. Maillard[1], L. Hippolyte[1], A. Mace[4,5], C. Ferrari[6], Z. Kutalik[4,5], J. Andrieux[7], R. Bernier[8], S. Bouquillon[7], B. Delobel[9], W. Andrew-Faucett[10], R.P. Goin-Kochel[11], L. Harewood[12], S. Lebon[13], D.H. Ledbetter[10], C. Lese-Martin[14], K. Mannick[12], D. Martinet[1], M.B. Ramocki[15], S.J. Spence[16], K. Steinmann[17], J. Tjernagel[18], J.E. Spiro[18], A. Reymond[12], W. Chung[19], J.S. Beckmann[1,4] on behalf of the Simons VIP Consortium, and the 16p11.2 European Consortium.* 1) Service de Génétique Médicale, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 2) Department of Neurology, University of California, San Francisco; 3) Department of Psychiatry, Children's Hospital Boston, Harvard Medical School, Boston; 4) Department of Medical Genetics, University of Lausanne, Lausanne; 5) Swiss Institute of Bioinformatics, University of Lausanne, Lausanne; 6) Department of Psychiatry, Centre Hospitalier Universitaire Vaudois, Lausanne; 7) Institut de Génétique Médicale. Hopital Jeanne de Flandre, CHRU de Lille, Lille; 8) Department of Psychology, Center on Human Development and Disability, University of Washington, Seattle; 9) Centre de Génétique Chromosomique. Hôpital Saint-Vincent de Paul, GHICL, Lille; 10) Genomic Medicine Institute, Geisinger Clinic, Danville; 11) Department of Pediatrics, Psychology Section, Baylor College of Medicine, Houston; 12) Center for Integrative Genomics, University of Lausanne, Lausanne; 13) Department of Pediatrics, Centre Hospitalier Universitaire Vaudois, Lausanne; 14) Department of Human Genetics, Emory University, Atlanta; 15) Department of Pediatrics, Section of Pediatric Neurology and Developmental Neuroscience, Baylor College of Medicine, Houston; 16) Department of Neurology, Children's Hospital Boston, Harvard Medical School, Boston; 17) Department of Neurology, Children's Research Institute & University of Washington, Seattle; 18) Simons Foundation, New York; 19) Departments of Pediatrics and Medicine, Columbia University, New York.

The recurrent ~600kb 16p11.2 BP4-BP5 deletion is among the most frequent known genetic etiologies of autism spectrum disorder (ASD) and related neurodevelopmental disorders. The diversity of published clinical features, together with the report of asymptomatic (but not fully evaluated) transmitting parents, highlighted the need to assess systematically the impact of this deletion. We collected clinical data on 285 deletion carriers and performed detailed evaluations on 72 carriers and 68 intrafamilial controls to define the medical, neuropsychological and behavioral phenotypes of carriers. When compared to intrafamilial controls, FSIQ is two standard deviations lower in carriers, and there is no difference between carrier probands referred for neurodevelopmental disorders and carriers identified through cascade family testing. Verbal IQ (mean 74) is lower than nonverbal IQ (mean 83); a majority of carriers require speech therapy. Over 80% of individuals exhibit psychiatric disorders including ASD, which is present in 15% of the pediatric carriers. Increase in HC during infancy is reminiscent of the HC and brain growth patterns observed in idiopathic ASD. Obesity, a major comorbidity present in 50% of the carriers by age seven, does not correlate with FSIQ or any other cognitive or behavioral trait. Seizures are present in 24% of carriers and also occur independently of the other symptoms. Malformations are an infrequent finding. Conclusion: The 16p11.2 deletion impacts in a quantitative and independent manner FSIQ, behavior and BMI, possibly through direct influences on neural circuitry. Although nonspecific, these features are clinically significant and reproducible. Lastly, this study demonstrates the necessity of studying large patients cohorts to characterize clinical consequences of rare variants involved in common diseases.

300

Ras/MAPK dysregulation caused by *MEK2* haploinsufficiency: A novel mechanism for a RASopathy phenotype. *M.J.M. Nowaczyk[1], B. Thompson[2], S. Zeesman[1], U. Moog[3], P.A. Sanchez-Lara[4], R. Falk[5], P. Magoulas[6], L. Brueton[7], S.M. Ahmudavalli[8], J.H. Fong[8], D. Batista[8], K. Rauen[2].* 1) Dept Pathology & Pediatrics, McMaster Univ, Hamilton, ON, Canada; 2) UCSF, San Francisco, Californa, USA; 3) Heidelberg Univeristy, Heidelberg, Germany; 4) Children's Hospital Los Angeles and USC Los Angeles, California, USA; 5) Cedars-Sinai Hospital, Los Angeles, California, USA; 6) Baylor College of Medicine, Houston, Texas, USA; 7) Birmingham Women's Hospital, Birmingham, United Kingdom; 8) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, USA.

Background The RASopathies are a class of human genetic syndromes caused by germline mutations in genes encoding components of the Ras/mitogen-activated protein kinase (MAPK) pathway which plays an important role in cell proliferation and embryonic development. Cardio-facio-cutaneous (CFC) syndrome, a RASopathy, is characterized by distinctive craniofacial features, congenital heart defects, and skin and hair abnormalities. CFC is caused by activating mutations of *BRAF*, *MEK1*, *MEK2*, and possibly *KRAS*. **Methods** We report 6 patients with de novo *MEK2*-containing deletions of 19p13.3 (0.06–1.8Mb). Functional assays were performed using primary fibroblast cell lines derived from patient 1 and a healthy age/sex matched control. At the desired confluency, cells were placed in complete or serum starvation media overnight. Serum starved cells were then treated with epidermal growth factor (EGF), a factor known to stimulate the MAPK pathway, for 0–120min. Cells were harvested and protein lysates were subjected to western blot analysis with antibodies specific for various MAPK pathway components including phospho-MEK and phospho-ERK. **Results** The patients present with a recognizable pattern of dysmorphic features that are similar to, yet distinct from, CFC syndrome. Clinical features include, but are not limited to, distinct craniofacial features, developmental delay, congenital heart defects, failure to thrive, obstructive sleep apnea, and skin anomalies. Studies in primary fibroblasts show that when the MAPK pathway is stimulated by EGF, phospho-MEK is less abundant in cells carrying the *MEK2* deletion compared to the control. Differences in the levels of phospho-ERK and Sprouty 1 were also observed when comparing the deletion and control cell lines post stimulation. This is the first report of a *MEK2* deletion altering regulation of the MAPK pathway and associated with a reproducible phenotype. **Conclusions** Individuals with deletions of *MEK2* share phenotypic features with known RASopathies; they represent a novel recognizable multiple malformation syndrome due to aberrant signaling of the MAPK pathway. Our studies show that dysregulation of the MAPK pathway can be caused by *MEK2* haploinsufficiency representing a novel mechanism that can cause a RASopathy.

301

Analysis of ESP5400 exomes for results of clinical utility in genes for conditions tested as part of newborn screening programs and age-related macular degeneration. *H.K. Tabor[1,2], S.M. Jamal[2], J.H. Yu[2], A.S. Gordon[3], W.S. Post[4], A.D. Johnson[5], T.A. Graubert[6], D.A. Nickerson[3], P.L. Auer[7], M.J. Bamshad[2,3] on behalf of the NHLBI Personal Genomics Project Team and the NHLBI Exome Sequencing Project.* 1) Treuman Katz Ctr Pediatric Bioethics, Seattle Children's Hospital, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Framingham Heart Study, National Heart Lung and Blood Institute, Framingham, MA; 6) Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO; 7) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.

Exome sequencing and whole genome sequencing (ES/WGS) are rapidly being deployed for use in clinical settings despite the paucity of data about the number and types of results with potential clinical utility that may be available for return to an individual. We analyzed deidentified ES data from 4826 participants from the NHLBI Exome Sequencing Project (ESP) to characterize the types and frequency of variants with potential clinical utility found in genes (n=50) for conditions that are tested as part of newborn screening (NBS) and genes (n=11) associated with age related macular degeneration (ARMD). We identified putatively pathogenic variants by review of OMIM, HGMD, locus-specific databases, and literature review using conservative criteria that required compelling evidence of causality. We found that nearly 60% of African Americans (AA; n=1821) and 55% of European Americans (EA; n=3005) were carriers of at least one known pathogenic variant in the NBS gene set; 25% of AA and 22% of EA had at least two such variants; and 8% of AA and 6% of EA were heterozygous for three or more pathogenic variants in the NBS gene set. In AA, the frequency of pathogenic variants that cause sickle cell disease (12%) and congenital adrenal hyperplasia (7%) were substantially higher than in EA (1% and 0.5%, respectively). In contrast, in EA, the frequency of pathogenic variants that cause galactosemia (19%) and organic acidemias (15%) was higher than in AA (6% and 7%, respectively). Several individuals were homozygous for pathogenic variants suggesting these variants may not be pathogenic or may be associated with reduced penetrance. In the ARMD gene set, 98% of individuals had at least 1 of 22 known variants associated with disease risk, and more than half (60%) had 5 or more risk variants. These estimates are conservative because some known pathogenic variants were not included in the analysis (i.e., indels) and/or the target definition (i.e., poor coverage). Accordingly, estimates based on WGS will be even higher. Overall, these findings suggest that variants with potential clinical utility will be detected at substantial frequency in ES data for both Mendelian and complex traits, yet the interpretation of results is not completely resolved. Identifying such variants is of considerable latent benefit and highlights the need to empirically test how to integrate ES/WGS into clinical practice and develop innovative tools that facilitate doing so.

302

High Congenital Malformation Rates in a Chornobyl Ionizing Radiation Impacted Population Isolate in Ukraine. *W. Wertelecki, L. Yevtushok, N. Zymak-Zakutnia, S. Lachenko.* OMNI-Net Ukraine Programs, Mobile, AL.

In 2000, we established two population registries in Ukraine adhering to methods and standards prevalent in Europe; a registry of neonates born in Rivne province and a registry and active monitoring of their congenital malformations (CM). In 2002, we noted elevated rates of neural tube defects (NTD) confirmed by an analysis of 2000–2006 data which also showed high rates of microcephaly (MIC) and microphthalmia (mOPH), anomalies that can be caused by ionizing radiation (IR). Analyses of 145,437 pregnancy outcomes from 2000 to 2009 also show persisting elevated population rates of NTD, MIC and mOPH, particularly in Rivne-Polissia (RP), a region of wetlands impacted by Chornobyl IR. Also known as Prypiat Marshlands, RP is inhabited by Polishchuks, a known ethnic sub-group in Ukraine which has characteristics of a population isolate. The Rivne rates of NTD, MIC and mOPH are among the highest in Europe and rates in RP are statistically significantly higher than elsewhere in Rivne. In contrast, rates of sentinel anomalies such as Down syndrome or facial clefts are similar in RP and non-RP. Also of note are statistically significant NTD-omphalocele and NTD-twinning associations. The birth of seven sets of conjoined twins and eleven infants with teratomas, both are rare anomalies, are likewise provocative. We analyzed 6026 recordings of whole body counts of incorporated IR obtained from pregnant women, 1157 of whom reside in four most northern counties of RP. The results show that 48% of those residing in the four most northern counties of RP had levels above the official norm for children (3700Bq) in contrast to none among those residing in the capital city. We also screened pregnant women for alcohol consumption, another cause of MIC. Among 852 and 566 pregnant women who reside in RP or the capital city, 1.53% and 6.36% respectively consumed alcohol. Analysis of the frequency of isonomy (an indirect index of consanguinity) showed that the highest rates were in northern RP counties where 8 to 18% of infants have one of five prevalent family surname in their county. We conclude that Rivne population monitoring for CM should continue and be expanded by prospective investigations with an emphasis on RP region with the aim to clarify long term health and reproductive impacts of Chornobyl IR in the context of other risk factors.

303

Somatic mosaicism is responsible for congenital melanocytic naevus syndrome, and underpins the associated risk of melanoma. *V.A. Kinsler[1,2], A.C. Thomas[2], N.W. Bulstrode[3], S. Abu-Amero[2], K. McKenzie[4], E. Chanudet[2], P. Stanier[5], E. Healy[6], N.J. Sebire[7], G.E. Moore[2].* 1) Paediatric Dermatology, Great Ormond Street Hospital for Children, London, United Kingdom; 2) Clinical and Molecular Genetics Unit, UCL Institute of Child Health, London, United Kingdom; 3) Paediatric Plastic Surgery, Great Ormond Street Hospital for Children, London, United Kingdom; 4) Paediatric Pathology, Royal Hospital for Sick Children, Edinburgh, United Kingdom; 5) Neural Development Unit, UCL Institute of Child Health, London, United Kingdom; 6) Dermatopharmacology, University of Southampton, Southampton, United Kingdom; 7) Paediatric Pathology, Great Ormond Street Hospital for Children, London, United Kingdom.

Congenital melanocytic naevus syndrome is the variable association of large or multiple congenital melanocytic naevi (CMN), neurological abnormalities and characteristic facial features. There is an associated increased risk of primary malignant melanoma during childhood, not restricted to the skin. The genetic basis has previously been unknown, but epidemiological studies suggest a somatic event, with a germline predisposition in some families. DNA was extracted from 36 samples of affected and unaffected tissue from 14 patients with accurately phenotyped large or multiple CMN, with or without neurological abnormalities. Blood DNA was obtained from six of these patients and six other affected patients. Tissue samples included cutaneous, neurological and malignant melanoma samples. Site-directed mutagenesis was used to generate a restriction enzyme site unique to the normal sequence of a candidate base pair, and repeated cycles of nested PCR and enzymatic digestion was used to selectively amplify putative mutant alleles. Sanger sequencing was performed after each cycle. Somatic mosaicism was found in 9/14 patients after only one cycle of digestion and two rounds of PCR, as demonstrated by the presence of the same abnormal allele in 26/28 affected but non-malignant tissue samples, and the absence in 6/6 unaffected tissue samples. The same missense mutation was seen in seven patients, with a different base pair change in the same codon found in a further two patients. No mutation was found at this codon in five patients' affected or unaffected tissue, and none in any of the blood samples. The commonest mutation was found in 2/2 melanoma samples without enzymatic digestion, implying a clonal expansion of an affected cell within mosaic tissue. In one case DNA from the affected tissue before and after the onset of malignancy (samples 5 years apart) showed loss of heterozygosity for the missense mutation at the onset of malignant change. This somatic mutation, which arose presumably in a single cell of the neural crest, could explain the cutaneous, neurological and facial features of affected children, and the predisposition to malignant melanoma. Mosaicism at other loci is likely to be responsible for the five cases who were negative for this mutation.

304

The type 2 diabetes (T2D) risk allele of rs11603334 increases *ARAP1* promoter activity and is associated with increased *ARAP1* mRNA in pancreatic islets. *J.R. Kulzer[1], M.L. Stitzel[2], M.A. Morken[2], F.S. Collins[2], K.L. Mohlke[1].* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Genome-wide association studies have identified many loci associated with T2D and pancreatic islet dysfunction, but for most, the functional variants and target genes have not been determined. We are investigating molecular and biological mechanism(s) underlying association of the *ARAP1* locus with T2D and fasting proinsulin. Index SNPs rs11603334 and rs1552224 are in perfect linkage disequilibrium (LD) ($r^2 = 1.0$) and located 113 bp apart within the 5'UTR of the short isoform of *ARAP1*. All SNPs in high LD ($r^2 > 0.8$) with the index SNPs are non-coding. A third SNP, rs7109575 ($r^2 = .86$ with the index SNPs), is located at the transcription start site of the long isoform of *ARAP1*. We hypothesized that one of these three SNPs influences the transcriptional regulation of *ARAP1*. We measured allele-specific mRNA levels from 87 human islet samples and determined that the index SNP T2D risk alleles are associated with increased *ARAP1* mRNA ($P \leq .01$ for each of two assays), suggesting that one or more risk variants increase *ARAP1* transcriptional activity or message stability. We then evaluated the allele-specific effects of rs11603334, rs1552224, and rs7109575 on transcriptional activity at the *ARAP1* promoters by performing dual luciferase reporter assays in the INS-1-derived rat beta cell line 832/13. The two-SNP haplotype containing the T2D risk alleles of rs11603334 (G) and rs1552224 (T) reproducibly increased promoter activity two-fold compared to the non-risk haplotype ($P \leq .001$). When the effects of rs11603334 and rs1552224 were separated by site-directed mutagenesis, the G allele of rs11603334 exhibited two-fold increased transcriptional activity ($P < .001$), while rs1552224 showed no effect. The DNA region surrounding rs7109575 demonstrated strong promoter activity, but that activity did not differ between SNP alleles. Taken together, these results suggest that the rs11603334 T2D risk allele increases *ARAP1* mRNA expression by upregulating transcriptional activity at the promoter of the short isoform. Ongoing studies focus on identifying transcription factors differentially bound to rs11603334 and examine the consequences of increased ARAP1 expression on insulin processing and secretion. Investigating the molecular mechanisms underlying T2D-associated loci is an important step toward identifying genes that contribute to T2D susceptibility.

305

***NOS1AP* is the major genetic electrocardiographic QT-interval regulator.** *A. Kapoor[1], R.B. Sekar[2], V. Pihur[1], M.K. Halushka[3], G.F. Tomaselli[2], A. Chakravarti[1,2].* 1) McKusick-Nathans Institute of Genetic Medicine; 2) Department of Medicine; 3) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD.

The major GWAS locus for electrocardiographic QT-interval variation, explaining ~1% of the population trait variation, includes the *NOS1AP* gene. To prove its role in cardiac electrical conductance/coupling and thus, regulation of QT-interval and arrhythmias we now demonstrate two key features: (1) overexpression of human *NOS1AP* in neonatal rat ventricular myocytes leads to shorter action potential durations and increased conduction velocity; and, (2) NOS1AP is localized at intercalated discs at cardiomyocyte junctions in adult human heart. Nevertheless, the identity of causal variant(s) and mechanism(s) by which they influence QT-interval variation remains unknown. At least 3 independent susceptibility variants map to a ~140 kb locus consisting of the 5' upstream region, *NOS1AP* exons 1 and 2, intron 1, and part of intron 2. Deep sequencing at the target locus, performed in 8 CEU HapMap and 46 KORA cohort samples, have generated a near-complete list of common variants. These variants, together with those from the CEU 1000 Genomes Project, were imputed in an independent set of 9,055 European ancestry samples from ARIC, to identify 210 variants that were genome-wide associated with QT-interval ($P<5\times10^{-8}$). Since none of the coding variants in *NOS1AP* explained the QT-interval association, we hypothesized that the causal variants must be non-coding, regulatory, residing within enhancers/silencers of *NOS1AP* expression. To this end, we selected 8 most-associated variants, added another 3 associated variants overlapping ENCODE annotated DNase I hypersensitive regions in human cardiac myocytes, and evaluated them using luciferase based enhancer/silencer assays in HL1 cells to assess transcriptional differences between variant alleles. At one of these variants, rs7539120, a small (1.5-fold) but consistent and significant ($P=6.7\times10^{-6}$, n=8) allelic difference was observed, and this differential allelic transcriptional activity was proven to be context (flanking sequence) dependent. Although, rs7539120 was predicted to be a binding site for MEF2, a major cardiac transcription factor, EMSA failed to demonstrate MEF2 binding. We find that a currently uncharacterized protein(s) present in HL1 nuclear extract binds to this site suggesting the possibility of novel major cardiac regulators. These results identify one of the many possible functional variants leading to QT-interval association with *NOS1AP* although direct effects on *NOS1AP* expression need to be demonstrated.

306

A regulatory polymorphism in Csk, a Lyp binding partner, associates with systemic lupus erythematosus and affects B cell signaling, maturation and activation. *N. Manjarrez-Orduño[1], E. Marasco[1], S.A. Chung[2], M.S. Katz[1], J.F. Kiridly[1], K.R. Simpfendorfer[1], J. Freudenberg[1], D.H. Ballard[1], E. Nashi[1], T.J. Hopkins[1], D.S. Cunninghame Graham[3], A.T. Lee[1], M.J.H. Coenen[4], B. Franke[4], D.S. Swinkels[5], R. Graham[6], R.P. Kimberly[7], P.M. Gaffney[8], T.J. Vyse[3], T.W. Behrens[6], L.A. Criswell[2], B. Diamond[1], P.K. Gregersen[1].* 1) The Feinstein Institute for Medical Research, Manhasset, NY; 2) Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco, CA. 94117; 3) Department of Medical and Molecular Genetics, Division of Genetics and Molecular Medicine, School of Medicine, King's College London, London, United Kingdom; 4) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Department of Laboratory Medicine, Laboratory of Genetic Endocrine and Metabolic Diseases, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 6) Genentech, South San Francisco, CA. 94080; 7) University of Alabama at Birmingham, AL 35294; 8) Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104.

Csk is an intracellular kinase that physically interacts with Lyp (*PTPN22*). A common Lyp variant, R620W, strongly associates with multiple autoimmune disorders, including systemic lupus erythematosus (SLE). The R620W variant of Lyp also alters molecular interaction with Csk; both proteins affect signaling thresholds in lymphocytes by modifying the activation state of src kinases. We identified an intronic polymorphism in *CSK* that influences SLE susceptibility (OR 1.32, $p=1.04 \times 10^{-9}$), increases *CSK* expression and augments inhibitory phosphorylation of Lyn, a target src kinase in B cells. In carriers of the *CSK* risk allele, B cell receptor-mediated activation of mature B cells and plasma IgM are increased. Moreover, the fraction of transitional B cells doubles in cord blood of carriers of the *CSK* risk allele due to an expansion of the late transitional cells, a stage targeted by multiple selection mechanisms. Our results suggest that the Lyp-Csk complex increases susceptibility to lupus by affecting B cells along multiple maturation and activation points. The authors thank the volunteers who participated in this study; the GaP registry (www.gapregistry.org), M. Keogh, M. DeFranco, C. Mason and C. Metz for recruiting subjects and collecting samples, H. Borrero for technical assistance, the Biostatistics Unit of the FIMR and M. Akerman for assistance. Work supported by NIH RC2AR059092; The Alliance for Lupus Research; the Kirkland Scholar Award; and NIH/NCRR 5 M01 RR-00079 (L.A.C.). The authors have no conflicting financial interests. Nijmegen Biomedical Study's principal investigators are L.A.L.M. Kiemeney, M. den Heijer, A.L.M. Verbeek, D.W. Swinkels and B. Franke.

307

ITGAM coding variant, rs1143679 (R77H) that is associated with systemic lupus erythematosus (SLE) susceptibility affects its own expression in monocytes and ligand binding activities in SLE patients. *AK. Maiti[1], X. Kim-Howard[1], P. Motghare[1], JM. Anaya[2], L. Loogers[3], SK. Nath[1].* 1) Gen Epidemiology Unit, A & CI, OMRF, Oklahoma City, OK; 2) Universidad del Rosario-Corporación para Investigaciones Biológicas, Bogota, Colombia; 3) Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, Virginia, USA.

ITGAM (CD11b) variant rs1143679 is one of the few SLE susceptibility loci which strongly associated with Systemic Lupus Erythematosus (SLE) and consistently replicated in multiple populations with European, African, Hispanic, and Asian ancestries. This nonsynonymous SNP causes amino acid changes from Arginine to Histidine (R77H) and also located at the 9th base from exon-intron boundary including a putative splicing enhancer/silencer sequences. We hypothesize that it could alter function at both the RNA and protein levels by modulating its own expression and altering its protein function. Expression analysis in SLE patient's monocytes showed that both mRNAs and proteins levels are significantly reduced in risk genotype (AA) patients in compare to non-risk (GG) genotype carrying patients. Furthermore, allelic expression suggests that the risk allele `A' carrying mRNA is 4 to 10 fold reduced in compare to G allele in monocytes of SLE heterozygous (GA) patients. This reduced level of risk allele specific mRNA level is not attributed to any risk allele splicing defect or mRNA degradation in the cell. Moreover, PCR products carrying the risk allele showed reduced efficiency of binding with a protein complex in vitro. This protein complex includes KU70/KU80 suggesting that this sequence could act as an enhancer during transcription mediated DNA repair. Risk allele specific oligo sequences also undergo structural modifications that could defy protein bindings. In addition, luciferase assay with wild type and risk allele indicate that these sequences have strong transcription regulatory element (TRE) activities and presence of risk allele carrying such sequences reduces enhancer activities. Subsequent experiments with purified protein and stable cells expressing wild type and mutant protein indicate that risk allele specific mutated protein binds inefficiently with its ligand, fibrinogen and vitronectin. Together with molecular modeling our results confirm that rs1143679 in ITGAM gene alters its function in mRNA transcription and ligand binding activities in vitro and in vivo and contributes toward developing SLE phenotype.

308

LOSS-OF-FUNCTION OF SEMAPHORINS 3C AND 3D IN HIRSCHSPRUNG DISEASE. *Q. Jiang[1], KP. Kilambi[2], T. Heanue[3], MX. Sosa[1], Q. Wang[4], JJ. Gray[2], AL. Kolodkin[4], DD. Ginty[4], A. Chakravarti[1]*. 1) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD; 3) MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK; 4) Department of Neuroscience, HHMI, Johns Hopkins University, Baltimore, MD.

Hirschsprung disease (HSCR, MIM# 142623), or congenital aganglionosis, results from the lack of enteric ganglia and innervation defects along varying lengths of the colon. Almost every patient harbors a loss-of-function allele at the gene encoding the receptor tyrosine kinase RET, which shows genetic interaction in mouse and zebrafish models, with type 3 C and D semaphorins (Semas) that are secreted axon guidance proteins. The latter genes were located within a locus identified by our genome-wide association study (GWAS) of 220 short segment HSCR probands with peak association in non-coding DNA between *SEMA3A* and *SEMA3D*. We performed targeted next-generation sequencing of *SEMA3A*, *SEMA3C* and *SEMA3D* among 254 isolated HSCR cases and identified 11 missense variants with the following features: (1) they were predicted to be deleterious; (2) were absent or rare in the NHLBI Exome Sequencing Project (ESP) ; (3) were overrepresented in short segment HSCR patients, compared not only to controls but also to long segment/total colonic aganglionosis HSCR cases; (4) mutant amino acid residues were almost exclusively conserved across 8 vertebrate species (human, chimp, mouse, rat, dog, cow, chicken and zebrafish) and all seven type 3 semaphorin family members (SEMA3A through SEMA3G); (5) variants were clustered in either the SEMA domain or the Ig-like domain, both demonstrated to be important for Sema structure and function; (6) variants were located in or close to the protein interfaces important for Sema homo-dimerization and binding to its receptors. We prioritized the variants with *in silico* protein structure modeling and structural predictions of substitutions and chose 5 variants (SEMA3C: S329G, V337M; SEMA3D: H424Q, V457I, P615T) for detailed functional analysis. Four variants (except SEMA3C S329G) affected the binding affinity of the cognate SEMA protein to both Neuropilin 1 and Neuropilin 2 but to different extents (40% - 80% of wild-type control). Moreover, 2 specific variants, V337M in SEMA3C and H424Q in SEMA3D, significantly decreased protein stability and led to reductions in protein expression and secretion. These data suggest that rare loss-of-function alleles at *SEMA3C*/*SEMA3D* can lead to HSCR in humans as in animal models, and variants affecting the genes encoding these two proteins are the causal factors leading to the GWAS results.

309

Functional Assessment of Human Coding Polymorphisms Affecting Skin Pigmentation Using Zebrafish. *Z. Tsetskhladze[1,6], V. Canfield[2], K. Ang[1,6], S. Wentzel[1,6], K. Reid[1,6], A. Berg[3], S. Johnson[4], K. Kawakami[5], K. Cheng[1,6]*. 1) Penn State College of Medicine, Hershey, PA; 2) Department of Pharmacology, Penn State Hershey College of Medicine, Hershey, PA, USA; 3) Department of Public Health Sciences, Penn State Hershey College of Medicine, Hershey, PA, USA; 4) Department of Genetics, Washington University Medical School, St. Louis, MO, USA; 5) Division of Molecular and Developmental Biology, National Institute of Genetics, Department of Genetics, Graduate University for Advanced Studies (Sokendai), Mishima, Shizuoka,; 6) Division of Experimental Pathology, Department of Pathology, Penn State Hershey College of Medicine, Hershey, PA, USA.

Functional validation of polymorphisms identified through genome-wide association studies (GWAS) and whole-genome sequencing has become increasingly important with the expansion of personalized medicine. Here, we focus on testing the effect of individual coding polymorphisms on phenotype. We have used pigmentation as a model polygenic human trait and the zebrafish as a vertebrate model system. Hypomorphic mutations in *SLC45A2* (also known as *MATP*) cause oculocutaneous albinism type 4 (*OCA4*), and genome-wide association studies have linked the *L374F* polymorphism in *SLC45A2* with lighter pigmentation in Europeans. A second *SLC45A2* polymorphism, *E272K*, is common in East Asian populations but its effect on skin color is poorly defined. The *A111T* polymorphism in a second pigmentation gene, *SLC24A5*, is also linked to European skin color. We tested each polymorphism in a novel whole-vertebrate-animal functional assay, “Humanized Zebrafish Orthologous Rescue” (HuZOR). HuZOR is based on the ability to rescue embryonic zebrafish mutant phenotypes by microinjection of mRNA into fertilized eggs. To test the functional impact of individual human polymorphisms in *SLC45A2*, we first identified the orthologous zebrafish mutant, albino, and demonstrated phenotypic rescue using wild-type *slc45a2* mRNA. Introduction of a single nucleotide change corresponding to the human *L374F* polymorphism abrogated rescue, confirming its functional significance. In contrast, the *E272K* polymorphism had no effect on rescue either alone or in combination with *L374F*. This finding, together with a lack of correlation between the *E272K* genotype and measured skin color in 59 East Asians, suggests that this polymorphism does not affect pigmentation in humans. Introduction of the *A111T* polymorphism abolished mRNA rescue of the zebrafish *slc24a5* (golden) mutant phenotype, consistent with its known contribution to European skin color. The results suggest that HuZOR and similar approaches using other small vertebrate models can provide insight into the impact of DNA polymorphisms on human biology and disease.

310

Dosage effects of 169 Chr21 genes on early development events in Zebrafish. *S. Edie[1], N. Zaghloul[2], D. Klinedinst[3], J. Lebron[3], N. Katsanis[4], R. Reeves[1,3]*. 1) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Medicine, University of Maryland, Baltimore, MD; 3) Department of Physiology, Johns Hopkins University, Baltimore, MD; 4) Center for Disease Modeling, Duke Institute for Genome Sciences & Policy, Durham, NC.

Down Syndrome results from trisomy for ca.300 genes on human chromosome 21 (Hsa21). We prepared a clone set of 169 Hsa21 genes based on a list of genes highly conserved between human and mouse (Gardiner et al. 2003). These were cloned into the Invitrogen Gateway entry vector to allow efficient subcloning into a variety of expression vectors. To evaluate individual gene dosage effects during early developmental stages, we performed an over-expression screen of Hsa21 genes in zebrafish. Large numbers of embryos, external fertilization and the transparent nature of these embryos made them amenable to this type of study. The clone set was subcloned into the pCS2+ vector and *in vitro* transcribed capped mRNAs were injected into zebrafish embryos to screen for phenotypes that may be associated with DS. mRNA was injected into wild type (Tübingen) zebrafish embryos at the 1–4 cell stage at two concentrations (either 10 pg and 50 pg or 50 pg and 100 pg). Embryos were then phenotyped at five days post fertilization (dpf). We noted those genes whose over-expression produced morphological phenotypes that may be associated with Down Syndrome, focusing on phenotypes related to the sonic hedgehog (Shh) and neural crest cell (NCC) pathways as these have been shown to be affected in mouse models of DS (Roper et al. 2006 and 2009). Genes that gave a phenotype were put on a short list of candidates. These were re-injected at 100 pg to confirm the phenotype, with 10 genes showing clear recapitulation of our initial screen. Of these, seven showed phenotypes commonly associated with down regulation of the Shh pathway, u-shaped somites and/or cyclopia; two showed phenotypes associated with NCC, one with craniofacial abnormalities and one with abnormal melanocytes; and one showed a heart based phenotype. These candidates are presently being evaluated to confirm the molecular basis for an over-expression effect on development in Down syndrome.

311

Two Birds, One Stone: Epistasis profiling of many single-nucleotide variants in a human gene. *O. Zill[1], J. Kitzman[1], J. Shendure[1], S. Fields[1,2].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Howard Hughes Medical Institute, Seattle, WA.

A major goal of personalized medicine is to understand which genetic variants predispose an individual to cancer and other heritable diseases. Although many disease-associated variants have been identified, rapid and scalable methods to determine which of these variants alter cellular functions are lacking. A formidable barrier to understanding which variants cause disease is that a disease state may be determined by genetic interactions (epistasis) between variants in an individual. We recently developed a "deep mutational scanning" method wherein the function of thousands of variants of a given gene can be measured in a single experiment. We are applying this approach in human cells to assign functional significance to thousands of variants of a human gene. The tumor suppressor p53, the protein most commonly mutated in cancer, provides a good test case due to the large amount of structure-function data available to benchmark our new technology. We developed a p53-responsive fluorescent reporter system in U2OS cells (*TP53*-wild-type) and Saos-2 cells (*TP53*-null) that is sensitive to *TP53* genotype. Transfection of certain cancer-associated variants (R175H, R273H) reduces reporter expression by dominant inhibition of endogenous *TP53*, whereas transfection of activating variants (e.g., K386R) induces the reporter above wild-type levels. Using a set of 17 p53 variants (including common variants, pathogenic variants, and presumptive rare variants of unknown clinical significance), we will use FACS-sorting to select those variants that activate the reporter, thereby enriching functional mutants in the population of cells. Illumina sequencing of the pre-selection and post-selection populations will allow us to distinguish inactivating variants, neutral variants, and activating variants by their relative abundances. To assay for genetic interactions, we will perform FACS-based selections in cells sensitized with siRNAs targeting genes involved in the p53 network (e.g., *MDM2*, *Ube4B*, *p300/CBP*). Eventually, we plan to perform selections on libraries encoding thousands of missense *TP53* variants in combination with a large set of siRNAs, targeting genes involved in cancer development. The results of these experiments should identify novel variants that are relevant to a person's likelihood of developing cancer. Additionally, they should provide empirical models for how interactions between variants in different genes contribute to heritable diseases.

312

Discovery and replication of pathway-based trans-eQTL associations. *L. Wiley, W. Bush.* Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN.

