### Abstract Numbers

#### Tuesday, November 6

2. 4:30–6:30pm: **Plenary Abstract Presentations**
   - **Room 134**
   - **#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**
    - **Ballroom 104**
    - **#43–#51**

15. **New Loci for Obesity, Diabetes, and Related Traits**
    - **Room 134**
    - **#34–#42**

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–12:45 pm: 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**

### Abstract Numbers

#### 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. Wolfink1, E. Poth1, N. Offenhäusera, A. Uzunmana, F.J. Kerstena, A.K. Rzadzińska2, O. Uyguner2, B. Lorento3, G. Nümberta, M. Emiroglu4, H. Kayserili5, I. del Castillo6, P. Nümberta, T. Moser2, C. Kubsich6,7, K.P. Steel8, P.P. Di Fiore9, H. Kremer4, Y. Li10. 1) Institute of Human Genetics, University of Cologne, Cologne, Germany; 2) IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, 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) Cologny Genetics, 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, Göttinngen 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 world-wide. Although highly effective, their use is restricted by side-effects such as ototoxicity in a significant subset of patients. However, underlying pathogenic and pharmacogenetic mechanisms 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 stereocilla length and that it interacts with whirin. The mutation severely impairs this interaction in vitro. Extensive biochemical studies showed that myosinVx can stabilize the ARP-whirin 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 a mutation variant found in the consanguineous family. Overall, our work suggests ARP loss as an underlying cause and targets the human ARP protein as a drug development target.


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 the genome-wide evolutionary changes 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-sequencing to obtain genome-wide profiles of H3K4me3, H3K4me1, H3K27me3 and H3K27ac histone modifications, as well as binding of RNA polymerase II. We also collected DNase-seq 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 changes 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 modifications.

3 Multidisciplinary and Translational Task Force for Neonatal Genomics. E.E. Davies4,5, A. Sabo6, N.C. Olenn, S.H. Katsanis5, H. Cope5, K. Sheets7, A. Sadeghpour5, K. McDonald6, M. Kousi6, J.R. Willer1, J. Kim1, S. Dugan-Rocha5, D.M. Muzny5, A. Ashley-Koch5, E. Hauser5, M. Hauser5, J. Sun2, J. Kurtzberg7, A. Murtha8, B. Boyd9, W.B. Gallentine9, R. Goldberg10, M.T. McDonald10, R.A. Gibbs1, M. Arngist5, C.M. Cotten10, N. Katsanis5,6,11. 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, 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 abnormalities are more likely to be within a timeframe for treatment. Anatomical defects are also amenable to optimizing surgical procedures in postnatal life. In this model, single nucleotide variants of unknown significance in physiologically relevant vertebrate and cell-based models to inform allele pathogenicity. Successful examples of a 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 chanelopathy 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 spondyloepiphysial 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.
Plenary Abstracts

4 Genome-wide Identification and Functional Analysis of Distant-Acting Craniofacial Enhancers. C. Attanasio1, Y. Zhu1, M.J. Blow2, A.S. Nord1, V. Atzal1, B. Hallgrimsson1, D. FitzPatrick1, H. Morrison1, E.M. Rubin1-2, L.A. Pennacchio1-2, A. Visel1-2. 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 IGM, University of 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), to delenion 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 reporter activity. In a subsequent study, we hypothesized that functionally allelic transcripts would show a detectable shift in this distribution. Methods: Cell lysates of lymphoblastoid cell lines (LCL) from 30 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 nonpolyribosomal 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 these genes, DCTN4, RPS26, and APOB demonstrated significantly correlated expression across 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 LDL, CRP, systolic 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 pneumonias in individuals with cystic fibrosis (n=91, p=2.5×10−5). We were also able to replicate several known associations across a number of traits [e.g., variants in the APOBgene are associated with LDL (p=2.5×10−6, n=3342) and the LEPR gene is associated with CR-Reactive protein (p=2.5×10−6, 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.
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. Risch1,2, T.J. Hoffmann1, M. Anderson2, L.A. Croen2, J. Grether2, G. Windham3, 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 (9.3%); 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. The interval effect was noted in maternal half sibs (6.2% for second born compared to 3.0% for later born sibs; 4.8% overall). This interval effect suggests that the risk is 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.

Identifying inherited autism mutations using whole exome sequencing.

T.W. Yu1, M.H. Chahroudi2, M.E. Coulter1, S. Jiraratspong2, K. Okamura-Ikedo2, K. Schmitz-Abe3, G.H. Mochida3, J.N. Parlow1, R.S. Hill1, M. Al-Saffar1, N.M. Mukaddess4, A. Hashimi5, S. Baikhy6, G.G. Gascon7, O. Oner8, S. Al-Saaid5, T. Ben-Omran9, L. Al-Gazali11, V. Eapen12, C. Stevens13, S. Gabriel13, K. Markianos1, H. Taniguchi1, N.E. Braverman9, E.M. Morrow12, C.A. Walsh1. 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, Telisizler, 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 observed a 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. The interval effect was noted in maternal half sibs (6.5% for second born compared to 3.0% for later born sibs; 4.8% overall). This interval effect suggests that the risk is 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.
9

Identical by descent filtering in extended families reveals novel autism genes detected by exome sequencing. H.N. Cukier1, N.D. Dueker1, S.H. Stifter1, J.M. Lee1, P.L. Whitehead1, E. Lalanne1, N. Levy1, I. Konidari1, R.C. Gentry1, W.F. Hulme1, D. Van Booven1, D.J. Hedges1, V. Mayo1, S.S. Ramsook1, B.A. Barrionuevo1, J.M. Jaworski1, M.A. Schmidt2, J.L. Haines3, M.L. Cuccaro3, J.R. Gilbert3, M.A. Pericak-Vance1, 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 in 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 (CSDM1). 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 as an essential factor in neuronal development. Furthermore, three genes, CXorf93, 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 implicated in autism risk. Notably, we identified three 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’Roak1, L. Vives1, A. Kumar1, I.B. Stanaway1, J.R. Gilbert1, E. Turetta1, C. Lee1, G.L. Canvill1, I.G. Phelps1, D.R. O’Day1, W. Fu1, J. Hiatt1, B. Martin1, N. Krumm1, B.P. Cooke1, R. Levy1, E. Borenstein1, D.A. Nickerson1, H.C. Mofford2, D.A. Doherty2, J.M. Akey1, R. Berrier1, E.E. Eicher1, J. Shendure1, 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) 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, AQP1, TBL1XR1, and CTNNB1. Notably, 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 disruptions form a novel network of which the most significant (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 harbored de novo nonsense, frameshift, or splice-site variants in CHD8. Probands were notable for increased head circumference (6.8/8 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. Lim1,2, M.J. Daily1,2,3, M.J. Daly1,2,3, 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 chemosensination. 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.
Exome-based discovery of CNVs in Simplex Autism Families. N. Krumm1, B. Nelson1, S. Girirajan1, M. Dennis1, C. Baker1, M. Malig1, NHLBI Exome Sequencing Project2, A. Quinlan3, D.A. Nickerson1, E.E. Eichler1,4. 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 myelination in Drosophila. Our previous work has suggested 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.

Delta Catenin (CTNND2): genetics and function of a novel autism gene. T. Turner1,9, E. Oh2, Y. Liu2, M.X. Sosa3,7, S. Sanders3,7, K. Sharma4, D. Morenci-De-Luca3,9, T. Plona3, K. Pike6, D. Soppe5, M.W. Smith5, M. State1,6, S.W. Cheung5, C. Lesse Martin1,9, R. Huganir4, E. Cook7, N. Katrasins2, A. Chakravarti1,9. 1) Center for Complex Disease Genetics, 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 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<2x10-12). 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 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 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. Ginraj,$^{1,}$ 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$^{1}$, E.E. Eichler$^{1,2}$  
$^{1}$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.

Repeatability 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 17p12) 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, KDM4D, and VWHA—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 MBDS—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) Department of Psychiatry, University of Washington, Seattle, WA.

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=1.50; p<0.00001). We also estimated genetic relationships and compared relatedness within versus across clusters (FST=0.15±0.26). 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 use of haplotype identity in population genetic data. D. Xifara1,2, I. Mathiesen1, G. McVean1. 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 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 small samples. Moreover, current methods fail to detect the shared ancestry typically confute 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 the shared generation to which a haplotype is 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, we show that this approach is able to identify shared haplotypes in a manner consistent with a long-range phasing method (Kong et al. 2008). In a small simulated data set, the approach is 1.8 cM, more than an order of magnitude 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 much older than 15% of the time, which has implications for the ability to impute rare variants.


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 effort, and development of 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 demographic 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 of similar size. However, the bottleneck size is small (<1000), 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 (CUM, 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.

18 Testing for rare variant associations in the presence of missing data. P. Livmerove Auer1, S. Leal2, F. Wang3, 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 specifically to 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, naive implementation of rare variants association (RVA) methods such as the CMC (Li & Leal 2008), BRV (Morris & Zeggini 2008) or transcript. When variant calls are missing, naïve implementation of rare variant association (RVA) methods such as the CMC (Li & Leal 2008), BRV (Morris & Zeggini 2008), QGV (Coscoy et al. 2009), CAST (C-ALPHA) and RVM (Torkamani et al. 2009) and reported in association studies of rare variants (RV). Most existing RV burden test methods, however, require individual-level data and therefore are inappropriate 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, EFT). 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 (CUM, 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.
20 Quantitative trait locus analysis for next-generation sequencing with the functional linear models. M. Xiong1, L. Luo2, Y. Zhu1. 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. Li1,2,3, S. Kwan1,4, H. Gu5, P. Sham1,4, 1) Department for Reproduction, Development and Growth, The University of Hong Kong, Hong Kong, HK; 2) Centre for Genomics Sciences, The University of Hong Kong, Hong Kong, HK; 3) Department of Medicine, The University of Hong Kong, Hong Kong, HK; 4) Department of Psychiatry, The University of Hong Kong, Hong Kong, HK; 5) Department of Epidemiology, Biostatistics and Preventive Medicine, University of New Mexico.

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 methods to extend the testability beyond the single-locus 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 I error rate and was generally more powerful than its “parents” and other alternative tests (including VEGAS, Bonferroni, 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 indicate 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 KG (http://bioinfo.hku.hk/kggweb/), with integrated protein and biological pathway resources for geneticists to insightfully explore the ‘missed’ susceptibility genes of complex diseases.

22 Statistics for X-chromosome association. U. Ozbek1, D.E. Weeks1,2, W. Chen1, J. Shaffer3, S.M. Purcell4,6, E. Feingold1,2. 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 Genomics 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 GWAS 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 genes 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 genetic architectures and phenotypic effects. 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. Salem1,2,3, J.N. Hirschhorn1,2,3, 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 in understanding genetic architecture of complex traits and also to characterize more completely the underlying biology and pleiotropic 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 pleotropic 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.
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.
Cancer Genetics I: Rare Variants

25 Exome sequencing of more than 6,700 samples and the study of genetic susceptibility to common cancer. A. Klezun, 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 delineate 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: ATR, BAP1, CHEK1, GEN1, RAD51B, and TP53BP1 — a 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 mutations seen in melanoma families. GEN1, a key Holliday junction resolution, 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 gland and ovary. Recurrent loss-of-function mutations in OBSL1 cause 3M, a prionmal 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 in parallel 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. Hibbers, M.C. Völker-Albert, W.W. Wiegant, M.P.G. Vreeswijk, H. van Attikum, P. DeVito, I.G. Campbell, J. Benitez, C.J. van Asperen, H. van Attikum, P. DeVito. 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 IRCC di Oncologia Molecolare (IFOM), Milan, Italy; 8) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, USA; 9) Cancer Genetics Laboratory, Memorial Sloan Kettering Cancer Center, New York, USA; 10) VCBRC 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 11 familial cases (0.8%) 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 over the genome. To examine the effect of 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 in IRS1 cells. The results suggest eight new candidate genes: ATR, BAP1, CHEK1, GEN1, RAD51B, ATR, BAP1, CHEK1, GEN1, RAD51B, and TP53BP1 — a 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 mutations seen in melanoma families. GEN1, a key Holliday junction resolution, 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 gland and ovary. Recurrent loss-of-function mutations in OBSL1 cause 3M, a prionmal 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 in parallel 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.

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%) and 19% of families positive for prostate cancer. Analysis of flanking markers based association test found that the G84E mutation was significantly overrepresented (51%) than those without (42 of 137, 30%), P=9.9×10^-8. The odds ratio for 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 families in harboring this rare allele will be needed to define the clinical utility of this observation.
30 Exome Sequencing in Families at High Risk for Lymphoid Malignancies. L.R. Goldin1, M.L. Mcmaster2, M. Rotunno1, K.B. Jacobson2, L. Burdette3, M. Malasky2, A. Hutchinson3, M. Cullen4, J. Boland2, M. Yeager2, M.A. Tucker5, S.J. Chanock2, N.E. Caporaso2. 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 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 HiSeq 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 (rs56400844) 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 rare variant account for familial clustering. Analogous follow up procedures are in progress for the other identified variants.

31 Rare allelic forms of PRDM9 associated with childhood leukemia. J. Hussin1,2, D. Sinnett1,2, F. Casals1, Y. Idaghdour2, V. Bruat2, V. Saillour5, J-C. Grenier2, J. Healy4, T. de Maillard5, J-F. Spinella1,2, M. Lariviere5, S. Busche2, G. Gibson2, A. Andersson2, L. Hofmefld2, J. Ma2, L. Wei1, J. Zhang4, G. Andelfinger1,2, J.R. Downing2, C.M. Mullighan5, P. Awadalla1,2. 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 significant 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 germ line potentially modifies the risk of acquiring aneuploidy or genomic rearrangements associated with childhood leukemia.

32 De novo mutation of the TGF beta family in early-onset ovarian cancer. I. Tournier1, F. Charbonnier1,2, S. Coutant1, K. Walton6, R. Martin1, M. Vezain1, J. Tinel5, E. Angot6, R. Seksou6, J-C. Sabourin1,4, D. Vaur1,5, C. Harrison2, T. Frebourg1,2. 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 possible 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 BRCA1 or BRCA2 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 database (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.
Somatic activating mutations in PIK3CA cause progressive segmental overgrowth. M.J. Lindhurst1, V.E.R. Parker2, J.C. Sapp1, S. Rudge4, J. Harris2, A.M. Witkowski1, Q. Zhang5, M.P. Groeneveld2, C.E. Scott3, A. Daly3, S.M. Huson5, L.L. Tosi5, M.L. Cunningham2, T.N. Darling8, J. Geer1, Z. Gucev10, P.A. Kreiger11, V.R. Sutton12, M.M. Thacker13, C. Tziotzios14, A.K. Dixon15, T. Helliwell16, S. O'Rahilly2,17, D.B. Savage2,17, M.J.O. Wakelam4, R.K. Semple2,17, I. Barroso2,3, L.G. Biesecker1. 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 \( \text{PIK3CA} \) 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 \( \text{PIK3CA} \) p.Glu542Lys variant in one patient and \( \text{PIK3CA} \) 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 \( \text{PIK3CA} \) 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. Welch1, C. Lee1, L.J. Scott2, R.A. Smith3, P. Imbriano3, M.A. Sarto1. 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 simulated regulatory regions, such as Transcription Factor binding sites and toprovidereasonablyhighqualitygenotypecallsatcommonvariantsacross

35 Enhanced exome sequencing to capture genome-wide common variants. I.C.R.M. Kolder1, K.I. Morley1, E. Binney2, I. Dunham2, J.C. Barrett2. 1) Dep. Human genetics, Wellcome trust sanger institute, Hinxton, Cambridge, 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 regions, 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 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 DNAse hypersensitivity sites; (ii) a similarly functionally focused pull down which also prioritizes regions not in LD with previously selected regions in order to maximize genomic coverage; (iii) use of pooled DNA 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.


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. 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’ ends 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 Bioconductor package DESeq with the final analysis results exported in a simple text format. 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 in sequencing libraries. The final output is an identification of unique bases incorporated in each read pair allowing 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 results explored 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 implemented an alternative 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/dect/files/.
40
Association of genetic variation affecting exon skipping to disease susceptibility. Y. Lee1, H. Im2, W. Hernandez2, N.J. Cox1,4. 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 of each of these ISE SNPs within the cancer genes $NFA15$, $GNA12$, 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 $NFA15$ 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 identification of SNPs associated with 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.

39
Discovering SNPs Regulating Human Gene Expression Using Allele Specific Expression from RNA-Seq data. E. Eskin1, E. Kang1, B. Han1, A.J. Luszti2, L. Martin1, S. Shiffman1, 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 allele-specific expression (ASE) in the human genome. Utilizing ASE has several advantages over traditional eQTL studies. First, it works for random sample data and not necessarily for families. Second, it can detect cis-regulation not necessarily for families. Third, it accommodates both univariate and multivariate traits. Fourth, it fosters both likelihood ratio tests and score tests. Both models are now implemented on our recently released software package Mendel for easy use by the genetics research community.

38

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 allow us to analyze 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 only the number of individuals but the number of allele specific expressed transcripts. We have the phased SNP data through HapMap release 22 data. For evaluation of the method, we used a RNA sequencing data of 54 lymphoblastoid cell lines derived from unrelated Nigerian individuals which have the phased SNP data through HapMap release 22 data. We consider 2267 number of human ensembl genes which has at least one ASE individuals. The number of 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.
The generation of this large set of MNP calls has enabled the observation of mutational pathways that generate mutiple 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.
A genome-wide association analysis of early-onset severe obesity: the SCOOP project. E. Wheeler\textsuperscript{1}, N. Huang\textsuperscript{1}, E. Bochukova\textsuperscript{2}, S. Lindsay\textsuperscript{1}, J. Keogh\textsuperscript{1}, R.J.F. Loos\textsuperscript{3}, N.J. Wareham\textsuperscript{1}, S. O’Rahilly\textsuperscript{2}, M. Hurles\textsuperscript{1}, I. Barroso\textsuperscript{1,2}, I.S. Farooqui\textsuperscript{2,3}, 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.

Mapping obesity traits using an integrated ‘omics’ approach in adipose tissue from female twins. A.K. Hedman\textsuperscript{1}, J.K. Sandling\textsuperscript{2}, E. Grundberg\textsuperscript{1,3}, K.S. Smajlović\textsuperscript{2}, E. Meduri\textsuperscript{1,3}, S. Keildson\textsuperscript{1}, A. Nica\textsuperscript{4}, A. Bull\textsuperscript{5}, T.T. Belt\textsuperscript{1,3}, J. Nisbet\textsuperscript{1}, M. Sekowska\textsuperscript{4}, A. Wilk\textsuperscript{5}, A. Barrett\textsuperscript{6}, N. Hassanali\textsuperscript{7}, T.-P. Yang\textsuperscript{2}, D. Glass\textsuperscript{2}, S.-Y. Shin\textsuperscript{2}, L. Parts\textsuperscript{5}, N. Soranzo\textsuperscript{2,3}, R. Durbin\textsuperscript{2}, K. Ahmadi\textsuperscript{2}, K.T. Zondervan\textsuperscript{1}, C.M. Lindgren\textsuperscript{1}, T.D. Spector\textsuperscript{2}, E.T. Dermitzakis\textsuperscript{8}, M.I. McCarthy\textsuperscript{9,6}, P. Deloukas\textsuperscript{2} for the MuTHER Consortium. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Wellcome Trust Sanger Institute, Wellcome 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 weight change.

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.

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 INS, 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 9.6 Gb of sequence was produced per patient, at a mean depth of 95 with 87% high quality coverage (mapping quality ≥ 30) at at 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.

Exome analysis in 8,232 Finnish men identifies novel loci and low-frequency variants for insulin processing and secretion. J.R. Huyghe1,2, R.K. Semple4, E. Raffan4, A. Thompson2, I. Barroso1,2. 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 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 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 the GWAS signal. A nonsense 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 the GWAS signal. A nonsense 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 the GWAS signal. A nonsense variant in SGSM2 (1.4% MAF) was independent of the GWAS-identified common signal.

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.

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 with HbA1c and insulin secretion. We found 10492 cis eQTLs 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 Lsoft. We replicated multiple previously identified T2D associations including TCF7L2 (p=1.3e-13) and KNG1 (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.
Discovery of new loci for type 2 diabetes susceptibility has been achieved through trans-ethnic meta-analysis. We identified signals at GWAS loci that are predominantly homogenous across ancestry groups, suggesting that these signals are unlikely to be driven by synthetic association with rare variation.

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 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 across 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<5x10^-8), mapping to 10q11.21 (variations in TCF7L2) which has been implicated in T2D susceptibility. The lead SNP in 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, HNF4A and PEK (p=7.1x10^-4). At all 3 of these loci, the East Asian ancestry group showed the least evidence of association. 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.

New Loci for Obesity, Diabetes, and Related Traits
Novel locus including FGF21 is associated with dietary macronutrient intake. A.Y. Chu1, T. Workalemahu2,3, N.P. Paynter1, L.M. Rose1, F. Guliani-1, CHARGE, Nutrition Working Group4, Q. Qi2,3, G.C. Curhan2, E.B. Rimm2,3, D.J. Hunter2,3, L.R. Pasquale5,6, P.M. Ridker1,7, F.B. Hu2,3, D.I. Chasman1,8, L. Qi2,3 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=58,355 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 (P=9.6E-8). At the FTO locus, a variant not in LD (r²=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. Ishiura1, W. Sako2, M. Yoshida2, T. Kawai3, T. Tanabe3, S. Goto3, Y. Takahashi4, H. Date1, J. Mitsu1, B. Ahsan1, Y. Ichikawa1, A. Iwata1, H. Yoshino4, Y. Izumi5, K. Fujita, K. Maeda4, S. Goto3, H. Koizumi6, R. Morikagi6, M. Ikemura6, N. Yamauchi7, S. Murayama8, G. Nicholson9, H. Ito4,5, G. Sobue11, M. Nakabayashi12, R. Kaji13, S. Tsug11,13, 1 Department of Neuropathology, The University of Tokyo, Tokyo, Japan; 2 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 and variable degrees of sensory involvement. The majority of the lower motor neuron involvement is related to deficiencies per copy number variation (CNV) of 300-100,000. 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 Bukharan 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 dysmorphic features. Despite the complete 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-Bukharan origin. This is a one basepair frameshifting deletion in the TRK-fused (TRK-F) gene, leading to a premature stop codon and significant degradation of the protein. TECPR2 has been recently reported as positive regulator of autophagic accumulation. We therefore examined the autophagy-related fate of the SMN complex and MAP1B (p62) and LC3 levels 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 significantly decreased the autophagic flux, 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 spastin, 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.

53 Mutation in the Autophagy-related TECPR2 Gene Causes Hereditary Spastic Paraparesis. D. Oz-Levi1,2, B. Ben-Zeev4,6, E. Ruzzo2, Y. Hitomi1, A. Gelman2, K. Pelak1, Y. Anikster1,6, H. Reznik-Wolf4,6, I. Bar-Joseph4,6, T. Olender1, A. Alkelai1, M. Weiss1, E. Ben-Asher1, D. Ge7, K. Shianna2, Z. Elazar1,2, D. Goldstein5,6, E. Pras3,5, D. Lancet1. 1 molecular genetics, Weizmann Institute of science, Rehovot, Israel; 2 Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel; 3 Institute of Human Genomics, 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 is related to mutations within a large number of genes. The majority of the lower motor neuron involvement is related to deficiencies per copy number variation (CNV) of 300-100,000. 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 Bukharan 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 dysmorphic features. Despite the complete 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-Bukharan origin. This is a one basepair frameshifting deletion in the TRK-fused (TRK-F) gene, leading to a premature stop codon and significant degradation of the protein. TECPR2 has been recently reported as positive regulator of autophagic accumulation. We therefore examined the autophagy-related fate of the SMN complex and MAP1B (p62) and LC3 levels 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 significantly decreased the autophagic flux, 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 spastin, 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. Melki1, K. Peliak1,2, M. Weiss1,2, D. Oz-Levi1,2, F. Tiziano2, J. Veillet2, M. Bayes3, F. Nolent4, V. Garcia1, S. Servide2, G. Carsi5, R. Costo-Giner6, F. Llorens7, G. Carpintero-Abadie1, L. Gut3, T. Levade3, H. Topaloglu2,1. 1 UMR-788, Inserm and University Paris 11, Biomedical Research Institute, Kremlin Bicetre, France; 2 Istituto di Genetica Medica, Università Cattolica del Sacro Cuore, Rome, 00168, Italy; 3 Centro Nacional de Análisis Genómico, University of 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, Rome, 00168, Italy; 6 Unit of Molecular Medicine, Dept. Neurosciences, Ospedale Bambino Gesù Research Institute, Rome, 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, homoygosity 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 chromosome 17 in the third case. Exome sequencing 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 ASAHI1 mutations. An association between the activity below 10% results in Farber disease, an early onset disease starting with subcutaneous lipogranuloma, 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.

Copyright © 2012 The American Society of Human Genetics. All rights reserved.
55 Genotypic variants in Chromatin Modifying Genes cause D4Z4 Hypomethylation, DUX4 Expression, and Contrruction-independent Facioscapulohumeral Muscular Dystrophy (FSHD2). D.G. Miller1, R.J.L. Lemmers2, L.M. Petek1, J. Balog2, P.J. van der Vliet3, G.J. Block1, J.W. Lim4, G.N. Filipova5, A.M. Armel6, G.W.E. Santen6, B. Bakker7, M.J. Bams8, S.J. Tapscott9, R. Tawil10, S.M. van der Maarel1, 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 variated 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. 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 the individuals 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 result in a decrease of repression 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 the development of therapeutic strategies as well as diagnosis and counseling of FSHD-affected family members.

