Burroughs-Wellcome Fund
Maryland Genetics, Epidemiology and Medicine (MD-GEM) Pre-doctoral Training Program

Abstract Book
Genetics Research Day
April 3, 2019

Contents

Dear Participants,

On behalf of the Maryland-Genetics, Epidemiology, Medicine Training Program (MD-GEM) it is our pleasure to welcome you to the fifth annual Genetics Research Day at Johns Hopkins University. MD-GEM includes faculty spanning the Mckusick-Nathans Institute of Genetic Medicine, the Johns Hopkins Bloomberg School of Public Health, the Johns Hopkins School of Medicine and the National Human Genome Research Institute, who join together to train doctoral students in population and laboratory sciences focused on genetics.

This Genetics Research Day provides the greater JHU community an opportunity to promote discussion and collaboration across JHU/NHGRI and to integrate students from different disciplines into the wide breadth of genetics research. We welcome all faculty, post-doctoral fellows and students, especially those new to the field of genetics. We look forward to continued partnerships and new relationships across the fields of Epidemiology, Biostatistics, Human Genetics, Biology, Computer Science, Mathematics and more.

The posters represent the Departments of Biochemistry and Molecular Biology, Biostatistics, Epidemiology, International Health and Mental Health of the Johns Hopkins Bloomberg School of Public Health; the Departments of Hematology, Neurology, Neuroscience, Oncology, Medicine, Molecular Biology and Genetics, Pathology, Psychiatry and Behavioral Sciences, and Surgery in the Johns Hopkins School of Medicine; the Center for Cell Dynamics, Center for Computational Biology, Center for Epigenetics, Lieber Institute for Brain Development, McKusick-Nathans Institute for Genetic Medicine, Welch Center for Prevention, Epidemiology & Clinical Research, and Wendy Klag Center for Autism of the Johns Hopkins University; the National Institute of Allergy and infectious Diseases; the National Cancer Institute; and the National Heart, Lung and Blood Institute.

A very special thank you to Dr. Judy Cho, Ichan School of Medicine at Mount Sinai, for joining us as our plenary speaker. Thank you to all faculty judges who have generously lent us their expertise and time and to whom we are indebted. We extend our sincere thanks to Sandy Muscelli, Jon Eichberger and Betsy Dee for all of their help in organizing and promoting this event. We are especially grateful for the tireless efforts of Jennifer Deal who graciously attended to every detail to bring this day together.

Thank you for participating.

Sincerely,

Priya Duggal, PhD, MPH
Director, MD-GEM
Johns Hopkins Bloomberg School of Public Health

David Valle, MD, PHD
Director, MD-GEM
McKusick-Nathans Institute of Genetics Medicine

Dani Fallin, PhD
Associate Director, MD-GEM
Johns Hopkins Bloomberg School of Public Health

Judy H. Cho, M.D.

DIRECTOR, INSTITUTE FOR PERSONALIZED MEDICINE
SENIOR ASSOCIATE DEAN FOR PRECISION MEDICINE
PROFESSOR | Genetics and Genomic Sciences
PROFESSOR | Medicine, Gastroenterology

Dr. Judy Cho earned her B.A. and M.D. at Ohio State University, where she graduated summa cum laude, and was a member of Phi Beta Kappa and AOA. After postdoctoral training at Northwestern University, she became a faculty member at the University of Chicago. In 2004, she was recruited as Associate Professor to Yale University where she became the Henry J. and Joan W. Binder Professor of Medicine, Genetics and Pediatrics. Dr. Cho was recruited in 2013 to the Icahn School of Medicine at Mount Sinai as the Ward-Coleman Professor of Translational Genetics and Medicine, Vice-Chair of Translational Genetics and Gastroenterology and Director of Ceported. She is currently the Principal Investigator and chair of the Steering Committee of the NIDDK IBD Genetics Consortium and is a member of the NIDDK Advisory Council. She is also on the council of the American Society of Clinical Investigation (ASCI) and is active in the Crohn's and Colitis Foundation and the American Gastroenterology Association, serving on its Research Policy Committee. Dr. Cho has extensive experience in defining genetic factors underlying susceptibility to inflammatory bowel disease (IBD). She was the senior investigator reporting the initial associations of NOD2 to Crohn's disease, the IBD GWAS first identifying the interleukin 23 receptor associations, and most recently, the IBD Immunochip manuscript identifying 163 IBD-assoicated loci. She is particularly interested in defining the genetic architecture underlying the higher IBD prevalence among Ashkenazi Jews. Her laboratory is interested in defining the genetic architecture underlying differentiation of distinct immune cell subsets, differences in epigenetic landscape of immune cell subsets, and their effects in IBD. Her research has been supported by various NIH institutes (NIDDK, NCRR, NIAID, and NIGMS), the Crohn's and Colitis Foundation of America, and The Eli and Edythe L. Broad Foundation and the Burroughs Wellcome Fund. Because Dr. Cho has served as Principal Investigator for the DCC of the 7 center NIDDK IBD Genetics Consortium (IBDGC) for the past twelve years, she also has extensive experience in leading multi-investigator research groups. During the present period of funding, the IBDGC has been charged with developing new collaborations in order to define the functional effects of IBD-associated variants. In this capacity, they have supported a variety of ancillary R01 applications from multi-disciplinary collaborators (epigeneticists, immunologists, systems biologists) in order to fully leverage the extensive IBD genetic discoveries.

Burroughs Wellcome Fund

The *Burroughs Wellcome Fund* is an independent private foundation dedicated to advancing the biomedical sciences by supporting research and other scientific and educational activities. Within this broad mission, BWF has two primary goals:

- To help scientists early in their careers develop as independent investigators
- To advance fields in the basic biomedical sciences that are undervalued or in need of particular encouragement

BWF's financial support is channeled primarily through competitive peer-reviewed award programs. A Board of Directors comprising distinguished scientists and business leaders governs BWF. BWF was founded in 1955 as the corporate foundation of the pharmaceutical firm Burroughs Wellcome Co. In 1993, a generous gift from the Wellcome Trust in the United Kingdom, enabled BWF to become fully independent from the company, which was acquired by Glaxo in 1995. BWF has no affiliation with any corporation.

http://www.bwfund.org/

Maryland Genetics, Epidemiology and Medicine (MD-GEM) Training Program

The *Maryland Genetics, Epidemiology and Medicine (MD-GEM)* is a pre-doctoral training program that comprehensively integrates Genetics, Epidemiology, and Medicine (GEM). Funded by the Burroughs-Wellcome Fund, the MD-GEM training grant brings together the expertise and training infrastructure of the Johns Hopkins Schools of Public Health and Medicine and the National Human Genome Research Institute. Together, these three institutions can provide laboratory, methodological and clinical expertise and coursework to train the next generation of scientists who can forge new avenues of research and address the rapidly changing field of human genetics. This program trains pre-doctoral students through integration of these important areas by partnering with established mentors and offering integrated learning. We envision a training program that will prepare scientists for the next generation of genetics research.

http://www.hopkinsgenetics.org/

MD-GEM Faculty

Priya Duggal, Co-Director
David Valle, Co-Director
M. Daniele Fallin, Associate Director
Dan Arking
Dimitrios Avramopoulos
Joan E. Bailey-Wilson
Terri Beaty
Aravinda Chakravarti
Debra Mathews
Ingo Ruczinski
Diane M. Becker
Lewis Becker
Larry Brody
Nilanjan Chatterjee
Josef Coresh
Jennifer Deal
Hal Dietz
Andrew Feinberg
Gail Geller
Loyal A. Goff
Ada Hamosh
Kasper Hansen
Julie Hoover-Fong
William Isaacs
Lisa Jacobson
Corrine Keet
Alison Klein
Christine Ladd-Acosta
Ben Larman
Jeffrey Leek
Justin Lessler
Brion Maher
Rasika Mathias
Shruti Mehta
Elaine A. Ostrander
Elizabeth A. Platz
Stuart Ray
Debashree Ray
Robert Scharpf
Alan Scott
Margaret Taub
David Thomas
Kala Visvanathan
Jeremy Walston
Xiaobin Wang
Alexander Wilson
Peter Zandi

Poster No.	Presenter	Title
1	M Differding	
2	L Chen	Genome-Wide Association Study of Astrovirus in Bangladeshi Infants with Diarrhea
3	MA Aksit	Common variants in CDKN2B-AS1 are associated with earlier onset of CFRD in females
4	E Bramel	GATA4 is a candidate regulator of angiotensin II signaling in Loeys-Dietz Syndrome
5	A Winters	Association of MALT1 with Peanut Allergy in the Learning Early About Peanut Allergy (LEAP) Study
6	S Clipman	Phylogenetic Evidence for Intercity Clusters of People Who Inject Drugs in India
7	A Kallman	Single-cell transcriptomic analysis of human NRL null organoids
8	R Mitchell	A Comprehensive Evaluation of the Genetic Architecture of Sudden Cardiac Arrest
9	S Robbins	Genes in the HIF-1 Pathway are Associated with Ollier Disease and Maffucci Syndrome
10	G Qi	Mendelian Randomization Analysis Using Mixture Models for Robust and Efficient Estimation of Causal Effects
11	K Weir	The Role of Somatostatin in Cell Type Choice in the Developing Retina
12	K Kanchan	Genomic integrity of human induced pluripotent stem cells across nine studies in the NHLBI NextGen Program
13	H Wick	The role of topoisomerase II beta in the formation of transcriptional hubs in prostate cancer cells
14	KR Iyer	The association between African ancestry and telomere length across the African diaspora: evidence from the CAPPA study
15	V Morrill	Polygenic Risk Scores of Gastrointestinal Symptoms Among Children with and without Autism Spectrum Disorder
16	S Venkataraghavan	Identifying genetic variants shared by orofacial cleft subtypes in GWAS of case-parent trios from POFC and GENEVA studies
17	RJ Longchamps	Mitochondrial DNA Heteroplasmy is Associated with Overall Mortality
18	G Pergola	From Gene Co-expression to Brain Circuits: Prefrontal Gene Co-expression Networks Predict Drug Response in Patients with Schizophrenia
19	A Narkar	Understanding the TP53 signaling cascade that regulates cell fate post-acute aneuploidy induction
20	JR Wang	Facilitating rapid precision oncology in anaplastic thyroid cancer: Clinical implications of next generation sequencing (NGS) mutation testing and impact on survival
21	D Duchen	
22	W Zhang	Identification and comparison of imputed and genotyped variants for genome wide association study of orofacial clefts disease
23	W Kim	Genome-wide gene-smoking interaction on COPD in UK Biobank study
24	V Busa	
25	S McClymont	Dopaminergic neuronal chromatin signatures reveal Parkinson Disease associated variation in a novel aminergic intronic enhancer at SNCA
26	A Liu	Methods for genetic analysis of secondary phenotypes in case-control GWAS with application to COPDGene data

27	V Pillalamarri	Transcriptome-wide association study of ASD genetic risk in the general population
28	A Price	
29	Z Yu	Kidney Function and Blood Pressure: A Mendelian Randomization Study
30	S Mohazzab-Hosseinian	The Epigenetic Developmental Clock and its prospective association with ASD in childhood
31	L Boukas	Co-expression patterns define chromatin regulators associated with neuropsychiatric phenotypes
32	C Wright	miRNA Isoform Quantification Method Development and Expression Analysis in Schizophrenia
33	T Nieuwenhuis	Highly Specific, Highly Expressed Genes Contaminate the GTEx Dataset
34	C Valencia	
35	O Sniezek	The Role of HMGA1 in Cellular Aging, Senescence, and Cancer

Genome-Wide Association Study of Astrovirus in Bangladeshi Infants with Diarrhea

Laura Chen[1], Dylan Duchen[1], Genevieve Wojcik[2], Rashidul Haque[3], William A. Petri Jr[4] and Priya Duggal[1]

1 Johns Hopkins Bloomberg School of Public Health
2 Stanford University
3 International Center for Diarrhoeal Disease Research
4 University of Vermont

Presented by Laura Chen

Background: Astroviral infections are a common cause of acute nonbacterial gastroenteritis in children globally. However, there is no treatment available for Astrovirus, and huge strain diversity within Astroviridae presents a challenge to potential vaccine development.