A logical mechanism by which SNPs affect the pathophysiology of disease is through altering the expression of genes. Several studies have explored how SNPs alter expression of nearby genes (cis-eQTLs), but far fewer studies have explored distant effects (trans-eQTLs). This is likely due to the dramatic expansion of statistical tests required and the limited interpretability of results. In this study, we performed a focused trans-eQTL analysis, hypothesizing that a single SNP can serve as a trans-eQTL by inducing expression changes through known molecular pathways. To test this hypothesis, we applied Signaling Pathway Impact Analysis (SPIA) to population-adjusted microarray expression values from HapMap II lymphoblastoid cell lines (Stranger 2007). We assessed the trans-acting effects of 771 significant ($p<7\times10^{-6}$) eQTL SNPs reported in Veyrieras et al 2008. For each eQTL, the effect of genotype on expression fold change was calculated for 11,466 genes under an additive model. These data were then processed by SPIA to assess pathway enrichment and perturbation effects for each SNP. From this analysis, we discovered 97 SNP-pathway combinations reaching genome-wide significance (Bonferroni corrected $p<6.5\times10^{-5}$), 14 of which have a significant perturbation component. Implicated pathways include RNA processing, notch signaling, and cell cycle functions as well as cancer, diabetes, and neurodegenerative diseases. Using an independent sample of HapMap III lymphoblastoid cell lines (Stranger 2012), we replicated 3 SNP-pathway effects at genome-wide significance ($p<4.55\times10^{-6}$), and find additional signals meeting nominal significance thresholds. Results of this work have potential disease implications. rs1609798 is a trans-eQTL for ~35 genes across the Alzheimer's, Parkinson's and Huntington's disease KEGG pathways ($p<6.8\times10^{-7}$) including multiple genes in the electron transport chain. This SNP is intronic to NFKB1 and a cis-eQTL for MANBA - a gene associated with the lysosomal storage disease beta-mannosidosis. Lysosomal disregulation has been implicated in all three of the neurodegenerative disease pathways associated with the eQTL SNP, and activity of beta-mannosidase has been shown to be reduced in the CSF of Parkinson's Disease patients. In summary, we identify trans-eQTL effects within the context of biological pathways that replicate across multi-ethnic populations.

313

Utilization of chromosomal microarrays in pediatrics. *A.H. Seeley[1], C.E. Keegan[1], C.S. Remmert[2], B.A. Tarini[3]*. 1) Pediatric Genetics, University of Michigan, Ann Arbor, MI; 2) Institute for Social Research, University of Michigan, Ann Arbor, MI; 3) Child Health Evaluation and Research Unit, University of Michigan, Ann Arbor, MI.

Background: Chromosomal microarrays (CMAs) have replaced G-banded karyotypes as the first tier test for evaluation of developmental delay/intellectual disability, autism spectrum disorders and multiple congenital anomalies in pediatrics. While other patient populations may benefit from CMAs, guidelines for ordering and following-up of CMAs are not clearly defined. In addition, the extent to which non-geneticists order CMAs and how they manage abnormal results is not well understood. Objective: To characterize the use for CMAs (ordering diagnosis, provider specialty, and follow-up) in pediatric patients at the University of Michigan (UM) from 2007–2010. Design/Methods: We searched billing claims in the UM Health System Data Warehouse from 2007–2010 to identify CMAs ordered for UM patients age 0–18 years. We excluded CMAs not performed by the Michigan Molecular Genetics Laboratory (i.e., sendouts). We obtained demographic data and medical record numbers for patients who met inclusion criteria. We then conducted a medical record chart review to obtain and summarize the following information: medical service ordering test, reason that the test was ordered (abstracted from lab slip or clinic note assessment), type of array, array result, and whether patient was subsequently evaluated by genetics. Results: A total of 1,714 CMAs were ordered from 2007–2010. There were 1,185 CMAs reviewed to date. The average age of the patients was 5 years (SD 4.8). Race/ethnicity of the sample was 74% White, 7% Black/African-American, 3% multi-racial, Asian and Hispanic respectively. Approximately 30% of CMAs at the University of Michigan were sent by non-geneticists, with Neurology being the main contributor (20% of total). Other subspecialists commonly sending CMAs included Behavioral Pediatrics, Child Psychiatry, and Cardiology. The majority of the ordering diagnoses were for developmental delay (48%). Aberrant results were noted in 23.5% of CMAs, and for 10% of these, there was no documentation of follow-up visit for genetic counseling. Conclusions: Many subspecialists other than geneticists utilize CMAs for evaluation of their patients. It will be important to work with these subspecialists closely to update them on new guidelines, changes in the type of arrays used, and notification practices for abnormal results so that we may optimize patient care and academic pursuits related to chromosomal abnormalities.

314

Maximizing detection and minimizing noise: the first report of large scale Whole Exome Sequencing data interpretation in a clinical laboratory. *F. Xia[1], J. Beuten[1], M. Bainbridge[2], Z. Niu[1], M. Vatta[1], M.R. Bekheirnia[1], R.E. Person[1], M. Hardison[1], J.G. Reid[2], D.P. Sexton[2], A.C. Hawes[2], P.A. Pham[2], M. Wang[2], N. Saada[2], W. Liu[2], H. Sun[1], M. Scheel[2], Y. Ding[2], A. Roy[4], J. Wiszniewska[1], A. Willis[1], D.M. Muzny[2], S.E. Plon[1,3], J.R. Lupski[1], A.L. Beaudet[1], R.A. Gibbs[2], C.M. Eng[1], Y. Yang[1]*. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 4) Department of Pathology, Baylor College of Medicine, Houston, TX.

Whole Exome sequencing (WES) is an efficient approach to detect variations in the coding region of the genome. WES has been translated into clinical settings very recently and demonstrated to be efficient and cost effective in diagnosing rare genetic disorders. Clinical interpretation of exome sequencing data is challenging due to the large and complicated data set generated for each patient and the types of variants (related to patient's phenotype, medically actionable, carrier status, etc.) expected for the clinical report.

We have been offering clinical WES since October 2011 and have performed interpretation for 60 clinical samples. The data were analyzed and annotated by the Mercury pipeline developed at the Human Genome Sequencing Center prior to clinical interpretation. Variant filtering by Mercury reduces the number from ~200,000 to ~1,400 per sample by excluding bad calls, common variants, synonymous or non-splicing variants while keeping previously reported mutations. Variants passed filtering are separated into deleterious mutations and variants of unknown clinical significance (VUS) categories, which are further sub-classified into mutations or VUSs related to patient's phenotype, mutations or VUSs unrelated to patient's phenotype, medically actionable mutations, carrier mutations, pharmacogenetic variants and mutations with no known disease associations. The category numbers are based on the numbering systems of ACMG sequence variation guidelines and expanded to include subcategory numbers representing subclasses of WES variants. Approximately 87,000 variants were interpreted for the 60 samples, including 23 deleterious mutant alleles related to patient phenotype, 164 VUS related to patient phenotype, 2,405 VUSs unrelated to patient phenotype, 12 medically actionable mutations, 84 carrier mutations, and ~800 deleterious mutations with no disease associations. The rest are VUSs in genes not associated with Mendelian disorder. Strategies to maximize detection rate of WES testing include maximizing indel call sensitivity, ensuring adequate WES coverage of candidate genes, family studies and weekly updating interpretation with new disease genes. Our experience of large scale WES data interpretation highlights the challenges of WES reporting and the importance of implementing well defined analysis and interpretation pipelines in a clinical setting.

315

Efficient Detection of Causative Mutations for Rare Diseases: Rethinking Clinical Practice. *H. Lee[1], J. Deignan[1], T. Toy[2], B. Harry[2], M. Yourshaw[2], P. Taylor[2], S. Webb[1], N. Dorrani[3], K. Das[1], F. Quintero[1], S. Kantarci[1], D.A. Wong[3], W.W. Grody[1], E. Vilain[2], S.F. Nelson[1,2]*. 1) Dept Pathology and Laboratory Medicine, David Geffen School of Medicine, Univ California, Los Angeles, Los Angeles, CA; 2) Dept Human Genetics, David Geffen School of Medicine, Univ California, Los Angeles, Los Angeles, CA; 3) Dept of Pediatrics, Division of Genetics, David Geffen School of Medicine, Univ California, Los Angeles, Los Angeles, CA.

The advent of routine and reliable human exome sequencing now permits ~95% of all protein coding bases to be reliably sequenced at a reasonable cost. Because ~85% of known disease causing mutations occur within the 1% of the genome well predicted to code for proteins, surveying the exome becomes a powerful clinical diagnostic tool for genetically heterogeneous disorders.The UCLA Clinical Genomics Center has launched the CLIA-certified Clinical Exome Sequencing (CES) Test with the purpose of identifying the causal mutations and facilitating early diagnosis of rare Mendelian disorders. This CLIA/CAP compliant test process includes DNA extraction, library preparation, capture, sequencing, data analysis, and interpretation. Once the data are processed through our bioinformatics pipeline that's been validated to be sensitive and reliable, the fully annotated variant list is reviewed at our Genomic Data Board comprised of an interdisciplinary team of physicians, pathologists, clinical geneticists, lab directors, genetic counselors, and informatics specialists to determine patient's final result.Since the launch, we've provided the service to both UCLA's and outreach patients. Most cases were those who had already exhausted all currently available diagnostics options. Even in these first 19 cases, we were able to establish new and clear genetic diagnosis in 53% that led physicians to perform more focused phenotyping supportive of the new genetic diagnosis. For instance, a diagnosis was changed from juvenile ALS to AAAS, drastically changing optimized patient care. Of note 4 patients have mutations in genes that are highly likely to be causal but not previously described highlighting the potential of CES for gene discovery and need for more efficient data sharing. Among the 47% of the cases in which a clearly pathogenic mutation was not identifiable, 56% were observed with potential pathogenic mutations that will require further research to confirm the causality. About 21% had no clear or even highly suspected variant detected. The enterprise is informing us best means to identify likely causal mutations and pointing the way forward for more efficient diagnosis using state-of-the art tools and algorithms and the innovative Genomic Data Board. We intend to implement CES as a first line genetic diagnostic tool while recognizing that better patient phenotyping or better tools that permit integration with patient data will be necessary for optimal implementation.

316

Intentions to receive individual results from whole-genome sequencing among participants in the ClinSeqTM study. *B.B. Biesecker[1], F.M. Facio[2], H. Eidem[2], T. Fisher[1], S. Brooks[2], A. Linn[2], K.A. Kaphingst[3], L.G. Biesecker[2]*. 1) Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD; 2) Genetic Diseases Research Branch, National Human Genome Research Institute, Bethesda, MD; 3) Department of Surgery, Washington University School of Medicine, St. Louis, MO.

Objectives: Genome sequencing has been rapidly integrated into clinical research, and is currently marketed to health care practitioners and consumers alike. The volume of sequencing data generated for a single individual and the wide range of findings from whole genome sequencing raise critical questions about the return of results and their potential value for end-users. Methods: We conducted a mixed-methods study of 311 sequential participants in the NIH ClinSeqTM study to assess general preferences and specific attitudes toward learning results and perceived opinions of valued others. We tested how these variables predicted intentions to receive results within four categories of findings ranging from medically actionable to variants of unknown significance. Results: Two hundred ninety-four participants indicated a preference to learn their genome sequencing results and six were unsure. Most often participants cited disease prevention as their reason, including intention to change their lifestyle behaviors. A third expressed a general desire to know, reflecting those who generally valued information and others who sought to understand the personal implications of findings. Participants had positive attitudes, strong perceived social norms and strong intentions to learn results overall, although there were significant mean differences among four categories of findings ($p<0.01$). Attitudes and social norms for medically actionable and carrier results were most similar and rated the highest. Attitudes and norms significantly contributed to the variance in intentions. Among these early adopters there was overwhelming enthusiasm to learn results. Conclusions: Participants distinguished among the types and quality of information they may receive despite strong intentions to learn all results presented. These intentions were motivated by confidence in their ability to use the information to prevent future disease and a belief in the value of even uninterpretable information. It behooves investigators to facilitate participants' desire to learn a range of information from genomic sequencing, while promoting realistic expectations for its clinical and personal utility.

317

Changes to Control Perceptions Following Disclosure of *APOE*-Coronary Artery Disease Associations during Genetic Susceptibility Testing for Alzheimer's Disease: Findings from the REVEAL Study. *K. Christensen[1, 2], J.S. Roberts[1], W.R. Uhlmann[3], P.J. Whitehouse[4], T.O. Obisesan[5], D.L. Bhatt[6], L.A. Cupples[7], R.C. Green[8]*. 1) Health Behavior and Health Education, University of Michigan School of Public Health, Ann Arbor, MI; 2) Genetic Alliance, Washington, DC; 3) Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI; 4) University Memory and Aging Center/Department of Neurology, Case Western Reserve University, University Hospitals Case Medical Center, Cleveland, OH; 5) Department of Medicine, Howard University, Washington, DC; 6) VA Boston Healthcare System, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA; 7) Departments of Biostatistics and Epidemiology, Boston University School of Public Health, Boston, MA; 8) Partners Center for Personalized Genetic Medicine, Division of Genetics, Brigham and Women's Hospital and Harvard Medical School.

Background: Prior analyses suggested that disclosing an additional association between *APOE* and coronary artery disease (CAD) during a genetic risk assessment for Alzheimer's disease (AD) may motivate improvements to health behaviors and reduce test-related distress. Changes to beliefs about whether AD outcomes can be altered through personal action ("AD control perceptions") that result from learning that the different diseases have shared genetic risk factors may explain this tendency. **Methods**: The Risk Evaluation and Education for Alzheimer's Disease (REVEAL) Study randomized participants into one of two disclosure arms. A control arm received *APOE* genotypes and an AD risk estimate (range: 6%–70%) based on gender, ethnicity, family history, and *APOE* genotype. Subjects randomized to a `pleiotropy' arm were also informed that *APOE* ε4 carriers were at increased risk for CAD. Control perceptions were assessed before and after testing on 5-point scales using items adapted from the Revised Illness Perceptions Questionnaire. Scores were compared to self-reported health behaviors 6 weeks after disclosure of *APOE* genotypes and AD risk, as well as test-specific distress as measured using the Impact of Events Scale (range: 0–75). **Results**: A total of 257 subjects (median age 58; 55% female; 16% African American; 71% with affected first degree relatives) received genetic risk estimates. Linear regression controlling for age, gender, race, and income showed no changes or slight drops in AD control perceptions among all participants in the control arm and ε4-negative participants in the pleiotropy arm (Δ: -.14-.01). ε4-positive participants in the pleiotropy arm, however, reported increases to control perceptions after testing (Δ: .36, $F=12.1$, $p=.001$). Stronger post-disclosure control perceptions were associated with a greater likelihood of reporting changes to diet (OR per 1-point increase in control perceptions: 1.64, $p=.03$), herbal supplements (OR: 2.62, $p<.001$), and stress reduction (OR: 2.09, $p=.01$), but were not associated with test-specific distress ($\beta=-0.54$, $p=.35$). **Conclusions**: Pleiotropic information shared during a genetic risk assessment for AD can motivate improvements in health behaviors by making AD feel more controllable, possibly because it makes individuals believe CAD and AD risk are interrelated. Findings suggest that disclosure of incidental genetic information can have unexpected benefits with respect to disease prevention.

318

Decrease prediction ability of common genetic variants on breast cancer risk with age: possible underlying models and impact on risk prediction. *H. Aschard[1,2], S. Lindstrom[1,2], P. Kraft[1,2,3].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Program in Molecular And Genetic Epidemiology, Harvard School of Public Health, Boston, MA; 3) Department of Biostatistics, Harvard School of Public Health, Boston, MA.

Background: Recent studies have shown that the prediction ability of common single nucleotide polymorphisms (SNPs) associated with multifactorial diseases such as prostate cancer may decrease with age. Despite the potential impact of these differences in risk assessment, the causes remain unknown. In this study we show that a similar pattern is observed in breast cancer and we compare potential underlying models that may explain these differences.

Method: Using 2287 breast cancer cases and controls from the Nurses' Health Study, we first examined the age-specific effects and predictive ability (as measured by the Area Under the ROC Curve (AUC)) of 15 common SNPs associated with breast cancer. We then perform simulation studies to explore how hypothetical underlying genetic models may fit the observed results.

Results: We observed a decrease in prediction ability of risk SNPs with age (AUC=0.623 and 0.577 for younger and older individuals respectively). Although no individual SNP by age interaction was significant after correction for multiple testing, we observed a significant (P=0.03) negative interaction effect between age and a genetic risk score define by the sum of risk alleles. Simulation study reveals models where SNP effects and AUC can differ by strata defined by strong risk factor without the presence of interactions on the odds ratio scale. However both empirical and simulated data suggest that the observed age-specific SNPs effect is likely to reflect genuine statistical interactions that explain the differences in AUC between age classes.

Conclusion: We show that the effect of common SNPs on breast cancer risk decreases with age. This age-specific effect of SNPs is likely to explain the lower prediction ability of a genetic risk model of breast cancer in older individuals. The identification of non-genetic factors that alter globally the effects of multiple genetic variants can help to understand the genetic architecture of multifactorial diseases and to identify sub-groups of individuals that may benefit more from genetic testing.

319

Large-sample size, comprehensive catalog of variants and advanced machine learning technique boost risk prediction for inflammatory bowel disease. *Z. Wei[1], W. Wang[1], J. Bradfield[2], E. Frackelton[2], C. Kim[2], F. Mentch[2], R. Baldassano[2,3], H. Hakonarson[2,3], The International IBD Genetics Consortium.* 1) Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey, USA; 2) The Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 3) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

Genome-wide association studies (GWAS) have been fruitful in identifying disease susceptibility loci for Crohn's disease (CD) and ulcerative colitis (UC), the two common forms of inflammatory bowel disease (IBD). However, it remains unclear if these advances can deliver sufficiently accurate predictions to make targeted intervention realistically possible. Like many other complex diseases, existing results for CD and UC risk prediction are generally modest, if not negative. Thus some practitioners believe that individual-based disease risk assessment may not be feasible any time soon. However, we argue that the discouraging risk prediction performance so far may be attributed to three major factors: insufficient sample size, incomprehensive catalog of variants and the curse of high dimensionality in modeling GWAS data. Here we performed risk assessment for IBD using data from the International IBD Genetics Consortium's Immunochip project. This dataset contains ~17,000 CD cases, ~13,000 UC cases and ~22,000 controls from 15 European countries typed on the Immunochip, a custom Illumina Infinium chip. This custom chip is designed both to perform deep replication of suggestive associations, and fine mapping of established GWAS loci of genome-wide significance. It provides a more comprehensive catalog of the most promising candidate variants by picking up the remaining common variants and certain rare variants that were missed in the first generation of GWAS. Given this unprecedented large sample size, we employed the most recent machine learning techniques to build optimal predictive models. Our final predictive models achieved AUCs of 0.87 and 0.83 for CD and UC, respectively, in an independent evaluation. To our knowledge, this is the best prediction performance ever reported for CD and UC. An examination of the final predictors shows that more than 20% of the selected predictors are not typed by current genotyping chips used for GWAS. Most of them are less common (MAF<0.05) or rare (MAF<0.01) variants, without which the AUCs of the predictive models are reduced to 0.81 and 0.76 for CD and UC, respectively. Our results for IBD lend support to an optimistic view that genotype-based genetic risk prediction may be feasible given sufficient sample size, wide variant spectrum and proper modeling.

320

A Comparison of Risk Estimates for Complex Diseases: Navigenics SNP-Based Testing and Family History Assessment. *L. Aiyar[1,2], C. Shuman[1,2], R. Hayeems[2,3], L. Velsher[4,5], S. Wodak[6], D. Chitayat[1,2,7], J. Davies[5].* 1) University of Toronto, Toronto, Ontario, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, Ontario, Canada; 4) North York General Hospital, North York, Ontario, Canada; 5) Medcan Clinic, Toronto, Ontario, Canada; 6) Centre for Computational Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Department of Obstetrics and Gynecology, The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, Toronto, Ontario, Canada.

In the past decade, new genetic testing technologies have allowed for the discovery of hundreds of genetic variants that contribute to the development of common complex diseases. Some of these tests are being offered directly to the public in a direct-to-consumer (DTC) manner. One area of concern regarding DTC personal genome testing is whether these common variants that contribute to disease can accurately assess individual risks for developing complex diseases. Risks generated from DTC genetic tests do not incorporate traditional measures of risk, such as family history and clinical assessments. Previous studies have provided some evidence that for specific common conditions (e.g. prostate cancer), personal genome testing can be combined with family history and clinical information to build predictive risk models. This study compared risk estimates generated by family history analysis and Navigenics SNP-based personal genome testing to determine risk correlations. The data will be valuable for assessing clinical utility of incorporating family history assessment with personal genome testing. In the present study, we compared risk estimates for 760 patients who purchased the Navigenics SNP-based test at the Medcan Clinic in Toronto, Canada, between 2009–2011. Risk categorizations for 20 complex conditions included in the Navigenics testing were compared with risk categorization estimates derived from a detailed family history assessment using kappa statistic. Risk estimates based on family history analysis were derived from a review of empiric risk estimates in the literature. Only conditions with an estimated population incidence of greater than 2% in either males or females were analysed. We expect that for some conditions (i.e. those with a higher lifetime risk and greater genetic contribution), a statistically significant positive correlation between risks will be identified. Data was also collected regarding additional genetic risks identified by genetic counselors at the Medcan Clinic, that are not assessed by personal genome testing. Of the cohort studied, 86 individuals (11.3%) were found to have additional potential genetic risks, including 38 individuals (5%) whose family histories were suggestive of hereditary cancer syndromes. Preliminary results suggest that eliciting family history adds value to the overall risk assessment for individuals undergoing personal genome testing.

321

Web-based Case Conferencing: An effective source of cancer genetics training for community-based clinicians. *K. Blazer, J. Weitzel.* Clinical Cancer Gen, City Hope Comprehen Cancer Ctr, Duarte, CA.

Rapid advances in genetic/genomic technologies are fueling demands for genetic cancer risk assessment (GCRA) services in the community setting, where few clinicians have GCRA training. City of Hope Cancer Center conducts a GCRA training course for community-based clinicians, comprised of 9 weeks of distance didactics, 5 days of face-to-face workshops and 12 months of post-course professional development. As designed, the course cannot meet increasing demands for training. Action research identified face-to-face workshops as a barrier to increasing course capacity. Review of adult learning theory led to identification of Web-based case conferencing as a potential situated learning environment for distance-mediated GCRA training. Previously-reported results of a quasi-experimental analysis comparing participation in a weekly Web-based GCRA case conference intervention to face-to-face case-based training demonstrated equivalent or better learning/skills outcomes in the intervention group across all disciplines (physicians, advanced-practice nurses, genetic counselors). Qualitative findings from the study are reported here. Methods: Multiple surveys and roundtable discussion groups were conducted with intervention group participants (n = 52) during and immediate-post distance-mediated training. Results: Participants reported that engagement in five weekly Web-based case conferences concurrent with distance didactic learning generated new learning, reinforced existing knowledge in a broad spectrum of interdisciplinary knowledge and skills domains, and helped participants identify, reflect upon and address individual GCRA-related knowledge and skills deficits. These findings, along with feedback on limitations, connectedness and communication during Web conferencing, informed the development of a new course design that incorporates Web conferencing as an accessible source of case-based GCRA training to promote best practices in the delivery of GCRA services across the United States and internationally.

322

Combination of modern and traditional techniques identify MCKD1 causal frameshift variants within the *MUC1* VNTR. *A. Kirby[1,2], A. Gnirke[2], D. Jaffe[2], V. Bareš ová[3], N. Pochet[2,4], B. Blumenstiel[2], C. Ye[2], D. Aird[2], C. Stevens[2], J. Robinson[2], M. Calibi[2,5], I. Gat-Viks[2], E. Kelliher[2], R. Daza[2], M. DeFelice[2], H. Hůlková[3], J. Sovová[3], C. Antignac[6,7,8], M. Guttman[2], R. Handsaker[2,9], K. Lindblad-Toh[2,10], S. Gabriel[2], P.S. Hart[11], A. Regev[2], C. Nusbaum[2], S. Kmoch[3], A. Bleyer[12], E. Lander[2], M. Daly[1,2]*. 1) Analytic and Translational Genetics Unit, Massachusetts General Hosp, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, Massachusetts; 3) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Ke Karlovu 2, 128 08 Prague 2, Czech Republic; 4) Department of Plant Systems Biology, VIB, Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium; 5) Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA; 6) Inserm, U983, Paris, France; 7) Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris, France; 8) Département de Génétique, Hôpital Necker-Enfants Malades, Assistance Publique-Hôpitaux de Paris, Paris, France; 9) Department of Genetics, Harvard Medical School, Boston, Massachusetts; 10) Science for Life Laboratory Uppsala, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 751 23, Sweden; 11) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health (NIH), Bethesda, Maryland; 12) Section on Nephrology, Wake Forest School of Medicine, Medical Center Blvd., Winston-Salem, North Carolina.

While the genetic lesions responsible for some Mendelian disorders can be rapidly discovered through massively parallel sequencing (MPS) of whole genomes or exomes, it is becoming clear that some diseases do not readily yield to such efforts. Here, we describe the illustrative case of medullary cystic kidney disease type 1 (MCKD1). MCKD1 is a simple Mendelian disorder that was mapped more than a decade ago to a 2-Mb region on chromosome 1. We performed whole-genome, whole-exome and regionally targeted sequencing on 12 individuals from 6 families showing linkage, but found no evidence of a causative mutation in any family. After a process of elimination (examining essentially every other nucleotide in the 2-Mb region), we concluded that the causative mutations likely resided in an extremely long (~1–7 kb), GC-rich (>80%) VNTR (variable number of tandem repeats) sequence in the coding region of the mucin 1 (*MUC1*) gene that was dramatically underrepresented in MPS data. By cloning, subcloning and reconstructing the region in many patients, we found that every MCKD1 family harbors an equivalent, but apparently independently arising, mutation, consisting of an insertion of an extra C within a run of seven Cs in one copy of the repeat; the mutation causes a frameshift and gives rise to a novel peptide repeat before premature termination. Existence of the predicted mutant protein was confirmed in the kidneys of MCKD1 patients using antibodies against the novel VNTR sequence. Further, we found no evidence for the insertion event in a subset of sporadic cases and smaller families, suggesting that other *MUC1* mutations and/or medullary cystic kidney disease types/loci may yet remain undiscovered. In addition to revealing the biology of MCKD1, the results provide a cautionary tale about the challenges in identifying the genes responsible for Mendelian, let alone more complex, disorders through MPS.

323

ARL13B, INPP5E, PDE6D and CEP164 form a functional network involved in Joubert syndrome and Nephronophthisis. *S. Seo[1], M.C. Humbert[1,2], K. Weihbrecht[3], C.C. Searby[2,4], Y. Li[5], R.M. Pope[5], V.C. Sheffield[2,4]*. 1) Dept of Ophthalmology and Visual Sciences, Univ of Iowa, Iowa City, IA; 2) Dept of Pediatrics, Univ of Iowa, Iowa City, IA; 3) Interdisciplinary Graduate Program in Genetics, Univ of Iowa, Iowa City, IA; 4) Howard Hughes Medical Institute; 5) Proteomics Facility, Univ of Iowa, Iowa City, IA.

Mutations affecting ciliary components cause a series of related genetic disorders in humans, such as nephronophthisis (NPHP), Joubert syndrome (JBTS), Meckel-Gruber syndrome (MKS), and Bardet-Bield syndrome (BBS), which are collectively termed `ciliopathies'. Recent protein-protein interaction studies combined with genetic analyses have revealed that ciliopathy-related proteins form several functional networks that build and maintain the primary cilia and its function. However, the precise functions of many ciliopathy-related proteins and the mechanisms by which these proteins are targeted to the primary cilia are still not well understood. Here, we describe a protein-protein interaction network of INPP5E, a prenylated protein associated with JBTS, and its ciliary targeting mechanisms. Through serial deletion mutagenesis, immunofluorescence microscopy, siRNA mediated gene knock-down approaches, tandem affinity purification and co-immunoprecipitation analyses, we determined that INPP5E is targeted to the primary cilia through a motif near the C-terminus independently of prenylation. Ciliary targeting of INPP5E is facilitated by another JBTS protein, ARL13B, but not by ARL2 or ARL3. We further demonstrate that INPP5E interacts with ciliopathy-related proteins CEP164 and PDE6D and that these proteins are required for ciliary targeting of INPP5E. These findings suggest that ARL13B, INPP5E, PDE6D, and CEP164 form a novel functional network that is involved in JBTS and NPHP but distinct from previously defined NPHP and MKS protein networks.

324

Mainzer-Saldino syndrome is a ciliopathy caused by mutations in the IFT140 gene. *I. Perrault[1], S. Saunier[2], S. Hanein[1], E. Filhol[2], A. Bizet[2], F. Collins[3], M. Salih[4], S. Gerber[1], N. Delphin[1], E. Silva[5], V. Baudouin[6], M. Oud[7], N. Shannon[8], M. Le Merrer[1], O. Roche[9], C. Pietrement[10], C. Bole-Feysot[11], P. Nitschke[12], M. Zahrate[11], P. Beales[13], H. Arts[7], A. Munnich[1], J. Kaplan[1], C. Antignac[2], V. Cormier-Daire[1], J-M. Rozet[1]*. 1) INSERM U781, Department of Genetics, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, Paris, France; 2) INSERM, U983, Paris Descartes University, Paris, France; 3) Department of Clinical Genetics,Westmead Hospital, Sydney, Australia; 4) Division of Pediatric Neurology, King Khalid University Hospital, Riyadh, Saoudi Arabia; 5) Department of Ophthalmology, Coimbra University Hospital, Coimbra, Portugal; 6) Derpartment of Nephrology, CHU Robert Debré, Paris, France; 7) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 8) Clinical Genetics Service, City Hospital, Nottingham, UK; 9) Department of Ophthalmology, Université Paris Descartes-Sorbonne Paris Cité, Paris, France; 10) Department of Pediatry, American Memorial Hospital, CHU Reims, Reims, France; 11) Genomics Plateform, IMAGINE Foundation and Paris Descartes University, Paris, France; 12) Bioinformatics Plateform, Paris Descartes University, Paris, France; 13) Molecular Medicine Unit, University College London (UCL) Institute of Child Health, London, UK.

Introduction: Ciliopathies is an emerging class of genetic disorders due to altered cilia assembly, maintenance or function. Syndromic ciliopathies affecting bone development have been classified as skeletal ciliopathies. Mutations in genes encoding components of the intraflagellar transport (IFT) complex A, that drives retrograde ciliary transport, are a major cause of skeletal ciliopathies. On the other hand, Mainzer-Saldino syndrome (MSS) is a rare disorder characterized by phalangeal cone-shaped epiphyses, chronic renal failure and early-onset severe retinal dystrophy. Methods and results: We collected 16 families presenting three diagnostic criteria of MSS. Through ciliome re-sequencing combined to Sanger sequencing, we identified IFT140 mutations in seven MSS families. The effect of the mutations on the localization of IFT140 was assessed using flagged-IFT140 mutant proteins which showed a partial to nearly complete loss of basal body localization associated with an increase of cytoplasm staining while the wild-type Flagged-IFT140 protein predominantly localized to the basal bodies in RPE1 cells. To assess the impact of IFT140 mutations on ciliogenesis, abundance and morphology of primary cilia were studied in cultured fibroblasts of patients and detected absent cilia in a high proportion of patient cells compared to controls. Ciliary localization of anterograde IFTs were altered in MSS patient fibroblasts supporting the pivotal role of IFT140 in proper development and function of ciliated cells. Conclusion: Here we report on compound heterozygosity or homozygosity for IFT140 mutations in seven MSS families. After Sensenbrenner and Jeune syndromes, MSS is the ultimate skeletal ciliopathy ascribed to IFT disorganization.

325

Mutations in ALDH1B1, which encodes a mitochondrial protein belonging to the aldehyde dehydrogenase family, result in hepatic failure and mitochondrial respiratory chain deficiency. *S. Salhi[1], V. Serre[1], M. Beinat[1], P. Nitschke[2], O. Bernard[3], A. Slama[4], A. Munnich[1], A. ROTIG[1]*. 1) Université Paris Descartes-Sorbonne Paris Cité, In, Hôpital Necker-Enfants Malades, PARIS, France; 2) Plateforme Bioinformatique Paris Descartes; 3) Service d'Hépatologie pédiatrique, Hôpital Bicêtre, 94275 Le kremlin Bicêtre; 4) Laboratoire de Biochimie 1, Hôpital Bicêtre, Le Kremlin-Bicêtre.

Hepatic failure is a frequent feature in respiratory chain (RC) deficiency. Patients usually present liver insufficiency, cholestasis, cirrhosis and/or cytolysis soon after birth or during the first weeks of life. Multiple respiratory chain (RC) enzyme deficiency with a severe decrease of liver mitochondrial DNA (mtDNA) copy number (mtDNA depletion) has been found in a significant fraction of patients and ascribed to mutations in genes involved in mtDNA replication or dNTP supply for mtDNA replication. Few cases of severe hepatic failure with normal mtDNA content in liver have been ascribed to genes involved in the mitochondrial translation machinery. We have performed next generation sequencing (exome) in a large series of patients with infantile liver failure and multiple RC deficiency and no quantitative or qualitative mtDNA anomalies. This allowed us to identify compound heterozygous mutations of the ALDH1B1 gene in one of these patients who presented liver failure and cardiomyopathy (p.Phe256X, Leu219Pro). ALDH1B1encodes a mitochondrial protein belonging to the aldehyde dehydrogenase family (aldehyde dehydrogenase 1 family, member B1). Normal amount of ALDH1B1 transcripts was detected in patient's fibroblasts. Only 30% of ALDH1B1 protein was observed in these cells by Western blot analysis. Moreover no abnormal sized ALDH1B1 protein could be detected. These data suggest that the truncated protein is unstable and the Leu219Pro modification alters the stability of the protein. RC enzyme studies revealed a combined R deficiency in fibroblasts of the patient but normal RC activities in cultured skin fibroblasts. Blue Native Polyacrylamide Gel Electrophoresis of the mitochondrial RC showed a complete absence of cytochrome c oxidase whereas all other complexes were normal in patient's fibroblasts. Aldehyde dehydrogenase is the second enzyme of the major oxidative pathway of alcohol metabolism but no enzymatic function related to ALDH1B1 has been ever described. Identification of ALDH1B1 mutations in a patient with abnormal assembly of mitochondrial complex IV suggests that this protein with yet unknown function is involved in RC assembly. Further studies will shed light on the function of this protein and its involvement in cytochrome c oxidase assembly.