56 Unexpected extension of the phenotype spectrum associated with SMA2 mutations in Aneurysms-Osteoarthritis Syndrome. M. Aubart1, 2, D. Gobert2, N. Hanna1, 3, C. Multi3, J. Roume4, V. Cusin5, B. Grandchamp6, L. Gouya2, 3, D. Detaint2, G. Jondeau3, C. Boileau2, 3, 5, 7, 1) INSERM U698, Bichat University Hospital, Paris, France; 2) Molecular Genetics Laboratory, Ambroise Paré University Hospital, Boulogne, France; 3) 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, Loes-Eydt syndrome). Genetic and clinical heterogeneity in isolated TAAD are important. Recently, mutations in the SMA2 gene have been identified, associated with the presence of diffuse arterial aneurysms and torted ribs, and early onset osteoarthritis (Aneurysms Osteoarthrotic Syndrome, AOS). To assess the prevalence of this new syndrome a cohort of 105 French patients with TAAD without mutations in other known genes (BFN1, TGFB1, TGFB2, 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 those 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 those seen in Loes-Eydt syndrome was observed. They involved TAAD in 66% (23/35), arterial tortuosity in 67% (10/15) and other aneurysms (including cerebral aneurysms) in 60% (9/15). Skeletal anomalies usually associated with MFS were more frequent than in the general population, similar to TGFB2R2 carrier families, 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 surprisingly, 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 OTOLG and SLITRK6 are novel genes. M. Tekin1, 2, O. Diaz-Horta1, 2, D. Duman3, J. Foster1, 2, A. Simiard1, 2, M. Gonzalez1, 2, N. Mahdieh5, M. Bonvay6, F. Cengiz7, R. Ulloa1, 2, S. Zuchner1, 2, S. Blanton1, 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) Kawar 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. Aiglent SureSelect Human All Exon 50 Mb kit and an Illumina HiSeq2000 instrument were used. WBA 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 coded RefSeq exons, respectively. Twelve homozygous mutations in known deafness genes, of which 9 are novel, were identified: MYO15A-p.Q142X5, -p.S1481P, -p.A1515N, p.R383X, p.E677X, GIPC3-p.H170N, ILDR1-p.Q230X, MYO7A-p.G2163S, TECTA-p.Y1737C, TM13-p.V529Lfs*2, TMRPSS3-p.F131fs*10, TRIOBP-p.R7799fs*6. Sanger sequencing confirmed co-segregation of the mutation with deafness in 11 families. In three families, those that are not targeted by the exome capture kit and regions with high GC content. Two novel deafness genes, OTOLG and SLITRK6 were identified. OTOLG encodes a protein that has structural similarities to epithelial-secreted mucin (MUC) family members. SLITRK6 encoded a protein with structural similarities to SLIT/RtK family, highly expressed in the inner ear. A SLITK6 KO mouse model is characterized by sensorioreneural hearing loss. Homozygous c.890C>A mutation introducing a stop codon (p.S297X) in SLITRK6 co-segregates with deafness in a consanguineous family with 4 affected and 3 unaffected children. SLITRK6 is a neuromodulating type I cell-surface transmembrane receptor that activates intracellular G protein-coupled signal transduction. Mapping and functional studies of the protein may provide new therapeutic strategies for deafness.

Copyright © 2012 The American Society of Human Genetics. All rights reserved.
correlations are emerging which are of utmost importance for prognosis, arNSHI and in 21% of those with presumed adNSHI. Genotype-phenotype 
tion of the genetic defect in 25% of unrelated patients/families with presumed 
identification of novel deafness genes. WES so far resulted in 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 per-
formed 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, TMCI, 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 
further for confirmation. In conclusion, the presented results demonstrate that 
WES only is suitable for diagnosis in genetically highly heterogeneous 
diseases and that employing a combination of strategies facilitates the 
identification of novel deafness genes. WES so far resulted in the identifica-
tion 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.

A mutation in Ca2+ binding protein 2, expressed in cochlear inner hair 
cells, causes autosomal recessive hearing impairment. i. Schrauwens, 
S. Helfmann, A. Inagaki, F. Wolf, M.A. Tabatabaiifar, M.M. Picher, 
M. Sommen, C. Zazo Seco, H. Kremer, A. Dheedene, A. Lee, T. Moser, G. Van 
Camp. 1) Department of Medical Genetics, University of Antwerp, Universi-
teitsplein 1, 2610 Antwerp, Belgium; 2) InnerEarLab. Department of Otolar-
yngology, Collaborative Research Center 889, University of Göttingen 
School of Medicine, Robert-Kochstr. 40, D-37075 Göttingen, 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, Radbou-
d University Nijmegen Medical Centre, 6500 HB Nijmegen, The Nether-
lands; 6) Donders Institute for Brain. Cognition and Behavior, Radboud 
University Nijmegen Medical Centre, 6500 HB Nijmegen, The Nether-
lands; 7) Nijmegen Centre for Molecular Life Sciences, Radboud University Nij-
megen Medical Centre, 6500 HB Nijmegen, 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, Shahrek-
ord, 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+ influx by antagonizing calmodulin-mediated Ca2+ dependent inactiva-
tion (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 consan-
quineous families from Iran with moderate-to-severe hearing loss. This muta-
tion, 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 moder-
ate-to-severe sensorineural hearing impairment. The mutation may cause 
a hypofunctional CaBP2 protein that alters Ca2+ signaling in hair cells.
Comprehensive diagnosis for hearing loss using personal genomics: The first 100 cases. E. Shearer1,2, E. A. Black-Ziegelbauer2, M. S. Hildebrand1, A. P. DeLuca3,4, R. W. Eppsteiner1, S. E. Scherer5, T. E. Scheetz3,4,6, T. L. Casavant3,4,7, R. J. H. Smith1,2,7
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.

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 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.
Characterization of de novo copy-number variations in two subjects with a constitutional "CNV mutator" phenotype. P. Liu1, K. Walter5, K. Writz6, V. Gelowani4, S. Lindsay4, C.M.B. Carvalho1, M. Withers1, J. Wiszniewska1, A. Patel1, B. Rautenstrauss6, M.E. Hurles4, J.R. Lupski1,2,3.

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 complexity 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 implicating that rearrangements were produced by replication mechanisms, consistent with 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.

Associations between gene expression and phenotypes in 16p11.2 rearrangements. E. Migliavacca1,2, K. Mannik1, F. Zufferey2, N.D. Beckmann1, L. Harewood1, A. Mace1,2, Z. Kutalik4,2, L. Hippolyte1, M. Maillard1, V. Siffredi4, R.M. Witwicki1, G. Didelot1, S. Jacquemont1,2, S. Bergmann3, N.D. Beckmann1,2, J.S. Beckmann1,2,4, 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 Genétique, 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 neuropsychological phenotypes 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 PRR2T were identified in patients with epilepsy and paroxysmal dystkinesia; phenotypes often reported in deletion patients. An atypical 118kb deletion from MVP to KCTD13 segregated with ASD but not epilepsy, whereas a 138kb duplication encompassing SPN and QRT 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=0.0053; 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 introduces reducing the complexity of large sets of expression data by decomposing it into transcription modules, subgroups 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. Geleová et al. (2011) Cell 145: 611–619. In the absence of the duplicated segment, individuals appear to be normal.

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. Powell1,2, R. Coulson1,2, F. Cray1, S. Wong1, D.H. Yassul1, J.M. LaSalle1. 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, hyperagia, 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 small non-coding RNAs (ncRNAs): Snord116 small nuclear RNAs (snRNAs) and the spliced exons of the host gene, 116HG. While priority 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 paternally-specific, neuron-specific chromatin decondensation of the PWS locus. Furthermore, 116HG remains tethered to its site of transcription and separates from the neighboring IncRNA 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 RBP5 as interacting proteins. As SAF-A has both RNA- and DNA-binding domains and is known to function as a tether for the IncRNA XIST, we hypothesize that SAF-A acts as a tether between the 116HG IncRNA and the Snord116 locus. RBP5 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 IncRNA in the PWS critical region important to understanding the pathogenesis of PWS and related neurodevelopmental disorders.

Supported by NIH-F31NS073164 and NIH-1R01NS076263.
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 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 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.
68

Discovery and interpretation of balanced chromosomal aberrations in neurodevelopmental abnormalities and prenatal diagnostics. M.E. Talkowski1,2,3, P. Pillalamarri1, I. Blumenthal1, C. Hanscom1, Z. Ordulu4, J. Rosenfeld6, L.G. Shaffer6, J.F. Gusella1,2,5, C.C. Morton1,4,5. 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, discovered 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 cytogentic diagnostic studies that is tractable, interpretable, and immediately available to the diagnostic community.

69

Predispisition of acrocentric short arm fusions due to nuclear location, nucleolar disorganization, and telomere-induced DNA damage. K.M. Stimpson1,2, L.L. Sullivan5, S. Chen1, B.A. Sullivan1,2. 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 the 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.
Lessons learned from Next-Gen Cytogenetics: Whole genome sequence-based prenatal diagnosis of apparently balanced de novo chromosome rearrangements. Z. Orduño1, M. Tałkowski2,1,4, V. Pillipa2,6, S. Pereira1, I. Blumenthal1, C. Hanscom3, A. Lindgren4, N. Hussain5, S. Connolly6, L.E. Wilkins-Haug1, J.F. Gusella2,3,7, C.C. Morton1,2,8 1) Obstetrics/Gynecology, Massachusetts General Hospital, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 3) Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, USA; 4) Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; 5) Department of Radiology, Children's Hospital Boston, Harvard Medical School, Boston, MA, USA; 6) Department of Genetics, 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, 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 a 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 of two pregnancies. DGAP239 presented as a third trimester pregnancy with multiple fetal anomalies and 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(6)(q13q13)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 prenatatal 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 unfavorable 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 demonstrates 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 realizing the benefits. This study foretells an empowered prenatal diagnostic environment in which DNA sequencing becomes the standard of care.
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, "high-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), 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 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 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. A risk-assessment approach based on a single amniocyte analysis up to 55%. The highest variations were found in intermediate mutant load ranges. A risk-assessment approach based on a single amniocyte analysis 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.

The challenge of preconceptional, preimplantation, and prenatal genetic diseases with high transmission risk, due to their maternal inheritance, is to improve the follow-up. Preimplantation and prenatal heteroplasmy assessments are therefore valuable tools to prevent the recurrence of the disease in families where a mtDNA mutation segregates. Long clinical follow-up is required, and are thus valuable tools to prevent the recurrence of the disease in families where a mtDNA mutation segregates.
75 Spina bifida risk is conferred by multiple polymorphisms in folate one-carbon pathway genes. D. Gilbert1, K. Lazaruk2, J. Stein1, J.Hardin1, J. Wilde2, G. Shaw1, E. Lammer3, M. Marin4, J. Rine3, 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 risk prediction for 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 for prediction of 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 1,166 controls were genotyped with the 824 SNPs to define risk alleles and develop the prediction model. We found 58 single nucleotide variants (SNPs) with p-values < 0.01 in our 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 an unexpected population signal, 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. Uzun1,2, I. Kurihara1, J. Tavormina3, R. Cabezas4, A. Laliberte5, A. Dewani6, E. Triche7, J. Padbury1,2, 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 indicated 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 expression 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.
Genetic normalization of day-3 embryos: Results from two independent preimplantation genetic screening (PGS) laboratories. P. Brezina, E. Littman, Y. Sun, V. Phan, R. Anchan, A. Barker, M. Hughes, G.R. Cutting, W.G. Kearns.

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, Phoenix, 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, 100% (269/269) 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.
Recent work has definitively implicated TGFβ signaling dysregulation in the pathogenesis of the 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 (TIRβ/II), signaling effectors (Smad3) or ligands (TGFβ2)) would intuitively infer decreased TGFβ signaling. In addition, some phenotypic features of LDS (i.e. cranio-nasal dysmorphia 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. Haploinsufficient TGFβ2+−/+ 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 morpholinoinduced silencing of ski and skb 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).

Loss-of-function mutations in TGFβ2 cause Loeys-Dietz syndrome: towards solving the TGFβ paradox in aortic aneurysmal disease. The clinical 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 vascularopathies.
81 Genetic dissection of aortic disease in the Marfan syndrome. F. Ramírez1, L. Carta1, M. Del Solar1, M. Lindsay2, H. Dietz3, J.R. Cook4. 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 rupture. Only a few human genes have previously been associated with BAV, COA and PDA. We assayed the embryonic expression of mouse Matr3 in the developing heart and great vessels, whereas Matr3 is strongly expressed in the developing heart and great vessels. We also assayed Matr3 expression in Matr3/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 (Cdhr5-Cre), mice without fibrillin-1 microfibrils in the medial layer of either the aortic root and the entire aortic tree (Sm22-Cre) or only in the aortic arch (Wnt1-Cre) died from ruptured TAA. The genetic data therefore identify the medial layer of the ascending aorta as the predominant determinant of TAA formation and progression. New findings highlight the importance of the early postnatal maturation of the aortic arch and establish that Matr3 is strongly expressed in the developing heart and great vessels, whereas Matr3 is not. Moreover, while homozygous Matr3 loss-of-function mice are embry 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, revealing a striking developmental sensitivity to Matr3 gene dosage, and establish that Matr3 loss-of-function in mouse causes aortic and arterial valve phenotypes strikingly similar to those associated with human MATR3 disruption.

83 Identification of the cause of Blue Rubber Bleb Nevus Syndrome. J. Sobiet1, N. Limaye1, M. Cordisco2, A. Dompmartin3, O. Enjolras4, S. Holden5, A.D. Irvine6, C. Labrèze7, A. Lanoe8, P.N. Rieu8, S. Syed9, C.J. van der Vleuten10, R. Watson11, S.J. Fishman12, J.B. Mulliken13, L.M. Boux14, M. Vikulla15. 1) Human Molecular Genetics, du Duve Institute, Université catholique de Louvain, Brussels, Belgium; 2) Department of Dermatology, Hospital General in 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 School, Boston, MA; 14) Center for Vascular Anomalies, Department of 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-mutated phenotypes. To test this hypothesis, we sequenced 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 control tissues. These changes occur in highly conserved residues, and are not found in dbSNP. In contrast to VCMCs and VMs, BRBNs predominantly show double (cis) mutations, suggesting a pheno-genotype correlation. They cause ligand-independent receptor hyperphosphorylation in vitro. These results unequivocally demonstrate that BRBNs are caused by post-zygotic activating TIE2 mutations.
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 Boston, USA (mass spectrometry); 6) Centre d’Imagerie BioMédical, INSERM, CNRS UMR8808, AP-HP, Hôpital Necker, Paris, France; 7) Cliniques universitaires Saint-Luc-ULC, 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) Genetics, Hospital Clinic, University of Barcelona, Barcelona, Spain; 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 genetic variants that are associated with CHD in DS. We identified two CNV regions of 4.8 and 1.8 Kba near genes RIP4 and ZNF295 on chromosome 21 (KRTAP7-1 gene) as CHD risk alleles (adjusted p-values < 0.05). Further analysis revealed that expression levels of these genes were significantly higher in DS compared to non-DS controls. We also searched for CNVs in other susceptibility regions and identified two CNV regions on chromosomes 5 and 8 that were significantly associated with CHD in DS, but not in non-DS controls. These findings suggest that genetic variation in specific genomic regions may contribute to the risk of CHD in DS.


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 patients and 8 non-trisomic individuals. We further 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 < 10^-10", 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 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, it could be a result of canalization in DS gene expression which can result from purifying prenatatal 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) Alternatively, decreased variance in 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 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 ALTP 527-2010.
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 Hof¹,², Y.M. Ruigrok¹,², A.F. Baas³, L.A.L.M. Kiemeney⁴,5,6, S.H. Vermeulen⁴, A.G. Uitterlinden⁷,8, F. Rivadeneira⁸, A. Hofman⁹, G.J.E. Rinkel¹,², P.I.W. de Bakker³,9,10,11.

¹) Department of Neurology, University Medical Center Utrecht, Utrecht, The 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.
The value of population-specific reference panels for genotype imputation in the age of whole-genome sequencing. C. Fuchsberger1, B. Howie2, M. Laakso3, M. Boehnke1, G. Abecasis1 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 Medical Genetics, 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, and 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 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 in 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 panels, which capture well the unique genetic structure within the population. Our results suggest that upcoming population-specific panels composed of 1000s of individuals will boost imputation quality substantially for rare variants.

Fast and accurate 1000 Genomes imputation using summary statistics or low-coverage sequencing data. B. Pasaniuc1,2, N. Zaitlen1,2, G. Bhatia1,2, A. Gusse1,2, N. Patterson2, AL Price1,2. 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 studies. Haplotype information is particularly useful for estimating the risk of complex diseases. However, imputation accuracy varies substantially, with the best accuracy achieved using a reference panel created by our group which consists of >10,000 individuals from a variety of different ancestry groups. We evaluated the value of population-specific reference panels for imputation accuracy for all SNPs, especially those with MAF 1–3% ($r^2 > 0.05$) between imputed dosages and masked genotypes.


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 we applied SHAPEIT to genotypes from heterozygous individuals 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 to mimic realistic allele frequency spectra and demography and then simulated reads based on these. The performance of SHAPEIT is assessed in terms of 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 at these conditions, with only 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.
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 possibility 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. Risch (2001) suggested that the control 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- 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 design that has done sequence 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 for the top 500 sites was 99.7%. We demonstrated that 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.
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) multi-ethnic 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 depressive 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.

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 10% in this cohort. Because the remaining 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.
Autism Traits in the RASopathies. I. Corbin1, G. Desachy1, K. Rauen2, L. Weiss3. 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 (ASD) 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 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.

Identification of novel recessive mutations in genes for intellectual disability. B. De Vries1, J.H.M. Schuurs-Hoeijmakers1, A.T. Vulto-van Silfhout1, L.E.L.M. Vissers1, J. de Ligt1, C. Gilissen1, I. van de Vondervoot1, M.T. Greally2, C.W. Ockelen1, M.H. Willemse1, E.M. Bongers1, G. Hira1, J.Y. Hehir-Kwa1, R. Pfundt1, T. Kleefstra1, K. Neveling1, S. Reinitate3, A. Vitello1, P. Failla3, D. Greco3, M. Fichera4, O. Galesi5, B.M.W. van Bon1, J.A. Veltman1, C. Romano3, M.A. Willemse1, H. van Bokhoven1, A.P.M. de Brouwer1. 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 mono- genic, 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 SLC6A6, both on the X-chromosome, were identified in two brother-pair families. Furthermore, seven novel candidate genes for ID were identified in two brother-pair families. Seven novel candidate genes for ID were identified in two family groups. One of these genes, PAMR1, was present in two affected brothers. Recessive missense variations affecting amino acids that are conserved amongst (at least) vertebrates and predicted to influence protein function 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 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 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 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 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 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 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 TDP2, a homozygous splice donor site variation inevitably resulting in aberrant mRNA splicing segregated in a family with three affected brothers.
Causal de novo SNVs, indels and CNVs in children with undiagnosed developmental disorders. M. Hurles1, M. van Kogelven2, K. Morley1, T. Fitzgerald1, S. Gerety3, A. Tivey1, S. Gribble1, S. Al-Turki1, S. Clayton1, C. Wright1, J. Barrett1, H. Firth1,2, D. FitzPatrick4, N. Carter1 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 CGMM, 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 84 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 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. Willemsen1, W.M. Wissink-Linde1, L.E.M. Vissers1,2, A.P.M. de Brouwer1, J.H.M. Rensen2, N. de Leeuw3, R. Pfundi1, H.G. Yntema1, J. de Ligt1,3, J.A. Veltman1,2, H.G. Brunner1,3, H.M.J. Lantman - de Valk4, B.C.J. Hamel1, H. van Bokhoven1,2, T. Kleefstra1,3, 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% monogenic defects, mostly fitting syndromes, for which the causative genes were identified recoding EHMT4, 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 PDHA1, GRIN2A, LR2P 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. Mullegama1, J.A. Rosenfeld1, C. Oreilana1, B.H. van Bon2, E.A. Repnikova3, L. Brick2, L. Dupuis2, D.J. Stavropoulos1,6, D.L. Thrush7, J.G. Foster8, K. Manickami9, A. Lin10, J.C. Hodge11, M.E. Talkowski12,13,14, J.F. Gusella10,12,14,15, S. Schwartz16, S. Aradhya17, R.E. Piatt18, B.E.A. de Vries19, R. Mendoza-Londono19, L.G. Shaffer2, S.H. Elisea21,1,1 Huinan 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, Division of Clinical and Metabolic Genetics, 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). A recent report highlighted the importance of common regulatory pathway in intellectual disability 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 phenotypic overlap of genetic and neurodevelopmental disorders (NDs) shares 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 that 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 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 NTRK2, 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


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 ascertainment 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 developed 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 variants present 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 KIAA2022 mutation segregated with the clinical phenotype 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 XLID families.
105

Biallelic mutations of a ubiquitin-ligase-encoding gene cause an Ohdo-like intellectual disability syndrome. G. Borck1, B. Dallapiccola2, R. Ramirez-Sola3, A. Segrel4,5, H. Thiele6, A. Edwards7, M.J. Arends8, X. Miro9, J.K. White10, J. Desir11, M. Abramowicz10, M.L. Dentici2, K. Hofmann4, A. Har-Zahav11, E. Ryder1, N.A. Karp3, N.J. Ingam5, G. Nuernberg6, S. Abdelhak12, M. Pasmanik-Chor13, O. Koen4, R.I. Kelley14, M. Shohat11,16, P. Nuernberg5,6, J. Flint7, K.P. Steel19, T. Hoppe4,5, C. Kubisch4, D.J. Adams3, L. Basal-Vanagaitė11,16. 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, Bellinson 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.
Androgenetic alopecia: identification of four new genetic risk loci and evidence for the contribution of WNT-signaling to its etiology. S. Heilmann, A. K. Kiefer, N. Kluck, D. Drichel, A. M. Hilmer, C. Herold, J. Y. Tung, N. Eriksson, S. Redler, R. C. Betz, R. Lü, H. Stefansson, D. R. Nyholt, K. Song, S. H. Vermeulen, S. Kanoni, G. Dedoussis, N. G. Martin, L. A. Kiemeney, V. Mooser, K. Stefansson, J. B. Richards, T. Beecker, F. F. Brockschmidt, D. Hinds, M. M. Nöthen. 1) Institute 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, Reykjavik, 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) Theodor carcinogen (DZNE), Bonn, Germany; 18) equally contributing senior authors.

Androgenetic alopecia (AGA, male pattern baldness) is the most common form of hair loss in 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 genes and genome-wide association studies have identified candidate genes and SNPs. 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 x 10^-5) but fell below the threshold of genome-wide significance (P-value > 5 x 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 meta-analysis data identified four additional genome-wide significant risk loci for AGA on chromosome (chr)2q35, chr3q25.1, chr6q33, and chr12p12.1. The strongest association signal was obtained for rs7349332 (P-value = 3.55 x 10^-8) on chr2q35 located intronically in WNT10A. WNT10A is a member of the family of WNT-genes that encode small secreted signaling proteins which play important roles during development and tissue homeostasis. Interestingly, WNT-signaling has 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.
GWAS from Head to Toe

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. A. S. Eskin1, A. R. Wood2, S. Vedantam3,4,5, J. Yang3, S. Gustafsson3, S.I. Berndt6, J. Karjalainen7, H.M. Kang8, A.E. Locke8, A. Scherag11, D.C. Croteau-Chonka12, F. Day14, R. Magi1, T. Ferreira15, J. Randall16, T.W. Winkler17, T. Fall17, Z. Kutalik17, T. Workmanohu18, G. Abecasis19, M.E. Goddard7, L. Franco4, R.J.F. Loos19, M.N. Weedon20, E. Ingelsson21, P.M. Visscher22, J.N. Hirschhorn3,4,5, T.M. Frayling23, 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, Medical Biotechnology and Genomics, 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 27595, 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) Department of Medical Biology, Umeå University, Umeå, Sweden; 17) Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland; 18) Department of Nutrition, Harvard School of Public Health, 655 Huntington Avenue, 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 18% of the heritable component (95%). 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. Within the GIANT consortium, we performed a GWAS of ~250,000 individuals. In conclusion, data from 250,000 individuals for distant relatedness across samples indicate that a large proportion of the observed enrichment for novel and biologically relevant pathways in these samples to predict the biological function of poorly annotated genes, we used a novel pathway analysis approach that usosco-expression data from 80,000 people from the GIANT consortium and altogether explained 1/4 of the inherited component in adult height.