Objectives: To identify host genetic risk factors associated with Astrovirus disease susceptibility

Methods: We performed a genome-wide association study of children with and without Astrovirus in the first year of life. Children were enrolled in one or two birth cohorts in Dhaka, Bangladesh: the Performance of Rotavirus and Oral Polio Vaccines in Developing Countries (PROVIDE) study or the Cryptosporidiosis and Enteropathogens in Bangladesh (CRYPTO) study. Each study was independently analyzed and then a meta-analysis was performed.

Results: From PROVIDE, complete genotypic and phenotypic data was available for 570 children. Using qPCR data, we determined a case definition of diarrheal event attributable to Astrovirus as Ct<30, resulting in 119 Astrovirus cases and 314 non-Astrovirus controls. The average age of children with Astrovirus is 108.97 days. A region on Chromosome 1 near the loricrin gene (*LOR*) was associated with higher Astroviral load in stool for individuals who carried 1 copy of the T:A allele at SNP rs75437404 (OR=2.598, p-value= 1.78×10^{-8}). Similarly, we identified a region on Chromosome 10 near the prolactin releasing hormone gene (*PRLHR*) that was also associated with higher Astroviral load in stool for individuals who carried 1 copy of the C:T allele at SNP rs72839103 (OR=5.335, p-value= 9.10×10^{-8}). For the CRYPTO study, we used the same case definition and this resulted in 58 cases and 96 controls. The average age of children with Astrovirus is 117.74 days. There were no significant independent findings in the CRYPTO study. The meta-analysis confirmed the significance of the *LOR* region (SNP rs75437404, p-value= 9.384×10^{-9}) and *PRLHR* region (rs115109981, p-value= 1.27×10^{-8}).

Conclusions: Highly significant SNPs in the *LOR* region on Chromosome 1 suggest that the *LOR* gene may play a role in susceptibility to infection by Astrovirus. Loricrin contributes to the protective barrier function of the stratum corneum of the epidermis. Highest levels of *LOR* are expressed in humid tissues, including newborn epidermis, the epithelia of oral and anal mucosa, and the esophagus. Previous studies suggest that its expression in non-keritinizing epithelia represents a protective mechanism of the body. *PRLHR* encodes a transmembrane domain receptor for prolactin-releasing hormone most commonly expressed in adrenal, endometrial, and brain tissue. The substrate of this receptor, prolactin-

releasing peptide, has been shown in mouse models to influence feeding patterns and energy balance. Specific analogs have been identified as potential candidates for antiobesity treatment. However, weight-by-age Z scores and height-by-age Z scores did not differ widely across case/control status and genotype for *LOR* and *PRLHR* alleles.

Content Area: Genetic Epidemiology
Keywords: Diarrhea, Astrovirus, genome-wide association study

Common variants in CDKN2B-AS1 are associated with earlier onset of CFRD in females

Rhonda G Pace, Briana Vecchio-Pagan, Hua Ling, Johanna M Rommens, Pierre-Yves Boelle, Loic Guillot Karen S Raraigh, Elizabeth Pugh, Peng Zhang, Lisa J Strug, Mitch L Drumm, Michael R Knowles, Garry R Cutting, Harriet Corvol, Scott M Blackman

Presented by Melis Atalar Aksit

Cystic fibrosis (CF) is an autosomal recessive disease caused by mutations in CFTR. CF-related diabetes (CFRD) is one of the most important complications of CF, affecting 19% of adolescents and 40–50% of adults with CF, and causes worse lung function, malnutrition and mortality. Wide variation in CFRD onset in people with severe CF was found to be heritable, i.e., attributable to genetic variation outside of CFTR (ref:1), and it develops earlier in females (Hazard ratio: 1.57). A combined genome-wide and candidate-based association study identified CFRD modifier variants intronic and 5' of SLC26A9, and in type 2 diabetes loci (TCF7L2, CDKAL1, CDKN2B-AS1 and IGF2BP2) (ref:2).
Since CFRD occurs earlier in females, we investigated sex-specific associations. Variants at CDKN2B-AS1 were associated with earlier onset of CFRD in females (e.g. rs1333045 p:2.2e-7, HR:1.46, 95% CI:1.27-1.69, N=3,040) but not males (e.g., rs1333045; male-only analysis p: 0.18 HR: 1.11, 95% CI:0.95-1.30, N=2,700). Females of C/C or C/T genotypes developed CFRD at significantly earlier ages than males (log rank p-values 5.04e-8 and 6.55e-14, respectively), while females with the T/T genotype had a later onset of CFRD which did not differ from that of males (p-value: 0.317). In a combined analysis of males and females, the sex*SNP interaction term was statistically significant (e.g., rs1333045 p-value: 7.3e-3; HR: 1.37, N=5,740), indicating that the association effect size differs significantly between males and females.
CDKN2B-AS1 is a well-studied locus that spans >50kb, and harbors at least 2 independent LD blocks. Different variants at this locus have been shown to influence many diseases including T2D, coronary artery disease (CAD), and many types of cancer (ref: 3). Notably, the CFRD-associated variants at CDKN2A/B (top variant: rs1333045; p-value: 2.3e-5; HR: 1.28; located in intron 17 out of 18 of CDKN2B-AS1) are the variants most significantly associated with CAD in the general population; and conditioning on rs1333045 abolishes the signal for association with CFRD. The variants most strongly associated with T2D at this locus (rs10965250; located 12kb 3' of CDKN2B-AS1) are not associated with CFRD (one-sided p-value: 0.9921; HR: 0.82), and conditioning on rs10965250 does not affect the association of rs1333045 with CFRD. The T2D association at both the rs1333045 and rs10965250 loci are not sex-specific in a European population (ref: 4), however variants at the rs10965250 locus have been reported to have a female-specific T2D association in North African Tunisian Arabs (ref 5). No study reported a sex-specific CAD association at this locus.
In conclusion, we identified a locus that is associated with earlier onset of CFRD in females, but does not affect CFRD in males. This could be one of the underlying mechanisms of the earlier onset of CFRD in females.

(1) Blackman, et al. JCEM. 2009.
(2) Blackman, et al., Diabetes. 2013.
(3) Hannou SA, et al. Trends Endocrinol Metab. 2015
(4) Morris, et al. Nat Genet. 2012.
(5) Turki, et al. Diabetes Res Clin Pract. 2014.

Content Area: Human Genetics
Keywords: Cystic Fibrosis, GWAS, CDKN2B-AS1, Diabetes

GATA4 is a candidate regulator of angiotensin II signaling in Loeys-Dietz Syndrome

Emily Bramel[1, 2],Tyler J. Creamer[2,3], Rustam Bagirzadeh[2,3] and Elena Gallo MacFarlane[2,3]

1 Predoctoral Training Program in Human Genetics and Molecular Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
2 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
3 Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Presented by Emily Bramel

Loeys-Dietz syndrome (LDS) is an autosomal dominant connective tissue disorder that affects multiple tissues including the bones and blood vessels. The most serious health risk in LDS is related to dilation and weakening of the vessel wall, with the aortic root being a site of high risk for both aneurysm and tear, which often have fatal consequences. Currently, the only effective treatment to prevent dissection is prophylactic surgery. LDS is caused by mutations that impair but do not completely abolish transforming growth factor-β (TGF-β) signaling, including heterozygous loss-of-function missense mutations in TGFBR1 and TGFBR2. At sites of aneurysm in the aortic root, these mutations have been shown to lead, indirectly, to increased activation of downstream mediators of this pathway in both patients and mouse models. The regional predisposition of the aortic root to both aneurysm and signaling alterations, despite the fact that causative mutations are expressed ubiquitously throughout the arterial tree, has been proposed to depend on the lineage-of-origin of the smooth muscle cells populating this region. The aortic root is composed primarily of secondary heart field-derived (SHF) vascular smooth cells (VSMCs), while the more distal, ascending aorta is composed of cardiac neural crest-derived (CNC) VSMCs. Previous work performed in LDS mouse models has shown that SHF-VSMCs are intrinsically more sensitive to the effects of an LDS-causing mutation and that this vulnerability associates with increased levels of Agtr1a, which encodes for angiotensin II (AngII) receptor I (AT1R), a receptor known to be critical to aneurysm pathogenesis in LDS and related disorders. We performed a gene expression analysis aimed at identifying transcripts differentially regulated in a lineage- and genotype-specific manner and found that LDS SHF-VSMCs express higher levels of the transcription factor Gata4, relative to both control SHF-VSMCs and CNC-VSMCs of either genotype. Differential expression of Gata4 mRNA corresponds to differences in protein levels, as assessed both in vitro and in vivo. A meta-analysis on existing GATA4 ChIP-Seq data shows that GATA4 binds putative Agtr1a cis-regulatory regions, including a candidate promoter and intronic enhancer region. To interrogate whether GATA4 binds to these regions in control and LDS SHF-VSMCs, we used chromatin immunoprecipitation (ChIP) to isolate GATA4-bound DNA and assessed the presence of candidate Agtr1a regulatory DNA sequences by qPCR. Our preliminary results show that the candidate intronic enhancer region for Agtr1a is enriched in LDS compared to control SHF-VSMCs. To functionally test whether GATA4 is necessary for Agtr1a upregulation, we also attempted to examine the effect of Gata4 RNAi-knock-down on Agtr1a expression. Surprisingly, while Gata4 mRNA levels were successfully reduced by RNAi, GATA4 protein levels increased as assessed with two unrelated antibodies. This paradoxical increase in GATA4 protein levels was also accompanied by increased Agtr1a mRNA levels. Taken together, these preliminary results suggest that GATA4 binds to Agtr1a regulatory elements and that this factor may play an important role in the development of aortic root aneurysms in LDS patients. Further investigation will be necessary to elucidate the functional consequences of GATA4 inactivation and the mechanisms by which its levels are regulated.

Content Area: Human Genetics
Keywords: Aortic aneurysm, GATA4, Loeys-Dietz Syndrome, TGF-β, Angiotensin

Association of MALT1 with Peanut Allergy in the Learning Early About Peanut Allergy (LEAP) Study

Alexandra Winter[1], Henry T Bahnso[2,3], Ingo Ruczinski[4], Meher P Boorgula[5], Claire Malley[1], Ali R Keramati[1], Sameer Chavan[5], David Larson[3], Karen Cerosaletti[2], Peter H Sayre[3,6], Marshall Plautm[7], George Du Toit[8], Gideon Lack[8], Kathleen C Barnes[5], Gerald T Nepom[3] and Rasika A Mathias[1]

1 Department of Medicine, Johns Hopkins University, Baltimore, MD, USA
2 Benaroya Research Institute, Seattle, WA, USA
3 Immune Tolerance Network, San Francisco, CA, USA
4 Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
5 Department of Medicine, University of Colorado, Anschutz, CO, USA
6 Department of Medicine, University of California, San Francisco, CA, USA
7 Allergy, Asthma and Airway Biology Branch, DAIT, NIAID, Bethesda, MD, USA
8 Department of Pediatric Allergy, Division of Asthma, Allergy and Lung Biology, King's College London, and Guy's and St. Thomas' NHS Foundation Trust, London, United Kingdom

Presented by Alexandra Winters

The Learning Early About Peanut Allergy (LEAP) study motivated a change in pediatric guidelines, as dietary introduction of peanut protein beginning in the first four to eleven months of life was shown to significantly decrease the frequency of peanut allergy later in childhood and modulate the immune response to peanuts among those at high risk of peanut allergy. Whole genome sequencing was performed to identify the genetic determinants of peanut allergy in the LEAP participants. We identified a strong association between peanut allergy and the MALT1 locus on chromosome 18 (rs57265082, OR = 10.99, p = 6.49x10-8) in the peanut avoidance group of LEAP participants (N=275). In addition, gene-based analysis of MALT1 showed a very strong association with p = 1.89x10-10. MALT1 carriers in the peanut avoidance group had the highest levels of peanut-specific IgE and risk for peanut allergy, with 58.6% carriers in the peanut avoidance group developing peanut allergy at 60 months as compared to 12.7% of non-carriers in the same group. We also demonstrated that the intervention of early peanut exposure works equally effectively to prevent development of peanut allergy in both carriers and non-carriers of the MALT1 risk allele. The MALT1 gene encodes a paracaspase that functions as a critical part of the CARMA1-BCL10-MALT1 (CBM) complex, causing NF-KappaB activation in response to antigen binding to the receptor and leading to T cell activation. It is possible that variants in MALT1 may predispose an individual to greater allergic disease by altering MALT1 expression or affecting the ratio of the two MALT1 isoforms, thus increasing Th2 differentiation after antigen presentation.