326

Targeted Exome Sequencing of 102 Patients with Clinical Evidence of Mitochondrial Disease. *D.S. Lieber[1,2,3,4], S.E. Calvo[1,2,3,4], K. Shanahan[5], N.G. Slate[1,2], S. Liu[1,2], S.G. Hershman[1,2,3,4], N.B. Gold[1,2], B.A. Chapman[1], M. Borowsky[1], D.R. Thorburn[6], G.T. Berry[7], J.D. Schmahmann[8], D.M. Mueller[5], K.B. Sims[2,8], V.K. Mootha[1,2,3,4]*. 1) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 3) Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA; 4) Broad Institute of Harvard and MIT, Cambridge, MA 02141, USA; 5) Department of Biochemistry and Molecular Biology, Rosalind Franklin University of Medicine and Science, The Chicago Medical School, North Chicago, IL 60064, USA; 6) Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, VIC 3052, Australia; 7) Harvard Medical School, The Manton Center for Orphan Disease Research, Children's Hospital Boston, Boston, MA 02115 USA; 8) Department of Neurology, Massachusetts General Hospital & Harvard Medical School, Boston MA 02114, USA.

The genetic diagnosis of mitochondrial disorders is challenging due to the large number of candidate genes that can underlie these phenotypically diverse disorders. We have assessed the utility of a “MitoExome” sequencing approach that targets the mtDNA and the exons of nearly 1600 nuclear-encoded genes implicated in mitochondrial disease, mitochondrial function, or phenotypically similar monogenic disorders. Previously, we benchmarked the approach in a cohort of severe infantile cases with biochemically-proven mitochondrial disease and estimated it could enable molecular diagnoses in up to 50% of such cases, including 24% of cases lacking a prior molecular diagnosis. Here, we extend the approach to a broader cohort of 102 phenotypically diverse patients with clinical evidence of mitochondrial disease, and show that it enabled molecular diagnoses in 22% of patients, including 17/18 (94%) with prior molecular diagnoses and 5/84 (6%) of previously unsolved cases. In three of five newly diagnosed patients we detected recessive mutations in genes underlying disorders in the differential diagnosis (*DPYD*, *KARS*, *WFS1*), underscoring the phenotypic overlap between mitochondrial disorders and other inborn errors. An additional 23 patients harbored likely deleterious mtDNA, recessive, or X-linked variants, a two-fold enrichment over the background rate. Experimental follow-up of one such variant in a yeast model suggests that recessive mutations in *ATP5A1* can cause mtDNA depletion and combined oxidative phosphorylation deficiency. Our results demonstrate the technical advantages of next-generation sequencing and highlight the challenges in variant interpretation for the molecular diagnosis of mitochondrial disease.

327

Genetic diagnosis of mitochondrial disorders by whole-exome sequencing. *C.J. Carroll[1], V. Brilhante[1], P. Isohanni[1,2], R. Pöyhönen[1], L. Euro[1], U. Richter[1], T. Lahtinen[1], A. Götz[1], H. Almusa[3], P. Ellonen[3], H. Pihko[2], B. Battersby[1], H. Tyynismaa[1,4], A. Suomalainen[1,5]*. 1) Research Program of Molecular Neurology, Helsinki University, Helsinki, Finland; 2) Department of Pediatric Neurology, Hospital for Children and Adolescents, Helsinki University Central Hospital, Helsinki, Finland; 3) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 4) Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland; 5) Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland.

Mitochondrial disorders are a heterogeneous group of genetic diseases that may affect any organ, disease may manifest at any age and have differing levels of severity. Mitochondrial diseases are caused by mutations in either the mitochondrial genome or in nuclear genes encoding mitochondrial proteins, and most patients still remain without specific diagnosis. We report here whole-exome sequencing results from 45 patients with suspected recessively inherited mitochondrial disease. We customized an analytic pipeline, enabling us to prioritize variants in genes encoding known or predicted mitochondrial proteins that were either homozygous or were compound heterozygous mutations and were predicted to be deleterious using SIFT and PolyPhen. We identified genetic background in over 30% of patients: novel mutations in 9 patients in known disease genes (7 genes) and mutations in 5 patients in genes not previously connected with a disease (5 genes). Of the unknown disease genes, a homozygous mutation in MRPL44 was uncovered as the underlying genetic defect in hypertrophic cardiomyopathy with complex I and IV deficiency in the heart and skeletal muscle. MRPL44 encodes a large ribosomal subunit protein of the mitochondrial translation machinery and its yeast homologue has been suggested to locate in close proximity to the tunnel exit of the ribosome. We found severely reduced MRPL44 levels in the patient's heart, skeletal muscle and fibroblasts suggesting that the mutation affected protein stability. The decreased MRPL44 resulted in low levels of assembled large ribosomal subunit and partial instability of the 16S rRNA, leading to complex IV deficiency in patient's fibroblasts. Retroviral expression of wild-type MRPL44 in patient cells rescued the large ribosome assembly defect and COX deficiency. The result suggests that MRPL44, which has no homologue in bacteria and has only emerged in the mitochondrial ribosome of eukaryotes, is essential for translation in human mitochondria. We conclude that exome-sequencing can reveal molecular background in a high proportion of recessive mitochondrial disorders, and suggest it as a valuable method for first-line genetic diagnosis in this disease group.

328

Constitutive activation of STIM1 causes tubular aggregate myopathy. *J. Laporte[1], F. Chevessier[2], A. Maues de Paula[3], C. Koch[1], S. Attarian[4], C. Feger[1], D. Hantaï[2], P. Laforêt[2], K. Ghorab[5], J.M. Vallat[5], M. Fardeau[2], D. Figarella-Branger[6], J. Pouget[4], M. Koch[1], C. Ebel[1], N. Levy[3], B. Eymard[2], M. Bartoli[3], J. Bohm[1].* 1) Translational Medicine and Neurogenetics, IGBMC, Illkirch, France; 2) Centre de Référence de Pathologie Neuromusculaire Paris-Est, Paris, France; 3) Faculté de Médecine, Aix-Marseille Université, Marseille, France; 4) Centre de Référence des Maladies Neuromusculaires et de la SLA, Hôpital d'Enfants de la Timone, Marseille, France; 5) Département de Neurologie et Centre National de Référence Neuropathies Périphériques Rares, CHU de Limoges, Limoges, France; 6) APHM, Service d&aposåatomie pathologique et neuropathologie, Hôpital de la Timone, Marseille, France.

Ca2+ is a major regulatory and signaling molecule in skeletal muscle, therefore the cellular Ca2+ dynamics need to be tightly regulated. Intracellular Ca2+ is mainly stored in the sarcoplasmic reticulum (SR) and is released to the cytoplasm upon stimulation, where it triggers muscle contraction and acts as a second messenger controlling growth and differentiation. Ca2+ stores are refilled through a process called store-operated Ca2+ entry (SOCE). Stromal interaction molecule 1 (STIM1) is the main Ca2+ sensor in the endoplasmic reticulum. We identified STIM1 mutations as the genetic cause of tubular aggregate myopathy (TAM), characterized by regular arrays of membrane tubules on muscle biopsies. The 4 families presented predominantly with mild and slowly progressive lower limb muscle weakness, upward gaze paresis and strongly elevated creatine kinase levels. The aggregates appeared in blue on NADH-TR staining, and were negative on muscle sections colored with SDH, indicating a reticulum and not a mitochondrial origin. Immunohistofluorescence revealed that aggregates were labeled with markers of the sarcoplasmic reticulum, while STIM1 localized in their periphery. Ultrastructural analysis demonstrated massive tubular aggregation with single or double-walled membranes of different diameter. All heterozygous mutations were found in the highly conserved intraluminal EF-hands, sensing and binding Ca2+. Upon Ca2+ store depletion, wild-type STIM1 oligomerizes and thereby triggers extracellular Ca2+ entry through Ca2+-release-activated Ca2+-channels (CRAC). In contrast, myoblasts transfected with the mutant constructs displayed constitutive STIM1 clustering, indicating that Ca2+ sensing was lost. We investigated the pathological mechanism underlying the disease and monitored the calcium response of patient myoblasts to SOCE. We found a significantly higher basal Ca2+ level in patient cells as compared to the control. Addition of high [Ca2+] medium induced a sudden and massive Ca2+ influx in the patient myoblasts, as compared to low gradual increase in control cell lines. These data support the constitutive activation of the Ca2+ entry channels in the myoblasts harboring a dominant STIM1 mutation. Recessive loss-of-function mutations in STIM1 have been associated with severe immune deficiency, demonstrating that a tight regulation of STIM1-dependent SOCE plays an essential role in T-cell activation as well as in normal skeletal muscle structure and function.

329

Mutation in PNPT1 gene, which encodes a mitochondrial polyribonucleotide nucleotidyltransferase, causes encephalopathy with choreo-athetotic movements. *V. Vedrenne[1], A. Gowher[2], P. De Lonlay[3], P. Nitschke[4], V. Serre[1], N. Boddaert[3], C. Altuzarra[5], A.M. Mager-Heckel[2], F. Chretien[3], N. Entelis[2], I. Tarassov[2], A. Munnich[1,3], A. Rotig[1].* 1) Université Paris Descartes-Sorbonne Paris Cité, In, Hôpital Necker-Enfants Malades, PARIS, France; 2) UMR No 7156 Université de Strasbourg - CNRS, Génétique Moléculaire, Génomique, Microbiologie (GMGM), 67084 Strasbourg, France; 3) Department of Pediatrics, Hôpital Necker-Enfants Malades, Université Paris Descartes, 75015 Paris, France; 4) Plateforme Bioinformatique Paris Descartes; 5) Service de Pédiatrie - Hôpital Saint-Jacques, 25030 Besançon, France.

Multiple respiratory chain deficiency represents an important cause of mitochondrial disorders. Hitherto however, mutations in genes involved in mitochondrial DNA maintenance and translation apparatus only account for a fraction of cases. Exome sequencing in two sibs, born to consanguineous parents, with severe encephalomyopathy, choreoathetotic movements and combined respiratory chain defect allowed to identify a homozygous missense mutation in the PNPT1 gene that encodes the mitochondrial polynucleotide phosphorylase (PNPase, p.Gln387Arg). Protein structure modeling on E. coli PNPase showed that the p.Gln387 is located in the trimerization interfaces of the protein. BN-PAGE analysis showed that no PNPase complex could be detected in patient fibroblasts confirming that the mutation disrupts the trimerization of the protein. PNPase is essentially localized in the mitochondrial intermembrane space and is implicated in RNA targeting to human mitochondria. Mammalian mitochondria import several small noncoding nuclear RNAs, namely 5S rRNA, MRP RNA, some tRNAs and microRNAs. By RNA hybridization experiments, we observed a significant decrease of 5S rRNA and MRP-related RNA import into mitochondria in cultured fibroblasts of our patients. Finally, as 5S rRNA import was recently demonstrated as essential for mitochondrial translation, we studied mitochondrial protein synthesis and found a reproducible decrease of mitochondrial translation rate in patients' fibroblasts compared to controls. In conclusion, we report here the first example of abnormal RNA import into mitochondria as a cause of respiratory chain deficiency. Moreover, our results support the view that exome sequencing is a powerful tool for the identification of disease gene mutations in small pedigrees.

330

Comprehensive analysis of 101 nuclear genes for molecular diagnosis of mitochondrial disorders. *R. Bai[1], J. Higgs[1], S.F. Suchy[1], S. Buchholz[1], D. Arjona[1], B. Boggs[1], C. Chinault[1], N. Smaoui[1], S. Benhamed[1], J. Compton[1], Y. Shevchenko[1], G. Richard[1], S. Bale[1], F.D. Kendall[2], S. Parikh[3], A.L. Gropman[4], W. Chung[5], A. Goldstein[6], S.E. Palmer[7], J.A. Panzer[8], S.W. Yum[8], M.J. Falk[9]*. 1) GeneDx, Gaithersburg, MD; 2) Virtual Medical Practice, LLC, Atlanta, GA; 3) Neurogenetics/Neurometabolism, Neurosciences Institute, Cleveland Clinic, Cleveland, OH; 4) Department of Neurology, Children's National Medical Center, Washington, DC; 5) Columbia University Medical Center, New York, NY; 6) Division of Child Neurology, Children's Hospital of Pittsburgh, Pittsburgh, PA; 7) University of Oklahoma HSC Dept. of Pediatrics, Section of Genetics Oklahoma City, OK; 8) Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA; 9) Department of Pediatrics, The Children's Hospital of Philadelphia & Perelman School of Medicine at the University of Pennsylvania Philadelphia, PA.

The majority of primary mitochondrial disorders (MtD) are caused by mutations in nuclear genes. To date, only about 100 nuclear genes have reported mutations associated with a primary MtD, about 25–30% of these genes have 1–35 pseudogenes/homologous sequences in the genome. Whole exome sequencing (WES) uses hybridization-based capture as the targeted enrichment method making it unable to avoid pseudogenes/homologous sequences. This creates many false positive and false negative results as well as many low/no coverage regions. By instead using droplet-based multiplex PCR and massive parallel sequencing, we comprehensively analyzed a panel of 101 nuclear genes thus avoiding the above problems while providing a technical sensitivity and specificity comparable to Sanger sequencing. Furthermore we combined this test with array CGH of these 101 genes to detect exon-level deletions or duplications. This comprehensive nuclear gene panel encompasses over 95% of known nuclear genes/mutations associated with a primary MtD. Blood samples from 70 unrelated patients clinically suspected of a MtD were tested. Of those, 13 fit into a discrete clinical syndrome or a definite MtD, 11 had probable MtD, 36 had possible MtD and for 10 was no clinical information provided. All variants identified were compared with mutation databases including HGMD, LSDBs, GeneDx variant DB and other online resources. All novel variants were thoroughly evaluated and classified as novel mutations (nMut), variants of unknown significance [likely mutation (VLM), undetermined (VUS), or likely benign (VLB)], or as a benign polymorphism. Mutations published as “pathogenic” but without solid supporting evidence were re-evaluated and reclassified. Of the 70 samples tested, disease-causing mutation(s) were identified and confirmed the diagnosis for 14 patients (20%), 10 of whom had a discrete clinical syndrome or a definite/probable MtD and two of whom also harbored a heterozygous VLM/VUS in other genes. Reportable variants (Mut/VLM/VUS) in 1 to 3 genes were identified in another 30 samples (42%), including 17 (24%) with likely disease-associated mutations and 13 (18%) with only one Mut/VLM/VOUS insufficient for making a diagnosis. This comprehensive test panel, with a positive rate of over 40% for patients with a definitive/probable MtD or with a discrete clinical syndrome, is a superb panel to identify pathogenic mutations in genes known to be associated with a primary MtD before considering WES.

331

A Map of Human Genetic Variation: Update from the 1000 Genomes Project. *F. Yu, the 1000 Genomes Consortium.* Molecular and Human Genetics Dept, Baylor College of Medicine, Houston, TX, United States.

To identify and describe the geographic and functional distributions of human genetic variation, we have now sequenced the genomes of >1,600 individuals, using a combination of whole genome shotgun sequencing (average depth > 8x) and targeted exome-resequencing (average depth > 80x). In the previous phase of the project, we combined information on variants identified by multiple read-mapping and variant-calling algorithms to construct an integrated haplotype map of 38 million SNPs, 1.4 million indels and over 14 thousand large deletions across 1,092 individuals. Our new map includes better coverage of geographic areas, such as South Asia, that were poorly represented in our previous set of samples. Another improvement is that, for the current analyses, we are incorporating de novo assembly of sequencing reads into variant discovery algorithms, resulting in a more unbiased view of human genetic variation and enabling us to better explore non-SNP variation. Our preliminary results show that individuals from different populations have different profiles of rare and common variants and that low-frequency variants show elevated geographic differentiation, which is further increased by purifying selection. We find that biological pathways and genomic elements vary substantially in the load of rare functional variation and that any individual harbors hundreds of rare, non-coding variants, such as transcription-factor-motif disrupting changes, with weak deleterious consequences. The 1000 Genomes haplotype resource is estimated to include over 99% of variants at a frequency of 1% across the genome and 0.2% in the exome. It enables imputation of geographic ancestry and common and low-frequency variants in sequenced or genotyped individuals from diverse populations.

332

Towards a whole genome map of heritable copy number variation. *S. Aradhya, L. Matyakhina, D. Pineda Alvarez, D. Riethmaier, A. Fuller, G. Richard, J. Meck.* GeneDx, Gaithersburg, MD.

Whole-genome copy number analysis by array CGH has led to the identification of numerous microdeletion/duplication syndromes. Some of the recently discovered pathogenic copy number variants (CNVs) cause phenotypes with significant clinical variability, and some carriers are only mildly affected or remain unaffected. Studies of large control cohorts have also identified benign CNVs across the genome and these data are available in public databases. We have examined 12,350 human genomes by array CGH and identified 4,115 CNVs that were pathogenic or suspected to have clinical relevance. Deletions and duplications were in roughly equal proportion. A total of 1,397 CNVs were tested for inheritance; 37% were maternally inherited, 23% were paternally inherited, 29% were de novo, and the remainder was inconclusive because only a single parent was tested. In contrast to de novo CNVs, the inherited CNVs were mostly less than 1 Mb in length. The ratio of inherited CNVs of maternal origin to that of paternal origin was 1.6. Our data showed a high frequency of pathogenic events at 15q11.2 (34), 15q13.3 (19), 16p11.2 (85), 1q21.1 (107), in concordance with previously published studies on these CNVs and the associated clinically variable syndromes. We identified more than 500 novel inherited CNVs, which constitute a group of rare variants typically found in genomic regions that lack structural complexity and contrast with the smaller list of high-frequency polymorphisms that tend to cluster at segmental duplication repeat sequences. A survey of gene content at these rare inherited CNV loci provides insight into the types of genes that can exist in abnormal copy number without causing an adverse outcome. The inheritance information also informs the interpretation of the clinical significance of these CNVs. These data contribute to ongoing efforts to build a CNV map of the human genome.

333

Charting the population-scale landscape of short tandem repeat variation in humans. *M. Gymrek[1,2,3,4], J. Chen[1,2], C. O'Dushlaine[5,6], M. Daly[3,7], D. Reich[3,8], Y. Erlich[2].* 1) Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; 2) Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Department of Molecular Biology and Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 5) Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry and Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 6) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 7) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA; 8) Harvard Medical School Department of Genetics, Boston, Massachusetts, USA.

Short Tandem Repeats (STRs) have a wide range of applications, including medical genetics, forensics, and population genetics. High throughput sequencing has the potential to type hundreds of thousands of STR loci. However, these markers have not been routinely profiled due to the gapped alignment challenge they pose to mainstream aligners. Recently, we developed an algorithm, called lobSTR, for rapidly profiling STRs from high throughput sequencing datasets. We validated lobSTR by measuring the consistency of calls from whole genome sequencing of biological replicates, by tracing Mendelian inheritance patterns in whole genome sequencing of a HapMap trio, and by analyzing its consistency against traditional molecular techniques. Applying lobSTR to the genomic dataset of a single individual reveals more than 50,000 STR variations, including loss-of-function alleles and major deviations from the NCBI reference. We developed a likelihood-based framework for calling STR alleles in population-wide samples. Unlike previous SNP and STR studies using loci ascertained in Europeans, a genome-wide STR study avoids ascertainment bias since all STRs are assumed polymorphic. We applied lobSTR to approximately 3,000 whole exome, 50 high coverage whole genome, and 2,000 low coverage whole genome sequencing datasets to generate a baseline map of human STR polymorphism. We determine for the first time the allele size range and allele frequency spectrum of tens of thousands of genome-wide STRs. From this map we derive insights on genome-wide STR variation patterns including distinct differences in sex chromosomes vs. autosomes and sequence features affecting STR variability. We used this resource to analyze population-specific STR trends and to draw population genetic inferences. This initial landscape of STR variation will provide an invaluable resource for future medical and population genetic studies.

334

Whole-genome sequencing analysis of iPSC lines uncovers lineage-manifested CNVs. *A.E. Urban[1], A. Abyzov[2], D. Palejev[3], L. Rosenberg-Belmaker[3], Y. Zhang[4], J. Mariani[3], L. Tomasini[3], A. Ferrandino[3], A. Szekely[4], M. Wilson[3], M. Haney[1], E. Grigorenko[3], A. Huttner[5], S. Weissman[4], M. Gerstein[2], F. Vaccarino[3]*. 1) Psychiatry and Genetics, Stanford University, Palo Alto, CA; 2) Molecular Biophysics and Biochemistry, Yale University, New Haven, CA; 3) Child Study Center, Yale University, New Haven, CT; 4) Genetics, Yale University, New Haven, CT; 5) Pathology, Yale University, New Haven, CT.

We have performed whole-genome sequencing based CNV analysis in 7 fibroblast samples and 20 corresponding induced pluripotent stem cell lines obtained from two families (Abyzov et al., in revision). We found that on average an iPSC line has two LM-CNVs (lineage-manifested CNVs). We defined the term LM-CNV to describe CNVs detected by genome-wide analyses in an iPSC line but not in the fibroblast culture from which the given iPSC line was derived but without making a statement as to the nature of the CNV-forming event (i.e. whether the CNV arose de novo during reprogramming from fibroblast to iPSC or whether it was present as a somatic variant in mosaic fashion in the fibroblast culture). After detecting LM-CNVs by sequencing based analysis in the iPSC lines (Illumina HiSeq, read-depth analysis plus paired-end mapping wherever possible) we were able to design PCR primers that would cross the CNVs' breakpoints allowing us to investigate the masked, mosaic presence of the same CNVs in the fibroblast tissue of origin. We determined that more than half of the LM-CNVs detected in iPSC lines were actually present as low allele frequency, mosaic somatic CNVs in the fibroblasts which then became unmasked by the clonal selection during iPSC-line creation, and that more than 35% of fibroblast cells could be carrying such medium-sized to large somatic CNVs. We also carried out correlative analyses between the detected LM-CNVs and gene expression determined by RNA-Seq from the same iPSC lines. When analyzing expression levels (by RPKM) of genes intersecting LM-CNVs in iPSCs we found a clear tendency (p-value of 0.02 by Fischer's exact test) of increase in expression for genes in duplications and decrease in expression for genes in deletions, respectively. However there are also genes in deletion CNVs with increased expression and genes in duplication CNVs with decreased expression, an observation that invites further analysis on the molecular level. Our analyses revealed the presence of somatic copy number mosaicism in fibroblasts that carries over into iPSC-lines and becomes unmasked in the process, with a resulting effect on gene expression being detectable.

335

SNP markers identify areas with restricted recombination suggesting structural variation across the human genome is widespread. *P.G. Hysi[1], B. Tamraz[2], A. Nag[1], C. Venturini[1,3], J.S. Rahi[3], T.D. Spector[1], C.J. Hammond[1]*. 1) Twin Research, King's College London, St Thomas' Hospital, Westminster Bridge Road, London SE1 7EH, United Kingdom; 2) Institute for Human Genetics, University of California at San Francisco, CA 94143, USA; 3) MRC Centre of Epidemiology for Child Health, Institute of Child Health, University College London, Guilford Street, London, UK.

Background. Human chromosomal structural variation is common. Most data on these variation comes from small scale experiments or indirectly via the consequences of chromosomal arrangements over probe intensity or alterations in LD near break points. Chromosomal re-arrangements frequently change the recombination patterns between markers, but little is known about the frequency and extent of these events. We investigated the genome-wide frequency of regions with markers showing atypical recombinations in the autosomes of three European panels and one of African descent. Methods. We systematically looked for common (>5%) pairs of markers where observed non-recombination differed to the expected recombination and genetic distances. Genotyping was performed with different chips: Human Hap610Quad for 1738 TwinsUK unrelated subjects, Affymetrix 6.0 for 2989 WTCCC panel subjects and 1378 GAIN subjects and the Human 1M-duo for 1000 subjects from the 1958 British Birth Cohort (partly overlapping with the WTCCC). Allelic combinations over each pair of loci were compared with the null hypothesis of free recombination and recombination rates expected from HapMap families. We looked for contiguous segments of significantly suppressed recombination between markers and studied phylogenic trees over these regions. Results. We found about 4,000 regions consistently showing anomalies in recombination patterns between markers in the three cohorts of European descent and about 7,000 such regions in African-Americans, which largely included the same regions observed in Europeans. These regions often span over more than one recombination block and LD patterns did not explain these properties. They included most known and predicted human inversions, but correlated poorly with known CNVs. Phylogenies of reconstructed haplotypes revealed well-defined, non-communicating, bifurcating, branches. Conclusion and discussion. The frequency of non-recombining markers and haplotypes in the genome raises the possibility of widespread structural anomalies, such as inversions. This would cause alignment of non-homologous regions during meiosis and selective recombination suppression between markers transmitted along separate phylogenetic branches. These non-recombining regions might reflect a large and yet unidentified number of structural variants in the population, which may have implication over gene mapping efforts. This possibility will require further experimental validation.

336

Mapping the L1 interactome reveals RISC-associated helicase MOV10 as a potent inhibitor of retrotransposition. *J. Goodier, L. Cheung, H.H. Kazazian.* Johns Hopkins University School of Medicine, Baltimore, MD.

LINE-1 retrotransposons constitute one-sixth of human DNA and have helped shape our genome. Despite the immense significance of L1s for genome evolution and organization, much about their biology remains unknown, including cellular factors involved in the complex process of retrotransposition. By immunoprecipitation (IP) of tagged L1 constructs and MS sequencing, we identified 95 non-ribosomal proteins associated with L1 ribonucleoprotein (RNP) particles. These included helicases, RNA transport proteins, chaperone proteins, post-translational modification enzymes, and splicing factors. We cloned 70 of these proteins and confirmed that 50 directly co-IP with a tagged L1 from 293T cells. We also showed that a majority co-IP endogenous ORF1p from 2102Ep cells. One-quarter of L1 RNP-associated cellular proteins colocalize with tagged ORF1 protein in cytoplasmic granules, further confirming their association with the L1 RNP. We have also assayed the effects of these proteins on cell culture retrotransposition. Most proteins inhibited retrotransposition or left it little changed. These studies suggest candidate cofactors that interact with the L1 to modulate its activity, and increase our understanding of the means by which the cell coexists with these genomic “parasites”. One factor, the RNA helicase MOV10, is a component of the RNA-induced silencing complex (RISC), and has recently been shown to inhibit HIV replication. We now show that MOV10 also severely restricts human L1, Alu, and SVA retrotransposons in cell culture. MOV10 associates with the L1 RNP along with other RNA helicases, including DDX5, DHX9, DDX17, DDX21, and DDX39A. However, unlike MOV10, these other helicases do not strongly inhibit retrotransposition in multiple cell lines, an activity dependent upon intact MOV10 helicase domains. MOV10 association with retrotransposons is further supported by its close colocalization with L1 ORF1p in stress granules and P-bodies, cytoplasmic structures associated with RNA silencing, and by the ability of MOV10 to reduce endogenous and ectopic L1 expression. With homologs in other vertebrates, insects and plants, MOV10 may represent an ancient and innate form of immunity against both infective viruses and endogenous retroelements.

337

FoSTeS/MMBIR replicative repair mechanisms are error prone: high frequency of nucleotide variation at the breakpoint junctions. *C.M.B. Carvalho[1], M.B. Ramocki[2, 3], D. Pehlivan[1], P. Fang[1], L.M. Franco[1], J.W. Belmont[4], P.J. Hastings[1], J.R. Lupski[1, 2, 3].* 1) Dept Molecular Human Genetics, Baylor Col Medicine, Houston, TX; 2) Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX; 4) Division of Cardiology, Department of Pediatrics, Baylor College of Medicine, Houston, TX.

Human genome analysis revealed that *MECP2* duplication rearrangements are nonrecurrent and present complexities, e.g. the presence of triplication embedded in duplications, or stretches of duplicated segments interspersed with unaltered copy number, in 27% of cases. Here we studied genomic rearrangements in 31 patients by determining the DNA sequence for each of the rearrangement breakpoint junctions. Surprisingly, DNA sequencing revealed the presence of complexities in up to 52% of the rearrangements. The most striking observation, however, was the high frequency of rearrangement-breakpoint junctions with small insertions and deletions (indels) (32%) mediated by michrohomologies, frameshifts in homonucleotide runs and point mutations (13%) observed at/or flanking the breakpoint junctions, none of which are present in dbSNP or the original haplotype on which the rearrangement occurred, suggesting that they were generated concomitantly. These observations strongly support a role for replication-based mechanisms underlying such rearrangements and suggest that the involved polymerase is error prone. In summary, our data further document a role for a DNA replication mechanism in complex genomic rearrangements associated with genomic disorders and show that the mutation load of the FoSTeS/MMBIR mechanism generating nonrecurrent duplications may not be confined just to the breakpoint junctions of the CNV alteration, but also includes an increased SNV burden, which may have implications for phenotypic manifestations.

338

Telomere position effect in patients with subtelomeric deletions. *J. Gerfen, M.K. Rudd, H. Mason-Suares.* Human Genetics, Emory University, Atlanta, GA.

Chromosomal aberrations are a well-known cause of intellectual disability and congenital anomalies that may lead to altered gene expression within deletions and duplications. However, the mechanisms by which genomic copy number changes alter expression of nearby intact genes are less well studied in the context of human disease. One possible mechanism is telomere position effect (TPE), the silencing of genes due to the spreading of heterochromatin from the telomere. Though TPE is known to be a feature of yeast and fly chromosome ends, the effects on human chromosomes remain to be determined. Subtelomeric repeats that make up the terminal hundreds of kilobases (kb) of chromosome ends could insulate genes from telomeric heterochromatin spreading, preventing silencing. However, in patients with a subtelomeric deletion, the addition of a new telomere adjacent to euchromatin has the potential to alter the nearby chromatin environment and gene expression. We investigated TPE in cell lines from seven patients with subtelomeric deletions of five different chromosomes by chromatin immunoprecipitation with microarray (ChIP-chip) using antibodies to heterochromatic histone marks. In a cell line from a patient with a 5.5-Megabase terminal deletion of chromosome 4p, we identified a region of histone H3 lysine 9 tri-methylation (H3K9me3) starting from the breakpoint and extending ~500 kb that was absent from a control cell line. To determine if H3K9me3 enrichment resulted in altered transcription in cis with the deletion, we analyzed allele-specific expression of genes located within and beyond the ~500-kb H3K9me3 region by RNA-seq. The data from this terminal deletion cell line revealed skewed allelic expression of the only gene within the H3K9me3 enriched region, located ~150 kb from the breakpoint. The average expression across six polymorphisms in the 5' UTR showed 78% from one allele and 22% from the other. RNA-seq data from two cell lines without the deletion showed allelic ratios of 55%/45% and 57%/43%, consistent with equal biallelic expression. All three cell lines exhibited equal biallelic expression from a gene outside of the H3K9me3 enriched region, about 650 kb from the breakpoint. These data suggest that telomeric heterochromatin spreading can induce human gene silencing that may be related to disease pathogenesis. Gene expression and H3K9me3 enrichment studies of other terminal deletions to measure TPE on additional chromosomes are ongoing.

339

***De novo* CNV formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end joining.** *M.F. Arlt[1], S. Rajendran[1], S.R. Birkeland[2], K.M. McSweeney[1], T.E. Wilson[1,2], T.W. Glover[1].* 1) Dept. Human Genetics, University of Michigan, Ann Arbor, MI; 2) Dept. Pathology, University of Michigan, Ann Arbor, MI.

Copy number variants (CNVs) are found throughout the human genome, where they contribute to genetic variation and phenotypic diversity. Spontaneous CNV mutations are also a major cause of genetic and developmental disorders and arise frequently in cancer cells. A major class of CNVs, termed nonrecurrent CNVs, is thought to arise by nonhomologous DNA repair mechanisms due to the presence of short microhomologies, blunt ends, or short insertions at junctions, features recapitulated experimentally in human fibroblasts when CNVs are induced by exogenous replication stress. To test whether the nonhomologous end joining (NHEJ) pathway of double-strand break (DSB) repair is involved in the formation of this class of CNVs, chromosome integrity was monitored in NHEJ-deficient *Xrcc4*$^{-/-}$ mouse embryonic stem (ES) cells following treatment with low doses of the DNA polymerase inhibitor, aphidicolin. Xrcc4 is a component of the DNA ligase IV complex and is required for canonical NHEJ. As previously seen in human fibroblasts, mouse ES cells exhibited a significant amount of replication stress-induced CNV formation. In wild-type cells, *de novo* CNVs were found in untreated and APH-treated clones at a frequency of 0.43 and 5.19 CNVs per clone, respectively ($p<10^{-14}$). CNVs were mainly distributed throughout the genome with several hotspot regions, including some that are syntenic to those seen in human cells, such as in the *Auts2* and *Wwox* loci. The frequency, size, and location of spontaneous and aphidicolin-induced CNV formation were not altered by loss of Xrcc4, as would be expected if canonical NHEJ were the predominant pathway of CNV formation. Moreover, de novo CNV junctions displayed a typical pattern of microhomology and blunt end use that did not change in the absence of Xrcc4. A number of complex CNVs were detected in both wild-type and *Xrcc4*$^{-/-}$ cells, including an example of a catastrophic, chromothripsis event. These results establish that nonrecurrent CNVs can be, and frequently are, formed by mechanisms other than Xrcc4-dependent NHEJ, and implicate aberrant or collapsed replication forks rather than two-sided DSBs as a principal intermediate.

340

A genome-wide association study for cerebrospinal fluid tau and amyloid beta 42 identify new candidate variants implicated in Alzheimer's Disease. *J.S.K. Kauwe[1], C. Cruchaga[2], O. Harari[2], K. Mayo[2], S. Bertelsen[2], M. Bailey[1], D. McKean[1], P.G. Ridge[1], T.J. Maxwell[5], E. Peskind[3], D. Galasko[4], A.M. Goate[2], ADGC, ADNI, GERAD.* 1) Dept Biol, Brigham Young Univ, Provo, UT; 2) Dept Psychiatry, Washington University School of Medicine, St. Louis, MO; 3) Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, USA; 4) Department of Medicine, University of Washington, Seattle, WA, USA; 5) Human Genetics Center, UT Houston Health Science Center, Houston, Texas.

Cerebrospinal fluid (CSF) tau, tau phosphorylated at threonine 181 (ptau) and Aβ42 are established biomarkers for Alzheimer's Disease (AD). These biomarkers can be used as quantitative traits for genetic studies of AD. We performed the largest GWAS for CSF tau and Aβ42 levels published to date (n=1,338), identifying three new loci with genome-wide significant P-values for CSF tau and ptau: rs9877502 (P = 4.89 × 10-9 for CSF tau) located at 3q28 between the GEMC1 and OSTN genes, rs514716 (P = 1.07 × 10-8 and P = 3.22 × 10-9 for tau and ptau respectively), located on 9p24.2 within GLIS3 and rs6922617 (P = 3.58 × 10-8 for CSF ptau) at 6p21.1 close to the FOXP4 gene. Rs9877502 also showed a strong association with risk for disease (P = 9.19 × 10-4). Only SNPs in the APOE region showed genome-wide significant p-values with CSF Aβ42 levels, confirming previous reports. Analyses stratified by APOE genotypes show evidence of another associated locus in this region. These findings illustrate an endophenotype-based approach can be used to identify and characterize AD risk loci.