109 Genome-Wide Association Studies (GWAS) meta-analysis for fracture risk points to loci related to hormonal and neurological pathways: the GIANT consortium. E. Ingelsson1, J. J. Frayling1, A. E. Ntziora2, C. Nielsen3,4,5,6,7, U. Styrkarsdottir8, P.M. Ridker9, K.K. Tsilidis10, K. Estrada11,2,12, A. Enneman13,14, A. Vernon-Smith14, R.D. Jackson15,11,12, S. Trompet16,13, T. Lehtimäki17, S. Kaptoge18, T.B. Harris19, E. Eriksson20, N. Amin21, A. Metspalu22, R. Magi23,1,21, P.A. Ioannidis21,23,24, G. Thorleifsson13,21, A.G. Uitterlinden12,2, S.A. Cummings25,26, T. Spector27, D.P. Kiel28,29, D. Chasman2, E. Orwell3, J.B. Richards23,29, F. Rivadeneira23,29, 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, Quebec, 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, Reykjavik, 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) Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands; 13) University of Tampere and Tampere University Hospital, Tampere, Finland; 14) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 15) 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 Medicine, 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, CA, USA; 22) California Institute of Technology, Pasadena, 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 analyzing 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, 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 test secondary effects of previously associated loci, and identified a total of 10 new loci. The most significant SNP in 7q21 (SHFM1) (P=2.4×10−10, odds ratio = 1.10, 95% CI: 1.07-1.13) and in 18p11 (FAM210C) (P=1.4×10−6, odds ratio = 1.09, 1.06-1.12). These loci were recently identified as bone mineral density (BMD) associated loci in the eMERGE cohort. Other effect GEFOS loci included 7q21, where an SNP in 7q21 is significantly associated with BMD and fracture. In regression-based analysis controlling for the 7q21 signal includes SHFM1; this region is characterized by genome rearrangements 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 of these genes are known to cause skeletal abnormalities. These markers were further investigated. The 7q21 signals reached GWS in 17q11 (P=1.3×10−8) close to ZIC1 and ZIC4, which is involved in Dandy-Walker malformation, where affected individuals have motor deficits such as delayed development, hypotonia and ataxia. Suggestive SNPs (P<5×10−5) mapped in or near RSPO3 (P=6.3×10−5), LRPS3 (P=4.1×10−5), CDC242E3/F3/MAB2A1 (P=2.1×10−5), ZFHX3 (P=2.8×10−5), NOVA1 (P=4.1×10−5), BFLN5 (P=5.0×10−5). These seven 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 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.
Genetic landscape of the red blood cell. J.C. Chambers1, P. van der Harst2, W. Zhang3, I. Mateo Leach4, J. Sehmi5, N. Verweij5, D.S. Paul8, A. Rendón1, U. Elling5, H. Aliyee6, A. Radhakrishnan6, J.S. Beckmann6, C.V. Dedoussis7, P. Deloukas7, A.A. Hicks8, S. Sanna10, M. Uda10, J. Penninger5, C. Gieger17, J.S. Koosker1, W. Ouwehand9, N. Soranzo3, 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<1x10-8 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 genome-wide significance (r2>0.8) with sentinel SNPs at the red blood cell loci (∼P<1x10-8). These loci are novel. Together these loci explain 5−10% of the 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 (∼P<1x10-8) and then performed a comprehensive search for sequence variants associated with one or more red blood cell phenotypes at P<1x10-8, 43 of these loci were novel. Together these loci explain 5–10% of the phenotypic variance in some of these traits. 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 sentinel SNPs at the red blood cell loci (∼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 NDRs 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 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.
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. Kvale1, K. Laplham1, T. Hoffmann1, S. Sciortino2, L. Walter2, Y. Banda1, I. Listerman1, J. Lin1, S. Hesselson1, P. Kwok1, E. Blackburn1, C. Scheider2, N. Risch1,2. 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=assayed on). 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^{-10}); 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^{-10}), 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^{-12}), on chromosomes 1 and 12. These SNPs encoded amino acid substitutions, one for a G-coupled receptor, and the other for a Migration 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 is made possible by the collaborative, large-scale nature of genetic studies, in which results are reported to the community and to the field in a timely fashion. We have identified novel, strongly associated loci with telomere length that provide valuable information to the field of aging research.

113 Heritability of the Variation in Aging in Two Longitudinal Family Cohort Studies: SardiNIA/Progenia Study and Framingham Heart Study. J. Bragg-Gresham1, S. Sanna1, C. Sidoro1,2, A. Mulas2, F. Busonero2, A. Mascio3, M. Urru4, F. Reiner5, R. Berutti2,4, M. Marcelli2, M. Oppo5, D. Pitzalis5, M. Zolodzhiewska5, A. Angius6, C. Jones4, A. Cao4, M. Udo2,2, S. Kardia1, D. Schlussinger7, F. Cucca2, G. Abecasis1. 1) Bioinformatics, Univ Michigan, Ann Arbor; MI; 2) Istituto di Ricerca Genetica e Biomedica (IRGB); CNR, Monserato, 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) Michi- gen Center for Genomics & 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 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 (r^2 = 0.81). We validated this set of predictors in the Framingham Heart Study and observed continued high ability to predict chronological age (r^2 = 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 biological age would help identify the most rapid aging 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 age and chronological age corresponded to a 40% increased mortality) and with significantly higher odds of a variety of aging related conditions (hypertension, metabolic disorder, kidney/uric acid disease, coronary artery disease, immunologic disorders, diabetes, and depression). Overall, we observed heritability of 30–40% for physiological age and with significant higher odds of a variety of aging related conditions. Given our results, we have reason to believe that physiological aging is highly heritable and should be amenable for GWAS analysis. Genome-wide association studies and replication are currently ongoing.


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 dysmorphic 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 Sardinia and Framingham Heart Study samples. Heritability was highest among 44–63 year olds (peaking at h^2 of ~0.60) and lowest among the youngest and the oldest individuals in each group. Our investigations result in quantitative measures of physiological aging that is highly heritable and should be amenable for GWAS analysis. Genome-wide association studies and replication are currently ongoing.
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. Kanoni1, C. Willenborg2,3, N. J. Samani4, P. Deloukas5, P. D. Franks6, S. Chen, J. Huang, H. Li, X. Yang, Y. Hao, China Atherosclerosis Genetics Consortium. State Key Laboratory of Cardiovascular Disease, Fujui Hospital, Chinese Academy of Medical Sciences, Beijing, Beijing, China.

Coronary artery disease (CAD) is the 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 CDKNA2A 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.

Discovery of 63 Novel Loci and Refinement of Known Loci Associated with Lipid Levels. C. Willer2,3, P. D. Franks6, S. Chen, J. Huang, H. Li, X. Yang, Y. Hao, China Atherosclerosis Genetics Consortium. State Key Laboratory of Cardiovascular Disease, Fujui Hospital, Chinese Academy of Medical Sciences, Beijing, Beijing, China.

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 LRPPAP1. Gene-set enrichment analyses identified genes near novel genome-wide significant index SNPs that appear in pathways with known lipid-related functions 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.
Cardiovascular Genetics: GWAS and Beyond

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. Hoffmann1–4, M.N. Kvale5, Y. Banda6, S.E. Hesselson7, L. Waite1, S. Sciorino3, D.K. Ranatunga2, D. Ludwig1, C. Iribarren7, R. Grant3, P. Kwok2,4, C. Schaefer1, N. Risch1,3, 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 genotype data (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 GWA hits (p < 5×10^-8) from the large meta-analysis of Teslovich et al. (2010), and found that at a p < 0.001, 36 SNPs replicated, 8 did not replicate, and one was uninformative. Further, we identified 7 novel loci at genome-wide significance (p < 5×10^-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 < 10^-17 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 < 0.001, 36 SNPs were replicated, and one was uninformative. Twelve novel loci for LDL were also identified at genome-wide significance. Finally, a few genes were 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 the 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. Surakka1,2, A.-P. Sarin1,2, M. Hömberg1,4, M. Horikoshi4,5, S. Wilthshire4,5, T. Eso1,2, F. Ferreira6, L. Marullo4,7, G. Thorleifsson8, A. Maha-jan1, S. Hägg9, A. Isaacs10,11, M. Beekman12,13, J.S. Ried14, T.W. Winkler15, C.P. Nelson14,15, C. Willenborg16,18, A. Morris15, M.L. McCarthy14,20, J. Proko- penko16, S. Ripatti17,18 for the ENGAGE Consortium. 1) The Institute of 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, Deutsche Forschungsgemeinschaft, Institute 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 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, University of Oxford, 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 cholesterol (LDL), total cholesterol (TC), and high-density lipoprotein cholesterol (HDL)), there is still a substantial number of genetic factors that have yet to be identified. This study demonstrates that existing large-scale cohorts with extensive familial and longitudinal information have tremendous potential for discovering new genetic variants associated with plasma lipids.

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 from five countries (4,962 individuals) and combined these results with genome-wide association and genome resequencing data for the low frequency range. We identified 19 novel risk loci and 18 newly identified variants with stronger effects in previously reported lead SNPs. As an example, the lead SNP for LDL-cholesterol at the PCSK9 locus had minor allele frequency (MAF) below 5%. For example, the lead SNP for LDL-cholesterol at the PCSK9 locus had MAF of 1.4% and the four-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 (ABC460/8 and MOC1, with MAFs 10.5% and 28.0%) that are lead SNPs at those loci. In 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 2Mb from known loci. For eight of these (near ZDHHC18, TMEM48, PS39, PRKAG3, ADH1A1B, ORAC13, 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 PCSK9 locus had MAF of 1.4% and the four-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 (ABC460/8 and MOC1, with MAFs 10.5% and 28.0%) that are lead SNPs at those loci. In 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 in genome-wide association studies 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 3,935 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 (85% 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:0.68) had a modest association (p=0.02). After excluding previously reported GWAS variants, we identified 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. Nguyen1, A.C. Morrison2, A. Li7, R. Gibbs5, E. Boerwinkle5, A. Chakravarti1. 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 phylpP conservation score (S) and was classified as synonymous, missense/exon splice junction, non-NMD nonsense, non synonymous) 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 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 phylpP 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.
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, Q. Yang, F. Chen, X. Liu, K. Keen, P. Jacques, W.M. Chen, G. Weinstein, F.C. Hsu, A. Beiser, L. Wang, K.F. Doheny, P.A. Wolf, M. Zilka, J. Selhub, B.B. Worrall, S. Seshadri, M.M. Sale, 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 Friedman 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 (ΔtHcy), have been associated with risk of incident stroke in several community-based cohorts; tHcy in the Framingham Heart Study (FHS) and ΔtHcy with recurrent stroke in the Vitamin Intervention for Stroke Prevention (VISP) study. We conducted genome-wide association (GWA) analyses of ΔtHcy 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 ΔtHcy with SNPs within, or flanking, the glycine N-methyltransferase (GNMT) gene (p = 1.60×10^{-6}) and the major regulatory differences that may directly lead to the differences in ΔtHcy. Meta-analysis also revealed 4 additional one carbon metabolism genes associated with ΔtHcy, ALDH1L1 (p = 7.3×10^{-13}), CBS (p = 3.15×10^{-26}), CPS1 (p = 9.10×10^{-19}), and PSH (p = 1.17×10^{-16}), indicating a strong genetic component to ΔtHcy 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<0.001) 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 ΔtHcy. 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.
Baraitser-Winter syndrome: delineation of the phenotypicspectrum in a large series of molecularly defined patients, A. Vennoes1, O.A. Abdul-Rahman2, J. Allison3, J.F. Atkin4, M. Baraitser5, H. Brunner6, N. Chassaing6, K. Devriendt6, V. Drouin7, A. Fry8, J.P. Frysns9, F. Giuliano10, K.W. Gripp11, D. Lacombe12, A. Lin13, G. Manni14, M. Marbè15, M. Nezarati16, M. Nowaczyk17, S. Osimani18, J. Rijs19, C. Rusu20, Y. Sznajer21, C. Van Ravenswaay21, M. Masliah21, J.B. Rivière21, B.W.M. van Bon3, A. Hoischen4, W. Dobyns5, P. Dilz6, 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 30 patients with BWS, emphasizing the clinical variability of the syndrome, which also encompasses Fryns-Aftimos type. Previously, three other specific missense mutations in MED12 have been described in patients with FG syndrome and Lujan-Fryns syndrome. 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 coloboma is present in many cases, as does deafness. Pachygyria with an anterio-posterior gradient is present in most cases. Progressive joint stiffness and postural microcephaly may develop. Intellectual disability and epilepsy are variable and correlate with CNS anomalies.

Three novel mutations in MED12 cause Ohdo syndrome Maa-t Kievtit-Brunner type, A.T. Vulto-van Silfhout1, A. Hoischen2, B.W.M. van Bon3, W.M. Illesn4, C. Gilsen5, F. Gao6, J.M. Spaeth7, B.C. Hamel8, T. Kleestra9, M.A.A.P. Willemsen10, H. van Bokhoven11, H.G. Yntema12, B.B.A. de Vries13, H.G. Brunner13, T.G. Boyer14, A.P.M. de Brouwer12, 1 Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The 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 (OMIM 249620) is characterized by intellectual disability and typical facial features including blepharophimosis. Clinically the blepharophimosis-intellectual disability syndrome has been classified in five distinct subgroups: del(3)(p)ter type, Ohdo type, Say-Barber-Biesecker-Young-Simpson (SBBYS) type, Veloform type, and Maa-t Kievtit-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 most clearly differ from these two syndromes as they have the Ohdo classical features piosis and blepharophimosis, while the tail forehead macrocephaly, and broad thumbs and hallucs 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 with 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. The identification of an X chromosomal gene in Ohdo syndrome has important implications for the recurrence risk in the families.
will hopefully lead to the identification of another disease gene presumably supporting impaired TGFβ driven transcriptional control in MS. We finally observed a decreased expression of downstream TGFβ signaling. Thereafter, LTBP4 mutations all affecting Isoleucine 500 which is 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. Using exome sequencing in 2 MS cases, we selected mothers against DPP4 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 SMAD4 ubiquitination in patient fibroblasts and increased level of SMAD4 mRNA and protein levels. This correlated with increased transforming growth factorbeta (TGFβ) signaling. Therefore, LTBP4 mutations have been implicated in a similar phenotype. Studying mutational hotspot in exon 32 (c.2262delA) of DERMATOPATHOLOGY University Hospital of Liege, Liege, Belgium; 15) 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. Autoimmune/inflammatory cutaneous manifestations, including mainly cutaneous vasculitis, are the predominant skin alterations. Rarely, mutations can be identified in FBLN4 or FBLN5. Recently, LTB4 mutations have been implicated in a similar phenotype. Studying mutations in FBLN4 and LTB4 in 12 families with ARCL type I, we found bi-allelic LTB4N mutations in 2 probands, whereas 9 probands harbored biallelic mutations in LTB4P. In one patient, no mutations were found. LTB5N and LTB4P mutations cause a very similar phenotype associated with severe pulmonary emphysema, in the absence of vascular tortuosity or aneurysms. Gastrointestinal and genitourinary tract involvement seems to be more severe in patients with LTB4P mutations. Functional studies showed that most premature termination mutations in LTB4P result in severely reduced mRNA and protein levels. This correlated with increased transforming growth factorbeta (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. Therefore, LTB4P mutations may cause disease through both loss of function and gain of function mechanisms.
Clinical features of individuals with Floating-Harbor syndrome due to mutations in SRCAP


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) Ambulatorio 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. Through our exacting approach, we identified two additional mutations. One individual (at the 3' end of the gene) was the 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 strictures, constipation, diarrhea), hyperuricemia (91%), renal anomalies (65%), facial anomalies (53%) and/or vomiting/reflux in 13 – only one individual with Celiac disease), ophthalmologic findings (hyperopia, strabismus, nystagmus in 10), and ear issues (24%) in nine individuals. Mutations associated with FHS were observed 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. Through our exacting approach, we identified two additional mutations. One individual (at the 3' end of the gene) was the 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 strictures, constipation, diarrhea), hyperuricemia (91%), renal anomalies (65%), facial anomalies (53%) and/or vomiting/reflux in 13 – only one individual with Celiac disease), ophthalmologic 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, hydrenephrosis, 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.
A prospective natural history study of DICER1-related familial pleuropulmonary blastoma (PPB) syndrome shows incomplete penetrance, pleiotropy and variable expressivity. D.R. Stewart1, L. Doros2, G. Glenn3, A. Bauer4, G. Williams4, A. Carr5, J. Ivanovich6, R. Kase5, L. Harney5, K.A. Schütz7, C.P. Kratz7, L.P. Dehner8, D.A. Hill9, Y. Messinger4, 1) Clinical 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 ("affecteds") 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; 2/6 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 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.
Clinical implementation of a cancer care model based on comprehensive molecular profiling of tumor-normal pairs. J.C. Taylor1, K. Kaur2, S. Henderson3, E. Domingo4, A. Cutts5, J. Woods5, C. Molley5, B. Dougerty5, M. Middleton1, B. Hassan6, Y. Wang6, E. Beasley6, M. Naley7, I. Tomlinson1,2, A. Schuh2, 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.

BACKGROUND: We are conducting a multicenter clinical trial to evaluate the feasibility of cancer genomic sequencing in patients with advanced cancer. 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 approach 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 which biologic plausible actionable mutations are identified. All patients receive a report on actionable mutations back to clinicians and patients with three weeks. RESULTS: 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 BRCA1 mutations in melanoma.

Identification of novel mechanisms of drug resistance in BRCA1-deficient cancer by exome and RNA sequencing. K.K. Dhillon1, T. Taniguchi2,1


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 due to secondary BRCA1/2 mutations in recurrent ovarian cancers is associated with cisplatin resistance. However, restoration of BRCA1/2 does not necessarily equate to cisplatin resistance. Therefore, we hypothesized 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 FANC D2, 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 rendered the resistant clones sensitive to cisplatin again and significantly reduced CtIP and FANC D2 foci formation. Overexpression of FANCi in cisplatin-sensitive parental HCC1937 cells resulted in increased CtIP and FANC D2 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.


As sequencing costs fall ever lower and large-scale gene, exome and genotyping sequencing in patients is becoming routine, many individuals come 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, including in silico tools. The aims of the study were to report the spectrum of genetic variants reported in a large population cohort. 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 variants 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 variants we identified 48 missense variants in BRCA2 and 28 in BRCA1 of which 40 and 15 respectively had frequencies <1%.

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 rendered the resistant clones sensitive to cisplatin again and significantly reduced CtIP and FANC D2 foci formation. Overexpression of FANCi in cisplatin-sensitive parental HCC1937 cells resulted in increased CtIP and FANC D2 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.
Targeted re-sequencing of 10 ovarian cancer candidate genes in 2,240 cases and 355 controls. H. Song1, M. Cleck1, J. Cunningham2, B. Fridley3, E. Dicks1, P. Harrington1, S. Ramus4, S. Gayther5, E. Goode6, P. Pharoah7.

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 BRCA1/2 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 (98%). 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 (0.1%) 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 (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.
Risk of colorectal cancer for monoallelic and biallelic MUTYH mutation carriers. A.K. Win¹, S.P. Cleary²,3, J.G. Dowty¹, D.D. Buchanan⁴, J.P. Young², N.M. Lindor⁵, R.W. Haile⁶, P.A. Newcomb⁷, L. Marchand⁸, J.L. Hopper¹, S. Gallinger²,3, M.A. Jenkins¹, 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, Banchroft 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 (proband) 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 (penetration) 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 the most 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.
Newborn screening for cystic fibrosis: Preliminary results on the false positive experience. C.J. Barg1, F.A. Miller1, R.Z. Hayeems1,2, P. Dune3, J.C. Carroll4, T.J. P. Chakrabory5, B.K. Potter4, Y. Bombard6,9, K. Tam7,11, L. E.Kerr12, C. Davies13, J. Milburn3, F. Rajtjen13, A. Guttmann2,11,12,1
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 NBS results has not been systematically investigated.

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 screened 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 89%) and 13 control mothers 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 (p < .05). Qualitatively, mothers of FP infants reported the time between initial notification and confirmatory testing was highly stressful, but identified personal, familial, and clinical sources of support. Groups did not differ on marital status and income (p > .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 (p < .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.
144

Do Research Participants Really Want to Know? The Seattle Colorectal Cancer Family Registry Experience on the Return of Research Genetic Test Results. M. Laurino1,2,3, D. Fisher4, W. Grady1,3,4, P. Newcomb1,3,4 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) Washington State 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 advisors 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 41/72 (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 assessments (GC 1 session, post) 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 and findings from this research study indicate that 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. S. Thoss1, S. Kupper1, J. Youngblom1, C. Hartshorne2 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 sudden death in student-athletes. 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 general practitioners 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. The purpose of this study was to assess current SCT knowledge and impact of genetic counseling in the context of mandatory genetic testing in student-athletes, and understanding of disease prevention and health promotion.

146

Impact of Direct-to-Consumer Pharmacogenomic Testing. C.S. Bloss1, N.J. Schork1,2, E.J. Topol1,2,3 1) Scripps Translational Science Institute and Scripps Health, La Jolla, CA; 2) Department 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 sucinylcholine, 5FU/capacabine, abacavir, azathopurine/6-MP, beta blockers, carbamazepine, fosinopril, 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: 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<0.05). 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 a 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 utility of the test as a function of current or previous use of any of the medications on the panel.
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 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.

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 accurate 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.
Dynamics, definitions and discrepancies: public perspectives on the systematic collection and use of family health history in routine health care. H. Etchegary, B.J. Wilson, S.M. Craigie, S.G. Nicholls, D. Castle, J.C. Carroll, J. Allanson, B.K. Potter, P. Chakraborty 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 target 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 may 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.
Dense fine-mapping study identifies novel disease loci and implicates coding and non-coding variation in primary biliary cirrhosis risk.

J.Z. Liu1, M.A. Almarni1, D.J. Geffray4, G.F. Mells5, L. Jostins1, P.T. Donaldson8, A. Bathgate6, A. Burroughs7, D.E. Jones4, G.J. Alexander10, J.C. Barrett1, R.N. Sandford6, C.A. Anderson1, 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) (r^2>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 eQT in B-lymphoblastoid cells. We found multi-population 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 B8B1*0301 and newly implicating DRB1*0401 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 based on the number of OC regions among candidate causal variants (r^2>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 1,000 permutations, and repeating the enrichment analysis for each cell type. 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=0.086). Indeed, of the 26 independent non-HLA signals tagged on Immunochip, 15 have SNPs in B-lymphoblastoid OC regions in high LD (r^2>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.
Fifteen novel psoriasis susceptibility loci: disease-specific signals highlight the role of innate immunity. L.C. Tsoi1, S.L. Spanij2, J. Knight3,4, E. Ehlinghaus5,6, P.E. Stuart7, G. van Hees6, D. van Hees2, C. Wijmenga1, P.I.W. de Bakker1,2,4,5,1) Department of Genetics, University Medical Center of Groningen, Groningen, The Netherlands; 2) Blizzard Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, London E1 2AT, United Kingdom; 3) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Department of Medicine, Division of Rheumatology, University of Toronto, Toronto Western Hospital, Toronto, Ontario M5T 2S8, Canada; 5) Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands; 6) Department of Dermatology, University Medical Center Groningen, Groningen, The Netherlands; 7) Division of Human Genetics, Department of Dermatology, University Hospital, Schleswig-Holstein, Christian-Albrechts-University, 24105 Kiel, Germany; 8) Department of Pathophysiology, Centre of Translational Medicine and Centre for Translational Genomics, University of Tartu, 50409 Tartu, Estonia; 9) Estonian Genome Center, University of Tartu, 51010 Tartu, Estonia; 10) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA; 11) Department of Dermatology, University Hospital, Schleswig-Holstein, Christian-Albrechts-University, 24105 Kiel, Germany; 12) Department of Medicine, Division of Rheumatology, University of Toronto, Toronto, Ontario M5T 2S8, Canada; 13) Department of Dermatology, University Hospital, Schleswig-Holstein, Christian-Albrechts-University, 24105 Kiel, Germany; 14) Department of Medicine and Dentistry, University of Oslo, Oslo, Norway; 15) Department of Biociences and Nutrition, Karolinska Institutet, Stockholm, Sweden, and Folkhälsan Institute of Genetics, Helsinki, Finland, and Department of Medical Genetics, University of Helsinki, Finland; 16) St James's University Hospital, Leeds, 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 generated on the ImPACTchip, involving 10,989 cases and 22,806 controls in total. Nineteen out of the 21 known psoriasis loci achieve genomewide significance 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 (rs3456443). The two significant interactions confirmed or extended previous findings in the ENCODE project. In total, this dataset presents a vivid summary of the current knowledge of genetic risk of a complex disease.

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 ontology 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 techniques identified single genes supported by multiple lines of evidence in 35 loci. We have used the dense fine-mapping data from these imputations, 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 previously partnered with databases of gene expression and other functional annotations such as ChIPseq from the ENCODE project. In total, this dataset presents a vivid summary of the current genetic knowledge of risk of a complex disease.
Sequencing-based and multiplatform Genome-Wide Association Study for Multiple Sclerosis and Type 1 Diabetes in Sardinians. I. Zara, E. Porcu, M. Zoledziewska, M. Pizzala, R. Benetti, R. Pili, F. Busonero, R. Atzeni, M. Oppo, F. Reineir, R. Benotti, F. Deidta, C. Sidoro, S. Piras, A. Loi, S. Sanna, E. Cocco, F. Poddle, G. Farina, G. Rosati, L. Llamas, G. Cuccuru, G. Zanetti, A. Angius, M. Giorgio, C.M. Jones, G.R. Abecasis, Serena Sanna, F. Cucca. 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 Multipia, Dipartimento di Scienze Neurologiche e Cardiologiche, Università di Cagliari, Cagliari, Italy; 6) Clínica 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 ∼550 trios, with the Illumina Immunochip (A6). After applying standard quality filters, we analyzed 3,994 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 (MIMIC database, 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, GWAS signals from individual sequencing were not 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 data became genome-wide significant in IC GWAS, and another one with p-value up to 1.2×10−17.

Remarkably, in the original GWAS, no p-value for genotypic association with asthma than with healthy controls (25% relative decrease, p < 1.2×10−4). 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 ethnically-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.

Deep exome sequencing of psoriasis identified new association signals contribute by INDELs, CNVs and rare SNPs. X. Jin, H. Tang, H. Jiang, D. Cao, H. Shao, Q. Li, J. Shen, L. Song, Y. Shi, J. Mei, X. Yang, L. Coin, Y. Li, X. Zhang, J. Wang. 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 a 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.

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 Netherlands; 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) comprises 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 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 Spanish 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 abnormalities and an unusual peak on brain magnetic resonance spectroscopy (MRS). The peak was observed by cerebral MRS imaging that we propose to represent an as-yet, unknown lipid molecule. In one family we were able to confirm the diagnosis of DDHD2-related spastic paraplegia by finding significantly elevated levels of sphingosine in both the liver and the brain. This inhibition of the NPC pathway correlates with increasing sphingosine in tissues. Human and mouse SLOS fibroblast alike demonstrate significant increases of glycosphingolipids (GSLs) and sphingosine as demonstrated by finding significantly elevated levels of sphingosine in both the liver and the cortex of the null SLOS mouse model at E18.5 and the cortex of the hypomorphic T93M null model at multiple ages. These findings also translate clinically, as SLOS patients have elevated sphingosine and GSL levels in their cerebrospinal fluid (CSF). This increase in CSF sphingosine correlates with both residual DHR7 activity and phenotypic severity. Thus, decreased bioavailability of cholesterol due to 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 sphingosine accumulation. 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 findings also translate clinically, as SLOS patients have elevated sphingosine and GSL levels in their cerebrospinal fluid (CSF). This increase in CSF sphingosine correlates with both residual DHR7 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 sphingosine accumulation. 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 findings also translate clinically, as SLOS patients have elevated sphingosine and GSL levels in their cerebrospinal fluid (CSF). This increase in CSF sphingosine correlates with both residual DHR7 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 sphingosine accumulation. 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 findings also translate clinically, as SLOS patients have elevated sphingosine and GSL levels in their cerebrospinal fluid (CSF). This increase in CSF sphingosine correlates with both residual DHR7 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.
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 primed, constant dose of NaH13CO3 and U-13C6 glucose followed by continuous infusion of U-13C6 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 8 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 have reduced clearance of glucose, 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 provide further insights into the pathogenesis of DM in mitochondrial diseases in general.