Content Area: Genetic Epidemiology, Human Genetics
Keywords: Peanut allergy, MALT1, GWAS, food allergy, immunogenetics

Phylogenetic Evidence for Intercity Clusters of People Who Inject Drugs in India

Steven J. Clipman[1], Mary Rodgers[2], Shanmugam Saravanan[3], Shruti H. Mehta[1], Aylur K. Srikrishnan[3],M. Suresh Kumar[3],Allison M. McFall[1], Gregory M. Lucas[4], Thomas C. Quinn[4,5], Gavin Cloherty[2] and Sunil S. Solomon[1,3,4]

1 Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
2 Abbott Diagnostics, Infectious Disease Research, Abbott Park, IL
3 YR Gaitonde Centre for AIDS Research and Education, Chennai, India
4 Johns Hopkins University School of Medicine, Baltimore, MD
5 National Institute of Allergy and Infectious Diseases, Bethesda, MD

Presented by Steven Clipman

Background: Little data exist on HCV phylodynamics and transmission networks among people who inject drugs (PWID), especially from low- and middle-income countries (LMICs). HCV epidemics in a city can be considered a series of sub-epidemics caused by phylogenetically distinct viral lineages. Mapping these lineages to generate transmission clusters and overlaying epidemiologic data can be used to identify factors associated with clustering.

Methods: PWID were recruited via respondent driven sampling in 2016-17. Participants completed a survey and blood draw. HCV 5'UTR-core sequencing was performed on 486 HCV RNA positive samples from 4 cities (Amritsar [n=126], Delhi [n=128], Kanpur [n=138], Imphal [n=94]). Sequences were aligned using Multiple Sequence Comparison by Log-Expectation. The most appropriate nucleotide substitution model was determined using jModelTest and phylogenetic inference was carried out using Maximum Likelihood methods in RaXML with 500 bootstrap replications. Clusters were identified using ClusterPicker with posterior support and genetic distance thresholds of 70% and 4.5%, respectively. Given the large number of covariates of interest, a machine learning model utilizing the Boruta wrapper of the random forest algorithm was constructed to identify features predictive of clustering, as well as differences between clusters.

Results: Median age was 33 years, 99% were male and HIV prevalence was 75%. Mean p-distance for all sequences was 0.075. A total of 251 sequences fell into 19 transmission clusters (Fig). Mean cluster size was 7.4 (range: 2-49); 8 clusters were dyads. There were 6 large clusters comprised of > 10 samples. 7 of the 19 clusters contained samples from multiple cities. Machine learning based analysis revealed that no history of HIV testing and living with friends were predictive of clustering (both $p<0.05$), and that state, residential zip code, injection zip code, time spent away from home, and buprenorphine injection could be predictive of membership in a given cluster (all $p<0.05$). Age, gender, and HIV status did not predict clustering.

Conclusions: These are among the first data from a LMIC setting to demonstrate clustering across multiple cities. The median size of the clusters identified were also larger than self-reported injection networks in India. Treatment as prevention efforts for HCV have emphasized network-based approaches for PWID, and these data suggest that networks may need to be defined by space (zip code) as opposed to egocentric injection networks.

Content Area: Pathogen Genetics
Keywords: phylogenetics, machine learning, HCV, PWID, India

Single-cell transcriptomic analysis of human NRL null organoids

Alyssa Kallman[1], Elizabeth Capowski[2], Anu Kaushik[3], Cynthia Berlinicke[1], Jie Wang[1], Jiang Qian[1]
Tza-Huei Wang[3], David Gamm[2] and Don Zack[1]

1 JHUSOM
2 University of Wisconsin
3 JHU

Presented by Alyssa Kallman

Retinal degenerative diseases remain largely untreatable, partly due to limited understanding of the complex gene-expression patterns that direct photoreceptor (PR) development. By performing single-cell transcriptomic analysis of human in vitro models of retinal development in the presence and absence of Neural Retina-specific Leucine zipper protein (NRL), a transcription factor vital to rod PR development and degeneration, we are working to systematically define the gene networks involved in PR differentiation and degeneration.

Human induced pluripotent stem cell control and NRL null lines were differentiated into retinal organoids and samples were collected at 100 and 170 days for single cell mRNA capture and analysis using Dropseq. We analyzed the transcriptomic data by principle component analysis and t-distributed stochastic neighbor embedding (t-SNE) to cluster cells by variably expressed genes. Cell populations were identified by marker genes and PRs were selected to reconstruct developmental trajectories. While control organoids had both rod and cone PRs, NRL null organoids had only cone PRs, with S-cones as the dominant subtype. 5144 control and 8173 NRL null PRs were subsetted from the rest of the retinal cell types in order to reconstruct the PR developmental trajectory specifically. The combined PR trajectory showed various possible cell fates, including two branches that contain control rods and S-opsin expressing NRL null PRs. Compared to the PRs along the cone developmental branch, these NRL null PRs have significantly lower expression of various cone-specific transducin and phosphodiesterase genes ($p<10^{-8}$). Consistent with the function of NRL in directing PRs towards a rod fate, retinal organoids lacking NRL developed cone-dominant PR populations. A subset of these cones are bioinformatically distinguishable due to their aberrant expression of transducin and phosphodiesterase genes, causing them to cluster with rod photoreceptors. Further analysis of these PR populations will better define the regulatory networks involved in PR differentiation and cell fate specification.

Content Area: Human Genetics
Keywords: organoids, scRNAseq, photoreceptors

A Comprehensive Evaluation of the Genetic Architecture of Sudden Cardiac Arrest

Mitchell R[1], Ashar F[1,2], Arking DE[1] and Sotoodehnia N[3]

1 Institute of Genetic Medicine, Johns Hopkins, Baltimore, USA
2 SCD working group of the CHARGE Consortium
3 Cardiovascular Health Research Unit, Division of Cardiology, Departments of Medicine and Epidemiology, University of Washington

Presented by Rebecca Mitchell

Aims. Sudden cardiac arrest (SCA) accounts for 10% of adult mortality in Western populations. We aim to identify potential loci associated with SCA and to identify risk factors causally associated with SCA.

Methods and Results. We carried out a large genome-wide association study (GWAS) for SCA (n=3,939 cases, 25,989 non-cases) to examine common variation genome-wide and in candidate arrhythmia genes. We also exploited Mendelian randomization methods using cross-trait multi-variant genetic risk score associations (GRSA) to assess causal relationships of 18 risk factors with SCA. No variants were associated with SCA at genome-wide significance, nor were common variants in candidate arrhythmia genes associated with SCA at nominal significance. Using cross-trait GRSA, we established genetic correlation between SCA and (1) coronary artery disease (CAD) and traditional CAD risk factors (blood pressure, lipids, and diabetes), (2) height and BMI, and (3) electrical instability traits (QT and atrial fibrillation), suggesting etiologic roles for these traits in SCA risk.

Conclusions. Our findings show that a comprehensive approach to the genetic architecture of SCA can shed light on the determinants of a complex life-threatening condition with multiple influencing factors in the general population. The results of this genetic analysis, both positive and negative findings, have implications for evaluating the genetic architecture of patients with a family history of SCA, and for efforts to prevent SCA in high-risk populations and the general community.

Content Area: Human Genetics
Keywords: Sudden cardiac arrest, genome-wide association study, mendelian randomization

Genes in the HIF-1 Pathway are Associated with Ollier Disease and Maffucci Syndrome

Sarah Robbins[1,2], Renan Martin[1,2], Olivia Sniezek[1,2], Yanzi Xiao[3,4], Terri Beaty[3], David Valle[1,2] and Nara Sobreira[1,2]

1 McKusick-Nathans Institute of Genetic Medicine
2 Johns Hopkins School of Medicine
3 Johns Hopkins Bloomberg School of Public Health
4 NIH, National Cancer Institute

Presented by Sarah Robbins

Ollier disease (OD) is a rare disorder characterized by multiple enchondromas. Maffucci syndrome (MS) is a related disorder characterized by enchondromas and vascular overgrowth. Patients with OD or MS are also at increased risk for cancer, mainly, chondrosarcomas and gliomas. Amary et al., (2011) and Pansuriya et al., (2011) identified specific gain-of-function somatic mutations in IDH1 or IDH2 in 80% of enchondromas, vascular anomalies, and chondrosarcomas isolated from these patients. We hypothesize that germline variants in these patients predispose them to tumor development. We performed exome sequencing of 25 probands with OD or MS and show that 7 probands have rare missense variants in 4 genes in the HIF-1 pathway: KDM4C (1), HIF1A (1), VHL (4), and EGLN1 (2). We hypothesize that dysregulated HIF-1 pathway activity is responsible for the development of enchondromas. To explore this hypothesis and identify other perturbed pathways, we performed RNA-seq of multiple tissues from one patient with a heterozygous variant in VHL. We confirmed that the VHL variant is present at heterozygous levels (45-70%) in RNA from all tissues tested: an enchondroma, a vascular anomaly, an exostosis, and unaffected skin fibroblasts. We compared the vascular anomaly to 3 control fibroblast lines and found that expression of glyoxylate cycle genes, IDH1 and IDH2, was significantly dysregulated, as well as expression of genes regulating eIF2-alpha dephosphorylation and ubiquitin ligase activity. On comparison of the enchondroma to 3 technical replicates of a control chondrocyte line, we found significant dysregulation of cartilage development and chondrocyte differentiation genes, in addition to genes related to pyridoxal phosphate metabolism and cyanate metabolism. We also performed RNA-seq of patient fibroblasts (4 patients including the patient with the KDM4C variants and the patient with the VHL variant described above) and controls (3) cultured at normoxia. Chondrocyte differentiation and cartilage development were the most overrepresented PANTHER GO terms in our list of differentially expressed genes (n=50), driven by abnormal expression of GDF6 (p=0.00723), PKDCC (p=0.00214), and HOXA11 (p=0.04). Under hypoxic conditions, these terms plus vascular development terms were also overrepresented in our differentially expressed list (n=109). We believe our analysis will give us insight into the mechanism behind HIF-1 pathogenesis in OD and MS and reveal novel candidate genes to be further investigated.

Content Area: Human Genetics
Keywords: Ollier disease, Maffucci syndrome, RNA-sequencing, exome sequencing, cancer

Mendelian Randomization Analysis Using Mixture Models for Robust and Efficient Estimation of Causal Effects

Guanghao Qi[1] and Nilanjan Chatterjee[1,2]

1 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
2 Department of Oncology, Johns Hopkins School of Medicine

Presented by Guanghao Qi

Mendelian Randomization (MR) has emerged as a major tool for the investigation of causal relationship among traits, utilizing results from large-scale genome-wide association studies. Bias due to "horizontal pleiotropy", however, remains a major concern. We propose a novel approach to MR analysis using potentially large number of genetic instruments based on a model for bivariate effect-size distribution in the form of normal-mixtures. The model assumes existence of a fraction of the genetic markers that are valid instruments, i.e. they have only direct effect on one trait, while other markers can have potentially correlated, direct and indirect effects, or have no effects at all. We propose estimating causal effect ($\boldsymbol{\theta}$) through a procedure for maximizing the probability concentration of the residuals, ($\boldsymbol{\beta y} - \boldsymbol{\theta \beta x}$), at the "null" component of a two-component normal mixture model. We further extend the method to conduct multivariable MR analysis (MV-MRMix) by incorporating the known heritable confounder (U) into the three-trait residual $\boldsymbol{\beta y} - \boldsymbol{\theta \beta x} - \boldsymbol{\theta u \beta u}$. Simulation studies showed that MRMix provides nearly unbiased or/and substantially less biased estimates of causal effects compared to alternative methods under various scenarios, including when the underlying model for effect-size distribution is not correct. Further, MRMix is sensitive to direction and can achieve much higher efficiency (up to 3-4 fold) relative to other comparably robust estimators. MV-MRMix further improves robustness and efficiency on top of univariate MRMix, with the difference diminishing with increasing sample size. We applied MRMix for investigating causal relationship across a number of risk-factors and health outcomes using publicly available datasets. Notable observations included identification of causal effects of genetically determined BMI and age-at-menarche, which have relationship among themselves, on the risk of breast cancer; no causal effect of HDL and triglycerides on the risk of coronary artery disease; a strong detrimental effect of BMI, but no causal effect of years of education, on the risk of major depressive disorder.