341

Analysis of Whole Transcriptome specific to the Temporal Pole of Late-Onset Alzheimer's disease. *C.E. Humphries[1,2], M.A. Kohli[1], P.W. Whitehead[1], W.F. Hulme[1], L. Nathanson[1], D.C. Mash[3], M.A. Pericak-Vance[1], J.R. Gilbert[1].* 1) Hussman Inst Human Genomics, Univ Miami, Miami, FL; 2) Dr. John T Macdonald Foundation, Department of Human Genetics, Univ Miami, FL; 3) University of Miami, Miller School of Medicine, Department of Neurology, Univ Miami, FL.

To investigate functional pathways involved in Late-Onset Alzheimer's disease (LOAD), RNA-Seq was performed on total RNA from autopsy confirmed human temporal pole samples using next-generation sequencing (NGS). NGS permits the identification and quantitation of known genes as well as novel isoforms, splicing events, coding RNAs, and non-coding RNAs. This allows a more complete analysis of differential patterns of gene transcription, and the elucidation of pathways and regulatory mechanisms specifically affected between LOAD and control brain. Neuropathological specimens were sampled from age, sex and race-matched temporal poles (BA 38) from 10 cases each of LOAD, Diffuse Lewy-Body disease and pathologically and clinically normal controls. RNA was extracted using Qiagen's miRNeasy kit and libraries were prepared with Epicentre's Script-Seq protocol. Samples were run at two per lane on Illumina's HiSeq2000, generating 40–65 million reads per library. Using the alignment program GSNAP, ~75% of the reads aligned to the reference genome. The Cufflinks program assembled transcripts and determined expression values. On average, we identified 24,000 unique transcripts per sample. Approximately 16,000 annotated reference sequence (RefSeq) genes were expressed and a total of 2,782 novel isoforms of RefSeq genes were discovered, with 48 of them belonging to genes associated with LOAD (e.g. MS4A4A, PICALM, etc). Over 270,000 novel transcripts were observed, most mapping to intronic regions of genes (59%), intergenic regions of annotated genes (12%), or to the antisense strand of known genes (7%). Results also identified approximately 19,000 anti-sense transcripts. Initial results have identified 47 novel antisense transcripts that overlap genes or regions (e.g. MS4A7, SORT1, PARP1, etc.) associated with LOAD. Thirteen antisense transcripts are only expressed in LOAD, of which 10 are involved in phosphorylation. Approximately 46% of the LOAD transcriptome consists of non-coding transcripts, suggesting considerable uncharacterized regulation. We also detected over 2,000 previously uncharacterized alternative splicing events, novel intergenic transcription, and antisense RNA transcripts to known LOAD candidate genes and/or regions. The elucidation of abnormal regulatory mechanisms and pathways in LOAD offers a powerful approach to identify and understand the role of individual genes and transcripts in the etiology of Late-Onset Alzheimer's disease.

342

Rare Variants from High-Density Exome Genotyping in Late-Onset Alzheimer's Disease (LOAD): Update from Alzheimer's Disease Genetics Consortium (ADGC). *L.-S. Wang[1], A.C. Naj[2], C. Cruchaga[3], S. Mukherjee[4], C.-F. Lin[1], O. Valladares[1], L.B. Cantwell[1], R. Graham[5], T. Behrens[5], P.K. Crane[6], A.M. Goate[3], M.A. Pericak-Vance[2], G.D. Schellenberg[1], Alzheimer's Disease Genetics Consortium.* 1) Dept Pathology and Laboratory Medicine, Univ Pennsylvania, Philadelphia, PA; 2) John P. Hussman Institute for Human Genomics, Univ Miami, Miami, FL; 3) Department of Psychiatry, Washington University School of Medicine, St. Louis, MO; 4) Div General Internal Medicine, University of Washington, Seattle, WA; 5) Genentech, Inc., San Francisco, CA; 6) Dept Medicine, University of Washington, Seattle, WA.

Two genome-wide association (GWA) studies last year increased the number of susceptibility genes for LOAD to 9. As in other complex diseases, common variants in these genes do not fully explain the known heritability of LOAD, suggesting that other genetic factors such as rare variants may explain some of the “missing” heritability. Exome chips use proven chip-based genotyping technology to genotype nearly 250,000 rare functional variants identified in whole-exome sequencing studies of more than 12,000 subjects, and represent a powerful tool for identifying rare functional variants involved in complex diseases like LOAD. The ADGC genotyped 7,091 samples from 29 NIA AD centers using the Illumina HumanExome BeadChip, which captured complete genotype data on 247,870 SNPs. We compared exome chip genotypes on 40801 markers from 57 samples with whole genome sequencing data and found that exome chips have excellent specificity and high sensitivity for minor allele detection in rare SNPs (92% for SNPs with minor allele frequency >0.1%). We applied several quality control steps including (a) removing mismatched samples using GWA SNP array data genotype concordance, (b), removing subjects and variants with low call-rate (>5% missingness), and (c) excluding monomorphic variants, after which we retained 5,801 subjects and 145,681 markers. We performed analyses with SKAT (Sequence Kernel Association Test), adjusting for age-at-onset (cases)/age-at-exam (controls), number of APOE ε4 alleles, and MDS principal components from GWA data to adjust for population substructure. A total of 46 variants were associated with LOAD with P<0.001, including variants in APOE ($P=4\times10^{-8}$), BIN1 (rs744373, $P=1.24\times10^{-4}$), and many previously unreported loci. We observed similar results using a traditional logistic regression approach in PLINK and LASSO penalized regression in MENDEL. We have identified preliminary evidence of several promising rare functional variant associations in both known LOAD genes (APOE, BIN1) as well as in several new candidates for which further validation is under way. These findings suggest that rare variants in the exome may contribute to LOAD risk. We will be able to further validate these associations when exome chip genotype data on an additional 9000 participants become available in summer. We will report on the replication analyses of these preliminary findings in this presentation.

343

Common variants in ABCA7 and GRIN3B, HMHA1 and SBNO2, are associated with late-onset Alzheimer's disease in African Americans. *C. Reitz[1], G. Jun[19], J. Buros[19], B. Vardarajan[19], L.-S. Wang[18], J.D. Buxbaum[16], E.B. Larson[5], N. Graff-Radford[13], D. Evans[10], N. Ertekin-Taner[13], M. Logue[19], C.T. Baldwin[19], R.C. Green[19,20,21], L.L. Barnes[12], L.B. Cantwell[18], M.D. Fallin[7], J. Manly[1], K.L. Lunetta[22], M.I. Kamboh[8], D.A. Bennett[11], K. Hall[4], A.M. Goate[17], G.S. Byrd[15], W.A. Kukull[2], T.M. Foroud[3], J.L. Haines[9], M.A. Pericak-Vance[14], L.A. Farrer[19,20,21,22,23,24], G. Schellenberg[18], R. Mayeux[1], ADGC Consortium.* 1) Taub Institute on Alzheimer's Disease and the Aging Brain and the Gertrude H. Sergievsky Center, Columbia University, New York, NY; 2) National Alzheimer's Coordinating Center and Department of Epidemiology, University of Washington, Seattle, Washington; 3) Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana; 4) Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana; 5) Group Health Research Institute, and Department of Medicine, University of Washington, Seattle, WA; 6) Departments of Biostatistics, Epidemiology, Medicine (Genetics Program), Neurology, and Ophthalmology, Boston University, Boston, Massachusetts; 7) Department of Epidemiology, Johns Hopkins University School of Public Health, Baltimore, Maryland; 8) University of Pittsburgh Alheimer's Disease Research Center and Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 9) Department of Molecular Physiology and Biophysics and Vanderbilit Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 10) Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois; 11) Rush Alzheimer's Disease Center and Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois; 12) Departments of Neurological Sciences and Behavioral Sciences, Rush University Medical Center, Chicago, Illinois; 13) Departments of Neuroscience and Neurology, Mayo Clinic, Jacksonville, Florida; 14) The John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida; 15) Department of Biology, North Carolina A & T University, Winston-Salem, North Carolina; 16) Departments of Neuroscience and Psychiatry, Mount Sinai School of Medicine, New York, New York; 17) Department of Psychiatry and Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University School of Medicine, St. Louis, Missouri; 18) Department of Pathology and Laboratory Medicine University of Pennsylvania School of Medicine; 19) Department of Medicine (Biomedical Genetics), Boston University Schools of Medicine and Public Health. Boston, MA; 20) Department of Neurology, Boston University, Boston, Massachusetts, USA; 21) Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, US; 22) Department of Biostatistics, Boston University, Boston, Massachusetts, USA; 23) Department of Epidemiology, Boston University, Boston, Massachusetts; 24) Departments of Ophthalmology, Boston University, Boston, Massachusetts.

Background. A series of large genome wide association studies identified several additional variants that affect disease susceptibility in Caucasians including CR1, CLU, PICALM, BIN1, CD2AP, CD33, EPHA1, MS4A6A/MS4E4 and ABCA7. However, a large part of the genetic contribution to late-onset Alzheimer's disease (LOAD) remains to be explained and it is unclear whether any of these recently identified loci modify LOAD risk in other ethnic groups. Objective. To identify genetic loci associated with LOAD in persons of African descent. Methods. The ADGC assembled a discovery dataset of 5,903 subjects (1,971 cases and 3,932 controls) from 16 independent sites that are part of the ADGC consortium. Associations of LOAD with directly genotyped and imputed SNPs were assessed using logistic regression models for case-control datasets and logistic generalized estimating equations (GEE) for family datasets as implemented in PLINK. Using the results of the association analyses of the individual datasets we then performed an inverse variance weighted meta-analysis using METAL. Results. The strongest association was observed for the APOE region (p= 3.54E-41). After additional adjustments for age, sex and APOE genotype, SNPs at four loci on chromosome 19 (ABCA7, GRIN3B, HMHA1 and SBNO2), which are not in LD, were associated with AD at p<=10-7. The strongest association with genome-wide significance was observed for a SNP in ABCA7 (p=2.2E-9), which is in LD with the two SNPs previously reported to be associated with AD in Caucasians. The recently in Caucasians identified variants in PICALM, BIN1, CLU, CR1, CD33, CD2AP, MS4A4/MS4A6E and EPHA1 were not associated. Conclusions. This study suggests that ABCA7, GRIN3B, HMHA1 and SBNO2 are associated with LOAD in African Americans.

344

Genome-wide Association Analyses of Onset Age in Late-Onset Alzheimer Disease (LOAD) Demonstrate No Strong Effect Outside of the APOE Region. *A.C. Naj[1], Y.S. Park[1], R. Rajbhandary[1], K.L. Hamilton[1], G.W. Beecham[1], E.R. Martin[1], R.P. Mayeux[2], J.L. Haines[3], L.A. Farrer[4], G.D. Schellenberg[5], M.A. Pericak-Vance[1], The Alzheimer's Disease Genetics Consortium.* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 4) School of Medicine, Boston University, Boston, MA; 5) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

LOAD risk loci may also contribute to variation in age of onset (AAO) of LOAD, as do the allelic variants in *APOE*. However, roles in AAO for the confirmed risk loci outside of *APOE* have not been explored. Daw et al. (2000) estimated 4 additional loci with effects as great or greater than *APOE* contribute to AAO variation. We examined variants at ten confirmed LOAD risk loci (*APOE*, *CLU*, *PICALM*, *CR1*, *BIN1*, *CD2AP*, *EPHA1*, *ARID5B*, the *MS4A* region, *ABCA7*, and *CD33*) to determine if they contribute to variation in AAO among 9,160 LOAD cases from the Alzheimer's disease Genetics Consortium (ADGC). Examining the variants most significantly associated with LOAD risk at each locus, we tested association with AAO using linear modeling assuming additive effects, adjusted for population substructure, and performed a random-effects meta-analysis across datasets. We also performed a genetic burden analysis using genotype scores weighted by effect sizes from association testing to examine the aggregate contribution of these loci to variation in the AAO phenotype. Preliminary analyses confirmed association of *APOE* regional variation with AAO (rs6857, $P=3.30\times10^{-96}$). Variants at several other LOAD risk loci also demonstrated statistically significant associations with AAO ($P<0.005$), including rs6701713 in *CR1* ($P=0.00717$), rs7561528 in *BIN1* ($P=0.00478$), rs561655 in *PICALM* ($P=0.00223$). *APOE* contributes to 3.1% of variation in AAO ($R^2=0.220$) whereas the other nine genes contribute to 1.1% of variation ($R^2=0.200$) over baseline ($R1^2=0.189$). We performed a GWAS to see if any other genomic variants with weak or no association to LOAD risk were associated with AAO, and found several interesting but not genome-wide significant associations in a chromosome 8q24.11 region including *RAD21* ($P=2.38\times10^{-6}$), chromosome 6q16.1 ($P=6.17\times10^{-6}$), *CIB4* ($P=7.70\times10^{-6}$), and *ABCC4* ($P=9.25\times10^{-6}$). In AAO analyses among LOAD cases, we confirmed that the association of *APOE* variants with AAO are by far the strongest in the genome, and in contrast to earlier hypothetical modeling, we show here that the combined effects of other loci do not exceed the effect of *APOE* on AAO. Associations with AAO at known risk loci *CR1*, *BIN1*, and *PICALM* were much more modest than *APOE*, with a combined contribution of these loci a third the contribution of *APOE* to AAO variation. If additional genetic contributions to AAO exist, they are individually very small or are hidden in gene-gene interactions.

345

Identification by exome analysis of the molecular bases of Familial Idiopathic Basal Ganglia Calcification not related to *SLC20A2* mutation.

G. Nicolas[1], C. Pottier[1,3], D. Maltête[1,2], S. Coutant[1,3], A. Rovelet-Lecrux[1,3], S. Legallic[1,3], Y. Vaschalde[4], L. Guyant-Maréchal[2,3], J. Augustin[2], O. Martinaud[2,3], L. Defebvre[5], P. Krystkowiak[6], J. Pariente[7], I. Le Ber[8], T. Frébourg[1], D. Hannequin[1,2,3], D. Campion[1,3,9]. 1) Inserm U1079, Faculty of Medicine, University of Rouen, Institute for Research and Innovation in Biomedicine (IRIB), Rouen, France; 2) Department of Neurology, Rouen University Hospital, Rouen, France; 3) CNR-MAJ, Rouen University Hospital, Rouen, France; 4) Department of Neurology, Le Havre Hospital, Montivilliers, France; 5) Department of Neurology A, Salengro University Hospital, Lille, France; 6) Department of Neurology, Amiens University Hospital, Amiens, France; 7) Department of Neurology, Purpan University Hospital, Toulouse, France; 8) CRCICM, IM2A, UMR-S975, AP-HP, Pitié-Salpêtrière University Hospital, Paris, France; 9) Department of Research, Rouvray Psychiatric Hospital, Sotteville-lès-Rouen, France.

Familial Idiopathic Basal Ganglia Calcification (IBGC), also known as Fahr's disease, is an autosomal dominant neuropsychiatric disorder with a high level of genetic heterogeneity. Loss of function mutations of *SLC20A2*, which encodes the inorganic phosphate (Pi) transporter Pit2, have been recently identified as a cause of IBGC but are encountered in only a subset of cases. In a three-generation family with no *SLC20A2* mutation, we performed whole exome sequencing in two affected first cousins, once removed. Using the EVA (Exome Variation Analyzer) software developed by our team, non-synonymous coding variants, splice acceptor and donor site variants, and frameshift coding indels (NS/SS/I) were filtered against dbSNP131, Hapmap, 1000 Genome project, and our in-house database including 72 exomes. Seventeen genes were affected by identical unknown NS/SS/I variations in the two patients. After screening the relatives, only one unknown mutation segregated with the disease in the family. This variation, which is predicted to be highly damaging, was present in 13/13 affected subjects and absent in 8 relatives without calcifications. Sequencing of the gene in 13 other unrelated IBGC cases allowed us to detect another potentially pathogenic variation also predicted to be highly damaging. None of the two mutations was present among the 5,379 control exomes of the ESP consortium. This gene encodes a protein involved in angiogenesis and in the regulation of Pi transport in vascular smooth muscle cells via Pit1, a Pi transporter encoded by *SLC20A1*. Our results further support the involvement of this biological pathway in IBGC pathophysiology.

346

Mutations in DNAJ cause autosomal dominant Parkinson disease in the Mennonite community.

C. Vilarino-Guell[1], A. Rajput[2], S. Appel-Cresswell[3], B. Shah[1], I. Yu[1], C. Thompson[1], C. Szu-Tu[1], J. Trinh[1], M. Encarnacion[1], D.W. Dickson[4], A.J. Stoessl[3], M.L. Rajput[2], M.J. Farrer[1], A.H. Rajput[2]. 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC V6T 2B5, Canada; 2) Department of Neurology, University of Saskatchewan and Saskatoon Health Region, Saskatoon, S7N 0W8 SK, Canada; 3) Pacific Parkinson's Research Centre, Division of Neurology, University of British Columbia, Vancouver, BC V6T 2B5, Canada; 4) Department of Neuroscience, Mayo Clinic Jacksonville, Florida, 32224, US.

Background: Mendelian forms of PD account for 10–15% of parkinsonism. Such families present a unique opportunity to identify pathologic mechanisms underlying disease. Genetic findings can immediately aid diagnosis, are crucial for the development of physiologically relevant models, and lead to novel therapeutic strategies.

Methods: DNA samples from 57 members of a Saskatchewan family were available for this study. Known genetic causes of parkinsonism were excluded. Exome sequencing was performed in three symptomatic patients with Agilent capture and Illumina sequencing. Sanger sequencing and haplotype analysis were used to validated the discovery of nucleotide variants and examine segregation with PD. Subsequent genotyping examined the frequency of any disease-segregating variant in a Canadian case-control series consisting of 1,750 PD patients and 1,531 healthy controls.

Results: PD with autosomal dominant inheritance was diagnosed in a Saskatchewan family of Mennonite descent. Three clinically affected family members had Lewy body disease at autopsy. Comparative exome sequencing analysis in three clinically affected family members identified fifteen novel variants that resulted in missense or nonsense amino acid substitutions. Genotyping of these variants in additional family members, patients with PD and healthy controls, excluded fourteen variants based on segregation analysis or frequency over 2% in control subjects. The only disease-segregating mutation, an asparagine to serine substitution, was located within a species-conserved protein domain of DNAJ. This mutation was not observed in controls but was identified in two sporadic PD patients and two families with multi-incident parkinsonism. All patients with a DNAJ mutation appear to share a common disease haplotype, indicative of a founder effect in the Mennonite community.

Conclusions: Exome sequencing analysis has identified a novel mutation in DNAJ predisposing to autosomal dominant PD. The mutation appears to be of special relevance to PD in the Mennonite community. The discovery highlights the utility of next-generation sequencing, especially exome analysis applied to multi-incident families, and further illuminates the genetic etiology of Lewy body PD. DNAJ appears to be a major player in an emerging molecular synthesis for PD.

347

***C9ORF72* repeat expansion is a risk factor for Parkinson Disease.** *K. Nuytemans[1,2], G. Bademci[1,2], M.M. Kohli[1,2], G. Beecham[1,2,3], V. Inchausti[1,2], A. Dressen[1,2], L. Wang[1,2,3], J.I. Young[1,2,3], F. Nahab[2,4], C. Singer[2,4], E.R. Martin[1,2,3], J.R. Gilbert[1,2,3], M. Benatar[4], J.L. Haines[5], W.K. Scott[1,2,3], S. Zuchner[1,2,3], M.A. Pericak-Vance[1,2,3], J.M. Vance[1,2,3]*. 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami Florida, USA; 2) Udall Center, University of Miami, Miami Florida, USA; 3) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami Florida, USA; 4) Department of Neurology, Miller School of Medicine, University of Miami, Miami Florida, USA; 5) Center for Human Genetics Research, Vanderbilt University Medical Center, 519 Light Hall, Nashville, Tennessee, USA.

INTRODUCTION: Recently, variants in the ribonuclease angiogenin gene (*ANG*) were reported to be significant risk factors for Parkinson disease (PD) and amyotrophic lateral sclerosis (ALS). Further, families with the hexanucleotide repeat expansion (>30 repeats) in the C9ORF72 gene were recently reported to have more occurrences of PD in family members than expected. This led us to test whether the *C9ORF72* repeat expansion found in ALS and Frontotemporal Dementia (FTD) is associated with PD as well. MATERIAL AND METHODS: The initial dataset was 396 unrelated PD cases, 12 Essential Tremor plus Parkinsonism (ETP) cases and 427 older controls (age at exam ≥60 years). Repeat size was measured using a repeat primed PCR assay. A second replication set consisted of 481 PD patients from the NINDS dataset and 717 additional older controls. RESULTS: Overall, 14 cases (13 PD, 1 ETP) and 3 controls had >20 repeats (Fisher's exact p=0.008, initial dataset ratio (9:1), replication dataset (5:2)). Seven cases and no controls had more than 23 repeats (p=0.05). Additionally, in the initial dataset we observed a significant rightward shift (increased numbers of *C9ORF72* repeats) in the overall repeat copy number distribution in cases compared to the controls (Kolmogorov-Smirnov (KS) test p=5.28e-07). While the distributions were not significantly different in the replication dataset, the distributions in the combined dataset (889 cases, 1144 controls) were still significantly different between cases and controls (KS=0.01). Interestingly, we identified ≥30 repeat copies in two samples (1 with ETP, 1 with PD). DISCUSSION: Our data suggest that intermediate copy numbers of the *C9ORF72* repeat contribute to the risk for PD and ETP. This also suggests that PD, ALS and FTD share similar pathophysiologic mechanisms of disease, specifically abnormalities of RNA metabolism. Further studies are needed to elucidate the contribution of the *C9ORF72* repeat in the overall PD population and if other common genetic risk factors exist between these neurodegenerative disorders.

348

Age-dependent penetrance of ALS+/−FTD due to *C9orf72* hexanucleotide intronic repeat expansion mutations. *B.N. Smith[1], S. Topp[1], J. Barnwell[1], A. Al-Chalabi[1], J. Kirby[2], P.J. Shaw[2], H. Pall[3], K.E. Morrison[3], V. de Jong[4], F. Bass[4], C.E. Shaw[1], C.M. Lewis[5]*. 1) Clinical Neuroscience, King's College London, London, United Kingdom; 2) Sheffield Institute for Translational Neuroscience, University of Sheffield, United Kingdom; 3) School of Clinical and Experimental Medicine, University of Birmingham, United Kingdom; 4) Departments of Neurology and Genome Analysis, Academic Medical Center, University of Amsterdam, Netherlands; 5) Medical and Molecular Genetics, King's College London, United Kingdom.

The most common genetic cause of familial and sporadic amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) is a massive GGGGCC hexanucleotide intronic repeat expansion mutation within *C9orf72*. This expansion mutation arose from a single common founder, and accounts for approximately 6% of sporadic ALS and 26% of familial ALS. The size of the expansion mutation as estimated by Southern blot is highly variable, and a role for genetic anticipation has been proposed. These factors make the counseling of affected families and predictive testing challenging. We estimated penetrance of the *C9orf72* expansion mutation in 121 predominantly British ALS+/−FTD probands found to carry the expansion using Southern Blots and PCR. Family history of ALS and FTD was determined using a sequential sampling scheme for first degree relatives (parents, siblings), and including parents' siblings if affected. Survival analysis was performed to estimate penetrance and to test for modifying factors. Additional ALS/FTD cases in kindreds were assumed to be expansion carriers; unaffected family members were weighted by their probability of being a carrier (0.5). In total, 128 affected relatives were identified (88.4% ALS, 11.6% FTD +/− ALS), and 282 unaffected relatives. The mean age of onset was 57.9 years for ALS cases and 63.6 years for FTD. Age-dependent penetrance rose steeply from 9% at 50 years to 56% by 70 years and 74% by 85 years (95% C.I. 63% - 82%). Penetrance was not influenced by sex of the proband, and was similar for parents and siblings of probands. Where parental carrier status was known, we tested for difference in proband's age at onset by sex of parental carrier; this provided no evidence of an imprinting effect. Evidence of genetic anticipation was tested in affected parent and offspring pairs. A significantly earlier age of onset was seen in the offspring compared to their parents with a mean difference of 6.9 years (p=0.0013). These data provide a framework for determining age-dependant risk for the *C9orf72* mutation expansion, which accounts for a substantial proportion of ALS/FTD cases, and will be of value in establishing genetic counselling for recurrence risk.

349

Empirical and theoretical studies on genetic variance of rare variants for complex traits using whole genome sequencing in the CHARGE Consortium. *C. Zhu[1], A. Morrison[2], J. Reid[3], C.J. O'Donnell[4], B. Psaty[5], L.A. Cupples[4,6], R. Gibbs[3], E. Boerwinkle[2,3], X. Liu[2].* 1) Department of Agronomy, Kansas State University, Manhattan, KS; 2) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) NHLBI Framingham Heart Study, Framingham, MA; 5) Cardiovascular Health Research Unit, University of Washington, Seattle, WA; 6) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

As the frontier of human genetic studies have shifted from genome-wide association studies (GWAS) towards whole exome and whole genome sequencing studies, we have witnessed an explosion of new DNA variants, especially rare variants. An important but not yet answered question is the contribution of rare variants to the heritabilities of complex traits, which determine, in part, the gain in power from rare variants to discover new disease-associated genes. Here we present theoretical and empirical results on this question.

Our theoretical study was based upon the distribution of allele frequencies incorporating mutation, random genetic drift, and the possibility of purifying selection against susceptibility mutations. It shows that in most cases rare variants only contribute a small proportion to the overall genetic variance of a trait, but under certain conditions they may explain as much as 50% of additive genetic variance when both susceptible alleles are under purifying selection and the rate of mutations compensating the susceptible alleles (i.e. repair rate) is high.

In our empirical study, we estimated the proportion of additive genetic variances (σ_g^2) of rare variants contributed to the total phenotypic variances of six complex traits (BMI, height, LDL-C, HDL-C, triglyceride and total cholesterol) using whole genome sequences (8x coverage) of 962 European Americans from the Charge-S study. The results show that the estimated σ_g^2 of rare variants (MAF≤1%) ranged from 2% to 8% across the six traits. However, the standard errors (s.e.) of the estimated variance components from rare variants are relatively large compared to those of common variants. Using HDL-C as an example, the estimated σ_g^2s are 0.08 (s.e. 0.10), 0.05 (s.e. 0.05) and 0.58 (s.e. 0.05) for rare, low-frequency (1%<MAF≤5%) and common (MAF>5%) variants, respectively.

350

Leveraging admixture analysis to resolve missing and cross-population heritability in GWAS. *N. Zaitlen[1], A. Gusev[1], B. Pasaniuc[1], G. Bhatia[2], S. Pollack[1], A. Tandon[3], E. Stahl[3], R. Do[4], B. Vilhjalmsson[1], E. Akylbekova[5], A. Cupples[6], M. Fornage[7], L. Kao[8], L. Lange[9], S. Musani[5], G. Papanicolaou[10], J. Rotter[11], I. Ruczinksi[12], D. Siscovick[13], X. Zhu[14], S. McCarroll[3], G. Lettre[15], J. Hirschhorn[16], N. Patterson[4], D. Reich[3], J. Wilson[5], S. Kathiresan[4], A. Price[1], CAC. CARe Analysis Core[5].* 1) Genetic Epidemiology, Harvard School of Public Health, Boston, MA; 2) Harvard-MIT Division of Health, Science and Technology; 3) Department of Genetics, Harvard Medical School, Boston, MA, USA; 4) Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA; 5) Jackson Heart Study, Jackson State University, Jackson, MS, USA; 6) Boston University, Boston, MA, USA; 7) Institute of Molecular Medicine and Division of Epidemiology School of Public Health, University of Texas Health Sciences Center at Houston, Houston, TX, 77030, USA; 8) Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, United States of America; 9) University of North Carolina, Chapel Hill, NC, USA; 10) National Heart, Lung, and Blood Institute (NHLBI), Division of Cardiovascular Sciences, NIH, Bethesda, MD 20892, USA; 11) Cedars-Sinai Medical Center, Medical Genetics Institute, Los Angeles, CA, USA; 12) Johns Hopkins University, Baltimore, Maryland, United States of America; 13) University of Washington, Seattle, WA, USA; 14) Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, USA; 15) Département de Médecine, Université de Montréal, C.P. 6128, succursale CentrePville, Montréal, Québec, Canada; 16) Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital Boston, Boston, MA, USA2.

Resolving missing heritability, the difference between phenotypic variance explained by associated SNPs and estimates of narrow-sense heritability (h2), will inform strategies for disease mapping and prediction of complex traits. Possible explanations for missing heritability include rare variants not captured by genotyping arrays, or biased estimates of h2 due to epistatic interactions [Zuk et al. 2012]. Here, we develop a novel approach to estimating h2 based on sharing of local ancestry segments between pairs of unrelated individuals in an admixed population. Unlike recent approaches for estimating the heritability explained by genotyped markers (h2g) [Yang et al. 2010], our approach captures the total h2, because local ancestry estimated from genotyping array data captures the effects of all variants—not just those on the array. Our approach uses only unrelated individuals, and is thus not susceptible to biases caused by epistatic interactions or shared environment that can confound genealogy-based estimates of h2. Theory and simulations show that the variance explained by local ancestry ($h2\gamma$) is related to h2, Fst, and genome-wide ancestry proportion (θ): $h2\gamma = h2*2*Fst*\theta*(1-\theta)$. Thus, we can estimate $h2\gamma$ and then infer h2 from $h2\gamma$. We apply our method to 5,040 African Americans from the CARe cohort and estimate the autosomal h2 for HDL cholesterol (0.39±0.11), LDL cholesterol (0.18±0.09), and height (0.55±0.13). As expected these h2 estimates were higher than estimates of h2g from the same data using standard approaches: 0.22±0.07, 0.16±0.07 and 0.31±0.07, consistent with previous estimates. The difference between h2 and h2g suggests that rare variants contribute substantial missing heritability that can be quantified using local ancestry information. Larger sample sizes will sizes will enable h2 estimates with even lower standard errors, so that the possible contribution of epistasis to previous estimates of h2 can be precisely quantified. We additionally use local ancestry to estimate the fraction of phenotypic variance shared between European and African genomes that is explained by genotyped markers, by estimating h2g in European segments, h2g in African segments, and h2g shared between European and African segments. Given that most GWAS to date have been carried out in individuals of European descent, these estimates shed light on the importance of collecting data from non-European populations for mapping disease in those populations.

351

Applying a quantitative genetics test of evolutionary neutrality to finger ridge-count, a classical model trait in humans. *S. Medland[1], PM. Visscher[2,3], GW. Montgomery[5], DM. Evans[6], NG. Martin[4]*. 1) Quantitative Genetics, Queensland Institute Medical Ressearch, Brisbane, Australia; 2) Diamantina Institute, University of Queensland, Brisbane, Australia; 3) Queensland Brain Institute, University of Queensland, Brisbane, Australia; 4) Genetic Epidemology, Queensland Institute Medical Ressearch, Brisbane, Australia; 5) Molecular Epidemology, Queensland Institute Medical Ressearch, Brisbane, Australia; 6) MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine University of Bristol,Bristol,United Kingdom.

Genome-wide association analyses of height and other complex traits and diseases, have now identified hundreds of loci of small effects and it has been estimated that ~30–50% of additive genetic variation can be explained by variants represented on standard GWAS genotyping platforms. The remaining `missing' heritability is most likely due to imperfect linkage disequilibrium between causal variants and genotyped SNPs, which will occur if causal variants are at low frequency due to natural selection or recent mutation. Evolutionary theory predicts that a trait that is not correlated with fitness will have high narrow sense heritability (mutation-drift balance) and that nearly all additive genetic variance will be due to common causal variants. Here we test this prediction for a model trait which is not known to be subject to natural selection. Fingerprint ridge-count (TRC) is an extremely heritable (h^2=97%) classically polygenic trait which develops around 12–13th week of gestation with the regression of the volar pads. Using GCTA analyses of 1159 unrelated individuals from the Queensland Twin Registry (QTwin) we estimated that common variants explain 80% of the variation in TRC and 73% of the variation in intensity count (TIC), a highly correlated proxy measure (r=.94, h^2=89%). We replicated this using data from 683 unrelated individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC), finding that 87% of the variation in TIC could be explained by common variants, with a joint analysis of the QTwin and ALSPAC TIC data yielding an estimate of 74%. In an attempt to locate some of these variants we conducted GWAS analyses for TRC using a discovery sample of 2,910 QTwin participants yielded two genome-wide significant and eight additional suggestive loci ($p<1\times10^{-6}$). Both the significant loci and one of the suggestive loci replicated in an independent sample of 996 QTwin participants and meta-analysis of the two samples yielded strongly significant signals at these loci (6q21: $p=6.4\times10^{-11}$, %Variance (replication sample) = 1.17; 8q23.1: $p=2.0\times10^{-8}$, %V=.66; 13q12.11: $p=4.1\times10^{-10}$, %V=1.67). Considered together, these results suggest that TRC is indeed evolutionary neutral, that the resemblance between relatives is due to many common causal variants of small effect that act additively, and that cumulatively rare variants account for little of the heritability.

352

Does common variation contribute to the shared genetic basis for schizophrenia and autism? *P.H. Lee[1,2,3], S. Ripke[1,3], S. Santangelo[1,2,4], M. Daly[1,3,5], The Psychiatric GWAS Consortium - Schizophrenia & Autism Working Group.* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 2) Department of Psychiatry, Harvard Medical School, Boston, MA, USA; 3) Stanley Center for Psychiatric Research, the Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; 5) Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA.