164 Phenylbutyrate therapy for pyruvate dehydrogenase complex deficiency. R. Ferriero1, E. Lamantia2, P.W. Stacpoole1,6, B. Lee1, M. Zeviani1, N. Brunetti-Pierri1,2, 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, TX; 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 fatty 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 C57Bl/6J wild-type mice and we detected a significant increase in PDHC enzyme activity and a reduction of phosphorylation of E1-alpha subunit in brains, kidneys, and livers of phenylbutyrate 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. Erez1, S. Nagamani1, P. Campau1, Š. Šchelichov1,4, J. Kho1, K. Bissig2, G. Sun1, S. Cameron1, P. W. Stacpoole1,6, 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, which is responsible for converting argininosuccinate to fumarate and ornithine. Among subjects with ASA are supplemented with arginine, some need additional therapy with nitrogen scavengers. In spite of early diagnosis and initiation of therapy, patients with ASA deficiency have a higher incidence of neurocognitive delays compared to those with other urea cycle disorders, for which the cause is unclear. Recently, we reported that ASA 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 arginosuccinic 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 generated argininosuccinate synthase knock-out (Acsb−/−) 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 decrease arginine levels persist in the brain after correction of hepatic ureagenesis defect supporting the tissue specific requirement of ASA. In addition, we showed that NO deficiency contributes at least in part to the cognitive delay observed in ASA, involving NO dependent and independent pathways. In 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-dependent and independent pathways. Neurocognitive phenotype. To systematically analyze the levels of arginine metabolites in the neuronal tissue, we generated argininosuccinate synthase knock-out (Acsb−/−) mice, and generated induced pluripotent stem cells from ASA patients and control subjects. Our results are significant and have significant 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. Christodoulou1, N. Al-Haffid2, X.-Z. Tong3, K. Carpenter1,2, V. Willey1,4, S. Cunningham1, E. K. Aylander1,2, 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 observations. Depletion of arginine metabolites. Our study is significant as it has clinical implications relevant to optimization of the current treatment modalities for ASA.
A new inborn error of manganese metabolism caused by mutations in SLC30A10, a newly identified human manganese transporter. K. Tuschl1, P.T. Clayton1, S.M.Jr. Gospe4, S. Gulab1, S. Ibrahim1, P. Singha5, R.T. Ribeiro5, M.S. Zaki6, M. Luz del Rosario7, S. Dyack8, V. Price9, R.A. Wevers2, P.B. Mills1. 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 hepatoemgaly and hypermanganeseemia 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 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 Δmpr1 using Gateway technology (Invitrogen).

**Results** Homozygosity mapping identified SLC30A10 as the affected gene, and homozygous mutations 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 Δmpr1 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 hypermanganeseemia. 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-Adams1,2, H.A. Bender3, A.M. Akerstedt4, E. Miles-Mason5, T.P. Naidich6, S. Lipson6, T. Bottiglieri5, S.P. Young7, G.A. Diaz1,2, 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 methylocobalamin and adenosylcobalamin, essential co-factors for methionine synthase and methylmalonyl-CoA mutase, respectively. Early-onset cblC disease is associated with disorder white matter development on MRI, and developmental delays and visual deficits are common. Proposed pathophysiological 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 guanidoacacetate (GAA). We developed a comprehensive clinical protocol to prospectively examine various factors related to neurodevelopment 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.
Whole-exome sequencing of 10,000 type 2 diabetes cases and controls from five major ancestry groups. T.M. Teslovich, A.P. Morris, P. Fontanillas, M.A. Rivas, K. Sim, J. Flannick, N. Burtt, H. Chen, A.G. Day, Williams, M.A. Ahmadian, G. Altmann, P. Cingolani, L. Moutsianas, H.M. Highland, T2D-GENES Consortium. 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) Welcome 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 Jamboree 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=1056), EA (N=1154), and SA (N=1500) are available. Common, low-frequency, and rare 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 Devel- opment in Young Finns; 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 < 1x10^-5). 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 = 3x10^-9), low frequency nonsynonymous variants in PCSK9 (MAF < 5%; p = 2x10^-8), and singleton loss of function variants in APOB (p = 8x10^-8). Among five other genes implicated in Mendelian LDL-C related phenotypes, two genes (NPC1L1 and ABCG5) had suggestive evidence for association (p < 4x10^-4). Our results suggest that no single rare variant association test is ideal for detecting all relevant genes, as the genetic architecture underlying associa- tion 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 com- plex traits.
Exome sequencing of extreme phenotypes identifies DCTN4 and CAV2 as modifiers of chronic Pseudomonas aeruginosa infection in cystic fibrosis. M.J. Emond1, T. Louis1, J. Emerson2, S. McNamara2, W. Zhao2, R.A. Mathias3, M.R. Knowles3, F.A. Wright3, M.J. Reider1, H.K. Tabor2, D.A. Nickerson2, K.C. Barnes4, R.L. Gibson9, M.J. Bamshad2, 7, 10, 1) Biostatistics, University of Washington, Seattle, WA; 2) Department of Pedia-
trics, 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=40 each). For each gene, we tested for a significant difference in the frequency of rare genetic variants 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 American controls. In the extreme phenotype study, one gene, DCTN4 (dynamin 4), was found to be significantly associated with phenotype, with p=0.05 after correction for multiple testing (naive p=2.2×10^-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=8.7×10^-6) after CFTR. We then sequenced DCTN4 and performed exome chip genotyping in 667 children (612 severe) and 613 controls 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=5) 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 rs9890 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.

A high resolution study of Type 2 Diabetes genetic architecture through whole-genome sequencing of 2850 European individuals: the Go-T2D Study. J. Flannick1, C. Fuchsberger2, K.J. Gaulton3, N.P. Burtt1, H.M. Kang4, C. Hartl1, R.D. Pearson2, 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.

Exome sequencing of extreme phenotypes identifies DCTN4 and CAV2 as modifiers of chronic Pseudomonas aeruginosa infection in cystic fibrosis. M.J. Emond1, T. Louis1, J. Emerson2, S. McNamara2, W. Zhao2, R.A. Mathias3, M.R. Knowles3, F.A. Wright3, M.J. Reider1, H.K. Tabor2, D.A. Nickerson2, K.C. Barnes4, R.L. Gibson9, M.J. Bamshad2, 7, 10, 1) Biostatistics, University of Washington, Seattle, WA; 2) Department of Pedi-
trics, 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=40 each). For each gene, we tested for a significant difference in the frequency of rare genetic variants 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 American controls. In the extreme phenotype study, one gene, DCTN4 (dynamin 4), was found to be significantly associated with phenotype, with p=0.05 after correction for multiple testing (naive p=2.2×10^-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=8.7×10^-6) after CFTR. We then sequenced DCTN4 and performed exome chip genotyping in 667 children (612 severe) and 613 controls 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=5) 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 rs9890 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.
Mapping quantitative traits with integrated whole exome/genome/array panel in individuals of European descent. X. Sim1, M.A. Rivas2, A.K. Manning3,4,5, A.E. Locke1, C.M. Lindgren2, GoT2D Consortium. 1) Department of Bioinformatics 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 harboring 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 (LDL-P=1.1e-08), with changes in metabolites. The completion of the integrated panel of SNPs, indels, and structural variants will provide a comprehensive suite of tools 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.
phenotypes is necessary to further our understanding of complex disorders. Our combined assessments of allele frequencies in independent case and control populations. Moreover, novel mutations are being modeled for their functional assessment. Further characterisation of ASD candidate risk genes will be achieved by clinical follow-up, such as early screening in families with cancer syndromes. Furthermore, incidental findings from our study have also precipitated NRXN1 with direct diagnostic applications for carrier individuals.

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, ARRD4C 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 characterization 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.

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. Ma 1,2,3, A. F. Wood 1, P. Fontani 1,2, G. Jun 6,7, C. Fuchsberger 9, J. Grunstad 8,9, J. Blangero 7, J. Maller 8, K. Gaulton 8, M. Boehnke 10, R. Grossman 11,2, J. M. Laramie 11, T. M. Teslovich 1,2,3, A. R. Wood 1, A. Prasad 6, S. W. Scherer 1, E. Fombonne 1,2,3. 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) Montreal 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; 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. Within this context 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: 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.

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 advantage 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. Bryc1,2,3,4, A. Williams1,2,3, N. Patterson1,2, M. Sale3,4, W. Chen3,4, J. Divers6, M. Ng7, D.W. Bowden3, J.G. Wilson3, D. Reich5.
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, School of Medicine, 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 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 and 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 a large number of 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 groups. 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 American and sub-Saharan 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 allows for ancestry inference 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.


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 African American genomes. 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 performs 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 gene pools beginning more than 1000 years ago. Using a sample 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. Yang1, J. Novembre2,4, E. Eskin1,2, E. Halperin1,5,6, 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 objective is 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 23andMe 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.
The applicability of the Balding-Nichols model to a dataset of over 100,000 Brazilian individuals. R.V. Rohlf1, A. Bhaskar1, V.R.C. Aguilar2, K. Lohmueller1, A.M. Castro3, A.C.S. Ferreira3, F.S.V. Malta3, Y. Song3, I.D. Louro3, R. Nielsen3. 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 increase 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. Keinan1, E. Gazave1, A. Coventry1, S. Gottipati1, D. Chang1, L. Ma1, D. Muzny2, E. Boerwinkle2, C. Sing2, R. Gibbs2, A. Clark1. 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; Tennesen 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 estimation 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 (Keinan et al. 2012). Here, we disentangled the 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 97 16 European Americans from the ARIC study using EIGENSOFT and selected 500 individuals that constitute a homogeneous cluster. For singlets to be reliably called, we sequenced the neutral regions in these individuals to a very high median coverage of 20X 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 singlets. 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. Sheehan1, K. Harris2, Y.S. Song2, 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 100,000 Brazilian individuals to obtain a detailed size change history since some recent times.


**185**


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, Steinruecken, & 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 sub-population, 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.
On the Sardinian ancestry of the Tyrolean Iceman. M. Sikora¹, M. Carpenter¹, A. Moreno-Estrada¹, B.M. Henn¹, P.A. Underhill¹, I. Zara², M. Pitzalis³, C. Sidoro³,⁴,⁵, F. Reinier², M. Marcelli², A. Angius³,⁴, C. Jones⁴, T.T. Harkins⁶, A. Keller⁷,⁸, A. Zink⁹, G. Abecasis⁴, S. Sanna³, F. Cucca³, C.D. Bustamante¹. 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.
Gene-based epistasis analysis in genome-wide association studies. L. Ma¹, A. Brautbar², E. Boerwinkle², C.F. Sing³, A.G. Clark¹-³, A. Keinan¹. 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 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 $P = 0.002$). Using a locus-based validation procedure, we successfully validated this gene-gene interaction in the EA cohorts from the Framingham Heart Study ($P = 0.002$) and the Multi-Ethnic Study of Atherosclerosis (MESA; $P = 0.006$). The interaction between these two loci is also significant in the African American sample from ARIC ($P = 0.004$) and in the Hispanic American sample from MESA ($P = 0.04$). 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 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 in multi-locus 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.

Building Human Phenotype Networks from Shared Genetic Risk Variants. C. Doros¹, K. Desai², M. Giacobini³, M. Lupien², J.H. Moore³. 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 increasingly important in the study of complex genetic disorders. Ultimately, understanding the interaction 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 rsSNP with all SNPs found in linkage disequilibrium (ldSNPs) using the HapMap project data. This impuritiesome (iRAV) allows us to establish connections between disease traits that share haplotype blocks, i.e. that share their rsSNPs or ldSNPs. The weight of the interaction is computed as the normalized number of SNP 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 erythematosus. 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.
Large-scale multi-phenotype meta-analysis evaluates pleiotropic effects at FADS1 and GIPR loci. V. Lagou\(^1\), R. Magi\(^2\), K. Fischer\(^3\), M. Akern\(^4\), I. Surakka\(^5\), H. Kaakinen\(^4\), J.S. Ried\(^6\), A. Mahajan\(^1\), M. Horm\(^7\), L. Marullo\(^1\), K. Strauch\(^8\), C. Gieger\(^9\), S. Ripatti\(^6,7\), A.P. Morris\(^1,2\), V. Lysyenko\(^1,12\), I. Prokopenko\(^2\) 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) Welcome 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 test common and independent effects for 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. We performed in up to 34,048 European individuals from six studies within the ENGAGE consortia. The smoothed functional principal component analysis (SFCA) 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 SFCA-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=55985, P\(_{\text{null}}=5.4\times10^{-17}\)) and for weight and stroke (BIC=4447, BIC\(_{\text{null}}=4469, P_{\text{null}}=9.35\times10^{-5}\)). For GIPR, BMI (BIC=79110, BIC\(_{\text{null}}=79181, P_{\text{null}}=1.3\times10^{-6}\)) and a combination of weight, FG, 2hGlu, total cholesterol (BIC=24421, BIC\(_{\text{null}}=24452, P_{\text{null}}=6.8\times10^{-14}\)) were prioritised. Results from this study are being tested in up to 200,000 individuals available within ENGAGE. The proposed design and methodology for assessing pleiotropic effects at genetic loci proves feasible and can be applied in large-scale studies.

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 $p<0.0001$) 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 ($p<5.4\times 10^{-5}$) and the eSNP tends to have a cis effect, modulating the expression of the source gene ($p<2.3\times 10^{-13}$). 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.


von Clauswitz 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. Conquered by many different functional genomic 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 practically simulated.


1) Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; 2) Department of Epimio-lology, 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, New Brunswick, 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., Emerville, 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 Discovery (VAAST) system uses a rare variant association test that combines amino acid substitution and allele frequency information using a composite likelihood ratio test (CLR). Here, we report two new algorithms that expand pVAAST to incorporate phylogenetic sequence conservation and pedigree information, improving across previous rare, Mendelian and complex genetic diseases. We tested our phylogenetic conservation algorithm, implemented in the package pVAAST 2.0, on variants from the Human Gene Mutation Database and rare variants from the BRCA1/2 genes. VAAST 2.0 prioritizes 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
drosophila. These findings fit with the genital abnormalities found in PCH mRNAsplicingisinlinewithTSEN,RARS2andEXOSC3mutationscausing
encodes for a SR kinase involved in mRNA splicing. The role of CLK2 in
LNA in situ hybridizations on human embryonic brain tissue showed CLK2
To verify this candidate gene, we performed knockdown experiments in
identified homozygous in both patients and heterozygous in the parents.
A missense mutation (p.A390S) in cdc2-like kinase 2 (CLK2) was
PCH type 1,2,4,5 and 6. Up to now, no locus for PCH7 has identified. We
mitochondrial arginyl-tRNA synthetase (RARS2), exosome component 3
abnormalities in males in addition to pontocerebellar hypoplasia. Mutations
acteristics of all subtypes include hypoplasia of pons and cerebellum, and
recessive neurodegenerative disorders with prenatal onset. Common char-
Academic Medical Centre, Amsterdam, The Netherlands.

CLK2 missense mutation in a family with pontocerebellar hypoplasia
type 7. V.R.C. Eggens, Y. Namavar, M.A. Haagmans, K. Fluitert, E.J.
Bradley, P.G. Barth, B.T. Poll-The, F. Baas. 1) Genome Analysis, Aca-
demic Medical Centre, Amsterdam, Amsterdam, Netherlands; 2) Depart-
ment of Clinical Genetics, Academic Medical Centre, Amsterdam, The Neth-
erlands; 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 char-
acteristics 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 RNA splicing endonuclease (TSEN) complex,
mitochondrial arginyl-tRNA synthetase (RARS2), exosome component 3
(EXOSC3) and vacinia 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.

Mutations in ITPR1 cause autosomal dominant congenital nonprogressive
spinocerebellar ataxia. J. Warman Chardon, L. Huang, M. Carter, K. Friend, T.
Dudding, J. Schwartzentruber, R. Zou, P. Schofield, S. Douglas, B. Bulman,
K. Boycott. 1) Genetics, Children’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 Inno-
vation Centre, Montréal, QC, Canada; 8) Ottawa Hospital Research Institute,
University of Ottawa, Ottawa, ON, Canada; 9) Centre for Translational Neu-
roscience 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. Link-
age 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 domi-
nant congenital nonprogressive spinocerebellar ataxia and identified mis-
sense mutations in the ITPR1 gene in both families. ITPR1 encodes inositol
1,4,5-trisphosphate receptor, type 1, a ligand-gated ion channel that

Copyright © 2012 The American Society of Human Genetics. All rights reserved.
199

Vps37A causes a novel form of complex Hereditary Spastic Paraparesis. T. Falik-Zaccay1,2,5, Y. Zivony-Elboum1,2, W. Westbroek2, D. Savitzki2, Y. Shoval1, Y. Anikster6, A. Waters7, R. Kletzien8,8. 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 quadriaparesis, 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 segregating with the disease in both families. Mobility was significantly impaired during writing, a much more common task. Interestingly, WD often occurs when a (professional) musician is playing his instrument, WD is more common in professional musicians. A possible correlation may be that the repetitive movements and abnormal postures of musicians may impart novel pathways for dystonia. Musician’s dystonia (MD) 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\(\times\)10\(^{-8}\)) was found for two intronic variants in genes on chromosome 17 (SNP1: P=1.45\(\times\)10\(^{-8}\); OR=3.12; 95% CI=2.16-4.51) and chromosome 9 (SNP2: P=2.39\(\times\)10\(^{-8}\); OR=1.66; CI=1.39–1.99), respectively. Both genes are neurorally 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.
Autosomal Recessive Axonal Neuropathy with Neuro-Myotonia: A novel disease entity caused by mutations in HINT1. J. Baets1,2,3, M. Zimovcic1,3,4, J. Nikolic5,6,7, Y. Parmentier8,9, V. Guergueltcheva10, I. Tourne2,3, M. Auer-Gruenbach10, T. Muller11, P. Van Damme12, W.N. Loscher13, N. Barisic14, Z. Mitrovic15, S.C. Previti16, H. Topaloglu17, G. Bemert18, A. Beleza-Meireles19, S. Todorovic20, B. Istedekova21, K. Peeters22,4, A.P. Hahn20, S. Züchner21, V. Timmerman6,5, V. Milic Rasic6, A.R. Janecke11,12,22, A. Jordanova2,5,23, P. De Jonghe1,2,3. 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, University Hospital Antwerp, 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 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, Children’s Hospital, University of Graz, Graz, Austria; 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 Antwerp, Belgium; 22) Department of Human Genetics, Innsbruck Medical University, Innsbruck, Austria; 23) Department of Medical Chemistry and Biochemistry, Molecular Medicine Center, Medical University-Sofia, Sofia, Bulgaria.

Inherited neuromuscular disorders are a group of 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 neuro-myotonia was previously recognized. Neuro-myotonia 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 electrophysiologic signs of neuro-myotonia or a combination of both. In a screening cohort of 262 unrelated index patients harboring HINT1 mutations, we found HINT1 mutations in 11%. In a more specific subset of patients presenting with AR neuropathy and clinical or electrophysiologic signs of neuro-myotonia, the HINT1 mutation frequency was 76%. Autosomal recessive axonal neuropathy with neuro-myotonia (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.

De novo gain of function KCNT1 channel mutations cause seizures and developmental delay in malignant migrating partial seizures of infancy. G. Barcia1,2, M.R. Fleming3, A. Deligniere4, V. Gazula1, M.R. Brown1, M. Langouet1, H. Chen1, J. Kronengold1, A. Abhyankar2, R. Cilio7, P. Nitschke8, A. Kaminska9, N. Boddart10, J.L. Casanova11, I. Desguerre12, A. Munnich1, O. Dulac13, L.K. Kazmerek14, L. Colleaux15, R. Nabbout16. 1) Department of Paediatric Neurology, Centre de Reference Epilepsies Rares, Department of Paediatric Radiology, Clinical Electrophysiology Unit, Hôpital Necker Enfants Malades, APHP, Paris, France; 2) Inserm U683, Université Paris Descartes, PRES Sorbonne Paris Cité, Hôpital Necker Enfants Malades, Paris, France; 3) Department 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 Gesù 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 polymorphic 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 missense variants (p.Arg428Gln and p.Ala934Thr) in 2 unrelated probands. Both mutations result in direct silencing 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. Electroencephalography revealed a characteristic paroxysmal EEG in 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 linked 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. Ruzzo1,2,3, E.L. Heinzen1,3, R. Wedel2, K.V. Shianna1, I.E. Scheffer4,5, S.F. Berkovic4, R. Ottman2,3, D.B. Goldstein1,3. 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 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 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 HgPlex 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 individual families. Genes of high interest include: seizure threshold 2 (SZT2); CCR4-NOT transcription complex, subunit 3 (CNOT3); CNOT1; 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. Erogullari1, P. Seibler2, D. Braunholz2, A. Grünewald3, R. Depping3, J. Eckhold1, A. Rakovic2, T. Lohmann1, K. Lohmann1, F.J. Kaiser4. 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 autoregulatory manner. As a preliminary step we characterized the THAP1 promoter region by using in-silico prediction and luciferase reporter gene assays. Interestingly, five THABs are localized in the THAP1 core promoter. The specific binding of THAP1 to its core promoter region was demonstrated by chromatin immuno-precipitation (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 expression changes caused by mutation-induced 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.
Large-scale genotyping identifies more than 40 novel breast cancer susceptibility loci. K. Michailidou1, P. Hail2, A. Gonzalez-Neira3, M. Ghous- saini4, J. Dennis1, R.L. Milne3, M.K. Schmidt4, J. Chang-Claude5, S.E. Bojesen6, M.K. Humphreys7, Q. Wang1, M. Garcia-Closas10, P.D.P. Pharoah1, G. Chenevix-Trench, A.M. Dunning2, J. Benitez12, D.F. Easton14, 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 Krebsforschungs- zentrum, 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 Epido- miology 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.

Prostate cancer (PrCa) is the most frequently diagnosed male cancer in developed countries. To identify common PrCa susceptibility alleles, we genotyped 2,511,151 SNPs on a custom-designed iCOGS array in 63,031 PrCa cases and 45,980 controls; 2,574 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, Canada or Mexico) and 3,232 SNPs selected by 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 OR > 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<0.01) 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), epithelial 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 1,500 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 4.7-fold increased risk compared with the average of the population. These results will facilitate population risk stratification for clinical studies.
207 Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. A.C. Antoniou1, X. Wang2, L. McGuuffog1, A. Lee1, M.M. Guedet1, K.B. Kuchenbaeck1, P. Soucy3, J. Simard4, K. Offit5, D.F. Easton6, G. Chenevix-Trench6, F.J. Couch7, 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 Quebec 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=1.4×10−8, per-allele HR=1.27, 95%CI:1.17–1.38) and 4q32.3 (P=3.4×10−8, 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=2×10−3), with several SNPs providing P=2×10−11 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=2.7×10−8, 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. Kocherus,5 M. Gaudet2, J. Vijai2, R. Klein3, T. Kirchhoff3, L. McGuffog1, D. Barrowdale1, A. Dunning4, A. Lee1, P. Hall1, F. Couch7, J. Simard4, K. Offit5, D. Altshuler10, D. Easton1, G. Chenenvix-Trench14, A. Antoniou1, K. Offit5, 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 Quebec 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 among BRCA2 mutation carriers. To date, all identified common genetic modifers 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. 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.
Three independent loci within the TERT-CLPTM1L locus associated with telomere length and risk of breast and ovarian cancer. G. Chenevix-Trench, S.E. Bojesen, K.A. Pooley, S. Johnatty, J. Beesley, K. Michailidou, J. Tyer, S.L. Edwards, H.C. Shen, K. Lawrenson, H. Pickett, M. Stutz, C. Smart, J. French, P.L. Mai, M.H. Greene, S. Gayther, R. Reddel, P.D.P. Pharoah, E.L. Goode, A. Berchuk, D.F. Easton, A.C. Antoniou, A.M. Dunning on behalf of CIMBA, BCAC 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 Genetics, Queensland Institute of Medical Research, Brisbane, Australia; 7) These authors contributed equally to the work.

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 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) expression. The three SNPs are the largest of any GWAS-discovered breast cancer variants in Europeans, providing further support that these SNPs may have directly mediated through shorter telomeres. For Estrogen receptor (ER)-negative but not ER-positive breast cancer, the strongest signal maps to a transcriptional enhancer element. SNP rs614367 on chromosome 17q21 increases the risk of breast cancer among BRCA1 carriers (p-value=1.1×10-5). Luciferase assays showed that a construct carrying the risk alleles of three highly associated SNPs in the TERT promoter, contained SNPs associated with TL (p=5.8×10-7), overall breast cancer cases and controls, and 11,705 BRCA1 mutation carriers. TL was determined in 10,959 normal individuals and 2,728 controls, and 11,705 BRCA1 mutation carriers. 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) and 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 enhancers and that one of the risk-associated SNPs in Peak 3 decreases relative luciferase signal by ~30%. Analysis of transcripts produced from a minigenic 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 mammaplasty 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. Barrett1, J.C. Taylor1, M. Brossard2, A.M. Goldstein3, P.A. Kanetsky4, E.M. Gillanders5, J.A. Newton Bishop1, D.T. Bishop1, F. Demenais1, M.M. Iles1, 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 Epide-
miology & Biostatistics, University of Pennsylvania, Philadelphia, Pennsyl-
vania, 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. Geno-
types 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 variant in MC1R, despite the fact that the original signal is closer to other genes. This result provided 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 (r²=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 experi-
ments.