Content Area: Statistical Genetics
Keywords: Mendelian randomization, causal effects, mixture models, MRMix

The Role of Somatostatin in Cell Type Choice in the Developing Retina

Kurt Weir[1] and Seth Blackshaw, PhD[1]

1 Solomon H. Snyder Department of Neuroscience Johns Hopkins School of Medicine

Presented by Kurt Weir

Cell-extrinsic signals are important regulators of retinal development, though their identity is still largely unknown. Single-cell RNA-seq analysis of developing mouse and human retinas performed in the Blackshaw lab has identified a number of strong candidate factors for this role, including the neuropeptide somatostatin. Somatostatin receptor type 2 (SSTR2), a GPCR that interacts with the Gi protein, is highly expressed in retinal progenitor cells at the same time that somatostatin is highly expressed by retinal ganglion cells. It is thus likely that somatostatin exerts its influence through SSTR2 in progenitor cells; downregulating adenylyl cyclase and the PKA pathway.
I test the hypothesis that somatostatin acts as a quorum signal that serves to promote or inhibit retinal progenitor cell proliferation, neurogenesis, and/or production of different retinal cell types using mouse embryonic retinal explants.

Retinal explants were grown from embryonic day 14 to postnatal day 0 in the presence of either a SSTR2 agonist or Forskolin, an activator of adenylyl cyclase, at saturating concentration or as a control grown in its absence. The expression of cell type marker genes was assayed by qPCR and immunostaining.

SSTR2 agonist-treated explants show a dose-responsive decrease in photoreceptor gene expression relative to controls while Forskolin-treated explants show an increase in photoreceptor gene expression.

Somatostatin appears to act as a quorum signal to inhibit production of photoreceptor cell types by downregulating the activity of adenylyl cyclase in retinal progenitor cells.

Content Area: Molecular Genetics
Keywords: retina, development, neuropeptide

Genomic integrity of human induced pluripotent stem cells across nine studies in the NHLBI NextGen Program

Kanika Kanchan[1], Kruthika Iyer[1], Lisa R Yanek[1],Margaret A Taub[2], Claire Malley[1], Kristin Baldwin[3], Lewis C Becker[1], Ulrich Broeckel[4], Linzhao Cheng[1], Chad Cowan[5], Matteo D'Antonio[6], Kelly A Frazer[6], Ivan Carcamo-Orive[7], Joshua W Knowles[7], Thomas Quertermous[7], Gustavo Mostoslavsky[8], George Murphy[8], Marlene Rabinovitch[9], Daniel J Rader[10], Martin H Steinberg[11], Eric Topol[12], Wenli Yang[13], Cashell E Jaquish[14], Ingo Ruczinski[2] and Rasika A Mathias[1]

1 Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA
2 Department of Biostatistics, Bloomberg School of Public health, Johns Hopkins University, Baltimore, MD, USA
3 Department of Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA, USA
4 Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, USA
5 Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA
6 Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA
7 Stanford University School of Medicine, Stanford Center for Inherited Cardiovascular Disease, Stanford University, Stanford, CA, USA
8 The Center for Regenerative Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA
9 Department of Pediatrics, Stanford University, Stanford, CA, USA
10 Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA,USA
11 Department of Medicine, Section of Hematology-Oncology, Boston University School of Medicine, Boston, MA, USA
12 Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA
13 Penn Center for Pulmonary Biology and Institute for Regenerative Medicine, University of Pennsylvania, Philadelphia, PA, USA
14 National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA

Presented by Kanika Kanchan

Prior studies suggest that human induced pluripotent stem cell (hiPSC) lines suffer from genomic instabilities (e.g. karyotypic abnormalities, chromosomal aberrations). Within the NHLBI sponsored NextGen program, hiPSC lines were generated from a range of sources (PBMCs, fibroblasts and lung epithelium) at nine sites to study the complex genetics of metabolic, cardiovascular and respiratory diseases. The objective of this work is to assess the genomic integrity of the NextGen hiPSCs. We examined the integrity of hiPSC lines by comparing to their matched donor DNA from blood/PBMCs leveraging 1.3 million genetic variants from the Illumina Multiethnic Genotyping (MEGA) array. Two levels of genomic integrity were investigated using: (1) genotype discordance between 1,060 parent donor DNA and hiPSC lines; and (2) structural variability of 506 hiPSC lines using copy number variants (CNVs). We detected low rates of genotype discordance (median = 0.002%) between the donor and hiPSC lines; and observed that 149 hiPSC lines acquired 258 CNVs relative to the donor DNA. The cumulative impact per hiPSC line was small; 85% showing less than 2 Mb of cumulative CNV coverage. Furthermore, we identified six recurrent regions of CNVs on chromosomes 1, 2, 3, 16 and 20 that overlapped with cancer associated genes. In general, hiPSCs acquired deletions included tumor suppressor genes whereas duplications included oncogenes. Overall, a low level of genomic instability was observed in the NextGen generated hiPSC lines as compared to prior reports. However, given the observation of structural instability in regions with known cancer associated genes, our results substantiate the importance of systematic evaluation of genetic variations in hiPSCs before using them as disease/research models.

Content Area: Human Genetics
Keywords: hiPSCs, CNV, Genomic integrity, GWAS, chromosomal aberrations

The role of topoisomerase II beta in the formation of transcriptional hubs in prostate cancer cells

Heather C Wick[1]*, Michael Haffner[2,3]*, David Esopi[2], William Nelson[2,3], Srinivasan Yegnasubramanian[2,3]** and Sarah Wheelan[1,2]**

*contributed equally; **co-mentors

1 Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
2 Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
3 Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Presented by Heather Wick

Cellular transcriptional programs requiring changes in expression of multiple genes may be more efficient when the target genes are brought into physical proximity by chromatin conformational changes. We hypothesize that topoisomerases, which induce transient single and double stranded breaks in DNA to relieve topological constraints, are required to facilitate the formation of such transcriptional hubs upon stimulation of transcriptional programs. Prior work from the lab has shown that one of these topoisomerases, topoisomerase II beta (TOP2B), binds and exerts catalytic activity at the genes TMPRSS2, an androgen receptor (AR) target gene, and ERG, an Ets transcription factor oncogene, recurrently fused in prostate cancer. Illegitimate repair of TOP2B-mediated double stranded DNA breaks incurred during the induction of androgen receptor signaling may contribute to gene fusions and other genomic rearrangements commonly seen in prostate cancer. Understanding the effects of AR and TOP2B binding on gene expression and chromosomal conformation in prostate cancer cells may give us insight about the formation and progression of prostate cancer.

The role of TOP2B in the coordination of transcriptional hubs was studied in LNCaP prostate cancer cells stimulated with and without the AR agonist dihydrotestosterone (DHT). The phenomena of interest are the multifaceted and coordinated cellular changes induced by DHT, requiring investigation of multiple processes: gene expression was measured with RNA-seq; AR and TOP2B binding were measured with ChIP-seq; and spatial genome interactions were measured with Hi-C. Integration of this data together with publicly available ChIP-seq data of RNA Polymerase II binding and occupancy of histone marks was used to create a more complete picture of the effect of TOP2B and AR on the formation of these androgen-induced transcriptional hubs.

Content Area: Human Genetics
Keywords: topoisomerase, transcriptional hub, prostate cancer, gene fusion, HiC

The association between African ancestry and telomere length across the African diaspora: evidence from the CAPPA study

Kruthika R. Iyer[1], Margaret A. Taub[2], Terri H. Beaty[1] and Rasika A. Mathias[1,3]

1 Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD
2 Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD
3 Department of Medicine, Johns Hopkins University, Baltimore, MD

Presented by Kruthika Raman Iyer

Telomeres are aglets on the end of our chromosomes that can be used as a measure of biological aging. Recent studies suggest telomere length (TL) is influenced by ancestry; African Americans on average have longer telomeres compared to European Americans. However, to our knowledge, the correlation between the proportion of African ancestry and TL has not been explored. We used existing whole genome sequence (WGS) data from a multi-centered case-control study - Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) to explore mechanisms influencing TL.

CAAPA recruited N=917 African-admixed individuals from North, Central, South American plus Caribbean populations in addition to Yoruba-speaking individuals from Ibadan, Nigeria. This geographic spectrum is clearly reflected in the dynamics of their genetic admixture. Global estimates of ancestry obtained using ADMIXTURE illustrate a high variability over a wide range of proportion African ancestry (%YRI) [range: 0.01% - 99.9%; mean(sd): 78.2% (19.6%)]. The age range of these individuals was also considerable [range: 6yrs – 89yrs; mean(sd): 32yrs (16 yrs)]. TL on each individual was estimated using TelSeq, a computational approach that leverages WGS data. This algorithm counted the number of reads with at least 12 contiguous repeats of the telomere-identifying hexamer TTAGGG in each genome [range: 2.89kb – 10.25kb; mean(sd): 5.75kb(1.28kb)]. We found WGS derived TL was negatively correlated with age ($r = -0.51$; $p<2.26\text{-}16$), and did not significantly differ between males and females sex [meanMALES(seMALES): 5.79kb (0.06kb); meanFEMALES(seFEMALES): 5.83kb(0.06kb); $p=0.85$]. We found CAAPA sampling site - Honduras from Central America had the highest TL average of 7.35kb followed by Nigerians who had an average TL of 6.77kb. This could plausibly be driven by two factors (1) age and (2) %YRI ancestry. Although Nigerians had high %YRI [mean: 99.2% vs. 80.2%], individuals from Central America were relatively younger [mean: 11yrs vs. 13yrs; median: 10yrs vs. 13yrs; mode: 10yrs vs. 11yrs]. We further assessed the nature of the relationship between and TL and %YRI across the African diaspora using linear regression models after adjusting for %NAT ancestry, age and sex, and found %YRI ancestry was significantly associated with TL [estimated beta(se): 1.09kb(0.29kb); $p=0.000198$].

As CAAPA genomes were collected across 16 different sites, we are also examining the possibility of technical sources of variation that may have been inevitably introduced in our TL data, and could confound any association with %YRI. In the near future, we aim to replicate 11 previously implicated loci using a candidate gene study design while concurrently probing for novel variants affecting TL in individuals of African descent from the CAAPA study.

Content Area: Computational Genetics, Genetic Epidemiology
Keywords: Telomere Length, Whole Genome Sequencing, African Ancestry

Polygenic Risk Scores of Gastrointestinal Symptoms Among Children with and without Autism Spectrum Disorder

Valerie Morrill[1], Laura Schieve[2], Norbert Soke[2], John Brinton[3], Dani Fallin[4,5], Ann Reynolds[3] and Christine Ladd-Acosta[1,5]

1 Department of Epidemiology, Johns Hopkins
2 National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention
3 Department of Developmental Pediatrics, University of Colorado
4 Department of Mental Health, Johns Hopkins
5 The Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins University, Baltimore MD

Presented by Valerie Morrill

Children with Autism Spectrum Disorder (ASD) have a greater prevalence of gastrointestinal (GI) illnesses and symptoms compared to children without ASD. Understanding the mechanisms that lead to GI symptoms among ASD patients can guide intervention and treatment strategies. Two main theories have been postulated to explain the increased prevalence of GI symptoms. First, increased GI symptoms could be a consequence of having ASD and is likely due to restricted and repetitive food interests, i.e. is behavioral in nature. A second theory is that children with ASD have more intrinsic biologic risk factors (e.g. genetic or microbiome) for GI symptoms that children without ASD. The purpose of this study was to investigate genetic mechanisms of gastrointestinal symptoms among children with and without ASD to provide evidence supporting or refuting the second hypothesis. Our analyses included a subset of individuals enrolled in the Study to Explore Early Development (SEED), a national multi-site case-control ASD study, with extant genome-wide genotyping, rigorous clinical phenotyping, and gastrointestinal symptom and condition data available (n=589 ASD cases and 725 controls). We tested for differences in genomic burden of gastrointestinal genetic risk factors between ASD cases and controls using a polygenic risk score (PRS) for each of 3 GI conditions including Crohn's disease (PRS-CD), Ulcerative colitis (PRS-UC) and Irritable Bowel Disease (PRS-IBD). Our adjusted logistic regression models revealed a significant association between PRS-UC and 6 of 9 GI symptoms in the controls (for $p<0.05$, OR=1.61-4.55), suggesting the PRS captures genetic risk factors related to GI phenotypes in children. However, we did not observe significant associations between the PRS-UC and ASD case control status or between the PRS-UC and differences in GI symptoms among children with ASD. Our results suggest that GI symptoms in children with ASD are not explained by genetic risk factors for GI symptoms or conditions. Future studies to examine the association between GI symptoms and severity of restricted repetitive behaviors and food intake in ASD are warranted.