Background: Autism spectrum disorder (ASD) and schizophrenia (SCZ) are pervasive neurodevelopmental illnesses with high heritability. The two disorders are distinct in terms of clinical presentation, age of onset, and developmental course. Nevertheless, recent advances in neuroimaging and epidemiologic studies suggest that ASD and SCZ may share a common neurobiological basis that involves social/cognitive deficits as central features. Furthermore, recent genetic studies have established an important role of rare and/or de novo copy number variations (CNVs) in both illnesses. **Methods**: We investigated whether common genetic variation might contribute to the shared genetic etiology of ASD and SCZ. By utilizing the GWAS data of five psychiatric disorders, we looked for common SNPs that confer disease susceptibility to ASD and SCZ but not to bipolar disorder, major depression, and attention deficit hyperactivity disorder. Multinomial logistic regression-based modeling was carried out on a GWAS data set of 61,220 individuals from the Psychiatric GWAS Consortium. Fixed-effects-based meta-analysis was done to estimate the combined effect size of each SNP. **Results**: The primary modeling analysis identified dozens of risk genes that confer ASD-SCZ-specific susceptibility. Of note, two well replicated schizophrenia risk variants in *TCF4* and *mir137* were predicted to carry disease susceptibility to both ASD and SCZ. Meta analysis results for the SNPs were OR=1.351, *p*=4.268e-10 (rs17512836/*TCF4*) and OR=1.125, *p*=4.43e-08 (rs1625579/*mir137*). Another gene of interest was *NOS1*, which has been shown to be involved in the developing human cerebral cortex that forms the basis of language and decision-making, and interact with *DISC1*, another popular candidate gene for SCZ and ASD. **Discussion**: This combined GWAS analysis is the first to show robust associations of common genetic variants across ASD and SCZ. These findings indicate that the genetic etiology shared by ASD and SCZ involves not only rare CNVs and point mutations, but also common variants. We plan to perform bioinformatics analyses to infer within which biological processes/networks these ASD-SCZ risk variants participate. This work also demonstrates new analytic strategies for conducting cross-disorder analyses of related disorders, and thus will inform ongoing efforts to construct an etiologically-based classification of these and possibly other disorders.

353

Ultrafast genome-wide interaction scan on case-control data implicates epistatic calcium channels in Bipolar Disorder. *S. Prabhu, I. Peer.* Computer Sci, Columbia Univ, New York, NY.

Long range gene-gene interactions are biologically compelling models for disease genetics which can provide insights on relevant mechanisms and pathways. Despite considerable effort, rigorous interaction mapping in humans has remained prohibitively difficult due to computational and statistical limitations. We introduce a novel algorithmic approach to find long-range interactions in common diseases using a standard two-locus test which contrasts the linkage disequilibrium between SNPs in cases and controls. Our ultrafast method overcomes the computational burden of a genome-xgenome scan by employing a novel randomization technique that requires 10X to 100X fewer tests than a brute-force approach. By sampling small groups of cases and highlighting combinations of alleles carried by all individuals in the group, this algorithm drastically trims the universe of combinations while simultaneously guaranteeing that all statistically significant pairs are reported. Our implementation can comprehensively scan large datasets (2K cases, 3K controls, 500K SNPs) to find all candidate pairwise interactions (p<1e-12, LD-contrast) in a few hours - a task that typically took days or weeks to complete by methods running on equivalent desktop computers. We applied our method to the WTCCC bipolar disorder data and found a significant interaction between SNPs located within genes encoding two calcium channel subunits: RYR2 on chr1q43 and CACNA2D4 on chr12p13 (p=4.6e-14, LD-contrast test). We replicated this pattern of inter-chromosomal LD between the genes in a separate bipolar dataset from the GAIN project, demonstrating an example of gene-gene interaction that plays a role in the largely uncharted genetic landscape of bipolar disorder.

354

Computational challenges in the analysis of low coverage sequence data in thousands of individuals. *Y. Luo, L. Jostins, C.A. Anderson, J.C. Barrett, UK10K, UKIBDGC.* Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Next-generation sequence data is beginning to be generated for the analysis of complex diseases, but it is becoming clear that, much like GWAS, very large sample sizes will be necessary to uncover low frequency risk variants. While costs are dropping rapidly, it is not yet feasible to generate complete high depth genomes in thousands of individuals. Two attractive, lower-cost alternatives are exome sequencing (focused on just the coding regions of the genome) and low coverage sequencing, where the entire genome is sparsely covered with sequence reads, and statistical imputation is used to fill in the gaps. While methods for analyzing exome sequence data are relatively mature, significant challenges remain in the analysis of low coverage sequence data, in particular when scaling up to 1000s of samples.

In this work, we evaluate the performance of existing methods for raw sites calling and haplotype phasing using 1,825 Crohn's disease (CD) cases from the UKIBDGC that are sequenced at an average depth of 3.8X and 2,432 controls from the UK10K project that are sequenced at an average depth of 7.1X. The sheer size of these data presents a substantial challenge to produce accurate genotypes and association statistics in a computationally tractable manner. We therefore systematically compare various genotype calling methods with respect to their accuracy and computational efficiency. To assess the genotypic accuracy, we estimate dosage r^2 between the inferred allele dosage from the sequence and the genotypes from GWAS chips typed on the same individuals.

Pipelines from large projects such as the 1000 Genomes and UK10K currently use multiple imputation engines (e.g. BEAGLE, IMPUTE2, THUNDER) to improve the accuracy of genotype calls from low coverage data. IMPUTE2 provides the highest accuracy calls, but at the cost of substantially larger computational burden. We show that calls improved by BEAGLE alone with an increased number of iterations still provide accurate genotypes at a fraction of the computational cost. Balancing this accuracy-compute trade off will be a crucial design decision for large scale low coverage sequencing studies.

355

Sparse sequencing of 6,000 cases and 6,000 controls from Chinese women for genome-wide association study of major depression. *X. Gan, R. Mott, J. Flint, CONVERGE consortium.* The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom.

Major depression (MD) is the commonest psychiatric disease. Within the next ten years it is predicted to become the second ranking cause of morbidity in the world, an honour it already possesses in developed countries, behind ischaemic heart disease[1]. MD is associated with considerable morbidity, mortality and substantial economic costs. Across the world, lifetime prevalence estimates of vary, from 3% in Japan to 16.9% in the US, but in all countries the disorder is common, with a frequency typically varying from 8% to 12%.

Unfortunately, attempts to find the molecular variants that contribute to this heritability have, so far, had limited success. Genetic analysis of MD was recently recognized to be among the greatest challenges facing mental health researchers. Despite considerable effort, the field lacks robust replicated molecular findings.

In this project, we are performing genome-wide association study of major depression using 6,000 cases and 6,000 controls from Chinese women. These genomes are being sequenced at about 1-fold coverage using Illumina short reads. This is the largest sequencing study being carried out at the world's largest sequencing Centre BGI in Shenzhen, South China.

Capitalizing on the low coverage (4-fold coverage) sequence data from Chinese population in the 1000 genome project[2], we first create a haplotype map tailed to the Chinese population by pooling sparse sequences, and then imputing near complete sequence for each individual. Sequencing 10 individuals at high coverage shows that the error rate for the detected variation can be kept very low. Using the resulting imputed sequence variants, it is possible to test association between multiple measures and actual causal variants.

We present the haplotype identification and sequence imputation pipeline we used, and initial results on the association of major depression to sequence variants.

1. Murray, C. J. L. & Lopez, A. D. (Harvard School of Public Health on behalf of the World Health Organization and the World Bank (Global Burden of Disease and Injury Series, Vol. I). Cambridge, MA,, 1996)

2. The 1000 Genomes Project Consortium (2010). A Map of Human Genome Variation from Population-Scale Sequencing. Nature 467, 1061–1073.

356

Deep targeted sequencing of 12 breast cancer loci in 4,700 women across four different ethnicities. *P. Kraft[1], S. Lindstrom[1], B. Chapman[2], G. Chen[3], C. Chen[1], O. Hofman[2], D. Mirel[4], C. Haiman[3].* 1) Dept Epidemiology, Harvard Sch Pub Hlth, Boston, MA; 2) Dept Biostatistics, Harvard Sch Pub Hlth, Boston, MA; 3) Dept of Preventive Medicine, Univ So Calif, Los Angeles, CA; 4) Prog in Medical and Population Genetics, Broad Institute, Boston, MA.

Genome-wide association studies (GWAS) have identified more than two dozen genetic loci associated with breast cancer risk. However, the underlying genetic structure in these regions is not fully understood and it is likely that the index GWAS signal originates from one or more as yet unidentified causal variants within the region. We used next-generation sequencing to characterize 12 GWAS-discovered breast cancer loci in a total of 2,313 breast cancer cases and 2,353 controls across four ethnic populations (937 women of African American ancestry, 1,260 women of Japanese ancestry, 910 women of Latino ancestry and 1,559 women of European ancestry). Our primary aims were to identify sets of putative causal low-frequency and common alleles and assess whether these regions are enriched for rare variants that are exclusive to cases (or controls). Region boundaries were defined by the nearest recombination hotspot downstream and upstream from the original GWAS signal. Region intervals spanned between 46 kilobases (kb) and 973 kb. In total we hybrid-captured and sequenced 5,500 kb. On average, we were able to capture 82% of the non-repetitive sequence in the targeted regions, and the average fraction of captured bases sequenced with a depth >20x was more than 98%. Single Nucleotide Variant (SNV) genotypes were called using the GATK pipeline and were over 99.5% concordant with GWAS SNP data. Data from the Illumina HumanExome Beadchip is also available on a subset of these samples, allowing us to present concordance rates for rare non-synonymous variants as well. We will present association results across ethnicities for rare as well as common genetic variants with emphasis on functional variants such as non-synonymous, regulating and splicing variants. We will also discuss practical issues in targeted sequencing and the importance of thorough quality control procedures.

357

Population stratification of human disease-associated SNPs, and their relevance to human disease networks[1]. *S.M. Raj[1], G.E. Hoffman[2], A.G. Clark[1,2].* 1) Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA; 2) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, USA.

Many disease-associated SNPs show differentiation among human populations worldwide[2]. It is not known, however, whether stratification of disease-associated SNPs show similar network associations as diseases of related etiologies. Here we apply Principal Components Analysis (PCA) to capture population variation in individuals of European descent from the publicly available 1000 Genomes data. We collected data on disease associations from human genetic studies conducted only in European individuals, including 444 clinical traits compiled in the HuGe database[3]. For each trait-associated SNP, flanking exonic SNPs having $r^2 > 0.8$ with the candidate SNP were deemed to be associated with the trait. We measured the degree to which a trait showed population stratification by correlating these trait associated SNPs with the first ten principal components. We then used Maximal Information Based Nonparametric Exploration (MINE)[4] to construct a network of disease relationships based on the frequency of each SNP allele and the correlation of each set of disease-associated SNPs with the principal components. To see if correlated diseases based on stratification of sets of SNPs are in the same phenotype-level network, we compared our stratification-based disease map to the network map[1]. Our results indicate that patterns of differentiation of SNP allele frequencies differed among traits, and these patterns are correlated with disease phenotype-based networks[1]. Population stratification of disease-associated SNPs within European populations suggests that these disparities may also occur among global populations. The extent to which this network is recapitulated in non-European populations will shed light on the heterogeneity of disease etiology.

1.CA Hidalgo et al. A Dynamic Network Approach for the Study of Human Phenotypes. PLoS Comput Biol 2009 2.R Chen et al. Type 2 Diabetes Risk Alleles Demonstrate Extreme Directional Differentiation among Human Populations, Compared to Other Diseases. PLoS Genetics 2012 3.W Yu et al. A Navigator for Human Genome Epidemiology. Nat Genet 2008 4.D Reshef et al. Detectng Novel Associations in Large Data Sets. Science 2011.

358

Loss of function mutations in known human disease genes in 572 exomes. *J. Johnston[1], K. Lewis[1], D. Ng[1], S. Gonsalves[1], J. Mullikin[2,3], L. Biesecker[1,2]*. 1) Genetic Disease research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Genome and exome sequencing costs continue to fall and many individuals are undergoing these assessments as research participants and patients. The issue of secondary findings in exome analysis is controversial and data are needed on their meaning in otherwise healthy individuals. The genetics literature includes a substantial fraction of papers describing causative variants based on minimal data. The default is to assume such a variant, especially if it falls in a gene with a function related to the patient's phenotype, is causative. To better understand the frequency of potentially causative mutations in apparently healthy persons, we have analyzed potential loss of function mutations including stop, frameshift and splice site alterations in 572 subjects of the ClinSeq™ project. A total of 9,421 potential loss of function variants were identified. As we were interested in clinical significance, further analyses were restricted to variants predicted to alter the protein-coding region of the transcripts annotated in the Human Gene Mutation Database (HGMD). This reduced our variant list to 835. Forty-five variants predicted to cause cancer susceptibility, lipid disorders, or cardiomyopathy/channelopathy phenotypes have been analyzed separately and previously reported and were removed from our dataset. Final filters included frequency, quality and predicted inheritance. As our goal was to understand the impact of these variants on health, we filtered for dominant or x-linked conditions. This filtering reduced our variant list to 54 variants. Eight of these variants were present in HGMD and the remaining 46 were novel. Five variants were present in multiple individuals. Phenotypes predicted to result from the identified variants included intellectual disability, developmental disorders including OFD1 and CHARGE syndrome, neuromuscular disorders including Becker muscular dystrophy, polycystic kidney disease, deafness and cataract. Proband and family history suggested a small minority of these are truly pathogenic. We will summarize the further proband and family investigations undertaken to evaluate the associated phenotypes. We recognize that an approximately 10% incidence of such predicted mutations in an otherwise healthy adult cohort is unreasonable. These data highlight challenges associated with interpreting apparently pathogenic null variants in exome and genome sequencing.

359

The Problem of Multiple Plausible Molecular Diagnoses in Next Generation Sequencing Data: The NIH Undiagnosed Diseases Program Experience. *D. Adams[1,2], C. Boerkoel[2], K. Fuentes Fajardo[2], P. Cherukuri[2], M. Sincan[2], C. Toro[2], C. Tifft[1,2], W. Gahl[1,2], T. Markello[1,2]*. 1) NHGRI, NIH, Bethesda, MD; 2) Undiagnosed Diseases Program, NIH, Bethesda, MD.

Introduction: Exome sequencing (ES) is a powerful diagnostic tool that is now available for clinical use. Methods and standards for interpreting the resulting data have yet to be established. Substantial practical challenges arise from clinical application of ES including genetic counseling, proof of disease association and secondary variants. The NIH Undiagnosed Diseases Program evaluates patients with complex medical syndromes. The use of ES for selected participants has revealed an additional characteristic of ES data—the presence of multiple likely-pathogenic variants in single individuals.

Methods: 380 exome sequences have been obtained in 80 families. The resulting variants are subjected to an extensive set of filters including population frequency, segregation consistency, consistent high-quality genotypability, alignment correctness, and predicted pathogenicity. Promising variants are Sanger validated then assessed utilizing data from extensive clinical phenotyping. The resulting “high quality” variants are referred for further experimental validation.

Results: To date, 54 families have generated high quality variants for further study. Of those, 39 have multiple high-quality variants. In an increasing number of individuals, we are finding multiple DNA mutations that are either known to be pathogenic or demonstrate severe pathogenic potential at the in vitro, cell biological and/or model organism level. In one example, two sibs were affected by an autosomal recessive disorder. In addition, each sib had one new-dominant mutation in a different gene; each sib had one shared, and one unshared, inherited condition.

Discussion: Our patient cohort comprises undiagnosed patients who have undergone extensive prior evaluation. Our early experience with genome and exome sequencing suggests that some patients will be affected by multiple interacting disorders, rather than by a single condition. An alternative hypothesis is that one of the detected variants explains the entire syndrome, and other environmental or inherited disease modifiers have complicated prior efforts at diagnosis. In either case, our data highlights the fact that the ES may reveal increasingly complicated relationships between DNA variation and medical disease.

360

Exome sequencing to identify the cause of Mendelian Diseases. *J. Lupski[1,2,3], C. Gonzaga-Jauregui[1], W. Wiszniewski[1], D. Pehlivan[1], E. Karaca[1], A. Stray-Pedersen[1], S. Jhangiani[4], J. Reid[4], D. Muzny[4], R.A. Gibbs[1,4], Baylor-Hopkins Center for Mendelian Genomics.* 1) Mol & Human Gen, Baylor Col Med, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

The diagnosis of rare mendelian diseases is being revolutionized by the application of high-throughput, massively parallel, next-generation genomic sequencing technologies. Targeted exome sequencing provides a rapid, inexpensive and genome-wide approach to the identification of highly-penetrant and potentially disease-causing mutations in a given individual. Application of exome sequencing to identify the genetic cause of mendelian diseases has proven to be successful for a variety of diseases, however many more rare or uncharacterized genetic diseases remain to be solved. In addition, the identification of genes involved in diseases with very specific phenotypes not only provides an answer to what mutations cause the disease, but provides clues and relevant information on the biological function of the product of the gene affected and the pathways it may be involved in. We have applied deep targetted exome sequencing in approximately 150 affected patients of different families with more than 30 mendelian genetic diseases of unknown or heterogeneous molecular cause or previously uncharacterized syndromes. Through extensive bioinformatic analyses, we have started to identify novel deleterious mutations in candidate genes involved in a variety of pathways such as cell cycle progression, neuronal migration, neurite outgrowth and differentiation, among others. We have confirmed these mutations in the affected patients and confirmed segregation of potential disease causing mutations in the family. Further functional studies await in some cases in order to assess their disease-causing potential and elucidate how these mutations are affecting the function of the protein products of these genes in these patients, the role they play in pathways or networks, and how perturbations result in the disease state. Exome sequencing has started to be applied in the clinical setting for molecular diagnosis, however much work remains to be done on a research basis studying rare mendelian diseases in order to understand and provide a better framework for the routine application of genomic sequencing technologies for personal genomes in medical genomics and routine healthcare.

361

Domain specific-mutations in CDKN1C cause two disorders with opposing phenotypes: The undergrowth disorder IMAGe syndrome or the overgrowth disorder Beckwith-Wiedemann Syndrome. *V. Arboleda[1], H. Lee[1,2], R. Parnaik[4], A. Fleming[1], A. Banerjee[1], B. Ferraz-de-Souza[5], E. Delot[3], I. A. Rodriguez-Fernandez[1], D. Braslavsky[6], I. Bergadá[6], E. C. Dell'Angelica[1], S. F. Nelson[1,2], J. A. Martinez-Agosto[1,3], J. C. Achermann[4], E. Vilain[1,3,7]*. 1) Dept Human Gen, Univ California, Los Angeles, Los Angeles, CA; 2) Dept Pathology and laboratory Medicine, Univ California, Los Angeles,; 3) Dept of Pediatrics, Univ California, Los Angeles,; 4) Developmental Endocrinology Research Group, Clinical and Molecular Genetics Unit, University College London Institute of Child Health, London, UK; 5) Department of Endocrinology, LIM-18, University of Sao Paulo School of Medicine, Sao Paulo, Brazil; 6) Division of Endocrinology, Hospital de Niños Ricardo Gutierrez, Buenos Aires, Argentina; 7) Department of Urology, David Geffen School of Medicine, University of California, Los Angeles, California.

IMAGe Syndrome (Intrauterine growth restriction, Metaphyseal dysplasia, Adrenal hypoplasia congenita, and Genital anomalies) is an undergrowth developmental disorder with life-threatening consequences. Identity-by-descent analysis in a family with IMAGe syndrome identified a 17.2 megabase (Mb) locus on 11p15 that segregated in affected family members. Targeted exon array capture of the disease locus, followed by high-throughput genomic sequencing and validation by dideoxysequencing, identified missense mutations clustered in the PCNA binding domain of in the imprinted gene CDKN1C (P57KIP2) in two familial and four unrelated patients. No mutations in this domain were identified in 23 patients with isolated adrenal hypoplasia. Familial analysis demonstrated an imprinted mode of inheritance where only maternal transmission of the mutation resulted in IMAGe syndrome. CDKN1C inhibits cell-cycle progression and targeted expression of IMAGe-associated CDKN1C mutations in Drosophila caused severe eye growth defects compared to wild type CDKN1C, suggesting a gain-of-function mechanism. Furthermore, IMAGe- associated mutations resulted in loss of PCNA binding. We have identified the gene CDKN1C to be the genetic cause of IMAGe syndrome. In contrast, loss-of-function mutations in the N-terminal cyclin dependent kinase domain of CDKN1C has been shown to result in an opposite syndrome, Beckwith-Wiedemann Syndrome, an overgrowth syndrome with adrenomegaly. Domain-specific mutations within the same gene (CDKN1C) can therefore lead to opposing phenotypic features.

362

SCID Newborn Screening and Exome Sequencing Identifies Ataxia Telangiectasia and Low T Cells Early in Life. *J.M. Mallott[1], A. Kwan[1], J. Church[2], D. Gonzalez[1], S. Rana[3], U. Sunderam[3], R. Srinivasan[3], S.E. Brenner[4], L.F. Tang[5], F. Lorey[6], J. Puck[1,2]*. 1) Department of Pediatrics, University of California San Francisco, San Francisco, CA; 2) Department of Pediatrics, University of Southern California, Los Angeles, CA; 3) Tata Consulting Services, Hyderabad, India; 4) Center for Computational Biology, University of California Berkeley, Berkeley, CA; 5) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 6) Genetic Disease Laboratory, California Department of Public Health, Richmond, CA.

Infants with Severe Combined Immunodeficiency (SCID), caused by defects in any of several genes, have few or absent T cells and succumb to infection unless promptly treated. Since newborns with SCID appear healthy, population-based screening is required to identify them before infections occur that would compromise their immune reconstitution by hematopoietic cell transplant, enzyme replacement or gene therapy. Dried blood spot newborn screening for SCID is now done in several states by measuring the number of T-cell receptor excision circles (TRECs), which are DNA byproducts of T cell receptor gene rearrangement. In addition to successful diagnosis of SCID cases, TREC screening also identifies additional infants who have low T cell numbers, but do not have defects in genes known to be associated with SCID.

We conducted whole exome sequencing for one such infant and her parents after the baby was identified by the California TREC newborn screening program. Though clinically well, she had low T and B cells, low IgG and IgM, undetectable IgA and no antibody responses to primary immunizatons. By filtering variants against a list of genes associated with T cell development, we found 2 damaging heterozygous ATM mutations. The mutations were also identified by an automated search of mutations in all genes under a recessive inheritance hypothesis that allowed for compound heterozygosity. The ataxia telangiectasia (AT) diagnosis was supported by a high alpha fetoprotein serum level though no neurological signs were evident up to age 18 months. We also tested archival CA dried blood spots of 14 confirmed AT patients born prior to implementation of the CA TREC newborn screening program. Of these, 7 had sufficiently low TRECs to have been called positive in the current screening algorithm, 4 had low-normal and 3 had normal TRECs. Thus individuals with AT can have such profound T lymphopenia at birth that they are detectable by TREC screening. Moreover, with 63 exons, ATM is efficiently analyzed by exome sequencing. TREC newborn screening offers early diagnosis not only for SCID, but also for secondary targets such as AT, providing an opportunity to define the spectrum of conditions associated with low T cells early in life.

363

Identification of a new melanocyte differentiation gene underlying human autosomal recessive albinism. *K. Grønskov[1, 5], C.M. Dooley[2], E. Østergaard[3], R.N. Kelsh[4], L. Hansen[5], M.P. Levesque[6], K. Vilhelmsen[7], D. Stemple[2], T. Rosenberg[8, 9]*. 1) Kennedy Ctr, Glostrup, Glostrup, Denmark; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 3) Dept. of Clinical Genetics, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark; 4) Dept. Biology and Biochemistry and centre for Regenerative Medicine, University of Bath, Bath, United Kingdom; 5) Dept. of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark; 6) Dept. of Dermatology, University of Zürich Hospital, Zürich, Switzerland; 7) Dept. of Ophthalmology, The National Hospital of the Faroe Islands, Torshavn, The Faroe Islands; 8) National Eye Clinic, The Kennedy Center, Glostrup, Denmark; 9) Gordon Norrie Center for Genetic Eye Diseases, Copenhagen, Denmark.

Oculocutaneous albinism (OCA) and ocular albinism (OA) are the most common conditions among hypopigmentation disorders. Although five causative genes have been identified (*TYR*, *OCA2*, *SLC45A2*, *TYRP1* and *GPR143*), a substantial fraction of patients remain genetically unresolved. We investigated five individuals from an inbred OA family from The Faroe Islands by homozygosity mapping, and found a region on chromosome 10 encompassing six genes. In one of them we found upon sequencing a homozygous nonsense mutation, truncating the protein by five amino acids. Further investigations of OA/OCA patients from The Faroe Islands showed that further five out of eight probands (total 11 out of 15 patients) were homozygous for the mutation. Analysis of unaffected family members showed segregation in accordance with autosomal recessive inheritance. Clinical examinations showed that nine patients had OA, while 2 sisters had OCA. We sequenced the gene in 48 genetically unexplained OCA/OA patients, and found a mutation in one of them. The patient originated from Lithuania. The mutation, a one base-pair insertion, leaves only 23 amino acids of the protein intact. A deletion on one allele in this patient cannot be ruled out. Analyses of bigger cohorts are needed to estimate the contribution of this gene to OCA/OA. A Zebrafish (*Danio rerio*) homolog shows 69% similarity on both nucleotide and amino acid level, and is expressed in neural crest cells. Knockdown of the gene using morpholinos resulted in substantially decreased pigmentation and a reduction of the apparent number of pigmented melanocytes. *In situ* hybridization with melanoblast marker dopachrome tautomerase (*dct*) was performed and showed a partial reduction in both dct expression levels and in melanoblast number, indicating that the gene is important for multiple aspects of melanocyte development, including differentiation. We have identified a novel melanocyte differentiation gene, that when mutated causes autosomal recessive albinism in humans.

364

Exome sequencing results in 230 patients with severe developmental disorders in the DDD project. *M. van Kogelenberg[1], K. Morley[1], T. Fitzgerald[1], S. Gerety[1], A. Tivey[1], S. Al-Turki[1], S. Clayton[1], C. Wright[1], J. Barrett[1], H. Firth[1,2], D. FitzPatrick[3], N. Carter[1], M. Hurles[1] on behalf of the DDD project.* 1) The Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK; 3) MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Western General Hospital, Edinburgh, UK.

The UK-based Deciphering Developmental Disorders (DDD) project aims to delineate the genetic architecture in children with undiagnosed severe developmental disorders. Recruitment to the project started in April 2011, with the intention of recruiting 12,000 families via all of the 23 UK National Health Service Regional Genetics Services in three years. This project employs an approach that combines multiple sources of genomic information, both array and sequencing technologies, with clinical phenotypic data from patients and their parents to investigate the full spectrum of genomic variants and their mode of inheritance. Here we intend to present an overview of the exome data in 230 patient-parent trios and the analysis pipelines developed to explore 10,000 anticipated exomes. The clinically diverse patient cohort in this study requires the consideration of both inherited and *de novo* disease models but here we will only focus on the later. To assess the considerable number of potential causal variants that will be discovered in this study we developed a systematic classification to appraise their potential causative nature. In addition we describe the validation of exome variants using the Illumina Miseq platform, which allows for a scalable and high throughput approach, as well as further investigations of variants in model organisms. Initial *de novo* exome analysis in 64 patient-parent trios identified the likely cause of the developmental disorder in 15% of cases, and highlight plausible causal variation that require further validation in an additional 20% of cases. Our diagnostic rate demonstrates the power of using patient-parent trios in combination with exome sequencing to detect causal events currently eluding diagnosis in most clinical settings.

365

Genetic etiology of isolated congenital asplenia. *A. Bolze[1], L. Abel[2], A. Puel[2], N. Trede[3], L. Selleri[4], J-L. Casanova[1,2]*. 1) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller University, New York, NY; 2) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Necker Medical School, Institut National de la Sante et de la Recherche Medicale, U980, Paris, France; 3) Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 4) Department of Cell and Developmental Biology, Weill Medical College of Cornell University, New York, NY.

Isolated congenital asplenia (ICA) is a rare developmental defect that is characterized by the lack of a spleen at birth with no other abnormalities. ICA is a severe condition predisposing patients to lethal bacterial infections. However ICA is underdiagnosed because awareness is insufficient, because ICA often strikes suddenly with rapidly lethal infections that prevent diagnosis opportunities and finally because it is not well understood in comparison with other inborn errors of immunity. We aimed to decipher the molecular genetic basis of human ICA. We hypothesized that ICA results from single-gene inborn errors of spleen development. We used a formal statistical approach described by Ionita-Laza et al. (AJHG 2011) to analyze exome-sequencing results and we identified novel heterozygous mutations in ICA-01 in 16 patients among a cohort of 33 patients. Surprisingly ICA-01 is a ubiquitous protein. We then tested the hypothesis that haplo-insufficiency of ICA-01 led to ICA by knocking down Ica-01 in the zebrafish model and by making an Ica-01+/− mouse. Finally, we tested the function of the mutants in the patients' fibroblasts and PBMCs, as well as looked at the transcriptome of these cells to identify a specific set of genes regulated by ICA-01. The discovery of the genetic etiology of half of the ICA patients paves the way for ICA genetic screening to allow early diagnosis and implementation of effective antibacterial preventative treatments. These findings also shed light on the mechanism of pathogenesis of ICA and on the development of the spleen in humans. Finally this is one more step towards a genetic theory of Infectious Diseases.

366

Whole Genome Sequencing in Two Brothers with Heterotaxy Reveals BCL9L as a Novel Gene Associated with Autosomal Recessive Heterotaxy (HTX6). *C.J. Saunders[1, 2, 3, 4], N.A. Miller[1, 2, 4], S.E. Soden[1, 2, 4], E. Farrow[1, 2, 3, 4], D.L. Dinwiddie[1, 2, 3, 4], N.P. Safina[1,4], S. Humphray[5], P. Saffrey[5], Z. Kingsbury[5], J.C. Weir[5], J. Betley[5], R.J. Grocock[5], J.E. Petrikin[1,2], K.P. Hall[5], S.F. Kingsmore[1, 2, 3, 4]*. 1) Department of Pediatrics, Children's Mercy Hosp, Kansas City, MO; 2) Center for Pediatric Genomic Medicine, Children's Mercy Hosp, Kansas City, MO; 3) Department of Pathology, Children's Mercy Hosp, Kansas City, MO; 4) School of Medicine, University of Missouri-Kansas City, Kansas City, Missouri; 5) Illumina Inc., Chesterford Research Park, Little Chesterford, Essex, UK.

Heterotaxy (HTX) is a developmental disorder with a prevalence of 1:10,000, characterized by abnormal arrangement of the thoracic and/or abdominal viscera. This is associated with multiple congenital malformations, with most morbidity and mortality due to complex cardiovascular defects. HTX is clinically and genetically heterogeneous, with a number of loci yet to be discovered. We used the HiSeq2500 to perform rapid whole genome analysis of two brothers with HTX negative for mutations in ZIC3, as well as their healthy parents. Genome sequencing yielded 3.9 million variants per individual, 833–901 of which were classified as possibly damaging and had allele frequencies <1%. A search for a likely pathogenic autosomal recessive genotype common to the affected brothers and heterozygous in their parents yielded two missense variants, c.2102G>A, (p.Gly701Asp) and c.554C>T, (p.Ala185Val) in the B-cell CLL/lymphoma 9-like gene (BCL9L). BCL9L is one of two human homologs of Drosophila legless (lgs), a segment polarity gene required for wingless (Wnt) signaling during development. Recently, the Wnt pathway was implicated in the left-right asymmetric development of vertebrate embryos with a role in regulation of ciliated organ formation and function. The key effector of the canonical Wnt pathway is beta-catenin, which activates Wnt-specific gene expression by binding to BCL9L, which serves a docking protein for other transcription coactivators. The p.Gly701Asp mutation found in our patient lies within the BCL9L nuclear localization signal (NLS). The failure to bind beta-catenin in the nucleus would lead to decreased Wnt-mediated gene expression, which was recently shown in zebrafish to result in disruption of left/right patterning, shorter/fewer cilia, loss of cilia motility, and decreased foxj1a expression. Foxj1a is a member of the forkhead gene family and is a master regulator of transcription controlling production of motile cilia. We propose that BCL9L mutations result in heterotaxy due to downregulation of Wnt/beta-catenin gene expression, particularly Foxj1a. On the basis of this evidence, the symbol HTX6 has been reserved for BCL9L-associated autosomal recessive heterotaxy. Additional studies are in progress to show definitive causality.

367

The complete GENCODE human annotation: new insights into the functionality of transcriptional complexity. *J.M. Mudge[1], A. Frankish[1], GENCODE. Consortium[2,3,4,5,6,7,8], T. Hubbard[1], J.L. Harrow[1].* 1) Wellcome Trust Sanger Institute, Wellcome Trust Campus, Hinxton, Cambridge CB10 1SA, UK; 2) University of California, 1156 High Street, Santa Cruz, CA 95064, USA; 3) Massachusetts Institute of Technology, 77 Massachusetts Avenue 750, Cambridge, MA 02139, USA; 4) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 5) Centre for Genomic Regulation (CRG) and UPF, Dr, Aiguader, 08003 Barcelona, Catalonia, Spain; 6) Yale University, 47 College Street, Suite 203, P.O. Box 208047, New Haven, CT 06520- 8047, USA; 7) Spanish National Cancer Research Centre (CNIO), C/Melchor Fernandez Almagro, 3, E- 28029 Madrid, Spain; 8) Washington University, Campus Box 1054, One Brookings Drive, USA.

The GENCODE consortium aims to identify all gene features in the human genome, using computational analysis, manual annotation and experimental validation. While the number of protein-coding genes in the GENCODE geneset has remained steady since the first release, the average transcript count for these loci has increased from 4.8 to 6.9. GENCODE also contains the most comprehensive annotation of long non-coding RNAs publicly available with 11,790 loci, although this number will likely increase as more RNAseq datasets are integrated. GENCODE thus has a higher total transcript count than other public genesets, containing 109,000 models (68%) that are not present in RefSeq or UCSC, almost 40% of which are protein-coding. However, since GENCODE attempts to capture all human transcripts, the genebuild will contain biologically spurious models, e.g. formed by in vivo splicing errors or `noisy' transcription. While it is not trivial to separate the functional portion of the genebuild, this process will be of great importance. Also, it is vital that we identify those transcripts encoding proteins. For example, our geneset was recently used to filter putative Loss of Function (LoF) variants identified in the 1000 Genomes pilot project. Initially, contemporaneous manual annotation of lncRNAs and pseudogenes was able to resolve misannotated protein-coding loci, reducing false positive LoF calls. Furthermore, ~36% of LoF variants were seen to be subjected to alternative splicing (AS), raising the possibility that such genes undergo only partial LoF. However, while we remain uncertain as to the biological relevance of these AS events, so our ability to predict the functional consequence of such variants is compromised. We will discuss how information on transcript functionality can be gained by integrating a wide range of data sources. For example, evolutionary conservation and tissue specific splicing can indicate AS functionality, while modern proteomics data and ribosome profiling can validate predicted CDS. We will also discuss the coupling of next generation promoter mapping data to transcript models. New GENCODE genebuilds are released quarterly; each contains an updated merge between manually produced HAVANA models and computational Ensembl models. The geneset can be downloaded from gencodegenes.org, while data can also be visualized via the Ensembl and UCSC genome browsers or accessed through the Ensembl databases, Perl API and BioMart.