212

Combining expression phenotypes with high density imputation to identify melanoma risk genes. M. H. Law1, G. W. Montgomery2, K. M. Brown3, A. E. Cust4, N. G. Martin3, G. J. Mann3, N. K. Hayward3, S. MacGregor3, 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, Australia; 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-impuation 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-analy-
ysis. 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 high density imputation association test in 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.
Meta-analysis identifies four new loci for testicular germ cell tumor.

C.C. Chung1, Z. Wang1,2, P.A. Kanetsky3, C. Turnbull5, K. McGlynn1, R.L. Erickson12, M.H. Greene1, M.A.T. Hildebrandt6, R.I. Skofheim7,8, C. Kratz1, M.B. Cook1, F. Schumacher9, R. Koster10, M. Yeager1,2, K.B. Jacobs1,2, S.M. Schwartz11,12, D.T. Bishop13, H.K. Gjessing14,15, V. Cortessis9, N. Rahman4, X. Wu6, S.J. Chanock1, K.L. Nathanson4, 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; 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’s 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 × 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 × 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 × 10^{-3}, 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 × 10^{-3}, per allele OR 0.79, 95%CI 0.75–0.85 and rs7221274; P = 4.04 × 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. Roscioli1,2, E-J. Kamsteeg1, K. Buyse1, I. Maystad3, M. van Reeuwijk1, C. van den Elzen1, E. van Beusekom3, M. Riemersma1,4, R. Pfundt1, L.E.L.M. Vissers1, M. Schraders1,5, M.F. Buckley1,5, H.G. Brunner6, H. Zhou7, J.A. Veltman1, C. Gilsen1, G.M.S. Mancini8, M.A. Willemse4, D. Petkovic Ramadzˇa9, D. Chitayat10, C. Bennett10, E. Shendrot10, E.A.J. Peeters11, G.M.B. Tan-Sindhunata7, H. Kayseril13, O. Abd El-Fattah El-Hashash14,15, D.L. Stemple16, D.J. Lefebvre17, Y-Y. Lin18, H. van Bokho-ven1. 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 Oto-rhinolaryngology, 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 of Toronto, 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 Faculty of Medicine, 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 (WWWS). Isoprenoid synthase domain containing protein 2 (ISPD), codes for an enzyme that provides the isoprenoid moieties for β-NAG biosynthesis in an α-dystroglycan synthase reaction. ISPD is a nuclear protein that is localized in the endoplasmic reticulum and provides a N-terminal transferase in this pathway. ISPD also has been implicated in RAS signaling and has a homologous human orthologous protein in chordates. However, the role of ISPD in this pathway is not clear. We identified a homozygous ISPD mutation in a region of shared homology in three affected members of a WWWS family. This e.647C>A transversion results in a novel protein product that is predicted to disrupt the β-NAG biosynthesis pathway. ISPD is expressed in multiple human tissues and has a significant role in RAS signaling. 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. Yntema1, W. Nillesen1, J. Paardekooper Overman1, M. Bonetti2, J. de Ligt3, H. Venselaar4, M. Tartaglia5, S.G.M. Frants6, L.E.L.M. Vissers1, J. den Hertog2,6, I. van der Burgt1. 1) Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 2) Hubrecht Institute-KNAW and University Medical Center, Utrecht, The Netherlands; 3) Nijmegen Centre for Molecular Life Sciences (CMBL), 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. Cardiac 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 homozygous mutation in a novel gene that encodes 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 observed 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


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 Centre, Cambridge, UK; 4) Department of Paediatrics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Department of Nephrology, Institute of Mother and Child Health Care, Belgrade, Serbia; 6) Department of Genetics, 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, DYN2CH1, 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, while diabetes in JATD has been shown to exist, the underlying genetic heterogeneity of the disease and the fact that DYN2CH1 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. IFT140 mutations were excluded prior to WES. Results: WES showed biallelic mutations in DYN2CH1 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 using subsequent Sanger sequencing of IFT140 in 46 JATD patients with severe renal involvement revealed 2 more cases of compound heterozygous mutations. Interestingly, no significant extraskelatal involvement was reported for our DYN2CH1 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: DYN2CH1 mutations account for approximately a third of all JATD cases and mutations in other known JATD genes are rare. While extraskelatal disease was rare in our DYN2CH1 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


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 (NCLM), 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, osteochondrodysplasia 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 ABC9 in a total of 14 out of 16 Cantú syndrome cases. ABC9 is part of an ATP dependent potassium (K\textsubscript{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 ABC9. Using electrophysiological measurements we show that mutations in ABC9 reduce the ATP-mediated inhibition, which results in opening of the channel. Moreover, similarities between the K\textsubscript{ATP} channel agonist minoxidil indicate that the mutations in ABC9 result in channel opening. Thanks to the availability of ABC9 antagonists our findings may have direct implications for the treatment of Cantú patients.

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 phospholipase D1 from a child with bicuspid aortic valve and right coroand a poor clinical outcome. Heterozygous mutations suggest 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 with the general population. Heterozygous mutations affecting 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 was associated with residual biological activity cause a phenotype similar to ERFR1/2, karyotyping the RAB23-TPK1, V.P.

219 Mutations in the multidomain protein MEGF8 identify a new subtype of Carpenter syndrome associated with defective lateralization. D.L. Lloyds8, S.R. Twigg, N. Eliotoglou7, D. Jenkins5, C.D.O. Cooper4, N. Akarsu2, E. Taskir4, N. Al-Sanna4, A. Narragon4, G. Gilleiren-Kaesbach1, I. Stefanova8, S.J.L. Knight4, J.A. Goodship5, B. Keavney5, P.L. Beales2, O. Gileadi7, S. Mcgowan1, A.O.M. Witte10,11 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, Dhaarun Health Center, Saudi Aramco Medical Services Organization, Saudi Arabia; 7) Division of Neonatology, Selup University, Meram Faculty of Medicine, Konya, Turkey; 8) Institute for Human-genetik 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 4 unrelated children with multisuture craniosynostosis. Candidate genes with pathogenic mutations include MEGF8, dextrocardia, clinically diagnosed with Carpenter syndrome but in which 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 identified 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 human syndrome. Here, we describe a patient with multisuture craniosynostosis, polysyndactyly, and heart defects resulting in death by E10.5. To explore the function of ERF during development, we engineered mouse harboring a conditional allele (ERFloxP/loxP). Both heterozygous (ERFloxP/+) and homozygous (ERFloxP/loxP) conditional mice were phenotypically normal, but compound conditional/null heterozygotes (ERFloxP/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, 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 membranous bones, but the osteogenic regulator Runx2 was reduced. Moreover, ChIPseq analysis showed that ERF binds preferentially to enhancers containing Runx2 motifs. Based on these data we propose that ERF normally antagonises positively acting Ets factors acting in concert with Runx2 to activate osteogenesis during membranous ossification.
220 Increased frequency of FBN1 variants in adolescent idiopathic scoliosis. J.G. Buchan1, D.A. Alvarado4, M.C. Willing2, M.B. Dobbs4, C.A. Gurnett3,4. 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. 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. 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 systemeric 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 patients. Targeted exome and whole exome 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. V. Sharma1, J.A. Harrington2, X. Gao3, D. Zhang4, C. Wise1,2. 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 high penetrance 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 AIS, or represent benign polymorphisms.

222 Recessive mutations in FKBP10, a PPlase 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. Barnes1, M. Weiss2, W.A. Cabral1, E. Makeareeva3, E.L. Mertz4, W. Paton4, G. Duncan5, E. Truillo6, S. Leikin7, D.R. Eyre8, S.J. Bale9, J.C. Marini1. 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 isozyme 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_1272del-CCinsA). 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 3’PPlase domain of FKBP65 which leads to normal to increased levels of FKBP10 transcripts and residual protein. Both pedigrees show minimal changes in the amount of matrix, yet the former 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.
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 ASW, 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.


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 [-hydroxy methylglutaryl] coenzyme A reductase (HMGG-CoA), a target site for statins, using insilico tools. Structural changes in mRNA due to synonymous and nonsynonymous SNPs were evaluated by SNPinfo. The 3D structures of normal and mutant proteins of P-gp and HMGG-CoA reductase were modeled by Molecular Operating Environment (MOE). A new lead molecule was designed from native structures 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 HMGG-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.
PhenoDB: a new web-based tool for the collection, storage and analysis of phenotypic features. A. Hamosh1; J. Hoover-Fong2; V.R. Sutton1, N. Sobrero1; C. Boehm1; P. Schiettell1; D. Valsesia1; 1) European Institute of Human Genetics, 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, phenotypic analysis has mostly 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 about the genetic architecture of traits has become 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 application that allows healthcare providers 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 mapable 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 user-templates can be created and accessed from user-specific and organization-specific 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 customize genes to report; and genetic counselors can view families and stage the 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 web framework that allows rapid development of secure, web-based applications 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).

N. Sobrero1; C. Boehm1; P. Schiettell1; D. Valsesia1; 1) European Institute of Human Genetics, Johns Hopkins Univ Sch Med, Baltimore, MD; 2) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 3) FS Consulting, Salem, MA.

We are interested in the association of genetic polymorphisms with metabolite levels. These intermediate phenotypes generally demonstrate larger effect sizes compared to phenotypes of interest, they can be measured repeatedly from the same individual, and they have the potential to provide mechanistic insights into the pathways that are associated with the trait of interest. In this study, we have measured the serum concentrations of 217 metabolites using the Human Metabolome Database (HMDB) in a cohort of 781 individuals from the Hormones, Diet, and Health study. The cohort includes individuals of European ancestry (n=506) and African ancestry (n=275). The study design was a cross-sectional design with the primary outcome being the association of single nucleotide polymorphisms (SNPs) with the metabolites. The study was powered to detect a 10% difference in metabolite concentration between individuals with different genotypes. The study was conducted at the University of California, San Francisco and the University of California, Davis. The study was approved by the institutional review board at the University of California, San Francisco. The study is registered on ClinicalTrials.gov (NCT03341106). The study was supported by the National Institutes of Health (R01 DK111947, K08 DK111776, T32 DK007411, and T32 DK007001). Future work will include the analysis of the metabolites in the context of other phenotypes and conditions, as well as the exploration of potential biological mechanisms underlying the associations observed in this study.
An Informatics Approach to Analyzing the Incidentalome. M.C. Adams¹, J.S. Berg¹, N. Nassar², C. Bizon², K. Lee¹, C.P. Schmitt², K.C. Wilhelmsen¹,², J.P. Evans¹. ¹Genetics, UNC-Chapel Hill, Chapel Hill, NC; ²Renaissance 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 to 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 Faukuyama muscular dystrophy. M. Taniguchi1,2, K. Kobayashi2, M. Kanagawa2, CC. Yu1, T. Oda1, A. Kuga1, H. Kurahashi1, H.O. Akem1, S. DiMauro1, T. Yokota1, S. Takeda1, T. Toda1. 1) Department 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, Fujiha University, Aichi, Japan; 4) Department of Neurology, Columbia University Medical Center, New York, USA; 5) Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Belgrade, Serbia. Design: We performed a comprehensive genetic study involving all patients diagnosed with FMD in Japan and kanagawa prefecture. Results: SVA-mediated diseases, including severe congenital muscular dystrophy (SVCMD), Becker muscular dystrophy (BMD), and fukutin-related disorders, have been identified as SVA-mediated diseases. We identified that SVA insertion disrupts the fukutin gene. Conclusions: Our present study suggests that the SVA insertion can disrupt the fukutin gene, which is a key regulator of muscle development and function.

234 Treating Pelizaeus-Merzbacher disease with clinically applicable compounds, curumin and chloroquine: preclinical studies. K. Inoue1, T. Matsukawa1, T. Morimura1, Y. Numata1, L.-H. Yu1, L. Gotoh1, R. Yamamoto1, N. Inoue1, B. Antalffy1, K. Deguchi1, H. Osaka2, Y. Goto3. 1) Dept. Mental Retardation and Birth 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 nervous system. Because PMD is an orphan disease of children, medicines with low entry barrier are required 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 curumin and chloroquine can independently mitigate the cellular phenotype of PMD in vitro and in vivo. We treated mice carrying A242V PLP1 mutation (Msdl mice) or HeLa cells transiently expressing PLP1 with either curumin or chloroquine. Curumin, a food compound from turmeric, was given to Msdl mice orally at 180 mg/kg day−1 from the postnatal day 3. Msdl 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 marker expression was observed. Meanwhile, chloroquine, a classic anti-malarial 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 mice treated with chloroquine injected intraperitoneally (0.5 mg/ml PBS, 10 μg/ml 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 Msdl mice. In conclusion, curumin and chloroquine, a food compound and an FDA-approved medication for the treatment of PMD, may serve as safe and clinically applicable medications for the treatment of PMD.

235 Systemic L-three-hydroxyphenylserine corrects neurochemical abnormalities in a mouse model of Menkes disease. S. Kolar1, A. Don-sante1, P. Sullivan2, D. Goldstein3, C. Holmes2. 1) Intramural Res Prog., NICHD/NINH, 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 adolescence due to increased expression of the dopamine-degrading enzyme, tyrosine hydroxylase (TH), which catalyzes the rate-limiting step in the synthesis of dopamine (DA) and norepinephrine (NE). DA and NE are critical for normal motor function, mood and autonomic control. In a mouse model of Menkes disease, mottled-brindled (mo-br), we previously showed that copper deficiency results in a decrease of DA in the prefrontal cortex and 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 mouse, although serum norepinephrine levels still increased significantly (P<0.001). To evaluate 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/6J-Atp7a mo-br breeding pairs were obtained from Jackson Laboratories (Bar Harbor, ME). For experiments, litters were culled to 4–5 pups. Shorty 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). In conclusion, curumin and chloroquine, a food compound and an FDA-approved medication for the treatment of PMD, may serve as safe and clinically applicable medications for the treatment of PMD.
Response to VPA therapy in SMA patients is concordant from blood to neurons and influenced by CD36. B. Wirth1, L. Heessen1,2, J. Höcker1, T. Bauer1, J. Schrem1, K. Zimmermann1, M. Thoenes1, M. Walter1, J. Dimoś2, M. Peitz2, O. Brüste3, R. Heller1, L. Garbes1. 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 lethality 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 anti-epileptic 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% of 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 promoter 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. Treatment of VPA non-responders and SMN1 knockout mice with CD36, response to VPA was abolished. Our data provide evidence on the ABIR test. Further studies are currently underway to analyze oxysterol and that neurotoxic oxysterols accumulate in the brain and retina 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 ABIR 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.

Beyond cholesterol: Antioxidant treatment for patients with Smith-Lemli-Optiz syndrome. E. Elias1, R. Braverman2, S. Tong1. 1) Dept Genetics & Pediatrics, Children's Hosp CO, Aurora, CO; 2) Dept of Ophthalmology, Children's Hosp CO; 3) Statistician, CREC, 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 retina 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.
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. Paez-Santiago-Sprouse1,2, J.P. Habashi1, N. Huso2, D. Bedja3, G. Rykiet1, J.J. Doyle1, H.C. Dietz1,2,3, E. Stapleton1,3, T. Sadeghin1, F. Mitchell1, T. Dixon3, M. Kingery2, A. Gropman1,2. 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=.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.

A mechanism and treatment strategy for pregnancy-associated aortic dissection in Marfan syndrome. J.P. Habashi1, N. Huso2, D. Bedja3, G. Rykiet1, J.J. Doyle1, H.C. Dietz1,2,3, E. Stapleton1,3, T. Sadeghin1, F. Mitchell1, T. Dixon3, M. Kingery2, A. Gropman1,2. 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(3.3±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. Ramos1, A. Doumatay1, H. Huang1, D. Shiner1, G. Chen1, S. Callier2, J. Zhou3, A. Adeyemo1, H. Moleod3, C. Rotimi1. 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 Genomic Data Science 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 and other treatments (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 for pharmacogenomics allowing for any relevant variations, 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 publicly 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, rs9292321, associated with altered drug metabolism (an anti-platelet drug) was found in 94% of Han Chinese populations whereas the variant was only observed in 0.2% of Yorubans from Nigeria. However, variation was not limited to just global comparisons. We also observed markers that varied in frequency within a given country illustrated by rs7673255, an SNP related to detoxification metabo(lism) in the Han samples (0.88 and 0.54 for Luhya and Masai, 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 counterpart 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.


Profiling genetic and genomic alterations with drug sensitivity is a way to develop a tailored approach to treating patients with cancer. To systematically identify drug resistant and 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. 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 compared 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 (F=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 transcriptional, and protein expressions, 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 (85% 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. Coenen1, C.J. van Marrewijk2, L.J.J. Derijks3, S.H. Vermeulen1, O.H. Klungel4, A.L.M. Verbeek5, H. Scheffer6, B. Franke1, H.J. Guchelaar7, D.J. de Jong8, TOPIC study. 1) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 2) Department of Clinical Pharmacy, Maxima Medical Centre, Veldhoven, The Netherlands; 3) Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 4) Department of Pharmacopoeidemiology 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. Thiopeurine 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 Pharmacogenometric testing in IBD (TOPIC) study. Patients were randomly assigned to undergo pre-treatment genotyping, 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 60% of the standard dose. To assess the effect of pre-treatment TPMT 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\textsuperscript{10}\textsuperscript{9} l\textsuperscript{-1}) in the first 5 months after treatment initiation. Of the 659 patients included in the TOPIC study, 422 (64%) were treated with thiopurine, 46 (10.6%) patients were genotyped for the tested variants at least once in the study period. TPMT genotyping was performed using gene-specific multiplex PCR allele specific microsatellite (MS) primers. We found that patients that were genotyped before treatment had a lower prevalence of leucopenia compared to patients that received standard treatment (2.4% (n=1) versus 21.4% (n=43), 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.
PGRNseq: a new sequencing-based platform for high-throughput pharmacogenomic implementation and discovery. A.S. Gordon1, J.D. Smith1, Q. Xiang2, M.L. Metzker2, R.A. Gibbs2, E.R. Mardis3, D.A. Nickerson1, R.S. Fulton2, S.E. Scherer2. 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 such as CYP2D6.* Using proprietary 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.

Genetic Variation in the GRK4 gene associates with susceptibility to hypertension and response to angiotensin receptor blocker blockade. M. Vohra1, Z. Wang3, H. Sanada2, M. Williams4, J. Bartlett5, L. Gordon1, S. Chen5, L. Asico5, E. Escano5, V. Villar5, C. Zeng5, L. Wong5, J. Jones5, R. Felder6, G. Eisner7, P. Jose2. 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) Department 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 patients with HT, we performed a genome-wide association study testing if any of the 534 genes were significantly associated with HT or ARB response. Single locus Chi-square analysis revealed that all three variants in the GRK4 gene were significantly associated with HT in both genotypic (6.56×10^-7<pvalues<0.02, OR=0.729) and allelic (0.34×10^-8<pvalues<1.34×10^-11) tests. A genetic variant located in the GRK4 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 of three GRK4 variants (2.69×10^-6<pvalues<0.0006, 2.33×OR<1.42) as well as the DRD1 variant (pvalue=0.039, 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.

Copyright © 2012 The American Society of Human Genetics. All rights reserved.
246 Genome-wide discovery of drug-dependent human liver enhancers. R.P. Smith1,2, K.M. Morissey1, X. Sun1, T.J. Hoffman1,3, K.M. Giacomini1, N. Ahituv1,2. 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 with 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 Driest1, T.L. McGregor1,2, Z. Lu1, S. Vear2, C.B. Creech3, P.J. Kannankeril1, K.B. Brothers1,2, A. Potts4, E. Bowton5, J.T. Delaney6, Y. Bradford2, S. Wilson2, L. Olson2, D.C. Crawford2, B. Saville3, D.M. Roden2,8, J.C. Denny7,9. 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 0 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 nephrotic 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=9.5x10^-9) and a 14% decrease in the geometric mean of the Ke (beta =-0.16, P=4.9x10^-9). The nearest gene, MIA4, is 3382bp downstream. HHIPL, TAF1A, AIDA, and BROX are within 100kb. At 5q13, rs10085144 was associated with a 20% increase in trough (+15%, beta = 0.14, P=3.1x10^-8) and a 14% decrease in the geometric mean of the vancomycin trough (beta =-0.16, P=4.9x10^-9). At 6q23, rs11347923 was associated with a 18% increase in trough (+12%, beta = 0.15, P=3.1x10^-8) and a 14% decrease in the geometric mean of the Ke (beta =-0.16, P=4.9x10^-9). The nearest gene, EDIL3, is 560kb upstream. For both candidate SNPs, there are no SNPs in known genes in linkage disequilibrium (r2>0.8) within 500kb based on 1,000 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.

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 (log2 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=4.5×10^-9, r²=0.26). We found that this QTL is also associated with transcriptional response at 161 genes (FDR<0.01). 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=1.2×10^-9). These results suggest that 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 pharmacological and key physiological regulator. Additional levels of phenotypic information could also be incorporated, ranging from molecular processes that modulate general effects of the treatment to those that are specific to participants' personal outcomes that depend on in vitro sensitivity, eventually uncovering all the links in the chain of mechanisms connecting genotype to phenotype.
250 Germline Mosaicism Does Not Explain the Maternal Age Effect on Trisomy. R. Rowsey1, B. Murdock1, P. Hunt1, C. Dickerson1, T. Woodruff1, T. Hassold1. 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, Hultén 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 errors occurring in the female ovary, specifically to pre-miotic 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.


The fact that human aneuploidy originates 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 cannot explain the association between maternal age and trisomy. 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 39.6% 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 etiology of the resulting zygote cannot be evaluated without testing for MI 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 MI 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.
The role of trans-acting factors on recombination in oocytes with nondisjoined chromosomes 21. C.D. Middlebrooks\(^1\), N. Mukhopadhyay\(^2\), S.W. Tinker\(^3\), E.G. Allen\(^4\), L.J.H. Bean\(^5\), F. Begum\(^2\), R. Chowdhury\(^2\), V. Cheung\(^1,2,5\), 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 publicly 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 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 that the function of these sequences in ESCs compared to later embryonic stages. While half of the tested sequences were previously shown to be 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 tool for the identification and characterization of enhancers active in a variety of disease states and developmental processes.

A single enhancer on human chromosome 11 directly controls >1,000 promoters and distal regulatory elements genome-wide. J.A. Stamatyannopoulos\(^1\), H. Wang\(^1\), G.J. Cost\(^2\), H. Quh\(^1\), Y. Santiago\(^2\), J. Belton\(^3\), R. McCord\(^3\), S. Orlando\(^1\), S. Neph\(^1\), Zhang\(^1\), T. Canfield\(^1\), E. Giste\(^1\), R. Sandstorm\(^2\), R.S. Hansen\(^2\), R.E. Thurman\(^1\), P.D. Gregory\(^1\), J. Dekker\(^1\), F.D. Urrum\(^2\).

1) Dept Genome Sci, Univ Washington, Seattle, WA; 2) Sangamo Biosciences, Point Richmond, CA; 3) University of Massachusetts, Worechester, 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 of 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. Busche1, B. Ge1, T. Kwan2, K. Wong3, S.-H. Chen4, M. Georges1, D. Ginzinger5, T. Pastinen6,7,1. 1) Human Genomics, 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 miRNAs. 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 miRNAs 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 miRNA alleles equally. We identified hundreds of miRNAs 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. Wang1, M.R. Campbell1, V.G. Cheung2, D.A. Bell1, 1) Laboratory of Molecular Genetics, National Institute of Environmental Health Science, Research Triangle Park, NC 27709, USA; 2) HowardHughes 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 the 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 agent 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 levels, 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. Gorkin1, S.K. Loftus2, D. Lee3, M.A. Beer4, W.J. Pevan5, A.S. McCallion1, 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 causal 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, freckling, hair color, and eye color (Han J. et al. 2011; Duffy D.L. et al. 2010; Han J. et al. 2008; Gallayh A.H. et al. 2009; Eriksson N. et al. 2010). The sequence containing rs12203592 drives strong reporter gene expression in cultured 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 melanocyte tissue specificity of this SNP implicating allele, as well as how knowledge of functional non-coding sequences can be systematically applied to identify variants that disrupt the function of these sequences.
Massively-parallel sequencing of the brain transcriptome reveals differential expression of novel genes in bipolar disorder. N. Akula1, J. Barb2, X. Jiang1, X. Duan2, K. Choi1,2, S. Sen1, B.K. Lipska1, J.E. Kleinman1, H.C. Bravo2, D.T. Chen1, P.J. Munson2, F.J. McMahon1. 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; 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-Ixl (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 analysis 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 on the Illumina 2.5M SNP array. The first three 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 × 10^-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.
Rare and common gain-of-function alleles of the serotonin transporter gene, SLC6A4, associated with Tourette disorder. P.R. Moy1, J.R. Wendland2, A.M. Andrews1, L.M. Richardson1, K.R. Timpano4, G.A. Heiman3, J.A. Tischfield2, T.R. Hein1, K.P. Thompson1, A.N. Nord1, T. Lencz1,2, T. Walsh1, M. Lee1, M. Gasparini1, F. Penzes3, J. McClellan1, M.-C. King1,2,1) Dept Genome Sciences, University of Washington, Seattle, WA; 2) Department of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Pharmacology, Northwestern University, Chicago; 4) Department of Psychiatry, University of Washington, Seattle, WA.

Functional analysis of rare chimeric genes in schizophrenia. C. Rippel1, C. Remmers2, M. Cahill3, A. Nord4, T. Walsh5, M. Lee1, M. Gasparini1, F. Penzes3, J. McClellan1, M.-C. King1,2,1) Dept Genome Sciences, University of Washington, Seattle, WA; 2) Department of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Pharmacology, 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 genes from different genomic regions to create 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 3' end of one gene and the 5' 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 known to be expressed in brain, as determined by RNAseq. Next, we used an approach 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 immunocytochemistry. For the chimeric genes, we observed both gain and loss of 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.