Content Area: Human Genetics
Keywords: Autism, GI, Polygenic Risk Score

Identifying genetic variants shared by orofacial cleft subtypes in GWAS of case-parent trios from POFC and GENEVA studies

Sowmya Venkataraghavan[1], Wanying Zhang[1], Jacqueline A. Bidinger[1], Margaret Taub[2], Terri H. Beaty[1] and Debashree Ray[1,2]

1 Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA
2 Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

Presented by Sowmya Venkataraghavan

Background: Non- syndromic Cleft lip w/wo Palette (CL/P) and Cleft Palate (CP) are Orofacial birth defects with both genetic and environmental risk factors. CL/P accounts for 70% of all orofacial clefts while CP is rarer and makes up the remaining 30%. The two phenotypic groups have shared environmental risk factors such as tobacco and alcohol consumption, and prescription drug use. There could be common biological pathways leading to shared genetic risk factors. Previous studies have indicated only one gene (FOXE1) that is associated with CL/P and CP. More recently, Leslie et al 2017 attempted to identify shared genetic variants by combining subjects with both CL/P and CP into one phenotypic group. Association signals identified by such approaches, however, are heavily influenced by the group with larger sample size or larger effect size. Additionally, such a 'pooled approach' may have limited power to detect variants that affect CL/P and CP in opposite directions. We developed a novel statistical approach called metapleio2 that can use summary statistics from two disjoint trait groups to identify genetic variants associated with both traits.

Methods: We performed extensive simulation experiments to assess the calibration and power of metapleio2 compared to some existing approaches. We combined the genome-wide transmission disequilibrium test statistics on case-parent trios from two large, multi-ethnic studies (POFC and GENEVA) for each of the two traits (CL/P and CP). The resultant meta-analyzed effect sizes and p-values for each of the traits were used by our method metapleio2 to jointly analyze at a SNP level and detect simultaneous association with both traits.

Results: Our simulation studies indicate appropriate type I error control of metapleio2 across several sample sizes and effect sizes (with non-null effect size for at most one trait) while the other methods were found to be very conservative. Specifically, we did not observe strong influence of the trait group with larger sample size or non-null effect size. From the GWAS of POFC and GENEVA studies, four loci reached genome – wide significance on chromosomes 1, 9 and 17. Our results vary from the previous 'pooled approach' analysis, which was based on a meta-analyzed data of POFC trios, POFC case-control samples and GENEVA trios. We observed that the metapleio2 association signals are not driven by the CL/P group – the larger group with many strongly associated SNPs. Additionally, a new locus was identified near the NOF gene on chromosome 17, which was not identified by the previous 'pooled approach' analysis.

Conclusion: Existing methods to discover shared genetic variants for >1 trait are often driven by the trait with larger sample size or magnitude of genetic effect. Our proposed approach metapleio2 can limit the influence of such factors in identifying shared genetic variants of disjoint trait groups. Results from the

GWAS of POFC and GENEVA studies using metapleio2 are in sync with findings from our simulation studies. We found a likely novel shared gene NOF for the two OFC subtypes. Further investigation is required to understand its role in affecting risk of both traits

Content Area: Genetic Epidemiology
Keywords: Orofacial birth defects, Pleiotropy, novel statistical methods

Mitochondrial DNA Heteroplasmy is Associated with Overall Mortality

RJ Longchamps[1], YS Hong[2], CE Newcomb[1], JA Sumpter[1], CA Castellani[1], ML Grove[3], JD Walston[4], BG Windham[5], J Coresh[4], E Boerwinkle[3], E Guallar[2] and DE Arking[1]

1 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
2 Departments of Epidemiology and Medicine, and Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD
3 School of Public Health, Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX
4 Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD
5 University of Mississippi Medical Center Department of Medicine and Center of Biostatistics

Presented by Ryan J Longchamps

Several biological processes have been hypothesized to explain the critical role of mitochondrial dysfunction in disease, such as declines in energy production, altered rates of apoptosis, and elevated free radical production. These processes may be exacerbated by the accumulation of mitochondrial DNA (mtDNA) mutations leading to increased levels of heteroplasmy - the presence of multiple distinct mtDNA genomes within an individual. While previous reports have shown heteroplasmy levels increase with age, little is known about the specific role of heteroplasmy in human disease and mortality. We hypothesized higher levels of heteroplasmy would be associated with overall mortality.

Heteroplasmy was measured from whole genome sequence data from 3,658 individuals of the Atherosclerosis Risk in Communities cohort using the mtDNA-server analysis pipeline. Individuals were removed if < 10,000 of the 16,569 mitochondrial bases achieved 250X coverage and haplogroup analysis indicated contamination. Heteroplasmies were identified with the following criteria: 1) Coverage ≥ 250X; 2) Minor allele frequency (MAF) ≥ 5%; and 3) MAF call differs < 4% between strands. Sites near known indel, transition, and transversion artifacts were excluded. After QC filtering we observed 1,219 heteroplasmies in 3,192 individuals.

During 61,520 person-years of follow up we observed 1,341 deaths. To assess the association of heteroplasmic mtDNA with overall mortality, we performed a Cox proportional-hazards model adjusting for age, sex, DNA collection site, smoking status, body mass index, systolic blood pressure, low-density lipoprotein, history of myocardial infarction and type 2 diabetes status. We observed a hazard ratio (HR) of 1.25 (95% CI 1.11 - 1.40; P = 0.0002) for heteroplasmic mtDNA carriers. To enrich for biologically relevant variants, we investigated the added effect of variants which cause deleterious mutations as defined by scaled CADD scores > 15. Although not statistically significant, the 145 Individuals with these variants showed greater mortality with a HR of 1.14 (95% CI 0.89 - 1.45; P = 0.30). Furthermore, the 466 individuals with recurrent variants had similar rates of mortality compared to individuals with singletons indicating recurrent sites are similarly tolerated (HR = 1.07; 95% CI 0.88 - 1.30; P = 0.50). Together, our findings highlight heteroplasmy as a strong predictor of overall mortality which warrants further investigation into disease-specific mortality.

Content Area: Genetic Epidemiology, Human Genetics
Keywords: Heteroplasmy, Mitochondria, Mortality, Genetics

From Gene Co-expression to Brain Circuits: Prefrontal Gene Co-expression Networks Predict Drug Response in Patients with Schizophrenia

Pasquale Di Carlo[1,2], Andrew E. Jaffe[2,3,4,5], Marco Papalino[6], Qiang Chen[2], Thomas M. Hyde[2,7,8], Joel E. Kleinman[2,8], Joo Heon Shin[2], Antonio Rampino[1,9], Giuseppe Blasi[1,9], Alessandro Bertolino[1,9] and Daniel R. Weinberger[1,8,10,11]

1 Group of Psychiatric Neuroscience, Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy
2 Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD
3 Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
4 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
5 Center for Computational Biology, Johns Hopkins University, Baltimore, MD
6 Group of Psychiatric Neuroscience, Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy
7 Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD
8 Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD
9 Azienda Ospedaliero-Universitaria Consorziale Policlinico, Bari, Italy
10 Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD
11 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD

Presented by Giulio Pergola

Background. Gene co-expression networks are relevant to functional and clinical translation of schizophrenia risk genes. We hypothesized that schizophrenia risk genes converge into co-expression pathways which may be associated with gene regulation mechanisms and with response to treatment in patients with schizophrenia.

Methods. We identified gene co-expression networks in two prefrontal cortex post-mortem RNA sequencing datasets (N=688) and replicated them in four more datasets (N=1295). We identified and replicated (p-values<.001) a single module enriched for schizophrenia risk loci (13 risk genes in 10 loci). In silico screening of potential regulators of the schizophrenia risk module via bioinformatic analyses identified two transcription factors and three miRNAs associated with the risk module. To translate post-mortem information into clinical phenotypes, we identified polymorphisms predicting co-expression and combined them to obtain an index approximating module co-expression (Polygenic Co-expression Index: PCI).

Results. The PCI-co-expression association was successfully replicated in two independent brain transcriptome datasets (N=131; p-values<.05). Finally, we tested the association between the PCI and short-term treatment response in two independent samples of patients with schizophrenia treated with olanzapine (N=167). The PCI was associated with treatment response in the positive symptom domain in both clinical cohorts (p-values<.05).

Conclusions. In summary, our findings in 1983 samples of human post-mortem prefrontal cortex show that co-expression of a set of genes enriched for schizophrenia risk genes is relevant to treatment response. This co-expression pathway may be co-regulated by transcription factors and miRNA associated with it.

Content Area: Human Genetics, Statistical Genetics
Keywords: Gene co-expression networks, dorsolateral prefrontal cortex, olanzapine, RNA sequencing, schizophrenia

Understanding the TP53 signaling cascade that regulates cell fate post-acute aneuploidy induction

Akshay Narkar[1,2], Pandurang Bharne[1], Anjali Nelliat[1], Jin Zhu[1], Debojyoti Biswas[1] and Rong Li[1,2]

1 Center for Cell Dynamics, Johns Hopkins University, School of Medicine, Baltimore, MD 21205, USA
2 McKusick-Nathans Institute of Genetic Medicine (IGM) Johns Hopkins University, School of Medicine, Baltimore, MD 21205, USA

Presented by Akshay Narkar

Genome integrity relies on the equal partitioning of replicated chromosomes to daughter cells during cell division. Errors in chromosome segregation lead to aneuploidy, a state where the number of chromosomes in a cell or organism deviates from multiples of the haploid genome (Compton 2011). In the case of a majority of normal somatic cells chronic aneuploidy and CIN adversely affect proliferation in culture, yet aneuploidy is also a hallmark of cancer, (Amon 2012, Li 2013) a disease of enhanced proliferative capacity. Aneuploidy can lead to TP53 a tumor suppressor gene activation and thereby alter proliferation, but the exact trigger for TP53 activation has remained controversial. This study investigates changes in the genome and kinase activity evoked in different mammalian aneuploid cells and their effect on cell-cycle progression. Acute aneuploidy was generated by inhibiting an essential SAC regulator MPS1, which resulted in cells passing through mitosis with irregular kinetochore and microtubule attachments and higher number of mis-segregated chromosomes in daughter cells. Genome wide CRISPR/Cas9 knockout screen comparing euploid and aneuploid populations revealed certain genes expression changes that may regulate the cells response to aneuploidy. Suppression of proliferation in different cell types depends on TP53 and its transcriptional target, CDK inhibitor p21 (CDKN1A). The amount of cell cycle arrest exhibited by acute aneuploidy induction varies which indicates different upstream regulators may be involved based on the random karyotype generated. A complementary Kinase inhibitor library screen was performed in colorectal cancer cells which revealed pathways having a set of kinases that can regulate the stability of TP53 directly or indirectly. These results can help piece together the complex molecular circuitry of upstream kinases and stress regulators that can facilitate a cell to sense chromosomal aberrations and regulate the TP53 response which in turn determines cell fate.