368

Genetic Analyses in the Genotype-Tissue Expression (GTEx) Project. *K. Ardlie[1], N. Cox[3], D. DeLuca[1], E. Dermitzakis[4], B. Foster[5], G. Getz[1], R. Guigo[12], S. Jewell[13], D. Koller[9], J. Liu[8], J. Londsdale[6], D. Mash[7], M. McCarthy[10], M. Moser[5], D. Nicolae[3], A. Nobel[2], J. Pritchard[3], I. Rusyn[2], J. Thomas[6], W. Winckler[1], F. Wright[2], J. Zhu[11], GTEx Consortium.* 1) Broad Institute, Cambridge, MA; 2) University of North Carolina, Chapel Hill, NC; 3) University of Chicago, Chicago, IL; 4) University of Geneva, Geneva, Switzerland; 5) Roswell Park Cancer Institute, Buffalo, NY; 6) NDRI, Philadelphia, PA; 7) Brain Endowment Bank, Miami, FL; 8) Harvard University, Boston, MA; 9) Stanford University, Palo Alto, CA; 10) Oxford University, Oxford, UK; 11) Sage Bionetworks, Seattle, WA; 12) Center for Genomic Regulation, Barcelona, Spain; 13) Van Andel Institute, Grand Rapids, MI.

While genome-wide association studies have identified thousands of novel loci for common diseases and complex traits, for the majority of these, the mechanisms underlying the disease susceptibility remain unknown. Most associated variants are not correlated with protein-coding changes, suggesting that polymorphisms in regulatory regions of the genome may influence many disease phenotypes. Hence the systematic examination of gene expression, and its relationship to genetic variation, has become a critical next step in the elucidation of the genetic basis of common disease. The Genotype-Tissue Expression project (GTEx) aims to create a public atlas for human gene expression and its regulation, enabling the research community to discover expression quantitative trait loci (eQTL) and establish associations with disease. In its pilot phase, GTEx is recruiting 190 low postmortem interval human donors, collecting blood and up to 35 different somatic tissues from each donor (with collection of an additional ~750 donors planned for the scale up phases). Blood DNA from each donor is genotyped at 5 million SNPs, and RNA from all tissues is characterized by both expression arrays and deep RNA sequencing. Given the unique challenges of postmortem samples, our Laboratory and Data Analysis Center has optimized both sample processing and RNA sequencing protocols, and our results indicate that we are able to obtain high quality nucleic acids from a wide range of postmortem tissues, with robust gene expression profiles. The current GTEx data release includes genotype data from 79 donors, with expression array and deep RNA sequence data from a total of 654 tissues. 122 donors and 752 tissues are projected to be completed by September. To maximize statistical power for eQTL analysis, a set of 6 tissues were sequenced from all donors. For a subset of donors from whom brain tissues were obtained, all qualifying tissues were analyzed to enable a preliminary look at expression variation across a wider range of tissues. Genome-wide analysis to detect cis-eQTLs, and to evaluate allele and tissue-specific expression patterns, was performed on the set of 6 tissues (blood, lung, thyroid, heart, muscle, skin), and validates both known eQTL's and reveals novel ones, suggesting that the GTEx resource will be a powerful tool to unravel patterns of genetic variation and gene regulation across diverse human cell types.

369

Characterizing the genetic basis of transcriptome diversity in a large RNA sequencing study. *A. Battle[1], S. Mostafavi[1], X. Zhu[2], S.B. Montgomery[3,1], J.B. Potash[4], M.M. Weissman[5], C. Haudenschild[6], C. McCormick[6], R. Mei[7], A.E. Urban[2], D.F. Levinson[2], D. Koller[1,3].* 1) Computer Science, Stanford University, Stanford, CA; 2) Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA; 3) Genetics, Stanford University, Stanford, CA; 4) Department of Psychiatry, University of Iowa, Iowa City, IA; 5) Department of Psychiatry, Columbia University, New York, NY; 6) Illumina, Inc. San Diego, CA; 7) Centrillion Biosciences, Mountain View, CA.

Understanding the effects of genetic variation on gene expression is an important step towards unraveling the genetics of complex phenotypes. Recently, RNA sequencing has enabled quantification of detailed expression traits, including isoform levels and allele-specific expression. We aim to characterize the landscape of regulatory genetic variants and their effects on the transcriptome, and further to model the functional mechanisms of regulation using genomic annotations. In this study, we have sequenced transcriptomes for 922 individuals from a homogeneous population, representing one of the largest available human RNA sequencing studies. For each individual, we collected genotype for 737,187 common SNPs, and sequenced mRNA from whole blood (over 60 million reads per person). This dataset yields a large catalog of cis-eQTLs (expression quantitative trait loci) associated with both total expression and isoform ratio. Given the number of individuals available, we also have sufficient power to uncover hundreds of trans-eQTLs affecting various aspects of expression. Finally, we are able to directly estimate the effects of genetic variation on allele-specific expression, utilizing the availability of many compound heterozygous individuals. From this extensive set of QTLs, we can begin to characterize the genetic variants that affect different expression traits, from basic cis and trans associations, to isoform ratio (splicing) and allelic effects. To this end, we have gathered a diverse set of genomic annotations, including regulatory sequence annotations derived from the ENCODE project such as DNAase hypersensitivity sites, Chip-seq data, and motif enrichments for a large set of transcription factors. We combine these annotations into a unified statistical model of regulatory potential, which predicts, for each SNP, the estimated effect on expression. Models of regulatory potential are learned for each type of QTL, revealing the similarities and differences between classes. For instance, several Chip-seq annotations are significantly associated with cis effects, while intronic snps often yield splicing QTLs. The learned models are significantly more predictive than individual annotations, and can be used to estimate effects of rare or un-genotyped SNPs, or to prioritize loci in smaller studies. Overall, the diverse QTLs and functional characterization described here will provide a useful resource in untangling the effects of genetic variation.

370

Genetic and molecular basis of RNA-DNA sequence differences in humans. *V.G. Cheung[1], A. Bruzel[1], L. McDaniel[1], A.L. Richards[2], J.M. Toung[3], I.X. Wang[1].* 1) Howard Hughes Medical Institute; Departments of Genetics and Pediatrics, University of Pennsylvania, Pennsylvania, USA; 2) Cell and Molecular Biology Graduate Program; 3) Genomics and Computational Biology Graduate Program.

In genetic studies, we examine mostly DNA sequence variants to explain disease risks and phenotypic manifestations. However, advances in sequencing technologies have enabled deep DNA and RNA sequencing which revealed that there are many more differences between RNA and corresponding DNA sequences than previously known. These RNA-DNA sequence differences, RDDs, include the known RNA editing events such as ADAR-mediated A-to-G editing but also other types including transversions, for instance A-to-C differences. The presence of RNA variants that are not reflected in the DNA suggests that there is genome variation beyond DNA sequence polymorphisms. This project aims to characterize RDDs and show its validity by demonstrating a genetic basis of RDDs.

To extend our initial study that uncovered tens of thousands of RDDs in 27 individuals, we carried out very deep DNA and RNA sequencing to obtain about 1 billion RNA-sequencing reads per sample and at least 30X DNA-sequence coverage for each of two individuals. The results showed that even with stringent inclusion thresholds, there are many thousands of RDD sites per individual and about 30% of the sites are shared across individuals. To confirm these RDDs, we sequenced the DNA (30X) and RNA (~ 150 million reads) samples of 50 unrelated individuals; these data validated the RDDs and revealed extensive individual variation in RDD levels. We also sequenced samples from different cellular compartments to study where during transcription and processing RDDs arise. To determine if there is genetic component to this variation, we sequenced paired DNA and RNA samples from monozygotic twins and siblings. While the level of each RDD site varies across unrelated individuals, this variability is significantly less among related individuals suggesting a genetic component to variation in RDD levels. We identified hundreds of RDD sites where intraclass correlation coefficients of monozygotic twins are significantly ($Pc<0.05$, permutation testing) less than other relative pairs. In this presentation, I will discuss results from the very deep sequencing of DNA and RNA of two individuals, and data from twins and siblings that demonstrate a genetic influence of RDD levels. I will also present some preliminary suggestions of mechanisms that underlie RDDs.

371

Characterizing gene expression variation across seven diverse human populations. *A.R. Martin[1], H.A. Costa[1], J.M. Kidd[2], B.M. Henn[1], M.C. Yee[1], F. Grubert[1], S.B. Montgomery[1], H.M. Cann[3], M.P. Snyder[1], C.D. Bustamante[1].* 1) Stanford University School of Medicine, Genetics Department, Stanford, CA, 94305; 2) University of Michigan School of Medicine, Department of Human Genetics, Ann Arbor, MI, 48109; 3) Foundation Jean Dausset, Centre d'Etude du Humain, Paris, 75010, France.

Recent large-scale cross-population sequencing projects have indicated that the vast majority of genetic variation is likely to be specific to continental groups, and even within closely related populations. While genetic variation has been studied across diverse human populations, our understanding of its impact on phenotypic variation is limited without extending these studies to determine the effect of genetic variation on cellular phenotypes, such as gene expression and methylation. Studies using RNA-sequencing to understand the genetics of gene expression in unrelated individuals from single populations have pointed out that a substantial amount of variation is novel and has elucidated genetic effects on transcript levels, gene expression level, miRNA levels, splicing variants and transcription termination points. An understanding of the full range of these effects on the human transcriptome requires examining individuals across the full range of the genetic diversity continuum. To this end, we have analyzed the genomes, transcriptomes, and methylomes of 4–7 individuals in the Human Genome Diversity Panel (HGDP) from seven worldwide populations, including the Namibian San, who retain the largest human effective population size yet observed, the Mbuti Pygmies of central Africa, Mozabites of northern Africa, Pathans of central Asia, Cambodians of east Asia, Yakut of Siberia, and Mayans, who have undergone several bottlenecks during their migration to Central America. This approach has allowed us to perform a comparative study using the single-nucleotide resolution of RNAseq to assess rare transcripts, novel gene structures, alternative splicing, and differential expression within and among populations. We have quantified reads for known exons, transcripts and whole genes and have employed a novel statistical approach via linear mixed model to identify systematically differentially expressed structures among populations. As expected, we also found a high degree of correlation between genetic and transcript diversity. Further, we have explored the role of genome selection on transcript diversity. Our dataset has allowed for a detailed investigation of the landscape of human transcriptome variation in diverse human populations.

372

Comparative eQTL analyses within and between seven tissue types suggest mechanisms underlying cell type specificity of eQTLs. *B. Engelhardt[1,2,3,4], C. Brown[1,5,6].* 1) Human Genetics, Univ Chicago, Chicago, IL; 2) Biostatistics & Bioinformatics, Duke University, Durham, NC; 3) Department of Statistical Sciences, Duke University, Durham, NC; 4) Institute for Genomics Sciences & Policy, Duke University, Durham, NC; 5) Institute for Genomics and Systems Biology, Univ Chicago, Chicago, IL; 6) Department of Genetics, University of Pennsylvania, Philadelphia, PA.

Genetic variants can impact gene expression in a cell type specific manner. However, little is known about the genome-wide replication patterns of cis-eQTLs across heterogeneous cell types. This problem is further complicated by the modest replication rate of cis-eQTLs ascertained from the same cell type in independent studies. EQTLs are increasingly being used to provide mechanistic interpretations of human disease associations. Because the phenotype-relevant cell type may be unknown or unavailable, understanding the frequency of and the mechanisms generating cell type specific eQTLs is critical. Recent eQTL replication studies have demonstrated that between cell type replication rates can be as low as 20%, that within cell type replication rates are often around 70% when accounting for study-specific differences, and that non-replicating eQTLs regulate tissue-specific biological processes. In this study, we leverage both within cell type and between cell type eQTL data to control for technical covariates when considering between-tissue replication. We demonstrate that both between-tissue and within-tissue replicating eQTLs are overrepresented near gene start sites. These comparisons show that eQTL SNPs that overlap predicted cis-regulatory elements have higher replication rates. Similarly, eQTL SNP-gene pairs that are not intersected by predicted insulator elements have greater reproducibility across cell types. Importantly, we demonstrate that eQTLs are more likely to overlap cis-regulatory elements (CREs) ascertained from the same cell type rather than from different cell types, and eQTLs are more likely to overlap cell type specific CREs if they are promoter distal. Further, eQTL SNPs that overlap cell type specific CREs are significantly enriched for cell type specific eQTLs, suggesting specific molecular regulatory mechanisms for those eQTLs. Lastly, to quantify the eQTL allelic heterogeneity (independent genetic variants associated with expression), we identify multiple, independent cis-eQTLs for a substantial number of genes in each study, and demonstrate that non-primary eQTLs are more likely to be promoter distal and cell type specific. These analyses provide new insight into the patterns of cis-eQTL replication across and within tissues while controlling for biological and technical variation. We anticipate that these results will improve scientists' ability to leverage eQTL data for the interpretation of human disease associations.

373

Identification of novel genetic determinants of induced innate immune responses and context specific eQTL in human primary monocytes. *B.P. Fairfax[1,2], S. Makino[1], J.C. Knight[1].* 1) Wellcome Trust Centre for Human Genetics, Oxford, Oxfordshire, United Kingdom; 2) Department of Oncology, Oxford Cancer Centre, Churchill Hospital, Oxford.

Differential regulation of gene expression provides a major mechanism whereby genetic variation can affect an organism's fitness. Such genetic variants may be resolved as expression quantitative trait loci (eQTL) and are often cell-type specific. We reasoned that further to cell-type, the activity state of a cell might modulate both the number and genomic location of eQTL. We anticipated sequence variants that are currently thought to be `non-functional' may associate with regulation of gene expression in discrete conditions. Furthermore, induction of innate immune response genes might reveal eQTL not evident in an unstimulated state when many such genes are unexpressed.

To investigate these hypotheses in early and late innate immune response genes we explored eQTL in $CD14^+$ primary monocytes exposed to either lipopolysaccharide (LPS) (2 or 24 hours) or IFNγ (24 hours, n=170–220 individuals per state). For eQTL analyses, individuals were genotyped at >730,000 loci. Expression was analysed in the unstimulated state and additionally after exposure to each stimulant using Illumina HT-12 whole genome expression arrays (n= 856). We performed eQTL analysis using linear and non-linear methodology, incorporating expression dataset principal components as covariates. A subset of context-specific eQTL were investigated in a second collected cohort with RNA-seq and quantitative PCR. In this large dataset, LPS was found to reproducibly regulate the expression of >5800 genes; ≈3000 uniquely to 2h LPS (early response), whilst ≥1800 were specifically regulated after 24h LPS (late response). Additionally, we identified a further ≈1200 genes that were regulated uniquely by 24h IFNγ. Each stimulation leads to the regulation of both unique and shared genes, over 60% of which do not have eQTL in the unstimulated state. Many novel, context specific eQTL are observed, some overlapping with GWAS loci, shedding further insight into the effect of genetic variation upon innate immune responses.

This study is the first to fine-map eQTL across different monocyte activation profiles. It highlights the plethora of mechanisms whereby genetic variation can affect gene expression and human health. We suggest many polymorphisms currently thought `non-functional' may indeed be functional in discrete states of cellular activity.

374

Gene-level and exon-level expression QTL signals in the UK Brain Expression Consortium dataset. *M.E. Weale[1], A. Ramasamy[1,2], D. Trabzuni[2,3], R. Walker[4], C. Smith[4], M. Ryten[2], J. Hardy[2], UK Brain Expression Consortium.* 1) Department of Medical & Molecular Genetics, King's College London, 8th Floor, Tower Wing, Guy's Hospital, London SE1 9RT, United Kingdom; 2) Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom; 3) Department of Genetics, King Faisal Specialist Hospital and Research Centre, PO Box 3354, Riyadh 11211, Saudi Arabia; 4) MRC Sudden Death Brain Bank Project, University of Edinburgh, Department of Neuropathology, Wilkie Building, Teviot Place, Edinburgh, EH8 9AG, United Kingdom.

BACKGROUND: Understanding the genetic regulation of gene expression and alternative transcripts across multiple brain regions is vital for advancing research in neurodegenerative diseases, psychiatric and behavioural fields. METHOD: A total of 1,266 RNA samples were extracted from 137 neuropathologically normal post-mortem brains and hybridized onto Affymetrix Human Exon 1.0 ST arrays. We analyzed up to 12 regions from each individual: frontal cortex, occipital cortex, temporal cortex, white matter, thalamus, putamen, substantia nigra, hippocampus, hypothalamus, medulla, cerebellar cortex and spinal cord. DNAs were extracted and analyzed using the Illumina Omni 1M and Immunochip arrays, and subsequently imputed to ~6 million SNPs using the 1000 Genomes Project (March 2012). Expression quantitative trait locus (eQTL) analysis was conducted to identify SNPs associated with expression levels. We used statistical interaction tests to distinguish between gene-level and exon-level eQTLs and between region-specific versus ubiquitous eQTLs. RESULTS: We find numerous and robust examples of ubiquitous and region-specific gene-level and exon-level eQTLs, even after stringent multiple testing correction. Weaker ubiquitous signals were identified by averaging expression values across brain regions for each individual. Interestingly, cerebellum and white matter show more unique eQTLs in comparison with other brain regions, suggesting different a genetic regulation architecture in these regions. A subset of our results coincide with SNPs identified through genome-wide association studies for relevant traits. CONCLUSION: This study yields insights into the regulation of gene expression in multiple brain regions, using a dataset with an unparalleled combination of multiple human brain regions, sample size and exon-level data. The dataset is a valuable resource for research into complex neurodegenerative diseases.

375

First complete haplotype of the human immunoglobulin heavy chain locus from a single individual and characterization of novel allelic and structural variation. *K. Meltz Steinberg[1], C.T. Watson[2], J. Huddleston[1], P.H. Sudmant[1], R.L. Warren[3], M. Malig[1], J. Schein[3], A.J. Willsey[2], J.B. Joy[2], J.K. Scott[4], T. Graves[5], R.K. Wilson[5], R.A. Holt[3], F. Breden[2], E.E. Eichler[1, 6].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada; 3) Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada; 4) Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada; 5) The Genome Institute, Washington University, St. Louis, Missouri, USA; 6) Howard Hughes Medical Institute, Seattle, WA.

The immunoglobulin heavy chain locus (IGH) is one of the most biomedically important regions of our genome, encoding genes responsible for antibody formation and vital to the adaptive immune response. IGH is enriched for segmental duplication and subject to extensive somatic rearrangement, complicating attempts to reconcile its organization from B-lymphocyte derived genetic material. Germline variation within IGH is not well understood; haplotype inference using traditional linkage disequilibrium methods has been difficult, and the reference sequence (hg19) is missing at least 11 expressed heavy-chain variable (V) genes. Here we present the first complete IGH haplotype sequence derived from a single individual, created by sequencing BAC clones from the hydatidiform mole library, CH17. A hydatidiform mole results from the fertilization of an enucleated egg by a single sperm, resulting in a genome that solely comprises a single haplotype. Using this resource, we generated 1 Mbp of high quality finished sequence and identified 47 functional IGHV, 27 functional diversity (D), and six functional junction (J) gene segments. A total of five large structural variants (SVs) and 15 nonsynonymous mutations compared to the hg19 reference haplotype involving 29 IGHV genes were characterized from the CH17 haplotype. Our work has added 100 kbp of novel sequence and four IGHV genes not represented in the hg19 reference. Using this new reference, we characterized germline SVs from a panel of nine diploid genomes of diverse origin. We resolved eight complete SV haplotypes comprising 125 kbp of additional novel sequence including four additional IGHV genes. The majority of these SVs resulted from non-allelic homologous recombination mediated by repetitive sequence containing IGHV genes or pseudogenes. Four of these SVs were genotyped using PCR in 425 individuals from nine diverse populations. We found that three of the four are polymorphic and highly population stratified (Fst values from 0.3 to 0.5), with the greatest differences observed between African and Asian populations. In addition, linkage disequilibrium was low between SNPs on two commercial arrays and the genotyped variants, revealing that these polymorphisms have likely been underrepresented in previous disease-association studies. Thus, we are currently exploring the role this newly found genetic variation may play in antibody expression and susceptibility to human autoimmune and infectious disease.

376

The Epitranscriptome Reveals Novel Mechanisms of RNA Regulation and Spatiotemporal Dynamics. *C.E. Mason[1], K. Meyer[2], Y. Saletore[1], P. Zumbo[1], O. Elemento[1], S. Jaffrey[2]*. 1) Physiology and Biophysics, Weill Cornell Medical College, New York, NY; 2) Department of Pharmacology, Weill Cornell Medical College, 1300 York Ave., New York, NY 10021.

The epigenetic modifications of DNA such as 5-methylcytosine (5mC) and 5-hydroxymethylcytosine, (5hmC) have attracted considerable attention due to their ability to affect chromatin. Although DNA has two well-known (and many other) base modifications, mRNA modifications are thought to be limited to 5' cap formation and only one internal modification, inosine, which forms as a result of A→I editing. We have made the remarkable finding that there is a widespread sixth base in mRNA, N-methyl-6-methyladenosine (m6A). m6A is not well known, and thought to be an oddity, enriched in various viral mRNAs, and only known in two mammalian mRNAs. Using an antibody specific to m6A, we developed a novel capture and deep sequencing approach called MeRIP-Seq (Methylated RNA and Immunoprecipitation) to gain insight into the potential biological functions and genome-wide distribution of m6A. We have found that m6A is present in mRNA from all tissues tested in human and mouse and is highly enriched in brain and neuronal tissues. Furthermore, m6A exhibits dynamic changes in its levels during development, suggesting important roles in cellular function. By using mapping the localization of over 11,000 m6A high-confidence sites in over 7,676 genes, we found a striking enrichment in the 3'UTR and near stop codons, suggesting roles for m6A in translation termination and other aspects of mRNA regulation. Further, in 3' UTRs, and we uncovered an association between m6A residues and microRNA-binding sites within 3'UTRs, which underscore this modification's spatial relationship to miRNA regulation. Finally, we also observe that this base modification can be detected on new single-molecule technology (PacBio) with direct RNA-sequencing. Our findings add another layer of complexity and regulation to basic RNA biology by identifying a new RNA base that is widespread in the transcriptome, and also opens exciting new avenues of research on the “epitranscriptome.”.

377

Epigenome-wide profiling of circulating DNA in colorectal cancer. *R. Cortese[1], Y. Li[2], A. Kwan[1], B. Zanke[3], Z. Zhang[2], A. Petronis[1]*. 1) The Krembil Family Epigenetics Laboratory, Centre of Addictions and Mental Health, Toronto, Ontario, Canada; 2) The Donnelly Centre. University of Toronto. Toronto, Ontario, Canada; 3) Ottawa Hospital Research Institute. Clinical Epidemiology. Ottawa, Ontario. Canada.

Early diagnosis represents the most effective way to reduce mortality in colorectal cancer (CRC). A potentially transforming development is the identification of epigenetic DNA alterations specific to CRC in blood plasma and their application to early diagnostics of the malignant disease. During the growth and expansion of tumors, some malignant cells die and their fragmented DNA is released into the bloodstream. Because tumor cells undergo major epigenetic changes, epigenetic DNA aberrations can be used to identify malignant growth in the organism. We performed a large-scale analysis of the circulating DNA (cirDNA) isolated from blood plasma of 193 stage I-II CRC patients and 200 unaffected controls, divided in two independent batches. We have previously developed a protocol for epigenome-wide analysis of cirDNA modifications using microarray technology. Universal DNA adaptors were ligated to the ends of cirDNA fragments, followed by digestion with DNA modification-sensitive restriction enzymes. cirDNA fragments that survived enzymatic hydrolysis were amplified by adaptor-mediated PCR and labeled using biotinylated nucleotides and a fluorescent dye. This enriched cirDNA-modified fraction was hybridized to tiling microarrays containing over 6 million probes that cover the entire chromosomes 4, 15, 18 and 20. This method represents a cost-effective approach for epigenome-wide marker discovery in cirDNA with immediate translatability to the clinical practice. Following extensive quality control and data normalization, univariate multiple-testing adjusted statistical analyses were used to identify regions exhibiting differential cirDNA modification between cases and controls. Moreover, we applied generalized linear model to identify age- and gender specific variations in the cirDNA modification profiles. We identified 2,236 and 2,327 differentially modified regions in the first and second sample batches, respectively. Amongst them, 645 regions were identified as differentially modified in both batches, showing high reproducibility across independent samples. Such regions are eligible candidates for formal biomarker validation protocols, either as single-locus markers or as an `epigenetic signature" for CRC. This first epigenome-wide screening for diagnostic markers in cirDNA may provide a new, non-invasive approach for early detection of CRC.

378

Alterations in genomically imprinted miRNA and snoRNA clusters in a mouse model of fetal alcohol spectrum disorders. *B. I. Laufer, K. Mantha, M. L. Kleiber, E. J. Diehl, S. M. F. Addison, S. M. Singh.* Molecular Genetics Unit, University of Western Ontario, London, Ontario, Canada.

Fetal Alcohol Spectrum Disorder (FASD) is a common and heterogeneous disorder caused by maternal drinking during pregnancy. While much of the research on FASD has focused on behavioural and neuro-structural changes, prenatal alcohol exposure also results in long-term alterations in gene expression; however, the mechanisms underlying the persistence of these changes are not known. In this study, we used four ethanol treatment protocols to model developmental ethanol exposure in mice: injections at 3 specific neurodevelopmental time points that model a “binge” exposure, and a voluntary maternal consumption model, which represents moderate chronic exposure throughout development. We then assessed small RNA brain gene expression in resulting adult offspring (PD 70) using miRNA expression arrays, mouse gene expression arrays, and quantitative PCR. The analysis revealed that a large number of microRNAs and snoRNAs are altered, both up and down, depending on treatment paradigm. Some of these expression profiles are unique to a treatment protocol while others overlap. Strikingly, approximately 20% of the altered noncoding RNAs (ncRNAs) localized to three imprinted clusters. The first two, S*nrpn-Ube3a* (Murine 7qC/Human 15q11-q13) and *Dlk1-Dio3* (Murine 12qF1/Human 14q32.2), are associated with processes involved in neuronal plasticity and several neurodevelopmental disorders. The third cluster contains *Sfmbt2* (Murine 2qA1) and an overlapping antisense transcript that is unique to mice and rats. We then assessed brain DNA methylation using methylated DNA immunoprecipitation followed by hybridization to DNA arrays (MeDIP-Chip), which revealed that fetal alcohol exposure has a genome-wide effect on DNA methylation with imprinted regions of the genome appearing to be particularly sensitive. Ultimately, our results suggest that imprinted ncRNAs, many of which play a critical role in neurodevelopment and brain function, may have a role in the long-term maintenance of altered gene expression and cognitive endophenotypes associated with FASD.

379

KDM6A escapes X inactivation and controls expression of reproduction-related homeobox genes in female ES cells and ovary: deficiency may explain embryonic and ovarian failure in Turner. *C.M. Disteche[1,2], J.B Berletch[1], X. Deng[1], D. Nguyen[1]*. 1) Dept Pathology, Univ Washington, Seattle, WA; 2) Dept of Medicine, Univ Washington, Seattle, WA.

The Rhox genes represent a type of developmentally regulated homeobox (HOX) genes implicated either in male or female reproduction. Within the Rhox cluster on the mouse X chromosome Rhox6 and 9 are expressed higher in ovary compared to testis. KDM6A, a histone demethylase that removes the repressive chromatin mark trimethylation at lysine 27 of histone H3 (H3K27me3), is encoded by a gene that escapes X inactivation and has higher expression levels in females than males. Here we report that KDM6A regulates Rhox6 and 9 expression during early female but not male ES cell differentiation. KDM6A is recruited to Rhox6 and 9 in female ES cells resulting in removal of H3K27me3 and increased Rhox expression, a process inhibited by KDM6A knockdown. In contrast, KDM6A occupancy at Rhox6 and 9 is low in male ES cells and knockdown has no effect on expression. In adult ovary where Rhox6 and 9 are highly expressed KDM6A occupancy strongly correlates with expression. Our study implicates for the first time a gene that escapes X inactivation and thus is deficient in Turner females, in the regulation of homeobox genes in female ES cells and in ovary, suggesting a potential role for this gene in the etiology of embryonic anomalies and ovarian dysgenesis associated with the presence of a single X chromosome.

380

Genome-wide scan of DNA methylation in the aging brain and its relation to Alzheimer's disease. *P.L. De Jager[1], G. Srivastava[1], M.L. Eaton[2], L.E. Chibnik[1], B. Keenan[1], N. Taner[3], A. Myers[4], B. Bernstein[5], A. Meissner[6], M. Kellis[2], D.A. Bennett[7].* 1) Neurology, Brigham & Women's Hospital, Boston, MA; 2) Massachusetts Institute of Technology, Cambridge MA; 3) Mayo Clinic, Jacksonville FL; 4) University of Miami, Miami Fl; 5) Massachusetts General Hospital, Boston MA; 6) Harvard Medical School, Boston MA; 7) Rush University, Chicago IL.

Background: Alterations in DNA methylation have been suggested to occur in the context of Alzheimer's disease. Here, we rigorously explore the role of brain DNA methylation in Alzheimer's disease (AD) on a genome-wide scale, Methods: We use a unique bank of frozen brains from two prospective studies of aging: the Memory and Aging Project and the Religious Order Study. Each subject is non-demented at the time of entry into the study. A sample of dorsolateral prefrontal cortex was obtained from each of 759 subjects, and, using the Illumina Humanmet450K platform, we generated data for 486,428 CpG sites distributed throughout the genome. The primary analysis uses a linear regression adjusting for sex and age. Results: In our primary analysis, methylation levels at 163 CpG dinucleotides are associated with a quantitative measure of AD-related amyloid pathology that is available for each subject (p<10^-7), including CpG in the validated BIN1 and ABCA7 AD susceptibility loci. In a secondary analysis, all 163 CpG also demonstrate strong evidence of association with AD, and 95% of them are hypermethylated in AD subjects. Further, when the analysis is limited to those subjects with normal cognition at the time of death, associations with neuritic plaque burden persists, suggesting that changes in methylation are an early feature of AD. We validate these results using RNA extracted from the same brain region of each subject and show that 35% of genes in the vicinity of the associated CpG have RNA expression levels correlated with our measure of AD pathology at a nominal level (P<0.05), and, for 9/171 of the tested genes, RNA levels are significantly correlated with AD pathology (p<0.0003). Further, in a second, independent collection of AD and control brains, 17% of the genes near these 163 CpGs demonstrate significantly altered RNA transcription levels in AD brains (p<0.00002). Finally, using a novel chromatin state map of the frontal cortex, we see that the 163 associated CpGs are distributed in a number of different chromatin states but appear to be depleted in strong promoters and enriched in polycomb-repressed and weakly transcribed loci. Conclusions: Robust changes in DNA methylation are found in the brains of subjects with AD. The associated sites are hypermethylated in AD, suggesting a coordinated AD-related chromatin remodeling in aging brains that is present even in individuals with early, asymptomatic pathology at the time of death.

381

RNA-mediated transcriptional silencing in Friedreich ataxia. *Y.K. Chutake[1], A.M. Castro[1], S.I. Bidichandani[1, 2].* 1) Department of Biochemistry & Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Friedreich ataxia (FRDA) patients have a deficiency of transcriptional elongation due to an expanded GAA triplet-repeat in intron 1 of the *FXN* gene and the surrounding repressive chromatin that ensues in response to the expanded repeat. We show that the depletion of CTCF, a chromatin insulator, near the transcription start site of the *FXN* gene (*FXN*-TSS), which occurs in FRDA patients, triggers the establishment of repressive chromatin that also contributes to *FXN* transcriptional deficiency. The repressive chromatin near *FXN*-TSS is associated with altered nucleosome positioning, deficiency of polII and H2A.Z occupancy, and deficient transcription upstream of the expanded repeat, all of which are consistent with an additional deficiency of transcriptional initiation. *FAST1*, an antisense transcript that overlaps the *FXN*-TSS forms double-stranded RNA with the *FXN* sense transcript. The *FXN-FAST1* double-stranded RNA template interacts with argonaute proteins, Ago 1 and Ago 2, which are key players in RNA-mediated transcriptional silencing. Knockdown of Ago 1 and Ago 2 in FRDA cells reversed the repressive chromatin near the *FXN*-TSS. CTCF depletion therefore functions as an epigenetic switch which permits RNA-mediated transcriptional silencing of the *FXN* gene in FRDA. Therapeutic strategies in FRDA would have to reverse the repressive chromatin in intron 1 and at *FXN*-TSS.

382

P53 regulates 5-hydroxymethylcytosine-mediated epigenetic landscape through GADD45A. *Y. Li[1], Y. Zhu[2], K. Szulwach[1], L. Lin[1], C. Street[1], H. Wu[3], D. Chen[2], P. Jin[1].* 1) Dept Human Gen, Emory Univ, Atlanta, GA; 2) 2State Key Laboratory of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, PR CHINA; 3) 3Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, Atlanta, GA 30322, USA.

Epigenetic changes, including DNA methylation and histone modifications, play a profound role in cellular differentiation and cancer cells. Recently a new modified DNA base, 5-hydroxymethylcytosine (5hmC), was found in mammalian DNA, raising questions as to its role in mediating epigenetic control of gene expression. 5hmC constitutes a significant portion of nucleotides in brain cells (~0.6% of total nucleotides in Purkinje cells and ~0.2% in granule cells) and embryonic stem cells (~0.032%). 5hmC can be derived from the oxidation of 5-methylcytosine (5-mC) in a reaction catalyzed by one of three TET family members, TET1, TET2, and TET3. Here we show that p53 is required for maintaining the 5hmC level and the loss of p53 leads to the significantly reduced 5hmC abundance in cancer cells. We find that the modulation of 5hmC by p53 is mediated through the multifunctional growth arrest and DNA damage 45 (Gadd45) protein family members, which have been shown previously to mediate DNA demethylation during cell differentiation and stress response. P53 could activate the expression of GADD45A, which interacts with TET proteins and enhance their activities both in vitro and in vivo. Furthermore, we provide the evidence that GADD45A could regulate the activity of Tet proteins in zebrafish, and Gadd45a genetically interacts with Tet protein to regulate the early development of zebrafish. These results together reveal an unexpected role of p53 in directly modulating cellular epigenetic landscape, and suggest that GADD45A participates DNA demethylation through modulating the hydroxylation activity of TET proteins.