Rare and common gain-of-function alleles of the serotonin transporter gene, SLC6A4, associated with Tourette disorder. P.R. Moy1, J.R. Wendland2, A.M. Andrews1, L.M. Richardson1, K.R. Timpano4, G.A. Heiman3, J.A. Tischfield2, T.R. Hein1, K.P. Thompson1, A.N. Nord1, T. Lencz1,2, T. Walsh1, M. Lee1, M. Gasparini1, F. Penzes3, J. McClellan1, M.-C. King1,2,1) Dept Genome Sciences, University of Washington, Seattle, WA; 2) Department of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Pharmacology, Northwestern University, Chicago; 4) Department of Psychiatry, University of Washington, Seattle, WA.

Functional analysis of rare chimeric genes in schizophrenia. C. Rippel1, C. Remmers2, M. Cahill3, A. Nord4, T. Walsh5, M. Lee1, M. Gasparini1, F. Penzes3, J. McClellan1, M.-C. King1,2,1) Dept Genome Sciences, University of Washington, Seattle, WA; 2) Department of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Pharmacology, 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 genes from different genomic regions to create 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 3' end of one gene and the 5' 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 known to be expressed in brain, as determined by RNAseq. Next, we used an approach 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 immunocytochemistry. For the chimeric genes, we observed both gain and loss of 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.

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 counseling. 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 (p=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 5.9% (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 instability, providing persons with IAs accurate genetic counseling and will assist them in their reproductive decision-making. This data also increases our knowledge on the dynamics of CAG repeat instability in HD.
De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. J. Lee1,2, M. Huynh3, G. Mathern3, J. Gleeson2. 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.
Minimal differences in single nucleotide variation calls between blood and cell-line derived DNA from the same individuals. C.M. Schaal1, N.G. Campbell2, G. Cai3, J.S. Sutcliffe2, J.D. Buxbaum4, K. Roeder1, 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 repliable 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 read depth and strong confidence in the individual calls. Further filtering, however, reduced candidate disagreements to 664. 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 confirmed several instances in which discordant genotypes appeared as low to very low level mosaicism 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 data in LCLs in the context of the loci to be mosaic 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 efforts. Moreover, while it supports the effectiveness of standard approaches for discarding low-quality variant calls, directions for improvement are evident and simple to implement.

The impact of genetic variation on diabetes-related quantitative traits from whole exome sequences: The T2D-GENES Consortium. H.M. Highland, J.S. Sutcliffe, J.D. Buxbaum, K. Roeder, C.M. Schaal, A.M. Manning, M. Rivas, G. Atzmon, S. Cho, B.K. Comes, J.J. Dupuis, J.C. Florez, J.T. Frayling, E.R. Gamazon, I.-S. Huh, H.K. Im, J. Kim, J.Y. Kim, C.M. Lindgren, A. Locke, J.B. Meigs, A.P. Morris, N. Palmer, I., Prokopenko, T.M. Teslovich, 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 Biochemistry, Boston University School of Public Health, Boston, Massachusetts, USA; 12) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA; 13) Department of Nutrition and Food Science, University of California, Berkeley, California, USA; 14) Massachusetts General Hospital, Boston, Massachusetts, USA; 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, 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 9HT, 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 groups were 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 Ashkenazi Diaspora), 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, Ashkenazi and South Asian samples for anthropometric measures, lipids, and blood pressure and, for the non-diabetic subjects, glycemetic traits. We identified 1,651,864 exome variants consisting of 599,841 missense variants, 12,073 nonsense variants; 375,110 synonymous variants; and 1,651,864 noncoding variants. We have detected genome-wide significant associations in covariate adjusted, single variant analyses at known GWAS loci: CETF (\(p = 3.9\times10^{-12}\)) and LPL (\(p = 1.7\times10^{-8}\)) for HDL, TOMM40 (\(p = 1.7\times10^{-15}\)) for LDL, and APOA5 (\(p = 1.2\times10^{-11}\)) for triglycerides. Other suggestive associations include RNF26 (\(p = 6.3\times10^{-11}\)) with fasting insulin, ZBTB41 (\(p = 5.9\times10^{-7}\)) with diastolic blood pressure, and PDE3A (\(p = 1.8\times10^{-5}\)) 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 gene GNT2 (\(p = 5.4\times10^{-5}\)) for fasting glucose, \(TTN (p = 6.9\times10^{-5})\) for HDL, THYN1 (\(p = 6.2\times10^{-5}\)) 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.
Whole-Exome Sequencing in Multiplex Families Identifies Novel Rare Variants in Multiple Sclerosis, A.H. Becheram1, J.L. McCauley1, A. Hadjixenofontos1, P.L. Whitehead1, I. Konidari1, A. Aviram1, Y. Pasco1, S.L. Hauser2, J.R. Oksenberg2, D.J. Hedges1, J.M. Vance1, J.L. Haines3, M.A. Pericak-Vance1. 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 COL11A2, 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 2 families, one being recessive, C1orf103 and Mcm10. We identified 2 new genes having a distinct variant segregating in 3 different families, one being recessive, COL11A2. COL11A2 is in the HLA complex and has shown association with outcome prediction of MS, thus being an excellent candidate locus.

A Mendelian randomisation study on vitamin D status and blood pressure: a meta-analysis in up to 89,042 individuals. K.S. Vimalaswaran1, D.J. Berry1, A. Cavado1, M.I. Järvelin1, E. Hyppönen1, D-CarDiacollaboration. 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 (for causal association) individually and in combination using separate allele scores for SNPs involved in synthesis (comprising DHC7R 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, gender, BMI, and region. 25(OH)D levels were associated with systolic (SBP) [−0.14 (95% CI: −0.20, −0.08), p = 2.4 × 10−6] 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 = 2.2 × 10−5). All four SNPs were strongly associated with 25(OH)D (DHC7R, p = 4.3 × 10−29; CYP2R1, p = 8.28 × 10−22; GC, p = 4.11 × 10−148; CYP24A1, p = 3.77 × 10−10). Of the four SNPs, CYP2R1 (rs12794714, p = 5.98 × 10−5) 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 SNPs were strongly associated with 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.
APoE modulates the relationship among triglycerides, cholesterol, and CHD through pleiotropy and gene-gene interactions. T.J. Maxwell, C.M. Ballantyne, J.M. Cheverud, C.S. Guild, C.E. Ntvumela, E. Boerwinkle. 1) Human Genetics, University of Chicago, Chicago, IL; 2) Genétique et Amélioration des Plantes, INRA, Versailles, France; 3) Biostatistics, University of Michigan, Ann Arbor, 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 analyze 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 over methods that analyze all tissues together, and that a meta-analysis that 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 [81%, 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 have considered 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.

273

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

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;disabling&rdquo; 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 the same as for 1q21.1 proximal duplications, 1q21.1 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 range from 0.4% 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.

274

Estimates of penetrance for common pathogenic copy number variations. J.A. Rosenfeld, B.P. Coe, E.E. Eichler, H. Cuckle, L.G. Shaffer. 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.

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.
Combining Illumina gene expression microarrays from different tissues: methodological aspects. K. Heim1, C. Schürmann2, A. Schillert3, C. Müller4, T. Zeller1, C. Horninger2, J. Kruppa3, T. Illig2, G. Homuth2, K. Strach2, A. Peters9, H. Wallaschofski12, M. Dörn7, T. Meitinger1,13, P.S. Wild14,15, S. Blankenberg4, U. Völker2, M. Roder5,16, A. Teumer2, H. Prokisch1,17, A. Ziegler1 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 Sleswig-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) Hanover 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, German Cancer 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 microarray data analysis challenging. To determine the impact of technical factors on the overall expression level, we investigated the influence of various factors that may influence the expression levels on the probe intensity. The first principal component analysis on the probes' 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 an age-related random distribution of 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 in whole blood, and in whole blood samples, the first principal component analysis on the probes' intensities was performed. The overall expression level was 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 BMI and the random phenotype. The strongest non-technical influences were observed for white blood cell composition parameters, sex effects and somatic 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 effects on the cell types the expression data was generated from.

A DNA variant caller adapted to assess mitochondrial DNA variation in lymphocytes from 1,000 Sardinians. J. Ding1, C. Sidore2,3, O. Meirelles4, M.K. Trott4, F. Busonero2,3, R. Nagaraja1, F. Cucca1, G.R. Abecasis5, D. Schlessinger2, 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., heteroplasmies). 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 Sardinian 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 higher variation. The overall heteroplasmies increase with age, but the slope is small, yielding an average increase of ~1 heteroplasmies 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 mtDNA haplogroups in several individuals; 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. Richards1,2, R. Li1, A. Montpetit, T.D. Spector1, C. Polychronakos1. 1 Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, Lady Davis Institute, Jewish Genital 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 improving fitness. Thus, understanding 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 human. 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 in the CG dinucleotide site in the 5UTR of the TCF7L1 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 x 10−10 and 1.20 x 10−9 per nucleotide per cell division. These data allow us to estimate that each individual carries on average the same number of 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 add to our knowledge in genome evolution. These 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. Campbell1, J.X. Chong1, M. Malig1, A. Ko1, B.L. Dumont1, L. Han1, L. Vives1, B.J. O’Roak1, P.H. Sudmant1, M. Abney2, C. Ober2, E.E. Eichler1,2. 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 sequencing. We took advantage of the extensive autozygosity (i.e., homozygosity by recent descent) 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 pedigrees, 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 x 10−8 (95% confidence interval 1.15 x 10−8 to 1.27 x 10−8) mutations per basepair per generation. We observed a nine-fold increase in the mutation rate for bases within CpG dinucleotides (10.3 x 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 nonsynonymous and synonymous) in the autozygous segments (p = 0.011) 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. Sebro1,2, G. Peloso1, J. Dupuis1, N. Rischa1,2, 1 Institut für Human Genetik, 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=2.6e-22) and also for PC2 (r=0.80, P=4e-29). In the Offspring cohort, the spousal correlation in PC1 was lower but still highly significant: r=0.35, P=3e-28 for PC1 and r=0.45, P =9e-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 several phenotypic traits, we found that age was one of the most conserved across generations, as were SNPs that contribute to disease predisposition (e.g., homoyzogote 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 assumptions 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.
Combined analysis of loss-of-function variants in protein-coding genes from over 16,000 individuals. D.G. MacArthur1,2, M. Lek1,2, K. Neale2, S. Balasubramanian3, E. Lin1,2, B.M. Neale1,2, L. Habegger3, S. Gabriel1, P. Sullivan4, S. Kathiresan1,2, M.I. McCarthy5, M. Boeheim6, S. Purcell7, S.A. McCarroll2, M.B. Gerstein1,2, D. Altshuler1,2, M.A. DePristo2, M.J. Daly1,2, 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 variants that could 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 consistency, we have presented, 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 artifacts, 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 the 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 phenotypic cohorts.

Abundant selection explains low diversity on human Y chromosomes. M. Wilson Sayres1,2, K. Lohmueller1,2, J. Lachance1,2, 1. Massachusetts General Hospital, Boston, MA; 2. Department of Population Genetics Genome-Wide Evolutionary history and adaptation inferred from whole-genome sequences of diverse African hunter-gatherers. J. Lachance1,2, B. Verostko3,4, B. Ferwerda5,6, A. Froment7,8, J. Bodo7,8, G. Lema7,8, J. Fyhrquist9, T. Nyambo5, T. Rebbeck8,9, K. Zhang2, J. Akey10, S. Tishkoff11, 1. Departments of Biology and Genetics, University of Pennsylvania, Philadelphia, PA 19104 USA; 2. Department of Genome Sciences, University of Washington, Seattle, WA 98195 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, 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 neutral evolution is an important mechanism in shaping hunter-gatherer populations, and that deleterious genetic variation is maintained in these hunter-gatherers. Based on the pattern of deleterious loci in these populations, we propose that neutral evolution and selective sweeps have shaped human genome evolution in hunter-gatherers. Our findings are consistent with the hypothesis that deleterious genetic variation is maintained in human populations through the action of positive selection that provides fitness benefits through local adaptation.

Mapping the human genome’s missing pieces using population admixture. G. Genovese1,2, R.E. Handsaker3,4, H. Li5,6, N. Altshuler7, A.M. Lindgren8, K. Chambert1,2, B. Pasaniuc9, A. Price1,4, D. Reich1,2, C.C. Morton1,2,3, M.R. Pollak1,2, J.G. Wilson1,2, S.A. McCarroll2,3, 1. Broad Institute, Cambridge, MA; 2. Massachusetts General Hospital, 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 genes but have been lost due to historical events. We find that these genome 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 sex-specific 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 mitochondrially. Our results suggest that sex-specific selection is necessary for explaining the evolutionary history of the human Y chromosome, and argue against the concept of the “junk” Y chromosome.

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 used to infer 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.
When ancestry runs deep: Trans-species polymorphisms in apes. L. Segurel¹,², E. Leffler², Z. Gao¹, S. Pfeifer³, A. Auton⁴, O. Venn², L. Stevenson², A. Venkat¹,², J.L. Kelley⁴, J. Kidd⁴, C. Bustamante⁶, R. Bontrop⁷, M. Hammer⁸, J. Wall⁵, P. Donnelly³,⁴, G. McVean³,⁴, M. Przeworski¹²,⁹.

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. 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: Annotationing 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 to 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 patientspecific generegulatory mutations very likely contributing to the 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 to 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 patientspecific generegulatory mutations very likely contributing to the relevant phenotype or disease and/or binding site disruption of regulating factors and target genes associated with relevant phenotypes or disease.

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. Gulsuner1, T. Walsh1, A.C. Watts2, M.K. Lee3, T. Oczolik2, M.C. King2. 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.


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 genome 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.J. Goldstein, C. Proctor, W. Morrison, R. Chudin, J. Gumbiner, D. Blumenstein, B. Kollman, J. Marra, K. Chamberlin, C. Stevens, P. Sklar, H. Culluman, S. Purcell, S. McCarthy, P. Sullivan, M. Daly, B. M. Neale, and D. Altshuler. 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) Genomic Analysis Platform, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Stanford 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 variants in Affymetrix Axiom Exome 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, K.R. Mersey, M. van Kogelenberg, E. Bragin, P. Vijayaranganan, A. Tivey, S. Clayton, S. Gribble, C. Wright, D. FitzPatrick, H. Firth, J. Barrett, N. Carter, M. Hurles. 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 quality varies. 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 inference and increased mapping read counts on numerous illumina human genome 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. Mokhtari, 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 annotations 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. We quantified the impact of an alignment-aware alignment tool, GSNAP, to recover missing reads at heterozygous sites compared to a commonly used aligner, BWA. First, we mapped DNa-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 = 0.007). Finally, we identified heterozygous sites throughout the genome at which GSNAP was better at identifying true positive peaks than BWA. 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 aligning sequencing data in identifying regulatory elements at heterozygous sites.
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 = 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. Lek1, N.F. Clarke2,4, L.B. Waddell2,4, B. Thomas1, M.A. DePristo2, M.J. Daly1,2, K.N. North2,4, D.G. MacArthur1,2, 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. Yourshaw1, S.F. Nelson1,3, M.G. Martin2. 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 malabsorptive diarrhea), ENPP1 (rickets), EPCAM (Tufting enteropathy), MYO5B (diarrhea with microvillus atrophy), NEUROG3 (diarrhea with microvillus atrophy, myopathy, and rickets), EPCAM (Tufting enteropathy), ADAM17 (inflammatory skin and malabsorptive diarrhea), PCSK1 (PC1/3 deficiency), SI (sucrase isoamylase deficiency), and SLCSA1 (glucose/galactose malabsorption). Notably, in some instances knowledge of the genetic mutation led to actionable improvements in patient care. Many of the remaining novel variants are likely 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.
Novel defect in kinetochore assembly causes short stature and microcephaly of postnatal onset. C.Y. Hung, J.E. Dallman, O. Rittinger, J.W. Bauer, M. Tekin, O.A. Bodamer. 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.

Mutations in PIGO, a member of the GPI anchor synthesis pathway, cause hyperphosphatasia with mental retardation syndrome. P.M. Krawitz, Y. Murakami, J. Hecht, U. Kräger, S.E. Holder, G.R. Mortier, B. Chiaie, E. Baere, M.D. Thompson, T. Roscioli, S. Kielsba, T. Kinoshita, S. Mundlos, P.N. Robinson, D. Horn. 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) Department 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 glycophasphatidylinositol (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. Germ-line 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 cobolomas, 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.
The recurrent 600 kb deletion syndrome at 16p11.2 is among the most frequent genetic causes of autism spectrum disorder. Cases with this deletion present with a recognizable and reproducible phenotype. This deletion can be caused by de novo MEK2 haploinsufficiency. A novel mechanism for a RASopathy phenotype.

Methods We report 6 patients with de novo MEK2-containing deletions of 16p11.2. Functional assays were performed using primary fibroblast cell lines derived from patient 1 and a healthy age/sex matched control. At the desired confluence, 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 0–120 min. Cells were harvested and protein lysates were subjected to western blot analysis with antibodies specific for various MAPK pathway proteins. The level of phospho-ERK1/2 and phospho-MEK2 was measured by band intensity, and the phosphorylation ratio for each patient was compared to healthy age/sex matched control.

Results Phospho-ERK1/2 and phospho-MEK2 levels were significantly lower in cells carrying the MEK2 deletion compared to the control. Differences in the level of phospho-ERK1/2 and phospho-MEK2 were observed when comparing the deletion and control cell lines. The level of phospho-MEK2 was less abundant in cells carrying the MEK2 deletion compared to the control.

Conclusions 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. The RASopathies are characterized by distinctive craniofacial features, congenital heart defects, and skin and hair abnormalities. CFC syndrome is caused by activating mutations of BRAF, MEK1, MEK2, and possibly KRAS.

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. The RASopathies are characterized by distinctive craniofacial features, congenital heart defects, and skin and hair abnormalities. CFC syndrome is caused by activating mutations of BRAF, MEK1, MEK2, and possibly KRAS.
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. Tabor1,2, S.M. Jamali1, J.H. Yu2, A.S. Gordon3, W.S. Post1, A.D. Johnson1, T.A. Graubert1, D.A. Nickerson1, P.L. Auer2, M.J. Bamshad2,3 on behalf of the NHLBI Personal Genomics Project Team and the NHLBI Exome Sequencing Project. 1) Trueman Katz Ctr for 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 4626 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 (ARMED). 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 galactosaemia (19%) and organic acidaemia (15%) was higher than in AA (6% and 7%, respectively). Several individuals were homozygous or compound heterozygous for these variants 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.


In 2000, we established two population registries in Ukraine adhering to methods and standards prevalent in Europe; a registry of newborns 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 Polischuks, a known ethnic sub-group in Ukraine which have characteristics of a population isolate. The Rivne registries 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 the 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 NTD. Among women from AA and EA, respectively 2% and 0.5% drank alcohol. The capital city, 1.53% and 6.36% respectively consumed alcohol. Analysis of the frequency of isonony (an indirect index of consanguinity) showed that the highest rates were in northern RP counties where 8 to 15% of infants have five or more prevalent family members. The overall rate in their age group is 5%. In the 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. King2,3, K.B. Thomas2,3, N.W. Bulstrode2, S. Abu-Amero4, E. Chanudet2, P. Stanier2, E. Healy6, N.J. Sebire7, G.E. Moore1. 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) Neonatal Unit, Royal Hospital for Sick Children, Edinburgh, United Kingdom; 4) Paediatric Dermatology, University of Southhampton, Southampton, United Kingdom; 5) Paediatric Pathology, Royal Hospital for Sick Children, Edinburgh, United Kingdom; 6) Paediatric Plastic Surgery, Great Ormond Street Hospital for Children, London, 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), nevus 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 present 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. Amplification 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 20/28 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 common mutation was found in 21/22 melanoma samples and in 22/22 naevus 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 (P. Katzer, M.L. Stiriba, M.A. Aiken, F.S. Collins, K.L. Mohiko1). 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 the association of the ARAP1 locus with T2D. Using in silico and experimental approaches, we identified candidate transcriptional and transcriptional activity was proven to be context dependent. We found that a currently uncharacterized ENCODE annotated DNase I hypersensitive region in human NOS1AP ing within enhancers/silencers of this gene. We hypothesized that the causal variants must be non-coding, regulatory, residing within enhancers/silencers of NOS1AP, and that the index SNPs are in strong 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 = 0.86 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 < 0.01 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 of the ARAP1 promoter by performing dual luciferase reporter assays in the INS-1-derived rat beta cell line 832/13. The snfN haplo-typing containing the T2D risk alleles of rs11603334 (G) and rs1552224 (T) reproducibly increased promoter activity two-fold compared to the non-risk haplotype (P < 0.001). When the effects of rs1160334 and rs1552224 were separated by site-directed mutagenesis, the G allele of rs11603334 exhibited two-fold increased transcriptional activity (P < 0.001), while rs1552224 showed no effect. The DNA region surrounding rs7109575 demonstrated strong promoter activity, but that activity did not differ between SNP alleles. To our knowledge, 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 rs1160334 and examine the consequences of increased expression and regulation and activation. N. Manjarrez-Orduno1, E. Marasco1, S.A. Chung1, M.S. Katz1, J.F. Kirdiy1, K.R. Simpfendorfer1, J. Freudenberg1, D.H. Ballard1, E. Nashi1, T.J. Hopkins1, D.S. Cunninghamham Graham1, A.T. Lee1, M.J.H. Coenen1, B. Franke1, D.S. Swinkels1, R. Graham1, R.P. Kimberly1, P.M. Gaffney1, T.J. Vyse1, T.W. Behrens1, L.A. Crowson1, B. Diamond1, P.K. Gregersen1. 1) The Feinstein Institute for Medical Research, Manhas-sett, NY; 2) Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco, CA. 94117; 3) 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 Epidemiology and Biostatistics, Erasmus Medical Center, Rotterdam, 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, Okla-homa 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 = 10^-6). A intronic polymorphism in Csk reduces Csk expression and influences the suppression of phosphorylation of Tyr555 in Csk. In B cells, 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 transitional B cell compartment. Nonetheless, the origin of this novel risk allele remains unknown. 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 transitional B cell compartment. Nonetheless, the origin of this novel risk allele remains unknown. Our results suggest that the Csk-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. (This work was supported by the GaP registry (www.gapregistry.org), M. Keogh, M. DeFrancesco, 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. Supported by NIH R01AR05992, The Alliance for Lupus Research, the Alliance for Lupus Research; the Kirkland Scholar Award; and NIH/NCRR 5M01RR-00709 (L.A.C.). The authors have no conflicting financial interests.)

305

NOS1AP is the major genetic electrocardiographic QT-interval regulator. T. Kapoor1, R.B. Sekar2, V. Pihur1, M.K. Halushka3, G.F. Tomaselli4, A. Chakravarti1,2,1) McKusick-Nathans Institute of Genetic Medicine, 2) 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 variance, includes the NOS1AP gene. To prove the functional role of this GWAS variant, we performed cardiac electrical conductions coupling study, 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) it acts on intercardiac signaling, not at cardiac tissue level, yet 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 a high density panel of ~9,055 European ancestry samples from ARIC, to identify 210 variants that were genome-wide associated with QT-interval (P<5x10^-8). Since none of the coding variants in NOS1AP explained the QT-interval association, we hypothesized that the causal variants need to 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 to assess transcriptional activity at these variant alleles. At one of these variants, rs7539120, a small (1.5-fold) but consistent and significant (P=6.7×10^-7; n=8) allelic difference was observed, and this differential allelic transcriptional activity was proven to be context (flanking TSS) dependent. Further, luciferase assays with wild type and risk allele indicate that these sequences have strong transcription regulatory element (TREE) activities and presence of risk allele carrying such sequences reduces enhancer activities. Subsequent experiments with purified protein and stable cell lines identified a putative RNA binding site in the protein-DNA-binding domain that is mutated in ITGAM-CRT. 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 to developing SLE phenotype.

306

An regulatory polymorphism in Csk, a Lyp binding partner, associates with systemic lupus erythematosus and affects B cell survival, maturation and activation. N. Manjarrez-Orduno1, E. Marasco1, S.A. Chung1, M.S. Katz1, J.F. Kirdiy1, K.R. Simpfendorfer1, J. Freudenberg1, D.H. Ballard1, E. Nashi1, T.J. Hopkins1, D.S. Cunninghamham Graham1, A.T. Lee1, M.J.H. Coenen1, B. Franke1, D.S. Swinkels1, R. Graham1, R.P. Kimberly1, P.M. Gaffney1, T.J. Vyse1, T.W. Behrens1, L.A. Crowson1, B. Diamond1, P.K. Gregersen1.

From SNP to Function in Complex Traits

127

Copyright © 2012 The American Society of Human Genetics. All rights reserved.
308 LOSS-OF-FUNCTION OF SEMAPHORINS 3C AND 3D IN HIRSCHSPRUNG DISEASE. Q. Jiang1, KP. Klambt2, T. Heanue3, MX. Sosa4, Q. Wang5, JJ. Gray2, AL. Kolodkin6, DD. Ginty4, A. Chakravarti1. 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/totally colonic aganglionosis HSCR cases; (4) mutant amino acid residues were almost exclusively conserved across 8 vertebrate species (human, chimp, mouse, rat, 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, 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 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.

309 Functional Assessement of Human Coding Polymorphisms Affecting Skin Pigmentation Using Zebrafish. Z. Tsetsikhladze1,2, V. Canfield2, K. Jones1, S. Wentzel1,6, K. Reid1,6, A. Berg3, S. Johnson4, K. Kawakami5, K. Cheng1,6. 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 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, Japan; 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 ocucutaneous 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 injecting homologous mRNA into fertilized embryos to confirm the functional impact of individual human polymorphisms in SLC45A2, we first identified the orthologous zebrafish mutant, albinon, 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 Asian families, 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. Edie1, N. Zaghloul2, D. Klinedinst3, J. Lebron4, N. Katsanis4, R. Reeves1,6. 1) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Medicine, University of Maryland, Baltimore, MD; 3) Johns University, 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, 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.
Two Birds, One Stone: Epistasis profiling of many single-nucleotide variants in a human gene. O. Zill1, J. Kitzman1, J. Shendure1, S. Fields2, 3.