Content Area: Human Genetics
Keywords: aneuploidy, cell cycle, p53

Facilitating rapid precision oncology in anaplastic thyroid cancer: Clinical implications of next generation sequencing (NGS) mutation testing and impact on survival

Jennifer Rui Wang[1], Gilbert J. Cote[2], Mark E. Zafereo[1], Priyanka C. Iyer[2], Ramona Dadu[2], Naifa L. Busaidy[2], Neil D. Gross[1], Erich M. Sturgis[1], Renata Ferrarotto[3],Charles Lu[3], G. Brandon Gunn[4], Jeffrey N. Myers[1], Stephen Y. Lai[1], Michelle D. Williams[5] and Maria E. Cabanillas[2]

1 Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center
2 Department of Endocrinology, The University of Texas MD Anderson Cancer Center
3 Department of Thoracic/Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center
4 Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center
5 Department of Pathology, The University of Texas MD Anderson Cancer Center

Presented by Jennifer R. Wang

Background: Anaplastic thyroid cancer (ATC) is a rare malignancy characterized by rapid progression and median overall survival (OS) of less than 6 months. With the goal of improving ATC outcomes, a multi-disciplinary program ("FAST") was developed at our institution, which provides rapid access to care, mutation testing, and personalized treatment. The objective of this study was to examine whether results of mutation testing are associated with OS in ATC patients.

Methods: ATC patients seen between 2012-2017 at our institution who had adequate tumor samples for NGS testing were included. DNA was extracted from pre-treatment core biopsies or surgical specimens of primary ATC. Targeted mutation testing was performed using NGS platforms for 46-409 genes. A common set of 26 cancer-associated genes tested across all NGS platforms was used for analysis. Cox proportional hazards models adjusted for ptotential confounders including age at diagnosis, M-stage, and systemic treatment were used to assess the associations between mutation status and OS.

Results: A total of 109 patients were included (median age: 65.5 years, 57% male). 84% presented with T4b disease. 50% were M1 at presentation. The most common mutations detected were p53 (54%), BRAFV600E (51%), PIK3CA(21%), and RAS (19%). RAS and BRAFV600E mutations were mutually exclusive. In multivariate models, RAS mutation was significantly associated with worse OS (adjusted HR 2.42, 95% CI: 1.12-5.24)compared to patients with BRAF-mutated ATC. Patients without BRAF or RAS mutations have similar survival to BRAF-mutated patients (adjusted HR 0.94, 95% CI: 0.46-1.90). Twenty-five of 52 (48%) BRAFV600E+ patients were treated with a BRAF inhibitor. In BRAFV600E+ patients, BRAF inhibitor treatment was significantly associated with improved OS at 1 year.

Conclusions: This is the largest study of mutation status and survival outcome in ATC to date. BRAFV600E and RAS are mutually exclusive drivers that impact prognosis in ATC. Mutation analysis at diagnosis offer prognostic evaluation and may lead to improvement in outcome by facilitating treatment with targeted therapies and enrollment in clinical trials.

Content Area: Genetic Epidemiology
Keywords: thyroid cancer, somatic mutations, targeted therapy

Identification and comparison of imputed and genotyped variants for genome wide association study of orofacial clefts disease

Wanying Zhang[1], Jacqueline A. Bidinger[1], Sowmya Venkataraghavan[2], Debashree Ray[1], Margaret Taub[2] and Terri H. Beaty[1]

1 Department of Epidemiology, Bloomberg School of Public Health Johns Hopkins University, Baltimore, Maryland, United States
2 Department of Biostatistics, Bloomberg School of Public Health Johns Hopkins University, Baltimore, Maryland, United States

Presented by Wanying Zhang

Background: Cleft lip with/without cleft palate (CL/P) and cleft palate (CP) are the most common craniofacial malformations among newborns, and both show strong familial aggregation with high estimated heritabilities. All previously identified genetic risk factors can only account for a modest proportion of the estimated heritability of orofacial clefts (OFCs), indicating additional genetic risk loci remain to be identified. The aim of this research thesis is to identify the genetic risk factors for OFCs and study the relative benefit imputed variants can provide over and above the directly genotyped variants in identifying genetic risk to OFCs.

Methods: We imputed genetic data on case-parent trios from the GENEVA study using the Michigan Imputation Server, and then conducted genome wide association analysis to identify genetic variants associated with CL/P and with CP. For a given cleft group, we performed genotypic transmission disequilibrium tests (gTDTs) using the trio package on the common single nucleotide polymorphism (SNP) markers (minor allele frequency [MAF] ≥ 5%) in all the trios together and then stratified into Asian and European sub-groups.

Results: We identified two genes not previously reported associated with risk to CL/P: 18q12 (TTR) and 4q22 (GRID2). The most significant SNP in the region of TTR (rs1375445) reached genome wide significance in the combined set of all trios ($p = 4.33 \times 10\text{-}8$) with RR=1.35 [95%CI: (1.21, 1.51)], but not in the European sub-group ($p = 2.94 \times 10\text{-}5$) or Asian sub-group ($p = 5.52 \times 10\text{-}5$) separately. To compare with, the most significant SNP of GRID2 (rs1471079) reached genome-wide significance only in the Asian sub-group ($p = 1.82 \times 10\text{-}7$) with estimated RR = 0.70 [95%CI: (0.60, 0.80)]. Both of these SNPs were imputed with high accuracy (rs1375445: R2 = 0.96; rs1471079: R2 = 0.97). Additionally, we confirmed 9 regions identified in previous studies, including 8q24, 1q32 (IRF6), 20q12 (MAFB), 17p13 (NTN1) and 1p22 (ABCA4). The most significant SNPs in six of these regions were imputed. The most significant SNP (rs17242358) in the region of 8q24 (recognized as gene desert) showed genome wide significance ($p = 1.75 \times 10\text{-}16$) in combined set of all trios. This imputed SNP showed over-transmission of A allele (over G allele) with estimated RR = 2.09 [95%CI: (1.76, 2.49)]. This imputed SNP showed quite different statistical significant in the European sub-group ($p = 7.11 \times 10\text{-}14$) and Asian sub-group ($p = 0.00073$) primarily because the minor allele frequency (MAF) differed across sub-groups, e.g. MAF = 23% and = 2% respectively. We did not detect genome-wide significant locus in the CP group.

Conclusions: Our findings confirm the complex architecture and genetic heterogeneity of OFCs. We replicated most previously reported genetic risk factors. We also discovered two new genetic risk factors

for CL/P that require further investigation. Stratification by ethnicity helps to detect OFC risk loci specific to certain ethnic groups. In addition, imputation helps improve the statistical power to detect the disease-associated genetic risk factors.

Content Area: Genetic Epidemiology
Keywords: orofacial clefts, genome-wide association study, imputation, 8q24

Genome-wide gene-smoking interaction on COPD in UK Biobank study

Woori Kim[1,2], Michael H. Cho[2,3], Phuwanat Sakornsakolpat[2,4], Brian D. Hobbs[2,3], Dmitry Prokopenko[5], Edwin K. Silverman[2,3] and Terri H. Beaty[1]

1 Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD
2 Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA
3 Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA
4 Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand
5 Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Boston, MA

Presented by Woori Kim

Background: The fact that not all smokers will develop COPD implies that the adverse effects of smoking on disease differ by an individual's characteristics such as genetic make-up. Genome-wide association study (GWAS) findings explain a small proportion of the genetic variance in phenotypes defining COPD. Our study aims to identify novel genetic variants by accounting for potential interaction with smoking on risk of COPD using UK Biobank study.

Method: We included 179,689 controls and 21,077 COPD cases from UK Biobank study. We conducted a joint test (2 degree of freedom) of SNP and SNP-by-Smoking interaction for the genome-wide interaction analysis. We considered ever/never smokers and pack-year as smoking exposures. We included markers of which minor allele frequency is over 0.01 and imputation quality score over 0.5. To examine the novel signals that were not identified in the previous GWAS, we 1) compared P-value of each SNPs between our analysis and previous GWAS results in UK Biobank and 2) conducted conditional analyses by adjusting for previously reported variants for each genetic signal and compared p-value and coefficients of SNP (main effects) and SNP*smoking (interaction effects) before and after conditioning. To investigate human leukocyte antigen (HLA) region in detail, we further conducted a joint test using 362 human leukocyte antigen (HLA) imputed markers.

Result: We identified 48 signals at the genome-wide significance level (P<5e-08) using a 1-Mb window (+/- 500kb) around a lead variant. Among 48 signals, six signals were genome-wide significant in our analysis but not in previous GWAS (rs2609255 (P=1.32e-08, Pprevious gwas=5.3e-08, nearest gene= FAM13A), rs11735046 (P=3.98e-09, Pprevious gwas=5.62-e08, nearest gene=GYPA), rs2844809 (P=1.68-e08, Pprevious gwas=5.99e-08, nearest gene=HLA-A), rs56326741 (P=2.04-e08, Pprevious gwas=6.05e-06, nearest gene=TNXB), rs9272426 (P=7.49e-10, Pprevious gwas=6.85e-08, nearest gene=HLA-DQA1) and rs12440014 (P=6.96e-18, Pprevious gwas=1.94e-06, nearest gene=CHRNB4)). In a conditional analysis for HLA region, we conditioned on 3 SNPs (rs9275068, rs3844313, rs7764819) located near/in HLA region previously reported in GWAS or genome-wide interaction analysis of lung function. There was low linkage disequilibrium between rs9272426 near HLA-DQA1 and 3 HLA SNPs (r2 ranging from 0.013 to 0.13). The P-value and coefficients of main effects and interaction effects were similar before and after conditioning on those 3 HLA SNPs. Seventeen HLA imputed markers were significant in a joint test of SNP and SNP-by-Smoking on COPD and were used for conditional analysis for rs9272426 near HLA-DQA1. Conditioning on DQA1*0501, the significance of main effects was attenuated (P=9.84E-04)

while the significance of interaction effects was maintained (P before conditioning= 7.58E-05 and Pafter conditioning = 8.12E-05).

Conclusions: Incorporating SNP-by-Smoking interaction in GWAS, we identified six genetic regions significantly associated with COPD, which did not appear from previous GWAS of COPD using UK Biobank study. Our results of the interaction of HLA-DQA1 with smoking suggest that COPD might be controlled by autoimmune mechanisms.

Content Area: Genetic Epidemiology
Keywords: Gene-environment interaction, COPD, Smoking, Genome-wide association study, HLA

Dopaminergic neuronal chromatin signatures reveal Parkinson Disease associated variation in a novel aminergic intronic enhancer at SNCA

Sarah McClymont[1], Paul Hook[1], William Law[1], Michael Beer[1], Owen Ross[2] and Andy McCallion[1]

1 McKusick-Nathans Institute of Genetic Medicine
2 Mayo Clinic College of Medicine

Presented by Sarah McClymont

GWAS predominantly implicate non-coding variants. Discriminating functional variants from those in LD and establishing the mechanisms modulating disease risk remain significant challenges. Assaying chromatin signatures in appropriate cell types/states is imperative to identifying pertinent functional elements and overcoming these challenges. To evaluate the role of non-coding variation in Parkinson Disease (PD), we assayed the open chromatin of dopamine (DA) neurons. We performed ATAC-seq on E15.5 Tg(Th-EGFP)DJ76Gsat FACS-isolated DA midbrain (MB) and forebrain (FB) neurons.

We identify >125,000 open chromatin regions, characterized by a high degree of functional sequence constraint and enrichment near neuron related genes. They are also enriched for sequences demonstrated in the VISTA database to direct expression in vivo in neuronal tissues (n=482/652, 74%) in contrast to negative and non-neuronal tissues (n=805/1735, 46%). We also assay five putative enhancers in transgenic zebrafish and mouse assays; all five regions display neuronal enhancer activity in one or both model organisms. Training a machine learning algorithm on the open chromatin intervals develops a robust regulatory vocabulary for MB and FB DA neurons, predicting transcription factors (TFs) acting on our putative enhancers (eg: Rfx, Fox families). TF footprinting of a deeply sequenced MB library confirms these predicted TFs' activity.

Importantly, we identify a 1kb enhancer within intron 4 of SNCA. This enhancer directs reporter expression in aminergic neuronal populations of transgenic zebrafish and mice throughout development. Further, sequencing this enhancer in ~1,000 PD patients and ~1,000 controls identifies two related risk SNPs, rs2737024 and rs2583959, significantly enriched among patients (OR=1.27, 1.25; $p<0.001$, 0.002). Haplotype analysis considering these and other GWAS-implicated SNPs at SNCA identifies a common haplotype which increases risk of PD; it includes the risk alleles of the two SNPs within the enhancer. These implicated SNPs are being evaluated for their effect on enhancer activity.