383

Maps of open chromatin highlight cell type-specific patterns of regulatory sequence variation at hematological trait loci. *C.A. Albers[1,2], D.S. Paul[1], A. Rendon[2,3], K. Voss[2], J. Stephens[2], P. van der Harst[4,5], J.C. Chambers[6,7,8,9], N. Soranzo[1], W.H. Ouwehand[1,2], P. Deloukas[1], HaemGen Consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) University of Cambridge, Cambridge, United Kingdom and National Health Service (NHS) Blood and Transplant, Cambridge, United Kingdom; 3) MRC Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom; 4) Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands; 5) Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands; 6) Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 7) Imperial College Healthcare NHS Trust, Hammersmith Hospital, London, United Kingdom; 8) Royal Brompton and Harefield Hospitals NHS Trust, London, United Kingdom; 9) Ealing Hospital NHS Trust, Southall, Middlesex, United Kingdom.

A major challenge in genome-wide association studies (GWASs) is to identify the functional sequence variants, effector cell types, and biological mechanisms underlying the associations. We explored these issues using data from two meta-analyses of GWASs in over 60,000 individuals of two quantitative platelet traits (count and volume) and six red blood cell (RBC) traits (including count, volume and hemoglobin). We used formaldehyde-assisted isolation of regulatory elements followed by next-generation sequencing (FAIRE-seq) to map regions of open chromatin (nucleosome-depleted regions, NDRs) in primary human precursors of platelets (megakaryocytes), and red blood cells (erythroblasts), as well as in monocytes.

We found 50,372 NDRs in megakaryocytes, 38,014 in erythroblasts, and 34,833 in monocytes. Genes near NDRs were characterized by different cellular processes, e.g. hematopoietic identity vs. house-keeping, depending on the number of sequence reads supporting the NDRs. We used bootstrapped p-value distributions of sequence variants imputed from the 1000 Genomes project to investigate enrichment of platelet and RBC trait associations in NDRs in each cell type. We observed strong enrichments (~2-fold to 10^5-fold lower p-value at the 0.005 quantile) for associations not reaching genome-wide significance, implying that NDRs can be used to reduce false positive rates for weaker signals. Considering the subset of NDRs present in two biological replicates of the same cell type, the strongest enrichment for platelet traits was found in megakaryocyte NDRs, and for RBC traits in erythroblast NDRs. Interestingly, platelet count showed enrichment in monocyte-specific NDRs, while RBC traits did not show enrichment in these NDRs. The strength and relative orderings of enrichments in NDRs revealed interesting patterns across cell types indicating distinct biological mechanisms underlying the traits. Using experimental and *in silico* methods, we tested 16 candidate regulatory variants in open chromatin at 12 known platelet quantitative trait loci, and found that 10 (62.5%) affected protein binding to the DNA, suggesting that this is a frequent mechanism by which sequence variation influences quantitative trait variation. We will report on the application of variance-partitioning methods based on whole-genome sequence data to partition the phenotypic variance into contributions from sequence variation in various subsets of NDRs.

384

Functional epialleles at an endogenous human centromere. *B.A. Sullivan[1, 2], K.A. Maloney[1], L.L. Sullivan[1], E.D. Strome[1], J. Matheny[1]*. 1) Duke Institute for Genome Sciences & Policy, Duke University, Durham, NC 27708. USA; 2) Department of Molecular Genetics & Microbiology, Duke University Medical Center, 27710, USA.

Human centromeres are defined by megabases of homogenous alpha satellite DNA that is packaged into specialized chromatin containing the centromeric histone CENP-A that triggers centromere assembly. Most endogenous human chromosomes have a single higher-order repeat (HOR) array of alpha satellite DNA. Dicentric chromosomes, those with two alpha satellite arrays that are physically linked on the same chromosome, are generally thought to be structurally abnormal products of meiotic or somatic rearrangement. However, several human chromosomes contain more than one alpha satellite array. *Homo sapiens* chromosome 17 (HSA17), for instance, has two juxtaposed HOR arrays, D17Z1 and D17Z1-B, suggesting that it is organized like a dicentric chromosome. Only D17Z1 has been previously studied in the context of centromere function on HSA17. We use human artificial chromosome assembly assays to demonstrate that either D17Z1 or D17Z1-B can independently support *de novo* centromere assembly. We extend these *in vitro* studies to show that in humans, the centromere can be assembled either at D17Z1 or D17Z1-B. Intriguingly, ~30% of humans that we studied are functional heterozygotes, meaning that within the same cell, the centromere is formed at D17Z1 on one HSA17 homologue and at D17Z1-B on the other HSA17 homologue. The site of CENP-A assembly on HSA17 is stable and transmitted through meiosis, as evidenced by inheritance of CENP-A location through multi-generational families. Chromatin modifications differ between active and inactive D17Z1 and D17Z1-B arrays, implying that functional differences between the arrays are regulated or maintained by distinctive chromatin structures. These results reveal the unexpected presence of centromeric epialleles in humans, and imply that at least one endogenous human chromosome normally behaves as a stable structurally dicentric, but functionally monocentric, chromosome. Our studies emphasize epigenomic complexities underlying mechanisms that determine centromere identity in humans and highlight the importance of complete genome assemblies of all human centromere regions.

385

Rare Insertion Polymorphisms Identified by Exome Sequencing May Be Associated With Age-Related Macular Degeneration. *I. farrer[1], j. kozubek[1], m. schu[1], j. farrell[1], m. morrison[2], k. mayne[2], d. morgan[2], r. robinson[2], a. swaroop[4], d. schaumberg[6], kh. park[9], ee. tsironi[8], g. silvestri[5], ik. kim[3], r. chen[7], c. huff[2], g. jun[1], m. deAngelis[2].* 1) Biomedical Genetics, Boston University, Boston, MA; 2) John A. Moran Eye Center, University of Utah, Salt Lake City, Utah; 3) Massachusetts Ear and Eye Infirmary, Boston, MA; 4) University of Michigan Kellogg Eye Center, Ann Arbor, Michigan; 5) Centre for Vision & Vascular Science Institute of Clinical Science, Queen's University, Belfast, Ireland; 6) Brigham and Women's Hospital, Boston, MA; 7) Baylor University, Houston, TX; 8) University Hospital of Larissa, Larissa, Greece; 9) University of Seoul, South Korea.

Age-related macular degeneration (AMD) is the most common cause of blindness in the elderly. Approximately 50% of the genetic variance for AMD is explained by common SNPs suggesting that rare and other types of variants contribute to AMD risk. We analyzed whole exome sequence (WES) data from 20 unrelated Caucasians with well characterized and documented neovascular ("wet") AMD. Data were obtained using an Illumina GA sequencer and subsequently mapped to the human genome version HG19 using BWA software. Sequence data were analyzed using VAAST (v1.2.0) which assigns a likelihood score to each non-synonymous variant. Variant scores within a gene were binned, thereby combining the power of multiple rare variants. The binned scores for the AMD sample and scores obtained from 1,092 control subjects in the 1000 Genomes Project database were permuted 1×1010 times to obtain an empirical P-value which is 2.78×10-6 after adjustment for multiple comparisons. Using this approach, we observed that the distribution of rare variants in RP1L1 was significantly different between cases and 1000 Genomes Project controls (p<3.7×10-10). Subsequent analysis revealed a novel 3-bp insertion in RP1L1 (T=>TTTC adding a lysine at 8:10467629–10467631) in 18/20 AMD cases which not present in the 1000 Genomes sample but recently added to dbSNP (rs146656804). RP1L1, the retinitis pigmentosa 1-like 1 gene, has an essential role in affecting photosensitivity and outer segment morphogenesis of rod photoreceptors. Mutations in this gene cause occult macular dystrophy. Analysis of the protein structure using Chimera software revealed that the insertion occurs precisely at the turn in a beta sheet. We also observed a novel 3-bp insertion in SAAL1 in 6/20 AMD cases. SAAL1 encodes the α-amyloid protein that undergoes phosphorylation in response to damage and has recently been implicated in the inflammatory response. These results suggest that analysis of WES data will uncover novel mechanisms and pathways leading to AMD.

386

The role of SIX6 in primary open-angle glaucoma. *M. Ulmer[1], B. Whigham[1], D. Parker[2], X. Qin[1], N. Katsanis[2], Y. Liu[1], A. Ashley-Koch[1], R. Allingham[3], M. Hauser[1], NEIGHBOR consortium investigators.* 1) Center for Human Genetics, Duke University, Durham, NC; 2) Center for Human Disease Modeling, Duke University, Durham, NC; 3) Department of Ophthalmology, Duke University, Durham, NC.

Primary open-angle glaucoma (POAG) is the second leading cause of blindness world-wide. Previous genome-wide association studies, in the combined NEIGHBOR/GLAUGEN dataset and others, have identified a region on chromosome 14q23 that is significantly associated with POAG risk (top SNP rs10483727,p-value=3.9 X 10–11, OR= 1.32) as well as POAG quantitative endophenotypes involving optic nerve measurements. This locus contains two homeobox genes, SIX1 and SIX6, which are known to play a role in ocular development. To identify potential causal variants, we sequenced the exons and flanking regions of the SIX1 and SIX6 loci in 262 POAG cases and 279 POAG controls. In SIX1, we identified two rare nonsynonymous changes in two control individuals (Asn193Ile, Pro249Leu). In SIX6, we identified six nonsynonymous changes including five novel variants in POAG cases and controls (Glu93Gln, Glu129Lys, His141Asn (rs33912345), Leu205Arg, Thr212Met, and Ser242Ile). Using an allele-based chi-squared test, rs33912345 was significant associated with POAG (p-value=0.0005, OR=1.54). We replicated this association, using a logistic regression model adjusted for age and gender, in a larger case-control dataset consisting of 482 POAG cases and 433 POAG controls (p-value= 0.005, OR=1.40). We performed ordered subset analysis case-control (OSACC), a method for performing a series of stratified analyses, using an important quantitative POAG risk factor, intraocular pressure. Rs33912345 showed increased evidence of association (permutation p-value=0.00008, OR=1.71) in a subset of 206 cases with elevated intraocular pressure (>27mm hg). In order to functional characterize the SIX6 gene, we have designed morpholino antisense oligonucleotides for targeted knockdown using microinjections in developing zebrafish. Preliminary results show structural abnormalities, although more work is needed to fully elucidate the effect the variants identified in this study have in the eye and what role they may play in POAG susceptibility.

387

Topical ocular sodium 4-phenylbutyrate rescues glaucoma in a mouse model of primary open angle glaucoma. *G.S. Zode[1,2], K.E. Bugge[1,2], E.M. Stone[1,3], V.C. Sheffield[1,2,3].* 1) HHMI, University of Iowa, Iowa City, IA.,52241; 2) Departments of Pediatrics,University of Iowa, Iowa, IA 52241; 3) Ophthalmology and Visual Sciences, University of Iowa, Iowa, IA 52241.

Primary open angle glaucoma (POAG) is a common chronic optic neuropathy, characterized by progressive loss of retinal ganglion cell (RGC) axons with the resultant irreversible loss of vision. Mutations in the myocilin gene (MYOC) are the most common known genetic cause of POAG. We developed a transgenic mouse model of POAG (Tg-MYOC^{Y437H}), which expresses human mutant MYOC in the trabecular meshwork (TM), a tissue that regulates intraocular pressure (IOP). Adult Tg-MYOC^{Y437H} mice have elevated IOP and display progressive loss of RGCs similar to POAG patients. In contrast to normal myocilin, we show that mutant myocilin is not secreted into the aqueous humor and accumulates in the endoplasmic reticulum (ER) of the TM, inducing ER stress. We further demonstrate that induction of ER stress is associated with elevation of IOP in adult Tg-MYOC^{Y437H} mice. We explored whether reducing chronic ER stress rescues the glaucoma phenotype of Tg-MYOC^{Y437H} mice by topically treating mice with the chemical chaperone, sodium 4-phenybutyrate (PBA). At 3 months of age, vehicle-treated Tg-MYOC^{Y437H} mice exhibited elevated IOP compared to WT littermates (n=24, p<0.0001). Topical PBA significantly reduced elevated IOP in Tg-MYOC^{Y437H} mice to the level of WT mice. PBA-treated Tg-MYOC^{Y437H} mice also preserved RGC function compared to vehicle-treated Tg-MYOC^{Y437H} mice. Analysis of the myocilin protein in the aqueous humor and the TM revealed that PBA significantly improved the secretion of mutant myocilin and reduced ER stress in the TM of Tg-MYOC^{Y437H} mice. Furthermore, we demonstrate that PBA treatment results in improvement of TM cell function. Of note, we demonstrate that ER stress markers are also increased in glaucomatous human TM tissues obtained from postmortem POAG donors. These combined mouse and human data suggest that topical ocular PBA is a potential treatment for POAG patients.

388

Meta-analysis of GWAS on corneal thickness identifies a total of 27 associated loci, including six risk loci for eye disease keratoconus. *S. Macgregor on behalf of the CCT Consortium.* Statistical Genetics Laboratory, Queensland Institute of Medical Research, Brisbane, Australia.

The human ocular biometric parameters comprise a set of highly heritable and often correlated quantitative traits. One notable example is central corneal thickness (CCT) which has ~95% heritability. CCT is associated with eye conditions including keratoconus and glaucoma. Previous genome-wide association studies (GWAS) have identified 11 CCT-associated loci. Among these loci, mutations in ZNF469, COL5A1 and COL8A2 are known to cause rare eye disorders. However, none was found to be associated with common eye diseases. We performed a meta-analysis on >20,000 individuals in European and Asian populations which identified 27 (16 novel) CCT-associated loci at genome-wide significant level (P<5×10-8). Most loci associated in Europeans also influenced CCT in Asians, with the 27 loci collectively explaining 8.3% and 7% of additive variance in Europeans and Asians, respectively.

To evaluate potential clinical relevance, we tested the 27 identified CCT loci in two studies on keratoconus risk (874 cases; 6085 controls) and three studies on primary open angle glaucoma risk (2979 cases; 7399 controls). We found that despite the modest effect on CCT, 11 SNPs showed nominal association with keratoconus, with 6 genes significant after correction for multiple testing. Two SNPs exceeded genome-wide significance (P<5×10-8) for keratoconus, with one of these also associated with primary open angle glaucoma (P=5.6×10-4). Our data clearly show that the endophenotype approach yields disease relevant loci. The 6 SNPs that were associated with keratoconus each had a relatively large odds ratio (range: 1.25–1.62). Since all 6 SNPs had moderately high risk allele frequencies, a risk profile based on even just these SNPs yields a reasonable risk prediction; e.g. ~1% of the population carry 1 or 2 risk loci at each of the 6 loci and these individuals are at 7.2 fold increased risk (assuming a multiplicative model) of keratoconus relative to the ~1% of the population homozygous for the protective allele at each SNP. Further evaluation of the clinical relevance of these SNPs is merited. Finally, using a novel pathway analysis tool we show that CCT loci converge to collagen and extracellular pathways. The collagen pathway remained significant even after the removal of the genome-wide significant collagen genes, suggesting additional collagen genes influence CCT.

389

Mouse models reveal an essential role for RERE in eye development. *B. Kim[1], Z. Yu[1], O. Shchelochkov[2], M. Justice[1], B. Lee[1], D. Scott[1].* 1) Molecular & Human Genetics, Baylor College Med, Houston, TX; 2) Department of Pediatrics, University of Iowa, Iowa IA.

Microphthalmia occurs in approximately 1 out of 10,000 individuals and can be caused by alterations in genes involved in the early development of the eye, most of which have not been identified. In an effort to identify novel genes involved in the development of microphthalmia, we carried out an autosomal recessive ENU screen. This screen resulted in the identification of a novel mouse stain (*eye3*) with microphthalmia. The mutation responsible for the *eye3* phenotype was mapped by linkage analysis to a region of mouse chromosome 4 that is synthetic to human chromosome 1p36.31-p36. Rere (arginine-glutamic acid dipeptide repeats) was selected as a positional candidate based on phylogenic profiling and sequencing revealed a homozygous c.578T>C change in *Rere*, which produces a single amino acid change in a highly conserved BAH domain of RERE (p.Val193Ala). To confirm that the microphthalmia seen in the *eye3* strain was due to a defect in *Rere*, these mice were crossed with mice carrying an RERE null allele (*om*). Homozygous *om* embryos ($Rere^{om/om}$) die around E9.5, but a portion of $Rere^{om/eye3}$ mice lived into adulthood but have microphthalmia and optic nerve atrophy. Next, we examined the expression pattern of RERE in the mouse retina. At E17.5, RERE expressing cells were primarily located in the ganglion cell layer and at postnatal day 14 (P14), RERE was detected in the outer nuclear layer (ONL), inner nuclear layer (INL), and ganglion cell layer (GCL). Further histological examination revealed that each layer of the retina was normally formed in $Rere^{om/eye3}$ mice at P21. However, in adult $Rere^{om/eye3}$ mice, the thickness of the retina was significantly reduced and the GCL was not identified. Apoptosis assays revealed increased numbers of apoptotic cells in the ganglion cell layer of the retina as early as E16.5 in $Rere^{om/eye3}$ embryos while the proliferation of precursor cells in the neuroblast layer was not altered. In addition, number of brn-3 positive ganglion cells was reduced starting at E17.5 and the number of Neu-N positive cells in the ganglion cell layer was also reduced in adult $Rere^{om/eye3}$ mice. We conclude that RERE is required for normal eye development and that the microphthalmia, retinal degeneration, and optic nerve atrophy seen in adult mice with RERE deficiency is triggered by abnormal loss of ganglion cells in the retina. It is possible that abnormalities in RERE may contribute to the development of similar phenotypes in humans.

390

Mutations in the Nuclear NAD synthesizing enzyme NMNAT1 cause autosomal recessive Leber congenital amaurosis with early-onset severe macular atrophy and optic atrophy. *J. Rozet[1], I. Perrault[1], S. Hanein[1], X. Zanlonghi[2], V. Serre[1], M. Nicouleau[1], S. Defoort-Delhemmes[3], N. Delphin[1], L. Fares-Taie[1], S. Gerber[1], O. Xerri[1], C. Edelson[4], A. Goldenberg[5], A. Duncombe[5], G. Le Meur[6], C. Hamel[7], E. Silva[8], P. Nitschke[9], P. Calvas[10], A. Munnich[1], O. Roche[11], H. Dollfus[12], J. Kaplan[1].* 1) INSERM U781, Department of Genetics, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, Paris, France; 2) Vision Functional Exploration Department, Clinique Sourdille, Nantes, France; 3) Service d'Exploration de la Vision et Neuro-Ophtalmologie, Hôpital Roger Salengro, Centre Hospitalo-Universitaire Régional, Lille, France; 4) Department of Ophthalmology, Fondation Ophalmologique Adolphe de Rothschild, Paris, France; 5) Department of Genetics, Centre Hospitalo-Universitaire, Rouen, France; 6) Department of Ophthalmology, Centre Hospitalo-Universitaire, Nantes, France; 7) Institut des Neurosciences, Hôpital Saint Eloi, Montpellier, France; 8) Department of Ophthalmology, Coimbra University Hospital, Coimbra, Portugal; 9) Bioinformatics Plateform, Université Paris Descartes-Sorbonne Paris Cité, Paris, France; 10) Department of Medical Genetics, Purpan Hospital, Centre Hospitalo-Universitaire, Toulouse, France; 11) Department of Ophthalmology, Université Paris Descartes-Sorbonne Paris Cité, Paris, France; 12) Department of Clinical Genetics, Strasbourg University Hospital, Strasbourg, France.

Introduction : Leber congenital amaurosis (LCA) is the earliest and most severe retinal dystrophy, responsible for neonatal blindness. In an effort to identify the yet unknown molecular bases of the disease, we subjected to whole exome resequencing (WER) the DNA of five unrelated families excluding known genes. Methods and Results: Through WER combined with Sanger sequencing, we identified NMNAT1 mutations in 22 unrelated LCA cases. NMNAT1 codes for the ubiquitously expressed nuclear isoform of nicotinamide mononucleotide adenyltransferases. In addition to its NAD+ synthesizing activity, NMNAT1 acts as a chaperone that protects against neuronal activity-induced degeneration. The review of ophthalmological data indicated that NMNAT1 mutations consistently caused LCA with a particular phenotype characterized by severe neonatal atrophy of the central retina leading to An aspect of posterior pole coloboma and early-onset optic nerve atrophy. These findings add further complexity to the physiopathological bases of the most common cause of blindness in childhood by giving support to the view that neuroprotection against light-induced stress is required to allow maintaining photoreceptor cells. Conclusion : In summary, we report NMNAT1 mutations in 22 LCA families clinically recognizable by the existence of a neonatal macular atrophy and early-onset optic atrophy.

391

RNA-DNA Differences in miRNA transcriptome of retina and retinoblastoma. *A. Ganguly[1], J. Leipzig[2], J. Richards[1], J. Purrazzella[1], T. Ganguly[3].* 1) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 2) Bioinformatics Core Center for Biomedical Informatics The Children's Hospital of Philadelphia; 3) DNA Sequencing Core, Department of Genetics, University of Pennsylvania.

Background: Deep sequencing of the transcriptome has revealed editing, and non-canonical RNA-DNA differences in non-coding RNA (ncRNA) sequences. This phenomenon is expected to add a previously unexplored layer of gene regulation and has been under-explored in tumorigenesis. Retinoblastoma (RB) is a childhood ocular malignancy that can lead to blindness and death. Previous studies implicated the let-7 family of miRNAs as ncRNAs involved in the etiology of RB. Methods: We performed deep sequencing of miRNA isolated from three normal retinas, five RB samples, and two RB cell lines on the Illumina HiSeq platform. We aligned the sequences to miRNA hairpin and genomic reference sequences using the Novoalign program with Needleman-Wusch global optimization, and selected the best hit with up to two mismatches. To mitigate cross-mapping issues, reads with a perfect alignment to other genomic regions and mature tRNAs were omitted from the analysis. Results and Discussion: Members of the let-7 family of miRNAs, among others, were differentially expressed with high significance in RB compared to normal retina. Differences in RNA sequences with respect to genomic sequences were detected within the mature miRNA sequences and at the 3'-end in let-7 family of miRNAs among others. For let{hyphen}7b, in one sample of normal retina, 3469 reads carried a nucleotide difference, T>G (1795), T>C (1387) and T>A (287) at position 6, while 238 and 488 reads carried A>G substitution (canonical editing) at positions 7 and 10 of the mature miRNA sequence. Thus, the majority of the nucleotide differences represent non-canonical RNA editing while a small fraction includes canonical events. The sequence results were confirmed on the Ion Torrent PGM platform using a different sequencing chemistry. Sanger sequencing of the corresponding genomic DNA region from the tissues tested, showed the consensus sequence ruling out presence of single nucleotide variations causing the differences in the miRNA sequences. Alterations in nucleotides involved around the miRNA seed sequence predicted altered specificity for their targets and were confirmed by real-time PCR assays. Importantly, the proportion of variant to normal miRNAs is significantly different in normal retina versus RB and is highly reproducible. Further analysis indicated inactivation of RB1 gene has a role in altering the proportion of RNA-DNA differences.

392

Knock-in of human *KIAA0649P* into the mouse *Rb1* locus: Modelling the mechanism of imprinted *RB1* expression in humans. *L. Steenpass, D. Kanber, M. Hiber, K. Buiting, D. Lohmann, B. Horsthemke.* Institute of Human Genetics, University Hospital Essen, University Duisburg-Essen, Essen, Germany.

The human retinoblastoma gene, *RB1*, is imprinted. Gene expression is skewed in favour of the maternal allele. This is due to parent-of-origin specific DNA methylation on the truncated and inverted pseudogene *KIAA0649P*. This pseudogene evolved after integration into intron 2 of the *RB1* gene. In its present form it harbors a CpG island, CpG85, which serves as promoter for an alternative *RB1* transcript, transcript 2B. CpG85 is methylated on the maternal allele only and, as expected, expression of transcript 2B is restricted to the paternal allele. Transcription of the paternal transcript 2B interferes with transcription of the regular *RB1* transcript on the same allele.

The mouse *Rb1* gene does not contain *KIAA0649P* and is not imprinted. To determine if the integration of *KIAA0649P* is sufficient to result in skewed expression of *Rb1*, we generated a knock-in of human *KIAA0649P* in intron 2 of the murine *Rb1* gene. To be able to distinguish expression of the two *Rb1* alleles, mouse embryonic stem (ES) cells were first modified such that they contain a single nucleotide variant in exon 3 of *Rb1*.

In the generated ES cell lines we analyzed the methylation status of the *Rb1* promoter, a second CpG island in *KIAA0649P*, CpG42, and CpG85 by next generation bisulfite sequencing. The *Rb1* promoter remains free of DNA methylation (mean level about 2%), whereas CpG42 and CpG85 acquire methylation at a mean level of about 17% and 8%, respectively. Using RT-PCR, expression of transcripts over CpG85 could be detected, but could not be connected to downstream exons of *Rb1* so far. Using the single nucleotide variant in exon 3 of Rb1, quantitative primer extension analyses (SNaPshot) revealed skewing of *Rb1* expression in favour of the wildtype *Rb1* allele, not carrying *KIAA0649P*.In summary, our data provide evidence that the retrotransposition of *KIAA0649P* into the *RB1* gene is sufficient to skew *RB1* expression. It remains to be determined how *KIAA0649P* acquired an imprint during evolution.

393

Gene therapy provides long-term visual function in a pre-clinical model of retinitis pigmentosa. *K.J. Wert[1, 2], R.J. Davis[1], S.H. Tsang[1].* 1) Bernard & Shirlee Brown Glaucoma Laboratory, Departments of Ophthalmology, Pathology & Cell Biology, College of Physicians & Surgeons, Columbia University, New York, NY; 2) Institute of Human Nutrition, College of Physicians & Surgeons, Columbia University, New York, NY.

Approximately 36,000 cases of simplex and familial retinitis pigmentosa (RP) worldwide are caused by a loss in phosphodiesterase (PDE6) function. In a preclinical model of this disease, the *Pde6*α^{nmf363} mouse, defects in the α-subunit of PDE6 results in a progressive loss of photoreceptors and visual function. We hypothesized that increasing PDE6α levels using an AAV2/8 gene therapy vector could improve photoreceptor survival and retinal function. We utilized a vector with the cell-type specific *opsin* promoter: AAV2/8(Y733F)-*Rho-Pde6*α, to transduce *Pde6*α^{nmf363} mutant retinas and monitor its effects over a three-month period. We used western analysis to measure PDE6α levels, histology to observe photoreceptor survival, and electroretinograms (ERGs) to measure both photoreceptor-specific and inner retina visual function. We found that a single injection of AAV2/8(Y733F)-*Rho-Pde6*α increases PDE6α levels, enhances survival of photoreceptors, and improves retinal function. At four months of age, a transduced mutant eye displayed an approximate three-fold greater number of photoreceptors in the outer nuclear layer (ONL) than mutant untreated retinae, which exhibited a loss of the ONL. At two months of age, ERGs displayed approximately ten-fold higher photoreceptor-mediated a-wave amplitudes in the transduced eyes compared to controls. This efficacy of the ERG response was observable through three months of age, and has persisted in one mouse for at least five and a half months. These results suggest that RP due to PDE6α deficiency in humans, in addition to PDE6β deficiency, is also likely to be treatable by gene therapy.

394

The genomic landscape of childhood pre-B acute lymphoblastic leukemia. *J. Spinella[1], R. Vidal[1], J. Healy[1], V. Saillour[1], E. Bareke[1], C. Richer[1], S. Busche[2,3], B. Ge[3], T. Pastinen[2,3], D. Sinnett[1,4]*. 1) Hematology/Oncology, Sainte-Justine UHC Research Center, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University, Canada; 3) McGill University and Genome Quebec Innovation Center, Canada; 4) Department of Pediatrics, Faculty of Medicine, University of Montreal, Canada.

Precursor B-cell acute lymphoblastic leukemia (pre-B ALL) is the most frequent pediatric cancer. Increased understanding of the pathobiology of B-cell ALL has led to risk-targeted treatment regimens and increased survival rates. However, the underlying causes of this pediatric cancer are still unclear. We are using next-generation sequencing technologies combined with a unique quartet design that involves sampling patient tumor (at diagnosis) and normal (remission) material, as well as DNA from both parents, to decipher the genomic landscape of pre-B ALL and build a catalogue of variations (germline and somatic) involved in pediatric ALL onset and/or progression. Here, we report the first deep-sequencing of the whole exomes of over 30 childhood B-cell ALL quartets and the identification of recurrent somatic mutations with the potential to drive leukemogenesis. Using Agilent SureSelect target enrichment and multiplexed paired-end sequencing (ABI SOLiD 4), we generated over 4.0Gb of sequence on average per sample with a mean coverage of over 30X. To identify driver mutations in the childhood ALL genomes that are causally implicated in leukemogenesis, we used an integrative computational biology approach to combine multiple sources of information and reveal putative drivers: crossing our DNAseq data with a RNAseq SNP calls led to the identification of pre-validated variants, genome-wide genotyping was used as quality control and public databases were used to rank tumor-specific variants. We were also able to incorporate parental sequence information to identify inherited rare/private variants, and to leverage the identification of recurrent leukemia-specific variants within the ALL cohort. Here we describe the somatic mutational landscape of childhood ALL and present a comprehensive catalogue of both recurrent and private leukemia-specific events, and highlight biologically-relevant pathways (chromatin remodelling, DNA damage control) and genes with increased mutational burden of loss-of-function variants. Functional validation of putative driver mutations is currently under way. Understanding the mechanisms through which these genes contribute to childhood ALL, although challenging, offers exciting and promising advances in biomedicine, by providing new insights into the genetic and molecular basis of leukemogenesis and by helping improve detection, diagnosis and treatment of this childhood cancer.

395

Genomic analysis of serial chronic lymphocytic leukemia samples suggests that epigenetic changes, rather than clonal evolution, drive progression of disease. *E.N. Smith[1,2], C. DeBoever[2,3], L. Rassenti[4], E. Ghia[4], S. Rozenzhak[1,2], P. Shepard[1,2], H. Alakus[1,2], O. Harismendy[1,2], C. Barrett[1,2], T.J. Kipps[4], K.A. Frazer[1,2,5]*. 1) Pediatrics and Rady Children's Hospital, University of California San Diego, La Jolla, CA; 2) Moores UCSD Cancer Center, University of California San Diego, La Jolla, CA; 3) Bioinformatics and Systems Biology, University of California San Diego, La Jolla, CA; 4) Department of Medicine, University of California San Diego, La Jolla, CA, 92093; 5) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, 92093.

Chronic lymphocytic leukemia (CLL) is the most common adult leukemia in the United States. CLL has a variable course ranging from long-term indolent to aggressive. To identify important genetic and epigenetic changes that specifically underlie the transition from indolent to aggressive disease, we are using genomic methods to analyze serial liquid tumor samples collected from 27 patients who required treatment from 9 months to 10 years post diagnosis. From the UCSD arm of CLL Research Consortium Biorepository, which is composed of close to 1000 patients, tumor samples with high tumor purity and sample dates within 1 year post-diagnosis and 1 year pre-treatment were selected. We have performed whole-exome sequencing (Agilent 50Mb capture, 100X), micro-RNA (miRNA) and RNA sequencing on the Illumina HiSeq 2000 as well as copy-number (Omni 2.5M) and methylation (450K) arrays. Unexpectedly, exome sequencing and copy-number analysis demonstrate that the acquisition of additional somatic mutations does not typically drive CLL progression. In most cases, the clonal population structure remains stable throughout the progression of the disease with multiple clones per patient. This suggests that clonal evolution is not a driving force in CLL progression and implicates epigenetic and regulatory alterations as important drivers of progression in CLL. To identify these changes, we are analyzing differential methylation patterns as well as miRNA and messenger RNA isoform expression differences across the serially collected tumor samples. In summary, our results suggest that CLL progression unexpectedly occurs in the absence of clonal evolution or genetic alteration but rather is a result of alterations in the epigenome altering RNA expression.

396

Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *A. Fujimoto[1], Y. Totoki[2], T. Abe[1], K.A. Boroevich[1], F. Hosoda[2], H.H. Nguyen[1], M. Aoki[1], N. Hosono[1], M. Kubo[1], F. Miya[1], Y. Arai[2], H. Takahashi[2], T. Shirakihara[2], M. Nagasaki[3], T. Shibuya[3], K. Nakano[1], K. Watanabe-Makino[1], H. Tanaka[3], H. Nakamura[2], K. Chayama[1,4], N. Kamatani[1], S. Miyano[3], H. Nakagama[5,6], Y. Nakamura[1,7], T. Tsunoda[1], T. Shibata[2], H. Nakagawa[1]*. 1) Ctr Genomic Med, Riken, Yokohama, Japan; 2) Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan; 3) Laboratory of DNA Informatics Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 4) Department of Medicine & Molecular Science, Hiroshima University School of Medicine, Hiroshima, Japan; 5) Division of Molecular Pathology, National Cancer Center Research Institute, Tokyo, Japan; 6) Division of Cancer Development System, National Cancer Center Research Institute, Tokyo, Japan; 7) Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related death worldwide. We sequenced and analyzed the whole genomes of 27 HCCs, 25 of which were associated with hepatitis B or C virus infections, including two sets of multicentric tumors. Although no common somatic mutations were identified in the multicentric tumor pairs, their whole-genome substitution patterns were similar, suggesting that these tumors developed from independent mutations, although their shared etiological backgrounds may have strongly influenced their somatic mutation patterns. Statistical and functional analyses yielded a list of recurrently mutated genes. Multiple chromatin regulators, including *ARID1A*, *ARID1B*, *ARID2*, *MLL* and *MLL3*, were mutated in ~50% of the tumors. Hepatitis B virus genome integration in the *TERT* locus was frequently observed in a high clonal proportion. Our whole-genome sequencing analysis of HCCs identified the influence of etiological background on somatic mutation patterns and subsequent carcinogenesis, as well as recurrent mutations in chromatin regulators in HCCs.

397

Breast cancer evolution revealed by deep whole-genome sequencing of early neoplasias and their concurrent carcinomas. *A. Sidow[1,3], D. Kashef-Haghighi[2], D. Newburger[2], Z. Weng[1], T. Sweeney[1], S. Batzoglou[2], R. West[1]*. 1) Pathology Dept, Stanford Univ, Stanford, CA; 2) Computer Science Dept, Stanford Univ, Stanford, CA; 3) Genetics Dept, Stanford Univ, Stanford, CA.