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, p500/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.

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<7×10^{-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<6.5×10^{-7}), 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<4.55×10^{-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<6.8×10^{-5}) 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 deregulation 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.
Utilization of chromosomal microarrays in pediatrics. A.H. Seeley1, C.E. Keegan1, C.S. Remmert2, B.A. Tarini2. 1) Pediatric Genetics, University of Michigan, Ann Arbor, MI; 2) Institute for Social Research, 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 follow-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 interpretation pipelines in a clinical setting.

Maximizing detection and minimizing noise: the first report of large scale Whole Exome Sequencing data interpretation in a clinical laboratory. F. Xia1, J. Beuten1, M. Bainbridge2, Z. Niu1, M. Vatta1, M.R. Bekheirnia1, R.E. Persson1, M. Hardison1, J.G. Reid1, D.P. Sexton2, A.C. Hawes2, P.A. Pham3, M. Wang4, N. Saada5, W. Liu2, H. Sun1, M. Scheel2, Y. Ding2, A. Roy1, J. Wiszniewska1, A. Willis1, D.M. Muzny2, S.E. Plon1,2, J.R. Lupski1, A.L. Beaudet1, P.A. Gibbs2, C.M. Eng1, Y. Yang1. 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 sub-classes of WES variants. Approximately 87,000 variants were interpreted for the 60 samples, including 23 deleterious mutant alleles related to patient phenotype, 2,405 VUSs related 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.
Efficient Detection of Causative Mutations for Rare Diseases: Rethinking Clinical Practice. H. Lee1, J. Deignan1, T. Toy2, B. Harry2, M. Younshaw3, P. Taylor4, S. Webb5, N. Dorrani6, K. Das7, F. Quinto1, S. Kantarcı1, D.A. Wong1, W.W. Grody1, E. Vilain2, S.F. Nelson1,2. 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 -96% of all protein coding genes to be at a 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 Clinical Exome Sequencing (CES) Test was the purpose of identifying the causative 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 causal genetic diagnosis in 53% that led patients to perform in medically actionable and 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.

Intentions to receive individual results from whole-genome sequencing among participants in the ClinSeqTM study. B.B. Biesecker1, F.M. Facio2, H. Eidem3, T. Fisher4, S. Brooks5, A. Limm4, K.A. Kaphingst6, L.G. Branch7, 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 benefits of receiving 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.

Changes to Control Perceptions Following Disclosure of APOE-Corony Artery Disease Associations during Genetic Susceptibility Testing for Alzheimer’s Disease: Findings from the REVEAL Study. J.S. Roberts1, W.R. Uihlein1, P.J. Whitehouse2, T.O. Obisesan3, D.L. Bhatt4, L.A. Cupples2, R.C. Green2. 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 randomization to control arm received an additional CAD risk estimate and were informed that carrier status for CAD and AD, 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 assessment for CAD in control arm. After receiving CAD risk results, participants were asked if they would change their diet, exercise, smoking, herbal supplements, or other health behaviors, and income showed no changes or slight drops in AD control perceptions among all participants in the control arm and non-negative participants in the pleiotropy analysis. 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 individuals believe CAD and AD risk are interrelated.
318 Decrease prediction ability of common genetic variants on breast cancer risk with age: possible underlying models and impact on risk prediction. H. Aichang1,2, W. Lindstrom1,2,3, P. Kraft1,2,3,1 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 defined 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. Under both models, we confirm that the observed age-specific SNP effects 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. Wei1, W. Wang1, J. Bradfield1, E. Frackelton1, C. Kim2, F. Mentch2, R. Baldassano2,3, H. Hakonarson1,2,3, The International IBD Genetics Consortium. 1) Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey, USA; 2) The Department of Human 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 individuals for whom conventional preventive measures have failed. Patients with 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 risk assessment is still necessary and we may have underestimated the potential of techniques that can target common variants that may not be captured by GWAS. We believe that advanced machine learning techniques can help to identify these variants and build predictive models that can identify subgroups of individuals at higher risk of developing IBD.

320 A Comparison of Risk Estimates for Complex Diseases: Navigenics SNP-Based Testing and Family History Assessment. L. Ayair1, C. Shuman2,3, R. Hayeems1,2,4, S. Wodak5, D. Chitayat1,2,7, J. Davies2,6, 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 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 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 empirical risk estimates in the literature. Only conditions with an estimated population incidence of greater than 2% in either males or females were included. We expect that for these 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 Navigenics SNP-based test. Of the 760 patients that are reported here, 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 education for health professionals. F. Mentch1,2,3, R. Hayeems1,2,4, S. Wodak5, D. Chitayat1,2,7, J. Witzel. 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. However, it provides challenges to provide GCRA services to all patients who pay or have insurance that covers GCRA. For example, some patients come from communities that do not have access to genetic counselors, and face-to-face workshops as a barrier to increasing course capacity. To address these challenges, the course incorporates family history assessment with personal genome testing. In the present study, we compared risk estimates for 760 patients who pursued a Web-based Case Conferencing: Aneffectivesourceofcancer genetics 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 suicidal 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 (including input from more than 200 participants) also identified multifaceted benefits of the approach, including more effective integration of genetic counseling and clinical genetics into medical practice. The course now includes distance-mediated 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. Girkke 3, Jaffe 1,4, A. Pochter 2,4, H. Van de Wiel 4, S. Pheasant 4, C. Stevens 4, J. Robinson 4, M. Calbi 4, 5, I. Gal-Viks 4, E. Kelliherr 4, R. Daza 4, M. De Felice 4, H. Hürung 4, J. Sovová 4, C. Antinac 4, 6, B. Ruttman 4, J. Handsaker 4, K. Lindblad-Toh 4, 1, S. Gabriel 4, K.S. Hart 4, 2, A. Regev 4, C. Nusbaum 4, S. Bieglmayer 4, 7, B. Lippuner 4, 7, P. Hehir-Kwa J and Translational Genomics 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 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, Gent, Belgium; 5 Department of Systems Biology, Harvard Medical School, Boston, MA; 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 I (MCKD1). MCKD1 is a simple Mendelian disorder that was mapped more than a decade ago to a 2-Mb region on chromosome 1q21.3-q21.1. We sequenced 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 results provide a cautionary tale about the challenges in identifying the genes responsible for Mendelian, let alone complex, disorders through MPS.

323
ARL13B, INPP5E, PDE6D and CEP164 form a functional network involved in Joubert syndrome and Nephronophisis. S. Seo1, M. Humbert2, C. Oud3,4, C. Kölzsch5,6, C.C. Searby7, Y. Li8, P.M. Pope9, V.C. Shefield2,4, 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 nephronophisis (NPHP), Joubert syndrome (JBS), Meckel-Gruber syndrome (MKS), and Bardet-Biedl 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 JBS, 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 JBS 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 JBS 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. Perrault1, S. Saunier2, S. Hanen2, E. Filhol3, A. Bizet2, F. Collin4, 5, M. Salih4, S. Gerber1, 2, H. Delph1, 4, E. Silva5, V. Baudouin6, M. Oud7, N. Shannon7, M. Le Merrer5, 6, O. Roche7, C. Pietrement7, 8, Bole-Feyssot11, P. Nitschke12, M. Zahrate11, P. Beales13, H. Arts1, A. Munnich5, J. Kaplan1, C. Antinac, 9, V. Cornier-Daire11, J.-M. Rozet11, 1) INSEM UT81, Department of Genetics, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, Paris, France; 2) INSEM, 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, Saudi Arabia; 5) Department of Ophthalmology, Coimbra University Hospital, Coimbra, Portugal; 6) Department 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 are 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 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 IFT140 fluorescence was restricted to the 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 control cells. In MSS cells, the effect of IFT140 mutations was 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.
Mitochondrial Disorders in ALDH1B1

Mutations in ALDH1B1, which encodes a mitochondrial protein belonging to the aldehyde dehydrogenase family, result in hepatic failure and mitochondrial respiratory chain deficiency. S. Sahli1, V. Somr1, M. Beinat1, P. Nitschke2, O. Bernard1, A. Slama1, A. Munnoch1, A. ROTIG1, 1

1 Université Paris Descartes-Sorbonne Paris Cité, In, Hôpital Necker-Enfants Malades, PARIS, France; 2) Plateforme Bioinformatique Paris Descartes; 3) Service d’Hépato-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 cytochrome c oxidase assembly, affecting the first years of life. Multiple respiratory chain (RC) enzyme deficiencies 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 DNP 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 deficiencies 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). ALDH1B1 encodes 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 suggested 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 the Leu219Pro as a homozygous 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.

Targeted Exome Sequencing of 102 Patients with Clinical Evidence of Mitochondrial Disease

D.S. Lieber1,2,3,4,5,6, S. E. Calvo1,2,3,4,5, K. Shanahan6, N.G. Slate1,2, S. Lu1,2, S.G. Herasham6,1,2,3, N.B. Gold1,2, B.A. Chapman1, M. Borowsky1, D.R. Thurbon6, G.T. Berry1, J.D. Schmahmann6, D.M. Mueller1, K.B. Sims1,2,6, V.K. Mothia1,2,3,4, 1) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA; 2) Center for Human Genomics and Bioinformatics, 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 cases, including 24% of cases lacking a prior molecular diagnosis. Here, we extend the approach to a broader cohort of 102 phenotypically similar 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 twofold 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.
Constitutive activation of STIM1 causes tubular aggregate myopathy. J. Laporte\textsuperscript{1}, F. Chevessier\textsuperscript{2}, A. Maues de Paula\textsuperscript{3}, C. Koch\textsuperscript{4}, S. Attarian\textsuperscript{5}, C. Feger\textsuperscript{1}, D. Hanter\textsuperscript{6}, P. Lafaté\textsuperscript{2}, K. Ghorab\textsuperscript{5}, J.M. Vallet\textsuperscript{2}, M. Fardeau\textsuperscript{1}, D. Figarella-Branger\textsuperscript{6}, J. Pouget\textsuperscript{2}, M. Koch\textsuperscript{1}, C. Ebel\textsuperscript{1}, N. Levy\textsuperscript{7}, B. Eymard\textsuperscript{6}, M. Bartoli\textsuperscript{7}, J. Bohm\textsuperscript{5}. 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) APHP, Service d'Anatomie pathologique et neuropathologie, Hôpital de la Timone, Marseille, France.

Ca\textsubscript{2+} is a major regulatory and signaling molecule in skeletal muscle, therefore the cellular Ca\textsubscript{2+} dynamics need to be tightly regulated. Intracellular Ca\textsubscript{2+} 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. Ca\textsubscript{2+} stores are refilled through a process called store-operated Ca\textsubscript{2+} entry (SOCE). Stromal interaction molecule 1 (STIM1) is the main Ca\textsubscript{2+} 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. Immunofluorescence 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 Ca\textsubscript{2+}. Upon Ca\textsubscript{2+} store depletion, wild-type STIM1 oligomerizes and thereby triggers extracellular Ca\textsubscript{2+} entry through Ca\textsubscript{2+}-release-activated Ca\textsubscript{2+}-channels (CRAC). In contrast, myoblasts transfected with the mutant constructs displayed constitutive STIM1 clustering, indicating that Ca\textsubscript{2+} 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 Ca\textsubscript{2+} level in patient cells as compared to the control. Addition of high [Ca\textsubscript{2+}] medium induced a sudden and massive Ca\textsubscript{2+} influx in the patient myoblasts, as compared to low gradual increase in control cell lines. These data support the constitutive activation of the Ca\textsubscript{2+} 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.

Multiple respiratory chain deficiency represents an important cause of mitochondrial disorders. 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 polyribonucleotide 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 non-coding 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.
Comprehensive analysis of 101 nuclear genes for molecular diagnosis of mitochondrial disorders. R. Bai¹, J. Higgs¹, S.F. Suchy¹, S. Buchholz¹, D. Arjona¹, B. Boggs¹, C. Chinault¹, N. Smaoui¹, S. Benhamed¹, J. Compston¹, Y. Shevchenko¹, G. Richard¹, S. Balle¹, F.D. Kendall², S. Parikh³, A.L. Gropman⁴, W. Chung⁵, A. Goldstein⁶, S.E. Palmer⁷, J.A. Panzer⁷, S.W. Yum⁸, M.J. Falk⁹, 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/VUS 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.
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 1,932 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.

Towards a whole genome map of heritable copy number variation. S. Aradhya, L. Matyakhina, D. Pineda Alvarez, D. Riethammer, 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.
Whole-genome sequencing analysis of iPSC lines uncovers lineage-manifested CNVs. A.E. Urban1, A. Abyzov2, D. Palejev2, L. Rosenberg-Belmaker2, Y. Zhang4, J. Mariani3, L. Tomasinio3, A. Ferrandino3, A. Szekeley4, M. Wilson5, M. Haney5, E. Grigorenko5, A. Huttenhower5, S. Weissman5, M. Gerstein5, F. Vaccarino3. 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 CNVs could be carrying such medium-sized to large mosaic 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 with decreased expression, an observation that invites further analysis when analyzing expression levels (by RPKM) of genes intersecting LM-CNVs with decreased expression, an observation that invites further analysis. When analyzing expression levels (by RPKM) of genes intersecting LM-CNVs with decreased expression, an observation that invites further analysis when analyzing expression levels (by RPKM) of genes intersecting LM-CNVs with decreased expression, an observation that invites further analysis when analyzing expression levels (by RPKM) of genes intersecting LM-CNVs with decreased expression, an observation that invites further analysis when analyzing expression levels (by RPKM) of genes intersecting LM-CNVs with decreased expression, an observation that invites further analysis when analyzing expression levels (by RPKM) of genes intersecting LM-CNVs with decreased expression, an observation that invites further analysis.

SNP markers identify areas with restricted recombination suggesting structural variation across the human genome is widespread. P.G. Hyslop1, B. Tamraz2, A. Nag3, C. Venturini4,5, J.S. Rahi,6 T.D. Spector1, C.J. Hammond1. 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 rearrangements over protein 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 recombination patterns 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 Hap100Quad 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 phylogenetic trees over these regions. Results. We found about 4,000 regions consistently showing reduced recombination 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. 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 retrotransposon transposons constitute one-sixth of human genome 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 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 of the L1 RNPs 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 co-opts 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 DD5X, 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 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.
terminal deletions to measure TPE on additional chromosomes are ongoing.

pathogenesis. Gene expression and H3K9me3 enrichment studies of other spreading can induce human genes silencing that may be related to disease. However, in 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 from the breakpoint junctions, without significant SNV burden, which may have implications for phenotypic manifestations.

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 their proximity to the telomere. Although 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 from the breakpoint junctions, without significant SNV burden, which may have implications for phenotypic manifestations.

De novo CNV formation in mouse embryonic stem cells occurs in the absence of Xrc4-dependent nonhomologous end joining. M.F. Arti1, S. Rajendran1, S.R. Birkeland2, K.M. McSweeney3, T.E. Wilson1,2, T.W. Glover1. 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 Xrc4−/− mouse embryonic stem (ES) cells following treatment with low doses of the DNA polymerase inhibitor, aphidicolin. Xrc4 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 Xrc4, as would be expected if canonical NHEJ was confined just to the breakpoint junctions of the CNV. Moreover, de novo CNV junctions displayed a typical pattern of microhomology and blunt end use that did not change in the absence of Xrc4. A number of complex CNVs were detected in both wild-type and Xrc4−/− cells, including an example of a catastrophic, chromosome-wide rearrangement. These results establish that nonrecurrent CNVs can be, and frequently are, formed by mechanisms other than Xrc4-dependent NHEJ, and implicate aberrant or collapsed replication forks rather than two-sided DSBs as a principal intermediate.

De novo CNV formation in mouse embryonic stem cells occurs in the absence of Xrc4-dependent nonhomologous end joining. M.F. Arti1, S. Rajendran1, S.R. Birkeland2, K.M. McSweeney3, T.E. Wilson1,2, T.W. Glover1. 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 Xrc4−/− mouse embryonic stem (ES) cells following treatment with low doses of the DNA polymerase inhibitor, aphidicolin. Xrc4 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 Xrc4, as would be expected if canonical NHEJ was confined just to the breakpoint junctions of the CNV. Moreover, de novo CNV junctions displayed a typical pattern of microhomology and blunt end use that did not change in the absence of Xrc4. A number of complex CNVs were detected in both wild-type and Xrc4−/− cells, including an example of a catastrophic, chromosome-wide rearrangement. These results establish that nonrecurrent CNVs can be, and frequently are, formed by mechanisms other than Xrc4-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. Kauwe1, C. Cruchaga2, O. Harari3, K. Mayo4, S. Bertelsen5, M. Bailey1, D. McKean1, P.G. Ridge1, T.J. Maxwell6, E. Peskind7, D. Galasko8, A.M. Goate1, 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 2p24.2 within GLIS3 and rs9922617 (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 Transcriptomes specific to the Temporal Pole of Late-Onset Alzheimer’s disease. C.E. Humphries1,2, M.A. Kohli3, P.W. Whitehead1, W.F. Huulme1, L. Nathanson1, D.C. Mash2, M.A. Pericak-Vance2, J.R. Gilbert1. 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 quantification 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 transcript. 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. Wang1, A.C. Naig2, C. Cruchaga1, S. Mukherjee3, C.-F. Lin4, O. Valladares1, L.B. Cantwell5, R. Graham5, T. Behrens6, P.K. Crane6, A.M. Goate1, M.A. Pericak-Vance1, G.D. Schellenberg7, 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 markers on 247,870 SNPs. We compared exome chip genotypes on 40,801 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 25 in APOE (P=4x10^−11, BIN1 (rs744373, P=1.24x10^−12), and many previously unreported loci. We observed similar results using a traditional logistic regression approach in PLINK and LASSO penalized regression in MÉNDEL. We have identified preliminary evidence of several promising rare variants associated with LOAD (both known LOAD genes such as 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.
Genome-wide Association Analyses of Onset Age in Late-Onset Alzheimer Disease (LOAD) Demonstrate No Strong Effect Outside of the APOE Region. A.C. Naj1, Y.S. Park1, R. Rajbhandary1; K.L. Hamilton1, G.W. Beecham2, E.R. Martin1, R.P. Mayeux2; J.L. Haines1, L.A. Farrer3, G.D. Schellenberg4, M.A. Pericak-Vance5, 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 6q16.1 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 (rs96757, P = 3.30×10^{-6}) with weak or no association to LOAD risk were associated with AAO, and found several interesting but not genome-wide significant associations in a LOAD dataset of 5,903 subjects (1,971 cases and 3,932 controls) from 16 independent studies.

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

G. Nicolas1, C. Potter1, A. Maltête2,2, S. Coutard1, A. Rovelet-Lecrux1,3, S. Legalic1,3, Y. Vaschalde4, L. Guyart-Maréchal2,3, J. Augustin1, O. Martinoud1,2, L. Defebvre5, P. Krystkowiak6, J. Pariente7, I. Le Ber5, T. Frébourg1, D. Hannequin1,2,3, D. Campion1,2,3.

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 dbSNP 131, 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 among the 5,379 control exomes of the ESP consortium. This variation also predicted to be highly damaging. None of the two mutations segregated with the disease in the family. This variation, which is predicted to be highly damaging, was present among the 5,379 control exomes of the ESP consortium.

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 among the 5,379 control exomes of the ESP consortium. This variation also predicted to be highly damaging. None of the two mutations segregated with the disease in the family. This variation, which is predicted to be highly damaging, was present among the 5,379 control exomes of the ESP consortium.

To exclude other genetic causes of IBGC. Ingenuity pathway analysis was performed for the 1,750 PD patients, 1,531 healthy controls, and 17 PD patients from the Mennonite community. The PD patients were indistinguishable from the controls for age, gender, and family history for parkinsonism. The only disease-susceptible population was the Mennonite community. The PD patients were indistinguishable from the controls for age, gender, and family history for parkinsonism. The only disease-susceptible population was the Mennonite community. The PD patients were indistinguishable from the controls for age, gender, and family history for parkinsonism. The only disease-susceptible population was the Mennonite community. The PD patients were indistinguishable from the controls for age, gender, and family history for parkinsonism. The only disease-susceptible population was the Mennonite community.
**347**

C9ORF72 repeat expansion is a risk factor for Parkinson Disease. K. Nuytemans1,2, G. Bademci1,2, M.M. Kohli1,2, G. Beecham1,2,3, V. Inchausti1,2, A. Dressen1,2, L. Wang1,2,3 J.J. Young1,2,3 F. Nahab1,2, C. Singer1,2, E.R. Martin1,2,3, J.R. Gilbert1,2,3, M. Benatar1, J.L. Haines1, W.K. Scott1,2,3, S. Zuchner1,2,3, M.A. Pericak-Vance1,2,3, J.M. Vance1,2,3 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 (388 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. Smith1, S. Topp1, J. Bamwell1, A. Al-Chalabi1, J. Kirby2, P.J. Shaw3, H. Pali3, K.E. Morrison3, V. de Jong4, F. Bass5, C.E. Shaw1, C.M. Lewis4. 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 (68.4% 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% CI. 63% - 82%). Penetration 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-dependent 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.
Empirical and theoretical studies on genetic variance of rare variants for complex traits using whole genome sequencing in the CHARGE Consortium. C. Zhu⁴, A. Morrison⁴, J. Reid⁴, C.J. O’Donnell⁴, B. Psaty⁴, L.A. Cupples⁵,6, R. Gibbs⁵, E. Boerwinkle²,3, L. Liu⁴. 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) NHGRI 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 (σ²g) 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 σ²g of rare variants (MAF≤5%) 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 BMI as an example, the estimated σ²g 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.

Leveraging admixture analysis to resolve missing and cross-population heritability in GWAS. N. Zaitlen¹, A. Gusev¹, B. Pasaniuc¹, G. Bhatia², S. Pollock³, A. Tandon¹, E. Stahl⁴, R. Do⁵, B. Vilhjalmsson⁶, E. Atyekshova⁶, A. Cupples⁶, M. Fornage⁷, L. Kao⁷, L. Lange⁷, S. Musan⁷, G. Papanciojao⁸, J. Rotter⁹, I. Ruczinski¹⁰, D. Siscovick¹¹, X. Zhu¹², S. Mccarroll¹³, G. Letter¹³, J. Hirschhorn¹⁰, N. Patterson¹⁰, D. Reich¹⁰, J. Wilson¹⁰, S. Kathiresan¹³, A. Price¹², CAC, CARe Analysis Core¹, 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 (NHBLI), Division of Cardiovascular Sciences, NIH, Bethesda, MD 20892, USA; 11) Mount Sinai School of Medicine, New York, NY, 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 CentreVille, Montréal, Québec, Canada; 16) Divisions of Genetics and Endocrinology and Program in Genomics, Children’s Hospital Boston, Boston, MA, USA.

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 h², 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 (h²γ) is related to h², Fst, and genome-wide ancestry proportion (γ): h²γ = h² - 2Fstγ(1-γ). Thus, we can estimate h²γ and then infer h² from h²γ. 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.
Applying a quantitative genetics test of evolutionary neutrality to finger ridge-count, a classical model trait in humans. S. Medland1, PM. Vitale2, WR. Montgomery3, DM. Evans4, NG. Martin4. 1) Quantitative Genetics, Queensland Institute Medical Research, Brisbane, Australia; 2) Diamantina Institute, University of Queensland, Brisbane, Australia; 3) Queensland Brain Institute, University of Queensland, Brisbane, Australia; 4) Genetic Epidemiology, Queensland Institute Medical Research, Brisbane, Australia; 5) Molecular Epidemiology, Queensland Institute Medical Research, 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 (h²=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, h²=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<1x10−8). 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 (bq21: p=6.4x10−11, %Variance (replication sample) = 1.17; bq21: p=2.0x10−8, % Variance (replication sample) = 1.4x10−8). Considering 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.

Does common variation contribute to the shared genetic basis for schizophrenia and autism? P.H. Lee1,2,3, S. Ripke1,2,3, S. Santangelo1,2,4, M. Daly1,5, 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, USA; 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 epidemiological 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. Multimodal 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 (rs152579/mir137). Another gene of interest was NOS3, 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 follow-up studies to confirm these associations and to identify candidate genes that confer ASD-SCZ specific susceptibility.

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 UKBIDGC 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 r² between the inferred allele dosage from the sequence and the genotypes from GWAS chips typed on the same individual.

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 in 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 is too 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.


356 Deep targeted sequencing of 12 breast cancer loci in 4,700 women across four different ethnicities. P. Kraft1, S. Lindstrom1, B. Chapman1, G. Chen1, C. Chen1, O. Hofmann1, D. Mir61, C. Haiman1. 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 recombinant 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 involved sequencing and the importance of thorough quality control procedures.

357 Population stratification of human disease-associated SNPs, and their relevance to human disease networks. S. M. Raj1, G.E. Hoffmann2, A.G. Clark3, 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. 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 human SNPs, and then map this variation to 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. For each trait-SNP pair, flanking exonic SNPs having r² > 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 network information based Nonparametric Exploration (MINE) 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. Our results indicate that patterns of differentiation of SNP allele frequencies differed among traits, and these patterns are correlated with disease phenotype-based networks.

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.

Exome Sequencing Uncovers Etiology of Mendelian Disease

358 Loss of function mutations in known human disease genes in 572 exomes. J. Johnston1, K. Lewis1, D. Ng1, S. Gonsalves1, J. Mullikin2, L. Biesecker1,2, 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. 