Content Area: Human Genetics
Keywords: Dopaminergic neurons, Parkinson disease, ATAC-seq, enhancers, chromatin

Methods for genetic analysis of secondary phenotypes in case-control GWAS with application to COPDGene data

Allen Liu, Terri H. Beaty, Debashree Ray

Presented by Allen Liu

Background: Current approaches for GWAS of secondary traits include adjusting the case-control status as a covariate in the analysis model, or considering a weighted regression based on estimated disease prevalence in the underlying population, or not considering any adjustment at all. The type I error calibration of the methods has been shown to heavily depend on the underlying causal pathway among genetic variants (X), secondary traits (Y) and the primary disease status (D). A new method, proportional odds model adjusted for propensity score (POM-PS), was developed and it exhibits proper type I error calibration in the scenarios where other methods do not. However, the validity of POM-PS was shown for nominal significance levels only and for scenarios where the direction of association of Ys and D is same. Considering the fact that Ys may be associated with D in different directions, we need to assess the validity of POM-PS in such scenarios at genome-wide levels.

Methods: Building on the publication that introduced POM-PS, we simulate three scenarios with two secondary traits: DAG A, where X affects D, and both Ys affect D; DAG B, where X affects D and D affects both Ys; and DAG C, where X and Y1 affects D while D affects Y2. We conduct type I error simulations to assess calibration of existing methods (e.g., MANOVA-unadjusted, MANOVA-adjusted, SMAT and POM-PS) at significance thresholds ≥10-5. We also consider a different DAG C simulation based on the COPDGene data. For COPDGene data, we are interested in testing genetic association of common SNPs with any of the secondary traits, log pack-years of smoking and log percent emphysema. We compare association signals – separately by ethnic groups – across afore-mentioned existing approaches, examine any systematic bias and investigate any new signal reported by any of these methods.

Results: In the simulation study, POM-PS demonstrate acceptable type 1 error control in both DAGs A and B, while other methods show heavily inflated type 1 error estimates in either DAG A or B. All methods experienced type 1 error inflation in the DAG C where direction of association between the secondary traits and the disease status are different. A subsequent simulation with parameters derived from COPDGene dataset indicated POM-PS has a better type 1 error control, while MANOVA (unadjusted and adjusted) and SMAT showed serious inflation. Among NHW in the COPDGene data, SMAT and MANOVA-unadjusted seem to show spurious signals, while POM-PS was conservative. Significant association of SNPs at MORF4L1 (most significant MANOVA-adj p-value = 2.48x10-8) on chromosome 15 is confirmed from previous publications. For the AA group, significant SNPs were not lying in genes that are correlated with COPD traits.

Conclusions: Our simulation experiments indicate better control of type I error for POM-PS under most scenarios compared to the other existing approaches. The genome-wide analysis of the COPDGene data using secondary traits log percent emphysema and log pack-years, seem to corroborate our findings from simulations. POM-PS remains uninflated in both AA and NHW groups but showed conservative performance in NHW.

Content Area: Genetic Epidemiology
Keywords: multiple secondary traits, multivariate analysis, propensity score, proportional odds model

Downstream Targets of Polygenic Risk for Autism Spectrum Disorder in a General Population of Neurotypicals

Vamsee Pillalamarri[1,2] and Dan Arking[1]

[1]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine
[2]Maryland Genetics Epidemiology and Medicine Training Program

Presented by Vamsee Pillalamarri

Autism spectrum disorder (ASD), a pervasive neurodevelopmental disorder that presents with coherent symptoms by the age of three, is of increasing public health significance with prevalence estimates of 1 in 59 in children diagnosed by age 8 years old in the United States. ASDs are strongly heritable disorders, with heritability estimates of ~60% to 90%, and MZ and DZ twin concordance rates of ~50% and ~20%, respectively. While ASD and ASD endophenotypes cluster in families, common inherited variation explains the majority of ASD liability. Hence, many ASD associated risk alleles are also found within the general population of unaffected individuals. Furthermore, core symptoms in individuals diagnosed with ASD include limited social responsiveness and restricted, repetitive behavioral patterns, yet these characteristic signs of ASD often underlie a wide diversity of comorbidities and individualized clinical manifestations. This phenotypic heterogeneity within the autism spectrum is likely due to the consequences of the common variant, polygenic genetic architecture that generates ASDs and can be understood as a multitude of common variants each having a small effect towards modulating overall disease liability under a quantitative threshold model. We therefore reasoned that studying the effects of common, polygenic risk for ASD in a general population of neurotypicals would allow us to uncover *shared* neurobiological disease mechanisms amongst the overlapping endophenotypes and comorbid conditions that make up an ASD diagnosis. Specifically, we hypothesized that there exist downstream targets of common, polygenic risk for ASD that will have altered gene expression in the general population within relevant cell and tissue types. To test this hypothesis of polygenic effects on gene expression, we carried out a transcriptome-wide association study (TWAS) to identify ASD-specific polygenic risk score (PRS) expression quantitative trait loci (PRS-eQTL) within the Genotype-Tissue Expression (GTEx) cohort of healthy neurotypicals. Using principal components analysis of filtered, linkage-disequilibrium pruned genotypes, we selected donors of European American (EA) ancestry (n=528) and constructed an ASD PRS for each of these donors using EA-specific effect size estimates from the latest ASD GWAS (Grove 2019; 18k cases and 28k controls). Normalized RNA-Seq gene expression estimates from post-mortem tissue samples were then treated as a quantitative phenotype and modeled as a function of the ASD PRS along with confounding technical and biological covariates, including latent factors captured through PEER. Collinearity was avoided through removal of covariates with high correlation. Independent tests were carried out per gene per tissue that had a minimum of 70 sampled donors (n=48). Across 48 tissues, nine tissues had ≥1 ASD-specific PRS-eQTL at a tissue-specific FDR ≤ 10%, including three tissues sampled from the brain (cortex, caudate basal ganglia, and substantia nigra). P-values for each test were well calibrated per rigorous permutation testing, and overall 17 genes were significantly associated with an ASD-specific PRS at a tissue-specific FDR ≤ 10% with seven PRS-eQTL originating from pancreas tissue (41.2%). Ongoing efforts include characterizing the effect of PRS on co-expressed gene networks using WGCNA, and extending this approach to larger sample sizes, including the PsychENCODE and CommonMind datasets.

Content Area: Human Genetics
Keywords: Autism spectrum disorder, polygenic risk score, GTEx, eQTL, TWAS

Kidney Function and Blood Pressure: A Mendelian Randomization Study

Zhi Yu[1], Josef Coresh[1], Morgan Grams[1], Guanghao Qi[1], Eric Boerwinkle[2], Harold Snieder[3], Alexander Teumer[4], Anna Kottgen[5], Cristian Pattaro[6], Nilanjan Chatterjee[1] and Adrienne Tin[1]

1 Johns Hopkins University
2 University of Texas Health Science Center at Houston
3 University Medical Center Groningen
4 University of Greifswald
5 University of Freiburg
6 EURAC Research

Presented by Zhi Yu

Background. Reduced kidney function has been considered a risk factor for hypertension but whether the relation is causal remains uncertain. We used summary statistics from large-scale genome-wide association studies (GWAS) to conduct two-sample Mendelian randomization (MR) analyses to estimate the causal effect of reduced kidney function on systolic blood pressure (SBP) and diastolic blood pressure (DBP).

Methods. Kidney function was represented by: i) estimated glomerular filtration rate (eGFR) based on serum creatinine calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation and ii) blood urea nitrogen (BUN). The European-ancestry GWAS results were from the CKDGen Consortium (eGFR, n=567,460; BUN, n=243,031) and the UKB-ICBP Consortium (SBP and DBP, n=~1 million). It has been known that both eGFR and BUN have marker-specific determinants that are unrelated to kidney function. To reduce marker-specific genetic influence that can bias the results, we used a filtering strategy requiring that the index SNPs of eGFR had opposite direction of association with BUN and association with BUN below the Bonferroni corrected threshold, and vice versa for BUN index SNPs.

Results. After filtering of the 256 eGFR index SNPs and 75 BUN index SNPs, 40 and 26 were kept for MR analysis, respectively. We observed significant causal effect estimates using MR methods that account for pleiotropy: weighted median, weighted mode, MR PRESSO, and MRMix. Estimates from the weighted median method: effect for 50% lower eGFR: 14.8 mmHg higher in SBP (p=9.7x10-11) and 7.7 mmHg higher in DBP (p=1.0x10-7), effect for doubling of BUN: 7.0 mmHg higher in SBP (p=3.4x10-8) and 3.7 mmHg higher in DBP (p=5.0x10-8).

Conclusions. MR analysis supports a causal effect of decreased kidney function on higher blood pressure. These findings help interpretation of observational epidemiologic associations.

Content Area: Genetic Epidemiology
Keywords: Mendelian randomization, kidney function, estimated glomerular filtration rate, blood urea nitrogen, blood pressure

The Epigenetic Developmental Clock and its prospective association with ASD in childhood

Sahra Mohazzab-Hosseinian, Jason Feinberg, Rebecca Schmidt, Lisa Croen, Irvz Hertz-Picciotto, M. Daniele Fallin, Craig Newschaffer and Christine Ladd-Acosta

Departments of Epidemiology and Mental Health, Johns Hopkins University
The Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins University
Department of Public Health Sciences and UC Davis MIND Institute, University of California Davis
Autism Research Program, Division of Research, Kaiser Permanente Northern California
A.J. Drexel Autism Institute and Departments of Epidemiology and Biostatistics, Drexel University

Presented by Sahra Mohazzab-Hosseinian

Background: Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects social interactions and communication as well as restrictive or repetitive interests. These impairments result in substantial personal and economic burdens at the individual, familial and societal levels. Identification of prenatal and/or early life risk factors can provide new targets for prevention and intervention efforts to reduce the burden of ASD and/or they could aid in detection of ASD detection. Multiple lines of evidence have implicated epigenetic factors in ASD risk. Indeed, site-specific changes in the epigenome have been observed among individuals with ASD although most studies have been limited by epigenetic measurement timing, i.e. after ASD diagnosis. Studies to assess the prospective associations between epigenetic measures, at birth or infancy, and later ASD diagnoses in early childhood, are needed. To address this gap, we examined DNA methylation levels at developmental aging loci, via "epigenetic clocks", in cord blood and prospective ASD diagnosis associations. The developmental age pathway is of particular interest given preterm birth is a risk factor for ASD and also because it maximizes power in this modest sample.

Methods: We used extant cord blood DNA methylation and rigorous clinical phenotype data from the Early Autism Risk Longitudinal Investigation (EARLI) study. EARLI is an ethnically diverse, familial enriched risk birth cohort. After full clinical evaluations, participants received a final diagnosis of typically developing (TD), non-typically developing (NTD), or autism spectrum disorder (ASD), at age 3, using diagnostic algorithms from the Baby Siblings Research Consortium. Cord blood DNA methylation measurements were obtained using the Illumina 450K platform. A rigorous quality control (QC) pipeline was applied to remove poorly performing samples and probes. Our post QC analytic dataset contained 171 samples. We applied the Knight et al gestational age algorithm to EARLI DNA methylation measurements to compute a single value, representing "developmental epigenetic age", per child. We tested for prospective associations between epigenetic age at birth and diagnosis at age 3 using linear regression models, while adjusting for cell composition, maternal age, maternal education, technical batch effects, and principal components to control for ancestry.

Results: We observed a significant difference in epigenetic developmental age among children with later ASD diagnoses relative to typically developing children (P=0.04). The ASD children showed epigenetic developmental age deceleration, i.e. were predicted to be biologically younger than their chronologic gestational age by approximately 5 days, or -0.68 weeks. No significant differences in epigenetic developmental age were observed for non-typically developing children relative to those with ASD or TD outcomes.

Conclusions: Our findings suggest that children who later go on to receive an ASD diagnosis show developmental deceleration at birth compared to typically developing children. Future directions should include replication of these results in other cohorts.