We deep-sequenced (50x whole genome coverage and more) concurrent precancerous lesions, normal breast tissue, and the clinically diagnosed breast carcinomas, from six mastectomy cases. From the constituent 31 whole genome sequences we identified somatic mutations and aneuploidies that (1) serve as markers for and (2) possibly drive cancer progression. Our studies begin to shed light on the earliest genomic events that occur in founder cells of eventual breast cancer, long before development of an invasive carcinoma. (1) Using the somatic mutations as lineage markers, we built lineage trees to establish the evolution of the cancer from normal breast tissue to precancerous lesions to carcinoma. Lineage trees are analogous to species trees in phylogenetics, as they allow the inference of phenotypes in ancestors and of the order and timing of evolutionary events as a function of lineage evolution. In a subset of cases, the last common ancestor of a neoplasm and carcinoma had undergone a genomic crisis that resulted in several aneuploidies and a burst of somatic mutations. (2) In contrast to highly advanced tumors that are the focus of much of current cancer genome sequencing, these early lesions, and the carcinomas as well, harbor a striking paucity of possibly functional somatic mutations. The earliest significant events that mark these genomes, those changes that occurred in a common ancestor of neoplastic and tumor cells, appear to be aneuploidies. Several aneuploidies are recurrent, suggesting that they are the earliest driver events.

398

Intra-Tumor Genetic Heterogeneity in Cancer Tissues: The Key to Assessing it's Significance is the Distribution Profile of Gene Variants not just their Presence in Tumors. *B. Gottlieb[1,2,3], C. Alvarado[1], C. Wang[5], B. Gharizadeh[5], F. Babrzadeh[5], L.K. Britel[1,2,3,4], M. Trifiro[1,2,3,4]*. 1) Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, PQ, Canada; 2) Segal Cancer Center, Jewish General Hospital, Montreal, PQ, Canada; 3) Dept of Human Genetic, McGill University, Montreal, PQ, Canada; 4) Dept of Medicine, McGill University, Montreal, PQ, Canada; 5) Stanford Genome Technology Center, Stanford University,, Palo Alto, CA.

The discovery of intra-tumor genetic heterogeneity in cancer tissues has had a significant effect on cancer genetics by making it even more difficult to assess, which gene alterations are drivers and which are passengers. To resolve this issue, we have developed a new approach to sequencing that has allowed us to create a distribution profile of mutant variants within tumors. First, we laser capture micro-dissect breast tumor cells, adjacent histologically normal breast epithelia cells, and matching peripheral blood leukocytes. Then, using next generation sequencing we have examined up to 37,000 reads of an androgen receptor gene (*AR*) CAG repeat length, which is a functional polymorphism associated with breast cancer risk. This has allowed us to measure the *frequency of distribution* of *AR* CAG repeat length variants within tumor tissues, rather than just their presence. Each of the tissue types had significantly different frequencies of distribution of CAG repeat lengths as measured by paired, twin t-tests. Our results, indicate that zero CAG repeat lengths that are prominent in normal breast tissues, are possibly protective against breast cancer and that preferential selection at menopause, of pre-existing, but minority populations of breast cells, that lack zero CAG repeat lengths, may contribute to breast carcinogenesis. This selection process possibly being the result of the changing breast tissues hormonal concentrations associated with menopause. Thus, while many studies have reported that longer *AR* CAG repeat lengths are associated with breast cancer, they have failed to reveal the actual process involved. Clearly, analysis which calculates the significance of CAG repeat lengths, by treating each length as equally important, gives little insight into the real relationship between *AR* CAG repeat length and breast cancer. We have also examined the frequency of actual mutations in *AR* in breast cancer tissues, and while we have yet to complete our analysis of the frequency of specific mutations and their association with breast cancer, initial data indicate that the relationship of *AR* mutations and breast cancer is complex and cannot be explained solely by the *presence* of specific mutations within breast cancer tissues. Clearly, understanding intra-tumor genetic heterogeneity and it's possible relationship to cancer ontogeny, will require analysis of the frequency of distribution of gene variants and not just their presence in tumors.

399

Next Generation Sequencing and Chromosomal Microarray Analysis Provide Novel Insight into the Genomic Landscape of Metastatic Breast Cancer. *M. Li[1], Y. Wen[1], E. Fang[1], Y. Li[1], P. Chen[1], G. Douglas[1], C. Carmack[1], K. Osborne[2]*. 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) The Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX.

Although most breast cancer cases are curable nowadays, approximately 25% of breast cancer patients develop metastatic disease and eventually die of breast cancer. To understand the mechanism of tumor metastasis and identify potential prognostic and therapeutic biomarkers, we studied 10 metastatic breast cancer tumors with paired blood samples using Next Generation Sequencing (NGS) and Chromosomal Microarray Analysis (CMA) technologies. The NGS mutation panel interrogates 739 common mutations in 46 key cancer genes including many clinically actionable mutations using a technology that merges multiplex PCR with ion semiconductor sequencing. The arrays used in the study are custom-designed CGH+SNP arrays that target approximately 2,300 cancer genes or cancer-associated genes with an average of 6 probes per exon and over 130 cancer-associated genomic regions. NGS identified 10 mutations in 7 tumor samples including mutations in TP53, PIK3CA, APC, ATM, MET, NRAS, and EGFR. CMA showed genomic aberrations in all tumor samples. The most common genomic alterations were 1q duplication/amplification observed in 9 of 10 cases followed by 8q duplication/amplification, part or full chromosome 5 duplication, and 17p deletion. Other common aberrations were 20q duplication/amplification, 9p deletion, and deletion, duplication, or amplification of long arm or short arm of chromosome 16. TP53 mutations appear to be associated with complex genomic alterations involving many chromosomes, especially 8q duplication/amplification, 7p deletion, and 20q duplication/amplification, and poor prognosis. Studies of primary tumor and tumor samples before and after treatment showed similar genomic alteration patterns, suggesting that the genomic alterations identified in primary tumors could be used to predict tumor metastasis and choose appropriate treatment regimens. In addition, mutations identified may help in selecting an effective targeted therapy.

400

The 3D topographic mapping of genetic variations in treatment naïve advanced ovarian cancer. *E. Cuppen[1], M. de Pagter[1], M. Hoogstraat[2], G. Cirkel[2], J. Kreeftmeijer[1], C. Lee[3], E. Levandowsky[3], T. Guy[1], K. Duran[1], R. 't Slot[1], T. Jonges[4], S. van Lieshout[1], M. Lolkema[2], R. Zweemer[5], M. Koudijs[2], I. Nijman[1], E. Voest[2], T. Harkins[3], W. Kloosterman[1]*. 1) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands; 2) Department of Medical Oncology, University Medical Center Utrecht, The Netherlands; 3) Life Technologies, Beverly, USA; 4) Department of Pathology, University Medical Center Utrecht, The Netherlands; 5) Department of Gynaecology, University Medical Center Utrecht, The Netherlands.

Ovarian cancer (OC) is a leading cause of death from gynaecological malignancies, with approximately 225,000 new cases and 140,000 deaths reported worldwide yearly. OC has a unique intra-abdominal metastatic pattern and is associated with high morbidity, over 70% of patients present with advanced disease (stage III/IV) at diagnosis. Advances in genome sequencing have expanded our understanding of the complexity of carcinogenesis, and have given insight in inter- and intra-individual heterogeneity of cancer genomes. Here, we present the first systematic and comprehensive study on 3D topography of structural variations and point mutations within tumours. Analysis is currently ongoing to demonstrate if these processes occur in parallel or independent of each other during tumour evolution. We obtained tumour tissue from 10–20 sites per patient from the primary lesion as well as metastatic lesions. Structural- and copy number variations are detected with high sensitivity by next-generation long mate-pair sequencing on AB SOLiD and SNVs are identified by targeted sequencing of a selected set of 2000 cancer-related genes. Furthermore, we used the IonTorrent comprehensive cancer panel TargetSeq assay to deep sequence 409 oncogenes for lower penetrance variation. By RNA sequencing the effect of these somatic variations on gene expression was determined. Bioinformatic cross-sample analysis revealed extensive genetic and functional heterogeneity between primary and metastatic lesions, in line with observations for other tumor types. Integrative analysis is currently ongoing and is expected to provide detailed insight into tumour evolution and the types of variation that drive tumour formation and metastatic spread.

401

Transcriptome sequence analysis of human colorectal cancer samples to reveal functional attributes. *H. Ongen[1], T.F. Orntoft[2], B. Oster[2], L. Romano[1], A. Planchon[1], C.L. Andersen[2], E.T. Dermitzakis[1]*. 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Geneva, Switzerland; 2) Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark.

In developed countries colorectal cancer (CRC) is the second leading cause of cancer death with a million new cases worldwide yearly and a mortality rate of ~50%. Among cancers, the incidence of CRC ranks fourth in men and third in women and will grow with the increase of westernized lifestyles. Systems Biology of Colorectal Cancer (SYSCOL) utilizes systems biology tools to develop a quantitative and comprehensive model of CRC tumorigenesis. Here we will report the RNA-seq analysis of matched tumor and normal mucosa from 56 CRC SYSCOL patients and 20 other normal tissues from unrelated healthy donors. The normal tissues are used to assess how CRC differentiates not only from its normal counterpart but also from other normal tissue types. RNA was purified from microdissected samples (tumour percentage ≥ 60%) and sequenced on the Illumina HiSeq (34–80 million reads). The DESeq package was used for differential expression. The correlation of gene RPKM values shows that a tumor sample clusters with other tumor samples instead of its originating normal tissue and is consistently similar to colon, small intestine and dissimilar to brain, testes. The correlation between the normal colon samples is higher and the variance lower than between the tumor samples. One patient had three synchronous tumors and their transcriptome signatures are distinguishable. The preliminary analysis of 24 samples indicates that between normal colon and cancer samples we observe 2282 differentially expressed genes ($P < 3.46 \times 10^{-6}$). Moreover, tumors, normal colons, and other tissue samples form three distinct clusters, and tumors cluster in between the other two, suggesting that cancer increases the variance of the transcriptome but in a predictable manner. On average there are 1558 significant allele-specific expression (ASE) signals ($P < 0.05$) per sample. The proportion of heterozygous sites that have an ASE effect is significantly more in tumors. About 8% of the ASE sites are tumor specific and ~24% shared ASE sites between normal and tumor exhibit reversal of the effect. We are investigating how functional variation affects splicing to discover variation resulting in cancer-specific transcripts and observe 14 somatic splice sites mutations causing cancer specific transcription. Cancer specificity of putative locus control regions, identified by co-expression of genes, is being assessed. Altogether these results will greatly benefit our understanding of colorectal tumorigenesis.

402

Regulatory regions are somatic mutation cold spots in cancer genomes. *S. Sunyaev[1,2,3], P. Polak[1,2,3], M.S. Lawrence[3], R.E. Thurman[4], N. Stoletzki[1,2,3], P. Stojanov[3], E. Rynes[4], L.A. Garraway[2,3,5,6], S. Mirkin[7], G. Getz[3], J.A. Stamatoyannopoulos[4].* 1) Div Gen, Dept Med, Brigham & Women's Hosp, Boston, MA; 2) Harvard Medical School, Boston, MA, USA; 3) The Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Genome Sciences, University of Washington, Seattle WA; 5) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA; 6) Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA; 7) Department of Biology, Tufts University, Medford, MA.

Carcinogenesis and neoplastic progression are mediated by the successive accumulation of somatic mutations. In cancer cells, the somatic mutation rate is highly heterogeneous along the genome. For example, transcribed regions exhibit lower mutation rates, which is commonly attributed to the action of transcription coupled repair. Here, we report that the somatic mutation rate is highly specifically reduced in regulatory regions of cancer genomes. This reduction is tightly linked to the degree of chromatin accessibility at regulatory DNA, is independent of any known factors influencing somatic mutation rate, and is observed in diverse cancer types suggesting a general mechanism. The hypomutational effect cannot be readily explained by differential DNA damage, replication fidelity, or purifying selection. Chromatin accessibility plays a major role in targeting nuclear proteins to regulatory DNA, and may provide a mechanism for preferential access by the repair machinery. Using analysis of 29 individual melanoma genomes, we show that reduced mutation rate within regulatory DNA requires intact global genome repair machinery. In some individual cancers, potentially deactivating mutations in multiple nucleotide excision repair components result in nearly complete abrogation of regulatory DNA hypomutation. Together, our results suggest a pervasive mechanistic connection between focal chromatin accessibility and mutation rates in cancer genomes. Reduced mutation in regulatory DNA may thus operate synergistically with transcription-coupled repair to stabilize cancer gene expression programs.

403

***TECTONIC3* mutations cause orofaciodigital syndrome type IV (Mohr-Majewski).** *S. Thomas[1,2], M. Legendre[1], S. Saunier[2,3], B. Bessières[4], C. Alby[2,5], M. Bonnière[4], A. Toutain[6], L. Loeuillet[7], K. Szymanska[8], F. Jossic[9], D. Gaillard[10], M. Tahar Yacoubi[11], S. Mougou-Zerelli[12], A. David[13], M-A. Barthez[14], Y. Ville[2,5], C. Bole-Feysot[15], P. Nitschke[16], A. Munnich[1,2,4], C.A. Johnson[8], F. Encha-Razavi[1,2,4], V. Cormier-Daire[1,2,4], C. Thauvin-Robinet[17,18], M. Vekemans[1,2,4], T. Attié-Bitach[1,2,4]*. 1) INSERM U781, Hôpital Necker-Enfants Malades, Paris, France; 2) Université Paris Descartes, Paris, France; 3) INSERM U983, Hôpital Necker-Enfants Malades, Paris, France; 4) Département de Génétique, Hôpital Necker-Enfants Malades, AP-HP, Paris France; 5) Service de Gynécologie-Obstétrique, Hôpital Necker-Enfants Malades, AP-HP, Paris, France; 6) Service de Génétique, CHRU de Tours - Hôpital Bretonneau, Tours, France; 7) Service d'Anatomie et de Cytologie Pathologiques, CHI Poissy, Saint Germain en Laye, France; 8) Section of Ophthalmology and Neurosciences, Leeds Institute of Molecular Medicine, St. James's University Hospital, Leeds, LS9 7TF UK; 9) Service d'Anatomie Pathologique, CHU de Nantes, Nantes, France; 10) Service de Génétique et Biologie de la Reproduction, CHU Reims, Université de Reims, France; 11) Service d'Anatomie et de Cytologie Pathologiques, CHU Farhat Hached, Sousse, Tunisie; 12) Service de Cytogénétique, Génétique moléculaire et Biologie de la reproduction, CHU Farhat Hached, Sousse, Tunisie; 13) Service de Génétique Clinique, Hôpital Hôtel Dieu, CHU de Nantes, Nantes, France; 14) Service de Neuropédiatrie, CHRU de Tours, Hôpital Clocheville, Tours, France; 15) Plateforme de génomique, Fondation IMAGINE, Hôpital Necker-Enfant Malades, Paris, France; 16) Plateforme de Bioinformatique, Université Paris Descartes, Paris, France; 17) Centre de Génétique, Hôpital d'Enfants, CHU de Dijon, Dijon, France; 18) EA GAD, IFR Santé - STIC, Université de Bourgogne, Dijon, France.

Orofaciodigital syndromes (OFDS) consist of a group of heterogeneous disorders characterized by abnormalities in the oral cavity, face and digits with associated phenotypic abnormalities that lead to the delineation of 13 OFDS subtypes.

By a combined approach of homozygozity mapping and exome ciliary sequencing, we identified truncating *TCTN3* mutations as the cause of an extreme form of OFD IV (Mohr-Majewski syndrome) associating orofaciodigital syndrome with bone dysplasia, tibiae defect, cystic kidneys and brain anomalies. Analysis of 184 patients/fetuses with various ciliopathies (OFD, Meckel, Joubert and short rib polydactyly syndromes) led us to identify four additional truncating *TCNT3* mutations in unrelated fetal cases with overlapping Meckel and OFDIV syndromes, and one homozygous missense mutation in a Joubert family.

By exploring roles of TCTN3 in human ciliary related functions, we found that TCTN3 is necessary for transduction of the sonic hedgehog (SHH) signaling pathway as revealed by the abscence of transcriptional activation of known SHH target genes in response to an agonist of the pathway in patient fibrablasts as compared to controls. We also found a diminution of the GLI3 full length to repressor form ratio in patient cells indicating that TCTN3 is essential for GLI3 processing and function. These results are consistent with the suggested role of its murine orthologue which forms a complex at the ciliary transition zone with Tctn1 and Tctn2, both of which are also implicated in the transduction of Shh signaling.

Overall, our data show the involvement of the transition zone protein TCTN3 in the regulation of the key SHH signaling pathway and that its disruption causes a severe form of ciliopathy, combining features of Meckel and OFDIV syndromes.

404

Abnormal development of NG2+PDGFRα+ neural progenitor cells causes neonatal hydrocephalus in a ciliopathy mouse model. *C.S. Carter[1], TW. Vogel[1], Q. Zhang[1,2], TO. Moninger[1,2], DR. Thedens[1], KM. Keppler-Noreuil[1], DY. Nishimura[1], CC. Searby[1,2], K. Bugge[1,2], VC. Sheffield[1,2]*. 1) University of Iowa, Iowa City, IA; 2) Howard Hughes Medical Institute, Chevy Chase, MD.

Hydrocephalus is a common neurological disorder leading to expansion of the cerebral ventricles and is associated with significant morbidity and mortality. The majority of neonatal cases are of unknown etiology. Identifying molecular mechanisms for neonatal hydrocephalus and developing non-invasive medical treatment modalities are high priorities. We employed a hydrocephalic mouse model of the human ciliopathy Bardet-Biedl Syndrome (BBS) to identify a role for neural progenitor cells in the pathogenesis of neonatal hydrocephalus. We found that hydrocephalus in this mouse model is caused by aberrant PDGFRα signaling, resulting in increased apoptosis and impaired proliferation of NG2+PDGFRα+ neural progenitor cells. Conditional knockout of Bbs1 in this progenitor cell population lead to neonatal hydrocephalus and confirmed the involvement of NG2+PDGFRα+ progenitor development in the pathophysiology of neonatal hydrocephalus. Notably, hydrocephalus in these conditional knockout mice developed in the presence of normal motile cilia challenging the conventional view that motile cilia are crucial in the pathogenesis of hydrocephalus. Targeting the defective PDGFRα signaling pathway early in development with lithium treatment rescued impaired NG2+PDGFRα+ progenitor cell development in BBS mutant mice resulting in reduced ventricular volume. Our findings indicate that abnormalities in a specific class of neural progenitor cells play a key role in the pathogenesis of neonatal hydrocephalus. Importantly, we identify novel therapeutic targets for this common neurological disorder. We also identify a novel role for BBS1 in mediating the PDGFRα signaling pathway and NG2+PDGFRα+ neural progenitor cell development.

405

Malformation of the brain cortex, as the only expression of a ciliopathy, results from mutation in human Rotatin. *G.MS. Mancini, F.W. Verheijen.* Department of Clinical Genetics. Erasmus Medical Center, Rotterdam, Netherlands.

Polymicrogyria (PMG) is a malformation of the developing cerebral cortex and has been sporadically described with cerebellar malformation (molar tooth) in Joubert syndrome. We have implemented the use of cilia staining in cultured skin fibroblasts as support to the diagnostic work-up of individuals with PMG and other malformations. This has allowed identification of structural cilia anomalies in patients without a known diagnosis of ciliopathy. Here we report mutations in a new gene in individuals with PMG as the only malformation. We have identified autosomal recessive mutations in the Rotatin gene, RTTN, in individuals with bilateral diffuse polymicrogyria from two separate families. Rotatin determines early embryonic axial rotation as well as anteroposterior and dorsoventral patterning in the mouse. Human Rotatin has recently been identified as a centrosome-associated protein. The Drosophila melanogaster homolog of Rotatin, Ana3, is needed for structural integrity of centrioles and basal bodies and maintenance of sensory neurons. We show that Rotatin colocalizes with the basal bodies at the primary cilium. Cultured fibroblasts from affected individuals have structural abnormalities of the cilia and exhibit down-regulation of BMP4, WNT5A and WNT2B,which are key regulators of cortical patterning and are expressed at the cortical hem, the cortex organizing center giving rise to Cajal-Retzius (CR) neurons. Interestingly, we have shown that in mouse embryos Rotatin expression co-localizes with CR neurons at the subpial marginal zone. Knockdown experiments in human fibroblasts and neural stem cells confirm a role for Rotatin in cilia structure and function. Rotatin mutations therefore link aberrant ciliary function to abnormal development and organization of the cortex in human individuals.

406

Whole exome resequencing identifies mutations in *LRRC6* as a novel single-gene cause of primary cliary dyskinesia. *M. Chaki[1], H. Y. Gee[1], E. A. Otto[1], K. Diaz[1], T. W. Hurd[1], J. Halbritter[1], S. J. Allen[1], M. B. Zariwala[2], M. R. Knowles[2], F. Hildebrandt[1,3,4].* 1) Pediatrics - Nephrology, University of Michigan, Ann Arbor, MI; 2) Cystic Fibrosis/Pulmonary Research and Treatment Center, University of North Carolina, Chapel Hill, NC; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 4) Howard Hughes Medical Institute.

Primary ciliary dyskinesia (PCD) is a heterogeneous group of rare autosomal recessive ciliopathies with a complex phenotype caused by impaired motility of cilia. The disease is characterized by abnormal function and/or ultrastructural defects of the ciliary mechanism. Mutations in genes encoding for parts of the ciliary ultrastructure have been recently implicated in the pathogenesis of the disease. To identify novel single-gene causes of PCD, we used homozygosity mapping followed by whole exome capture and massively parallel sequencing in a cohort of 44 PCD families. We identified a homozygous frameshift mutation (p.W210CfsX12) in the gene *LRRC6* (Leucine rich rrepeat containing 6) in a Pakistani family with two affected siblings with outer dynein arm (ODA) and inner dynein arm (IDA) defects. We identified the same mutation in 3 additional families of similar origin which indicates a founder effect. *LRRC6* is a gene involved in ciliary assembly and maintenance similar to *LRRC50*, a gene with known association with IDA and ODA ultastructural defects. This is congruent with the phenotype of the affected individuals. Additionally, there is a *D. rerio* homolog model (*Lrrc6l*) as well as a *D. melanogaster* homolog model (*tilB*) available for *LRRC6* that exhibit dynein arm defects and impaired ciliary motility. We thereby show that *LRRC6* is a novel single-gene cause of PCD. Our findings validate the importance of whole exome capture analysis for broadly heterogeneous single-gene disorders like PCD.

407

Temporally and spatially resolved catalogues of *in vivo* forebrain enhancers. *A.S. Nord[1], L. Taher[2], J. Akiyama[1], M.J. Blow[3], A. Holt[1], R. Hosseini[1], S. Phouanenavong[1], I. Plajzer-Frick[1], M. Shoukry[1], V. Afzal[1], E.M. Rubin[1,3], I. Ovcharenko[2], J.L.R. Rubenstein[4], L.A. Pennacchio[1,3], A. Visel[1,3].* 1) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA; 2) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD; 3) U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA; 4) Department of Psychiatry, Rock Hall, University of California at San Francisco, San Francisco, CA.

Mammalian forebrain development and maturation is a dynamic process that begins in early embryogenesis and continues long after birth. Growing bodies of evidence strongly support that this genetic control relies heavily on distant-acting regulatory circuits involving non-coding DNA elements. One particular class of element, enhancers, plays a critical role in determining temporal and spatial specificity of gene expression. Thus, characterizing distal enhancer activity across development has the potential to inform understanding of tissue ontogenesis, evolution, and disease. We generated a genome-wide time-series map of forebrain enhancer activities *in vivo* through epigenomic profiling of forebrain tissue collected across mouse development. Mouse forebrain tissue was collected at eight time points between mid-gestation and adulthood (e11.5, e14.5, e17.5, p1, p7, p14, p21, and p56) and analyzed by chromatin immunoprecipitation followed by sequencing (ChIP-seq) targeting H3K27ac. Through differential activity analysis, we identified enhancers active only at certain stages in development and enhancers that are constitutively active in mouse forebrain. Stage-specific forebrain enhancers regulate relevant developmental and disease-related pathways and processes, such as cell morphogenesis and differentiation at early time points and learning and memory later in development. Stage-specific enhancers exhibit enrichment for specific binding motifs and show decreasing evolutionary conservation as development progresses, suggesting an increased proportion of lineage-specific regulatory elements after basic forebrain patterning has occurred. Using mouse transgenic enhancer assays, we validated ChIP-seq predictions for a set of disease-relevant enhancers, showing reproducible *in vivo* forebrain expression patterns that matched stage-specific activity predictions. In a complementary analysis, we generated serial sets of histological brain sections for 145 e11.5 forebrain enhancers to catalogue spatial activity at high resolution and characterize enhancers active in different subregions of the developing forebrain. Results from our transgenic and histological studies are available through a public web database at [au0,2]http://enhancer.lbl.gov[xu. Together, these datasets provide an unparalleled view into the spatial and temporal activity of mammalian forebrain enhancers, and will be valuable towards understanding their role in developmental and disease processes.

408

SRY Regulation of the RET Gene Suggests a Potential Role of the Y-chromosome Gene in Sexual Dimorphism in Hirschsprung Disease. *Y. Li[1], Z.L. Tabatabai[1], M.M. Garcia-Barcelo[2], P.K.H. Tam[2], Y.-F.C. Lau[1].* 1) Medicine, VAMC-111C5, Univ California, San Francisco, San Francisco, CA; 2) University of Hong Kong, Hong Kong.

The SRY is the testis determining factor (TDF)on the Y chromosome, which initiates the testis determination by suppressing ovarian determining genes, and activating testis differentiating genes, particularly SOX9, which propagates the developmental process beyond SRY action. SRY and SOX proteins share significant homology at the DNA-binding HMG boxes, which are functionally interchangeable in transgenic mouse studies. Hence, SRY and SOX transcription factors could bind to the same DNA targets or interact with similar proteins via their HMG boxes. Recently, we have identified the target genes, whose promoters are bound by SRY and SOX9 at the time of sex determination using a genome-wide strategy. Gene ontology shows that 9 of 13 disease genes involved in Hirschsprung disease (HSCR) are SRY targets. HSCR, or aganglionic megacolon, is a complex congenital disorder, arising from abnormalities in enteric nervous system (ENS) development. There is a gender disparity among the patients, with male to female ratio as high as 6:1. Loss-of-function mutations of HSCR genes and haploinsufficiency of their gene products are the primary pathogenic mechanisms for disease development. Among the SRY targets involved in HSCR, RET represents the most important disease gene, whose mutations account for 50% genetic and 7–35% sporadic forms of HSCR. We show that SRY expression is detectable in HSCR samples, but not in normal enteric tissues. Promoter characterization shows RET is regulated by a distal and a proximal enhancer at its promoter, in which PAX3 and NKX2-1 are the resident transcription factors respectively. SOX10 interacts with PAX3 and NKX2-1 at its HMG box, and exacerbates their respective transactivation activities. SRY competitively binds to PAX3 and NKX2-1, and represses their transcription and SOX10 stimulation on the RET promoter. Chromatin immunoprecipitation and gene-specific PCR amplification demonstrate that SOX10 is capable of interacting simultaneously with PAX3 and NKX2-1, thereby drawing the distal and proximal enhancer in a knotted structure at the RET promoter. SRY competitively displaces SOX10 interaction and transcription functions, thereby repressing the RET promoter activities. Hence if SRY is ectopically expressed during ENS development, such RET repression could result in RET protein haploinsufficiency and HSCR development, thereby contributing to sexual dimorphism in HSCR.

409

MAP3K1 MUTATIONS IN 46,XY DGDs ALTER CROSSTALK IN DOWNSTREAM SIGNAL TRANSDUCTION PATHWAYS TO CAUSE ABNORMAL HUMAN GONADAL DEVELOPMENT. *J. Loke, A. Pearlman, H. Ostrer.* Pathology, Albert Einstein Sch Med, Bronx, NY.

Genetic analysis of human disorders of gonadal development (DGDs) has played an important role in understanding the genes involved in normal gonadal development. Recently, our laboratory demonstrated that mutations in the MAP3K1 gene account for ~20% of familial and sporadic cases of 46,XY DGD (AJHG 2010; 87: 898–904). All of the mutations occurred in-frame in the amino third of the MAP3K1 protein, a region that interacts with RHOA and other binding factors. New in-frame, non-terminating MAP3K1 mutations have since been identified. Here, we show that members of the MAP kinase gene family mediate the balance between the male and female sex determining pathways by affecting the activities of the testis-promoting SOX9 and ovarian-promoting CTNNB1 (β-catenin) genes. Such explanations do not require sexually dimorphic expression of MAPK signaling components, because transcription of SOX9 and other downstream factors is already sexually dimorphic. These predictions were tested directly in a series of analyses of primary lymphocytes and transfected NT2 cells by Western blot, flow cytometry phosphorylation and flow cytometry-based immunoprecipitation. First, we demonstrated that transfection of mutant MAP3K1 cDNAs produced phenotypes similar to what we observed in mutation-bearing primary lymphoblastoid cells, increased phosphorylation of p38 and ERK and increased binding of RHOA and other co-factors. By tandem mass spectrometry on IP samples from a mutant and normal male wild-type pair, we identified the WNT canonical pathway regulator, FRAT1, to be significantly more abundant in the mutant. Next, we demonstrated that the effects of these mutations enhanced β-catenin transcript stability by AXIN sequestration, mediated through the known WNT canonical pathway, and inhibited SOX9, FGF9 and FGFR2 transcripts mediated through RHOA hyperactivation. We also observed MAP3K4-increased binding to mutant MAP3K1. Binding of MAP3K1 to MAP3K4 has been shown to synergistically activate phosphorylation of ERK1/2 and p38, which in turn downregulate SOX9 activity. We further showed a phenotype-genotype correlation in patients with MAP3K1 mutations, normal gonads and either adrenal failure or immunodeficiency. Thus, we report a direct effect of MAP3K1 mutations to mediate the balance between β-catenin and SOX9 at multiple levels involving interaction with co-factors and downstream phosphorylation, and demonstrate a new method to rapidly screen the effects of mutations.

410

Soft tissue aspects of the Williams-Beuren syndrome facial phenotype can be attributed to *GTF2IRD1*. *S.J. Palmer[1], C.P. Canales[1], P. Carmona-Mora[1], P. Kaur[2], P.W. Gunning[3], E.C. Hardeman[1]*. 1) School of Medical Sciences, Neuromuscular and Regenerative Medicine Unit, University of New South Wales, Sydney, Australia; 2) Epithelial Stem Cell Biology Laboratory, Peter MacCallum Cancer Centre, Melbourne, Australia; 3) School of Medical Sciences, Oncology Research Unit, University of New South Wales, Sydney, Australia.

Williams-Beuren syndrome (WBS) results from a hemizygous microdeletion within chromosome 7q11.23 involving up to 28 genes. Its features typically involve a characteristic set of physical, cognitive and behavioural abnormalities. Genotype/phenotype correlations in patients with atypical deletions have mapped many of these features, including characteristic craniofacial dysmorphologies, to a pair of genes that encode the evolutionarily-related transcriptional regulators GTF2IRD1 and GTF2I (TFII-I). Skull measurements on WBS patients have revealed some hard tissue defects but many of the typical WBS facial features, including periorbital fullness, a long philtrum, full lips, a wide smile and full cheeks can be attributed to soft tissue abnormalities. Intriguingly, patients with the 7q11.23 microduplication syndrome, carrying 3 copies of the WBS region, show reciprocal soft tissue defects, including a broad nose, thin lips and a short philtrum. Clearly, these soft tissue abnormalities are caused by altered dosage of a specific gene or genes, and *GTF2IRD1* and *GTF2I* are the likeliest candidates. We have generated *Gtf2ird1* knockout mouse lines that show some striking similarities to aspects of the human disease, including soft tissue facial abnormalities. X-ray computed tomography revealed no obvious defects in the hard tissue components. Knockout mice have a defect around the nose and lip region that is caused by an extreme thickening of the epidermal layer. Immunohistochemistry has revealed a significant expansion in the proliferative compartment of the epidermis and the differentiated derivatives of this layer. We are investigating whether this is due to a defect in the stem cells or the transit amplifying cell population. The patterning of the facial defect bears a close resemblance to the pattern of *Gtf2ird1* expression in the developing facial epidermis, suggesting that the `fullness' phenotype is restricted to zones demarcated by *Gtf2ird1* activity. In addition, our ongoing studies on the molecular function of GTF2IRD1, its protein partners, mechanisms of auto-regulatory control, post-translational modifications and potential gene targets have allowed us to develop important insights into its cellular role. Together, these data allow us to explain the basis of the WBS soft tissue facial phenotype.

411

Notch Gain of Function Inhibits Chondrocyte Differentiation via Rbpj-Dependent Suppression of Sox9. *S. Chen[1,2], J. Tao[1,2], Y. Bae[1,2], M. Jiang[1,2], T. Bertin[1,2], Y. Chen[1,2], T. Yang[1,2], B. Lee[1,2]*. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) HHMI, Houston, TX.

Notch signaling plays a critical role during development by directing the binary cell fate between progenitors and differentiated cells. Previous studies have shown sustained Notch activation in cartilage leads to chondrodysplasia. Genetic evidence indicates that Notch regulates limb bud mesenchymal stem cell differentiation into chondrocytes via Rbpj-dependent Notch pathway. However, it is still unknown how Notch signaling governs chondrogenesis of the axial skeleton where Notch signaling also serves a patterning function. We hypothesize that both Rbpj-dependent and Rbpj-independent Notch signaling mechanisms might be involved. We generated cartilage specific Notch gain-of-function (GOF) mutant mice, which displays chondrodysplasia accompanied by loss of Sox9 expression in the vertebral body. To evaluate the contribution of Rbpj-dependent Notch signaling to these phenotypes, we deleted Rbpj in the Notch GOF background. These mice continued to develop spine abnormalities characterized by “butterfly vertebrae” suggesting that removal of Rbpj does not fully rescue the axial skeleton deformities caused by Notch GOF. However, Sox9 protein level was restored in the Rbpj deficient Notch GOF mice compared to Notch GOF mutant, suggesting regulation of Sox9 expression is canonical or Rbpj-dependent. To further understand the molecular basis of this control, we performed ChIP assays and demonstrate direct recruitment of Rbpj/NICD transcription complex to the Rbpj-binding sites upstream of Sox9 promoter. This association of Rbpj/NICD complex to Sox9 promoter is associated with transcriptional repression in a cellular model of chondrocyte differentiation. Hence, we show that Notch negatively regulates chondrocyte differentiation in the axial skeleton by suppressing Sox9 transcription. In addition, Rbpj-independent Notch signaling mechanisms may also contribute to axial skeleton patterning.