Genome and exome sequencing costs continue to fall and many individuals are undergoing these assessments as research participants and patients. In contrast to secondary findings in genome wide association studies, 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 one that is predicted to alter protein function, is causative. To better understand the frequency of potentially causative mutations in apparently healthy persons, we have analyzed potential loss of function variants 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. Filtered variants included frequency, quality and predicted inheritance. As our goal was to understand the role of these variants in health and disease, we selected cases defined as Mendelian, phenotypically and genetically. Pertinent phenotypes and known diseases were prioritized. Our resulting data have yet to be established. Substantial practical challenges arose from clinical application of ES including genetic counseling, proof of disease causality, and prioritization. Our exome sequencing to identify the cause of Mendelian Diseases.

359 The Problem of Multiple Plausible Molecular Diagnoses in Next Generation Sequencing Data: The NIH Undiagnosed Disease Program Experience. D. Adams1,2, C. Boerkoel3, K. Fujentes Fajardo4, P. Cherukuri4, M. Sincan5, C. Toro5, C. Tift5,2, W. Gahl2, T. Markello5,2, 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 causality, and prioritization. Results: Our exome sequencing (ES) data revealed and previously undiagnosed possible disease-causing variants in affected family members. 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 resultant 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 “o’bives” that are dispersed across the genome, 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 likely inherited disease modifying genes remain to be identified 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. Lupsik1,2, C. Gonzaga-Jauregui1, W. Wszniewski1, D. Pehravan1, E. Karaca1, A. Stray-Pedersen1, S. Hjørring1, J. Reid1, D. Myhr1, R.A. Gibbs1,2, 1) Baylor-Hopkins Center for Mendelian Genomics, 2) Mol & Human Gen, 358. 1) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 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 (NGS) technologies. Exome sequencing, 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 targeted 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 our dataset through putative patient family members.

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. Arboleda1, H. Lee2, F. Panaik3, A. Fleming4, B. Banerjee4, B. Ferraz-de-Souza4, E. Delot1, J. A. Rodriguez-Fernandez1, D. Braslavsky4, J. Berepad4, E. C. DellAngelica5, S. F. Nelson6, J. A. Martinez-Agosto2, J. C. Achermann7, E. Vilain1,2, 1) Dept Human Gen, Univ California, Los Angeles, CA; 2) Dept Pathology and Laboratory Medicine, Univ California, Los Angeles, CA; 3) Dept Human Gen and Medical Genetics, Univ California, Los Angeles, CA; 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) National Institute of Endocrinology, Hospital de Niños Ricardo Gutierrez, Buenos Aires, Argentina; 7) Department of Urology, David Geffen School of Medicine, University of California, Los Angeles, CA.

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 mega-base deletion that maps on 1p12-p13 in affected family members. Targeted exon array capture of the disease locus, followed by high-throughput genome sequencing and validation by deoxysequencing, identified missense mutations clustered in the PCNA binding domain of the imprinted gene Cdkn1c (P53KIP1) in IMAGe 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 Cdkn1c in Drosophila caused embryonic lethality and severe eye and 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.

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.
Genetic etiology of isolated congenital asplenia. A. Bolze1, L. Abel2, A. Fue1, N. Trede2, L. Selleri2, J-L. Casanova1,2. 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 rapid 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.

Whole Genome Sequencing in Two Brothers with Heterotaxy Reveals BCL9L as a Novel Gene Associated with Autosomal Recessive Heterotaxy (HTX6). C.J. Saunders1,2,4, N.A. Miller1,2,4, S.E. Soden1,2,4, E. Farrow1,2,3,4, D.L. Dinwiddie1,2,3,4, N.P. Safina1,2,4, S. Humphray3,4, P. Saffrey5, Z. King5, J.C. Weir6, J. Betley5, R.J. Grocock5, J.E. Petri5, K.P. Hall7, S.F. King8more1,2,3,4. 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., Chesterfield 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.
The complete GENCODE human annotation: new insights into the functionality of transcriptional complexity. J. M. Mudge1, A. Frankish1, GENCODE Consortium2,3,4,5,6,7,8, T. Hubbard1, J. L. Harrow1. 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 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 the large number of such variants, we define a LoF variant as one that is predicted to be functionally 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 flow cytometry can indicate transcript function. The set of such variants is compromised. 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 be visualized via the Ensembl and UCSC genome browsers or accessed through the Ensembl databases, Peri API and BioMart.

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 regulatory potential of genetic variation, revealing the similarities and differences between classes. For this end, we have gathered a diverse set of genomic annotations, including regulatory sequence annotations derived from the ENCODE project such as DNAse 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 that different QTLs arise from cis- and trans associations, to isoform ratio (splicing) and allelic effects. To this end, we have characterized gene expression variation across seven diverse human populations. A. Battle, A. Cheung, A. Bruzell, L. McDaniel, A.L. Richards, J.M. Young, T. Young, I.X. Wang, 1) Howard Hughes Medical Institute; 2) Cell and Molecular Biology Graduate Program; 3) Genomics and Computational Biology Graduate Program.

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 regulatory potential of genetic variation, revealing the similarities and differences between classes. For this end, we have gathered a diverse set of genomic annotations, including regulatory sequence annotations derived from the ENCODE project such as DNAse 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 that different QTLs arise from cis- and trans associations, to isoform ratio (splicing) and allelic effects. To this end, we have characterized gene expression variation across seven diverse human populations. A. Battle, A. Cheung, A. Bruzell, L. McDaniel, A.L. Richards, J.M. Young, T. Young, I.X. Wang, 1) Howard Hughes Medical Institute; 2) Cell and Molecular Biology Graduate Program; 3) Genomics and Computational Biology Graduate Program.

Characterizing gene expression variation across seven diverse human populations. A.R. Martin, H.A. Costa, J.M. Kidd, B.M. Henn, M.C. Yee, F. Grubert, S.B. Montgomery, H.M. Cann, M.P. Snyder, C.D. Bustamante, 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’Étude du Humin, 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-seq 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 levels, 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 remain 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 of RNA-seq gene expression resolution of RNA-seq data, 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 a 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.
Comparative eQTL analyses within and between seven tissue types suggest mechanisms underlying cell type specificity of eQTLs. B. P. Fairfax1,2,5, S. Makino1, J. C. Knight1. 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 either 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 stimulator 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.
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, C.T. Watson, J. Huddleston, P.H. Sudmant, R.L. Warren, M. Malig, J. Schein, A.J. Willsey, J.B. Joy, J.K. Scott, T. Graves, R.K. Wilson, R.A. Holt, F. Breden, E.E. Eichler. 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. Mason1, K. Meyer2, Y. Salete3, P. Zumbo1, O. Elemento1, S. Jaffrey1. 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 mRNA 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. Cortes1, Y. Li1, A. Kwan1, B. Zanke1, Z. Zhang1, A. Petronis1. 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 (circDNA) 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 epigene-wide analysis of circDNA modifications using microarray technology. Universal circDNA were ligated using adaptors and amplified followed by digestion with DNA modification-sensitivity restriction enzymes. circDNA fragments that survived enzymatic hydrolysis were amplified by adaptor-mediated PCR and labeled using biotinylated nucleotides and a fluorescent dye. This enriched circDNA-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 circDNA 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 circDNA modification between cases and controls. Moreover, we applied generalized linear model to identify age- and gender specific variation in the circDNA modification profiles. We identified 2,536 and 2,705 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 circDNA may provide a new, non-invasive approach for early detection of CRC.


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 behavioral 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 mRNA 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, Snrnp-Ube3a (Murine 7qC/Human 15q11-q13) and Dlk1-Dio3 (Murine 12fF1/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 mRNA methylation using methylated DNA immunoprecipitation followed by hybridization to DNA arrays (MeDIP- Chips) that show both short- and long-range effects on DNA methylation and 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 middevelopment and brain function, may have a role in the long-term maintenance of altered gene expression and cognitive endophenotypes associated with FASD.

Copyright © 2012 The American Society of Human Genetics. All rights reserved.
380 Genome-wide scan of DNA methylation in the aging brain and its relation to Alzheimer’s disease. P.L. De Jager¹, G. Srivastava¹, M.L. Eaton¹, L.E. Chibnik¹, B. Keenan¹, N. Taner², A. Myers¹, B. Bernstein¹, A. Meisner², M. Kellis², D.A. Bennett¹. ¹Neurology, 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 HumanMethyl450K 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 9/17 of the tested genes, RNA levels are significantly correlated with AD pathology (p<0.0003). Further, we evaluated the independent contribution 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 distant in 75% of different chromatin states, 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 role for CTCF motif recruitment 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 axiata. Y.K. Chutake¹, A.M. Castro¹, S.I. Bidichandani¹, ², ³, ¹Department of Biochemistry & Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK; ²Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; ³Friedreich axiata (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 polyt 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, and 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 that 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. Li¹, Y. Zhu¹, K. Szulwach¹, L. Lin¹, C. Street¹, H. Wu², D. Chen², P. Jin¹. ¹Dept Human Gen, Emory Univ, Atlanta, GA; 2) State Key Laboratory of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, PR CHINA; 3) Department 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 the possibility of its role in regulating epigenetic control of gene expression. 5hmC constitutes a significant portion of nucleo- tides 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 (5mC) 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 (GADD45A) 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 develop- ment of zebrafish. These results together reveal an unexpected role of p53 in regulating cytosine modification, suggesting that the GADD45A participates DNA demethylation through modulating the hydroxyla- tion activity of TET proteins.

383 Maps of open chromatin highlight cell type-specific patterns of regulatory sequence variation at hematological trait loci. C.A. Albers¹,², D.S. Paul¹, A. Rendon⁻¹,², K. Voss¹, J. Stephens¹, P. van der Harst², J.C. Chambers¹,²,³, N. Soranzo¹, W.H. Ouwehand⁴,², P. Deloukas⁴, HaemGen Consortium¹. ¹Wellcome 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 Clinical Genetics, 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 genetic association studies (GWAS) 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 cell types, monocytes and platelets, for 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-occupied regions, NDRs) to identify the primary human precursors of platelets (megakaryo- cytes), 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 cell processes, e.g. hematopoietic identity vs. housekeeping, depending on the number of sequence reads supporting the NDRs. We used bootstrap-tailed 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, using observed strong enrichments (~2-fold to 10²-fold) as a 0.05 quantile lower bound p-value at the 0.005 genome-wide significance level to identify NDRs that can be used to reduce false positive rates for weaker signals. Considering that 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 order of enrichments in NDRs revealed interesting patterns across different 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 a role in the regulatory effect 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.
Functional epialleles at an endogenous human centromere. B.A. Sullivan1,2, K.A. Maloney1, L.L. Sullivan1, E.D. Strome1, J. Matheny1. 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. l. farrer1, j. kozubek1, m. schu1, j. farrell1, m. morrison1, k. mayne1, d. morgant1, r. robinson2, a. swaroop3, d. schaumberg4, kh. park5, ee. tsirion6, g. silvestrin7, ik. kim8, r. chen9, c. huff9, j. johnson10, m. morrison1.

386 The role of SIX6 in primary open-angle glaucoma. M. Ulmer1, B. Whigham1, D. Parker2, X. Qin3, N. Katsanis4, Y. Liu3, A. Ashley-Koch4, R. Allingham2, M. Hauser1, NEIGHBOR consortium, 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.

387 Topical ocular sodium 4-phenylbutyrate rescues glaucoma in a mouse model of primary open angle glaucoma. G. S. Zode1, k. e. Budge1, e.m. stone1, v.c. steele1, j. f, h. hale2, r.h. h. hollands2, r.h. hollands2, 1) HHH, University of Iowa, Iowa City, IA, 52241, 2) Departments of Pediatrics, University of Iowa, Iowa, IA 52241.

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 3,979 keratoconus cases and controls (Glau93Glau, Glu129Glau, Glu193Glau, Rs339112345, Leu205Arg, Thr212Met, Ser242Leu). Using an allele-based chi-squared test, rs339112345 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 dose response related of association with intraocular pressure (p=2×10-7, 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 microinjection in developing zebrafish. First, we observed retinal phenotypes in retinitis pigmentosa 1-like 1 gene, has an essential role in affecting the retinitis pigmentosa 1-like 1 gene, has an essential role in affecting the 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-MYOCY437H), which expresses human mutant MYOC in the trabecular meshwork (TM), a tissue that regulates intraocular pressure (IOP). Adult Tg-MYOCY437H 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 elevated IOP in adult Tg-MYOCY437H mice. We explored whether reducing chronic ER stress rescues the glaucoma phenotype of Tg-MYOCY437H mice by topically treating mice with the chemical chaperone, sodium 4-phenylbutyrate (PBA). At 3 months of age, vehicle-treated Tg-MYOCY437H mice exhibited elevated IOP compared to WT littermates (n=24, p<0.0001). Topical PBA significantly reduced elevated IOP in Tg-MYOCY437H mice to the level of WT mice. PBA-treated Tg-MYOCY437H mice also preserved RGC function compared to vehicle-treated Tg-MYOCY437H 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-MYOCY437H 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.
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 strain (eye3) with microphthalmia. The mutation responsible for the eye3 strain maps 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 phylogenetic 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) die around E9.5, but a portion of Rereom/+ mice live 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−/− mice at P14. Further histological examination revealed that each layer of 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−/− mice at P14.

The review of ophthalmological data LCA cases. NMNAT1 codes for the ubiquitously expressed nuclear isoform severe retinal dystrophy, responsible for neonatal blindness. 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 strain (eye3) with microphthalmia. The mutation responsible for the eye3 strain maps 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 phylogenetic 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) die around E9.5, but a portion of Rereom/+ mice live 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−/− mice at P14.
Gene therapy provides long-term visual function in a pre-clinical model of retinitis pigmentosa. K.J. Wert1, 2, R.J. Davis1, S.H. Tsang1. 1) Ber-
nard & Shirlee Brown Glaucoma Laboratory, Departments of Ophthal-
moogy, Pathology & Cell Biology, College of Physicians & Surgeons, Columbia
University, New York, NY; 2) Institute of Human Nutrition, College of Physi-
cians & 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 Pde6αnmf363 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-Pde6α, to transduce Pde6αnmf363 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-Pde6α 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 retinas, 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.
The genomic landscape of childhood pre-B acute lymphoblastic leukemia. J. Spinella1, R. Vidal1, J. Healy1, V. Saillour4, E. Bareikis2, C. Richer4, S. Busche3,1, B. Ge1, T. Pastinen2,3, D. Sinnett4,1 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 mechanisms leading to this leukemia 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 database checks. Whole genome sequencing (WGS) was used 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.

Genomic analysis of serial chronic lymphocytic leukemia samples suggests that epigenetic changes, rather than clonal evolution, drive progression of disease. E.N. Smith1,2, C. DeBoever2,3, L. Rassenti1, E. Ghia1, S. Rozenzhak1, P. Shepard1, H. Alakus4, O. Harismendy1,2, C. Barrett1,2, T.J. Kipps1, K.A. Frazer1,2,1) 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, 92039-3, S. Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, 92039.

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 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, we treated this cohort of approximately 1,000 patients, tumors 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 was 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 used analyzing differential methylation, 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.
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 use micro-dissect breast tumor cells, adjacent historically 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 tumors, rather than just their presence. Each of the tissue types had significantly different frequencies of distribution of CAG repeat length variants measured by paired tumor-normal cell pairs. Our results, indicate that zero CAG repeat lengths that are present 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 tissue 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, 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 cancers. The identification of AR-tumor-specific mutations and their association with breast cancer, will require analysis of the frequency of distribution of gene variants and not just their presence in tumors.

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. Gottlieb1,2,3, C. Alvarado1, C. Wang1, B. Ghazizadeh5, F. Babrzadeh5, L.K. Brito1,2,3, M. Trifiro1,2,5,4, 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.

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 use micro-dissect breast tumor cells, adjacent historically 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 tumors, rather than just their presence. Each of the tissue types had significantly different frequencies of distribution of CAG repeat length variants measured by paired tumor-normal cell pairs. Our results, indicate that zero CAG repeat lengths that are present 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 tissue 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, 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 cancers. The identification of AR-tumor-specific mutations and their association with breast cancer, will require analysis of the frequency of distribution of gene variants and not just their presence in tumors.

The 3D topographic mapping of genetic variations in treatment naive advanced ovarian cancer. E. Cuppen1, M. de Pagter2, M. Hoogstraat3, G. Cirkel2, J. Kreetmajer1, C. Lee1, E. Levandowsky1, T. Guy1, K. Duran4, R. T Slot1, T. Jonges5, S. van Lieshout1, M. Lolkema6, R. Zweemer2, M. Koudijs7, I. Nijman8, E. Voest9, T. Harkins10, W. Kloosterman1, 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 carcinoma-genesis, 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 ≥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 PGM to generate comprehensive base calling sequence 40% of the cancer genomes for lower penetration variance. 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.

Transcriptome sequence analysis of human colorectal cancer samples to reveal functional attributes. H. Ongen1, T.F. Omtoft2, B. Oster2, L. Romano3, A. Pianchon1, C.L. Andersen1, E.T. Dermitzakis1, 1) Department of Genetic Medicine and Development, University of 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-related death worldwide. 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. 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 ≥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 PGM to generate comprehensive base calling sequence 40% of the cancer genomes for lower penetration variance. 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.

The preliminary analysis of 24 samples indicate that between normal colon and 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.

The 3D topographic mapping of genetic variations in treatment naive advanced ovarian cancer. E. Cuppen1, M. de Pagter2, M. Hoogstraat3, G. Cirkel2, J. Kreetmajer1, C. Lee1, E. Levandowsky1, T. Guy1, K. Duran4, R. T Slot1, T. Jonges5, S. van Lieshout1, M. Lolkema6, R. Zweemer2, M. Koudijs7, I. Nijman8, E. Voest9, T. Harkins10, W. Kloosterman1, 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 carcinoma-genesis, 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 ≥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 PGM to generate comprehensive base calling sequence 40% of the cancer genomes for lower penetration variance. 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.

Transcriptome sequence analysis of human colorectal cancer samples to reveal functional attributes. H. Ongen1, T.F. Omtoft2, B. Oster2, L. Romano3, A. Pianchon1, C.L. Andersen1, E.T. Dermitzakis1, 1) Department of Genetic Medicine and Development, University of 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-related death worldwide. 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. 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 ≥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 PGM to generate comprehensive base calling sequence 40% of the cancer genomes for lower penetration variance. 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.
Regulatory regions are somatic mutation cold spots in cancer genomes. S. Sunyaev1,2,3, P. Polak1,2,3, M.S. Lawrence4, R.E. Thurman4, N. Stolc2,3, P. Stojanov3, E. Rynes4, L.A. Garraway2,3,5,6, S. Mirkin7, G. Getz3, J.A. Stamatoyannopoulos4. 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. Thomas1,2, M. Legendre1, S. Saunier2,3, B. Bessières1, C. Alb0,2, M. Bonnière1, A. Toutain1, L. Loeuillet1, K. Szymanska1, F. Jossi2, D. Gaillard3, M. Tahir Yacoub1, S. Mouguou-Zerelli1,2, A. David1,2, M-A. Barthez1, Y. Ville2, C. Bole-Feyssot5, P. Nitsche1, A. Munnich1,2,4, C.A. Johnson3, F. Encha-Razavi1,2,4, V. Cormier-Daire1,2,4, A.T. Thouin-Robinet1,2,3, M. Vekemans1,2, T. Attié-Bitach1,2,4,1 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, Tunisia; 12) Service de Cytogénétique, Génétique moléculaire et Biologie de la reproduction, CHU Farhat Hached, Sousse, Tunisia; 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 Saint-Cyr, AP-HP, Paris France; 5) Service de Gynécologie-Obstétrique, Hôpital Necker-Enfants Malades, Paris, France; 4) Département de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 3) INSERM U983, Hôpital Necker-Enfants Malades, Paris, France; 2 Université Paris Descartes, Paris, France; 1) INSERM U781, Hôpital Necker-Enfants Malades, Paris, France; 18) EA GAD, INSERM U983, Hôpital Necker-Enfants Malades, Paris, France; 16) Plateforme de génomique, Fondation IMAGINE, Hôpital Necker-Enfant Malades, Paris, France; 15) Plateforme de Bioinformatique, Université Paris Descartes, Paris, France; 14) Service de Neuropédiatrie, CHRU de Tours, Hôpital Clocheville, Tours, France; 13) Service de Cytogénétique, Génétique moléculaire et Biologie de la reproduction, CHU Farhat Hached, Sousse, Tunisia; 12) Service de Cytogénétique, Génétique moléculaire et Biologie de la reproduction, CHU Farhat Hached, Sousse, Tunisia; 11) Service d’Anatomie et de Cytologie Pathologiques, CHU Farhat Hached, Sousse, Tunisia; 10) Service de Génétique et Biologie de la Reproduction, CHU Reims, Université de Reims, France; 9) Service d’Anatomie Pathologique, CHU de Nantes, Nantes, France; 8) Section of Ophthalmology and Neurosciences, Leeds Institute of Molecular Medicine, St. James’s University Hospital, Leeds, LS9 7TF UK; 7) Service d’Anatomie et de Cytologie Pathologiques, CHI Poissy, Saint Germain en Laye, France; 6) Service de Génétique, CHRU de Tours - Hôpital Bretonneau, Tours, France; 5) Service de Génétique, CHRU de Tours - Hôpital Bretonneau, Tours, France; 4) Département de Génétique, CHRU de Tours - Hôpital Bretonneau, Tours, France; 3) INSERM U983, Hôpital Necker-Enfants Malades, Paris, France; 2) Université Paris Descartes, Paris, France; 1) INSERM U781, Hôpital Necker-Enfants Malades, Paris, France; 163

Overall, our data show the involvement of the transition zone protein Rotatin in ciliary function to abnormal development and organization of the cortex. Rotatin in cilia structure and function. Rotatin mutations therefore link aberrant ciliary function to abnormal development and organization of the cortex in human individuals.

404 Abnormal development of NG2+PDGFRα neural progenitor cells causes neonatal hydrocephalus in a ciliopathy mouse model. C.S. Carter1,2, W. Vogel1, Q. Zhang1,2, TO. Moninger1,2, B.R. Théodens1, KM. Kepper-Noreille1,2, DY. Nishimura1, CC. Seary1,2, K. Buggese1,2, VC. Sheffield1,2,1) 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.M. Mancini, P.W. Verheijen. Department of Clinical Genetics. Erasmus Medical Center, Rotterdam, Netherlands.

Polygyria (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 manifestation. 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 cilium 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.

Primary ciliary dyskinesia (PCD) is a heterogeneous group of rare autosomal recessive ciliopathies with a complex phenotype caused by impaired ciliary 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 repeat 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 LRRC30, a gene with known association with IDA and ODA ultrastructural defects. This is congruent with the phenotype of the affected individuals. Additionally, there is a D. rerio homolog model (LR6a) 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.

Temporally and spatially resolved cataloging of in vivo forebrain enhancers. A.S. Nord1, L. Taher2, J. Akyama1, M.J. Blow1, A. Holt1, R. Hosseini1, S. Phouanenavong1, I. Plajzer-Frick1, M. Shoukry1, V. Afzal1, E.M. Rubin1,3, I. Ovcharenko1, J.L.R. Rubenstein1,4, L.A. Pennacchio1,3, A. Visel1,2, 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 the balance between sexual and temporal and spatial specificity of gene expression. Thus, characterizing distal enhancer activity across development has the potential to inform the balance between 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 others that are cooperatively active across forebrain development. Stage-specific enhancer activities enrich for transcription factors 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.tbi.edu/xu]. Together, these datasets provide an unparalleled view into the spatial and temporal activity of enhancers through a public web database at [au0.2][http://enhancer.tbi.edu/xu].

SRY Regulation of the RET Gene Suggests a Potential Role of the Y-chromosome Gene in Sexual Dimorphism in Hirschsprung Disease. Y. Li1, Z. E. Tabatabai1, M. Garcia-Barcelo2, P. H. Tam1, Y. F. C. Lau1, 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 sex determination by suppressing ovarian determinant 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 also interact with a variety of co-factors and downstream regulatory elements. 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 haplinsufficiency 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 mutation accounts 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 SRY is regulated by a distal and a proximal enhancer at its promoter, in which PAX3 and NKX2-1 are the resident transcription factors. Interestingly, in the context of HSCR, NKX2-1 interacts with its HMG box, and exacerbates their respective transcription activities. SRY competitively binds to PAX3 and NKX2-1, and represses their transcription and SOX9 stimulation on the RET promoter. Chromatin immunoprecipitation with gene-specific PAX3 and NKX2-1 antibodies 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 attenuating the RET promoter. Retinoblastoma and HSCR are both sex-linked genetic disorders, and we show that 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.

MAP3K1 MUTATIONS IN 46,XY DGDs ALTER CROSS-TALK IN DOWNSTREAM SIGNAL TRANSDUCTION PATHWAYS TO CAUSE ABNORMAL HUMAN Gonadal development. J. Loke, A. Pearlman, H. Dadon, Pathology, Albert Einstein College of Medicine, VAMC-111C5, Univ California, San Francisco, San Francisco, CA.

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 described in DGD patients. 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 ovari-an-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 blotting for low cytostasis phosphorylation and flow cytometry-based phosphorylation, and subsequent signaling. 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 transcriptional stability by AXIN sequestration and that the mutant SMD mediated the cAUG cAGG cAUG, and inhibited SOX9, FGF9 and FGF22 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 MAPK signaling 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.tbi.edu/xu]. Together, these datasets provide an unparalleled view into the spatial and temporal activity of enhancers through a public web database at [au0.2][http://enhancer.tbi.edu/xu].
410 Soft tissue aspects of the Williams-Beuren syndrome facial phenotype can be attributed to GTF2IRD1. S. Palmer1, S.P. Canales1, P. Carmona-Mora1, P. Kaur2, P.W. Gunning3, E.C. Hardeman1, 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 cranofacial 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 are 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 GIf2ird1 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 GIf2ird1 expression in the developing facial epidermis, suggesting that the `fullness' phenotype is restricted to zones demarcated by GIf2ird1 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. Chen1,2, J. Tao1,2, Y. Bae1,2, M. Jiang1,2, T. Bertin1,2, Y. Chen1,2, T. Yang1,2, B. Lee1,2. 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.