Content Area: Genetic Epidemiology
Keywords: autism, epigenetics

Co-expression patterns define chromatin regulators associated with neuropsychiatric phenotypes

Leandros Boukas[1], James Havrilla[2], Peter Hickey[3], Aaron Quinlan[2], Hans T. Bjornsson[1] and Kasper D. Hansen[1]

1 Johns Hopkins University
2 University of Utah
3 The Walter and Eliza Hall Institute of Medical Research

Presented by Leandros Boukas

Coding variants in chromatin regulators (CRs) have emerged as causes of neuropsychiatric phenotypes and cancer. Many studies have yielded insights into the individual functions of these genes; however, despite their obvious functional similarities, CRs have never been studied collectively as a class.
Here, we used protein domain annotations to define 295 human CR genes. We first observed that 170/295 are highly intolerant to coding loss-of-function variation (pLI>0.9 in ExAC, $p<0.001$, OR=7.7 vs other genes), with 102/170 having no current disease associations; CRs are even more variation-intolerant than transcription factors ($p<0.001$, OR=4.4 for enrichment in the pLI>0.9 category). Additionally, brain-specific enhancers surrounding CRs were enriched for the common-variant heritability signal of 5 neuropsychiatric traits (median enrichment = 5.05).
Subsequently, we observed that CRs have multiple protein domains. We derived a local, domain-specific constraint score using variation in gnomAD, and showed that 82% of the domains responsible for their epigenetic function are constrained. This provides direct genetic evidence that chromatin state dysregulation is mediating the disease phenotypes of these (rare coding or common regulatory) gene-dosage disrupting variants.

Finally, leveraging the GTEx dataset, we searched for expression signatures that distinguish between variation-intolerant and variation-tolerant CRs. Absolute expression levels or tissue specificity could not explain this difference. However, we discovered a large subset of 74 CRs which are co-expressed within multiple tissues; this is not observed with random genes or transcription factors. The co-expressed subset almost exclusively contains variation-intolerant genes (pLI>0.9 for 72/74, $p<0.001$). It shows enrichment for CRs causing Mendelian diseases with neuropsychiatric features ($p=0.006$, OR=3.3), even when accounting for high pLI ($p=0.003$, OR=3), but not for those linked to cancer ($p=0.41$). Moreover, partitioning heritability for IQ and generalized epilepsy (chosen since low IQ and seizures are the most common neurological symptoms of such Mendelian disorders), reveals that only the co-expressed CRs show significant enrichment.

Our results provide a comprehensive genetic analysis of human chromatin regulators and uncover a systems-level property (co-expression) associated with their variation-intolerance and involvement in brain disease.

Content Area: Computational Genetics, Human Genetics
Keywords: epigenetics, population genetics, neuropsychiatric disease, mendelian disease

miRNA Isoform Quantification Method Development and Expression Analysis in Schizophrenia

Carrie Wright[1,2], Anandita Rajpurohit[1], Courtney Williams[1], Nicholas J. Brandon[3], Thomas M. Hyde[1,4], Joel E. Kleinman[1,4], Alan J. Cross[3], Andrew E. Jaffe[1,4-11], Joo Heon Shin[1,5] and Daniel R Weinberger[1,4,5-7]

1 Lieber Institute for Brain Development, Johns Hopkins Medical Institutions (JHMI)
2 AstraZeneca Postdoc Program
3 AstraZeneca
4 Dept. of Psychiatry and Behavioral Sciences, JHMI
5 Dept. of Neurology, JHMI
6 Dept. of Neuroscience, JHMI
7 Institute of Genetic Medicine, JHMI
8 Dept. of Mental Health, JHSPH
9 Dept. of Biostatistics, JHSPH
10 Center for Computational Biology, Johns Hopkins University (JHU)
11 Dept. of Epidemiology, JHU Bloomberg School of Public Health (JHSPH)

Presented by Carrie Wright

Recent studies have confirmed that altered microRNA (miRNA) isoforms, termed isomiRs, are not merely sequencing artifacts but are in fact functional regulatory RNAs associated with a variety of diseases and conditions. Many of these altered miRNAs show differences in their stability, localization, and targeting relative to each other and relative to the canonical isoforms. Some isomiRs have been shown to regulate a different repertoire of genes relative to the canonical form. Method assessment for the quantification of miRNA isoforms has been largely overlooked. Thus we compared and evaluated miRNA isoform quantification estimates of synthetic miRNAs (an equimolar pool of 962 synthetic sequences corresponding to human, rat, mouse, and virus miRNAs) and a single homogenate human brain sample in replicates using a variety of methods. In addition, we also compared these methods to a modified version of one of these methods, in which we also utilized the randomized adapter sequences as unique molecular identifiers (UMIs) to collapse the reads to the sequences that were present before amplification to mitigate bias due to amplification. The synthetic samples were designed to not contain isomiRs, allowing for an assessment of the level of false positive isomiR quantifications obtained with each method, while the human brain sample allowed for evaluation of the capacity of each method to quantify isomiRs. We determined that there were significant differences in the performances of widely used methods, and that some methods may falsely result in highly abundant expression estimates for isoforms that may not be originally present in a sample. However, other methods performed quite well. Using a method that resulted in the fewest false isomiRs, while still capturing a large number of potentially true isomiRs, we evaluated miRNA isoform expression in postmortem dorsolateral prefrontal cortex brain samples of 30 samples with schizophrenia and 55 neurotypical samples. Using a logistic regression model in which we accounted for confounding factors of expression, such as gender, age, RNA integrity, ethnicity, and surrogate variables to account for latent confounding factors, we determined that 21 miRNAs (out of 276 miRNAs expressed above a lower expression limit and considering miRNA expression as counts for all isoforms including reference together) were differentially

expressed, while 11 (out of 576 expressed above a lower limit threshold) individual isoforms were differentially expressed. Several of the isoforms that were differentially expressed were not the reference sequence, but instead represented isomiRs with alterations that might alter stability or targeting, suggesting that these expression differences may indicate alterations in regulatory networks in schizophrenia. We also performed global differential evaluations of the various types of miRNA isoforms created through different enzymatic processes. We did not identify any global differences in the abundance or diversity for any of the types of isomiRs. Further research is warranted to evaluate how the expression changes in the miRNAs and miRNA isoforms identified may influence gene expression in schizophrenia.

Content Area: Human Genetics
Keywords: microRNA, RNA sequencing, schizophrenia, isomiRs

Highly Specific, Highly Expressed Genes Contaminate the GTEx Dataset

Tim O. Nieuwenhuis[1,2], Vamsee Pillalamarri[2], Avi Z. Rosenberg[1], Dan E. Arking[2], Matthew N. McCall[3] and Marc K. Halushka[1]

1 Department of Pathology, Johns Hopkins University School of Medicine
2 Institute of Genetic Medicine, Johns Hopkins University School of Medicine
3 Department of Biostatistics and Computational Biology, University of Rochester Medical Center

Presented by Tim Nieuwenhuis

Background: One of the challenges of next generation sequencing (NGS) is contaminating reads from other samples. We used Genotype Tissue Expression (GTEx), a large, diverse, and robustly generated dataset, as a useful resource to understand the factors that contribute to contamination.

Results: We obtained 11,340 RNA-seq samples, DNA variant call files (VCF) of 635 individuals, and technical metadata from GTEx as well as read count data from the Human Protein Atlas (HPA) and a pharmacogenetics study. We analyzed 48 tissues in GTEx. Of these, 24 had variant co-expression clusters of known highly expressed and pancreas-enriched genes (PRSS1, PNLIP, CLPS, and CELA3A). Nine additional highly expressed genes from other tissues were also indicative of contamination (KRT4, KRT13, PGC, CPA1, ACTA1, ALB, CTRB1, CPB1 and PRL). Sample contamination by non-native genes was highly associated with a sample being sequenced on the same day as a tissue that natively has high levels of those genes. This was highly significant for both pancreas genes ($p = 2.7E-75$) and esophagus genes ($p = 8.9E-154$). We used genetic polymorphism differences between individuals as validation of the contamination. Specifically, 11 SNPs in five genes shown to contaminate non-native tissues demonstrated allelic differences between DNA-based genotypes and contaminated sample RNA-based genotypes. Low-level contamination affected 1,841 (15.8%) samples (defined as ≥500 PRSS1 read counts). It also led to 328 eQTL assignments in inappropriate tissues among these 13 genes. In support of this type of contamination occurring widely, pancreas gene contamination (PRSS1) was also observed in the HPA dataset, where pancreas samples were sequenced, but not in the pharmacogenomics dataset, where they were not.

Conclusions: Highly expressed, tissue-enriched genes basally contaminate the GTEx dataset impacting on some GTEx data analyses. This contamination is not unique to GTEx, being shared with other datasets. Awareness of this process will reduce improperly implicating contaminating low-level gene expression in disease processes.

Content Area: Computational Genetics, Human Genetics
Keywords: GTEx, RNA-Seq, Contamination

The Role of HMGA1 in Cellular Aging, Senescence, and Cancer

Olivia Sniezek[1,2], Emily Bramel[1,2], Lionel Chia[3], Li Luo[4], Lingling Xian[4] and Linda Smith Resar[2,4]

1 Predoctoral Training Program in Human Genetics and Molecular Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
2 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
3 Pathobiology Graduate Program, Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
4 Department of Hematology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Presented by Olivia Sniezek

HMGA1, or high mobility group A1, is a low molecular weight chromatin remodeling protein that binds to AT-rich regions of chromosomal DNA. High levels of HMGA1 have been shown to promote plasticity and proliferation in stem cell populations. However, Hmga1 is silenced postnatally and is effectively absent from differentiated cells. We find that the global constitutive knock-out of Hmga1 results in partial embryonic lethality. More than half of all embryos are lost late in gestation. Hmga1 null mice develop cachectic dwarfism as they age, with loss of white fat and multiple progeroid features, including shortened lifespan, early thymic atrophy, lordokyphosis, osteopenia, and hair loss. We suggest that Hmga1 is essential for normal aging, as Hmga1 null mice also die significantly earlier than their wildtype littermates. Decreased lifespan in Hmga1-/- mice was more pronounced in males (10.05 months in Hmga1-/- versus 25.9 months in wild-type (WT) males compared to 18.95 months in Hmga1-/- versus 24.15 months in WT females). HMGA1 is also known to play a role in cancer development in humans. In refractory hematologic malignancies and poorly differentiated solid tumors, HMGA1 is reactivated, mimicking embryonic levels, where it induces transcriptional networks active in stem cells and tumor progression. When Hmga1-/- mice were crossed to mice capable of in vivo reprogramming via overexpression of the Yamanaka factors (Oct4, Sox2, Klf4, cMyc), teratoma formation was abolished in mice with complete loss of Hmga1. Surprisingly, even mice with heterozygous Hmga1 deficiency had a decreased capacity for in vivo reprogramming. To elucidate potential mechanisms underlying Hmga1's role in aging and cancer, we performed RNA-sequencing of mouse embryonic fibroblasts (MEFs) from WT and Hmga1-/- mice. We found global dysregulation of transcriptional networks involved in DNA repair, mitosis, Wnt signaling, and other developmental pathways in MEFs lacking Hmga1. Further, expression of a subset of genes involved in cellular senescence (Lmnb1, Mmp2, Cdkn1b) were dysregulated in Hmga1-/- MEFs versus WT MEFs. In published databases, Hmga1/HMGA1 expression decreases in HSC from WT mice and humans with natural aging. As Lmnb1 was found to be dysregulated in Hmga1-/- MEFs, we suspected that nuclear architecture was disrupted by the loss of Hmga1 expression. Mutations in genes encoding the nuclear lamina, such as lamin B, result in folding of the lamina and nuclear envelope causing the cells to have a wrinkled phenotype. Premature aging, as seen in the Hmga1-/- mice, is the hallmark of laminopathies such as Hutchinson Gilford progeria. Imaging analyses confirm that Hmga1-/- nuclei are frequently irregular and "wrinkled". These results reveal a potential role for Hmga1/HMGA1 in tissue regeneration in adult stem cells. Our findings suggest that lower steady-state levels of Hmga1/HMGA1 may account for poor regenerative function with advancing age and that modulating Hmga1 expression or function could mitigate or reverse aging phenotypes.

Content Area: Human Genetics
Keywords: HMGA1, Aging, Senescence, Stem Cells, Lamina