Department of Epidemiology
Institute of Genetic Medicine
Maryland-Genetics, Epidemiology and Medicine Training Program

2016 Genetics Research Day

Symposium and Poster Session

"Precision Medicine: Opportunities and Challenges for Human Geneticists and Epidemiologists"

Presented by

Pui-Yan Kwok, MD, PhD

Professor, Dermatology
University of California, San Francisco

Friday, February 26, 2016
12:15 – 1:15 p.m.
Sommer Hall (E2014)
Wolfe Street Building

Poster session immediately following, 1:30 – 4:00 p.m., Feinstone Hall

For more information, contact Jennifer Deal at jdeal1@jhu.edu

Sponsored by the Burroughs-Wellcome Fund

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

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

Abstract Book
Genetics Research Day
February 26, 2016

Contents

Dear Participants,

On behalf of the Maryland-Genetics, Epidemiology, Medicine Training Program (MD-GEM) it is our pleasure to welcome you to the third 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. Departments of Biostatistics, Epidemiology, and Mental Health in the Johns Hopkins Bloomberg School of Public Health; the Departments of Human Genetics, Medicine, Oncology, and Surgery in the School of Medicine; the Lieber Institute for Brain Development, the McKusick-Nathans Institute for Genetic Medicine, and the Wilmer Eye Institute, Johns Hopkins University; and the Computational and Statistical Genomics Branch and Medical Genetics Branch of the National Human Genome Research Institute.

A very special thank you to Dr. Pui-Yan Kwok, University of California, San Francisco, 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, Jungen Yi and Tracie Wyman 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

Dr. Pui-Yan Kwok is a dermatologist and a human geneticist with extensive experience in DNA sequencing, SNP discovery, and SNP genotyping. His group was one of the original contributors of SNPs to the SNP Consortium and the HapMap Project. His group was part of 5 genome-wide association studies, including a recently completed project that produced genome-wide SNP data on 100,000 subjects in an NIH ARRA funded project (a collaboration between Kaiser Permanente and UCSF) using the new Affymetrix Axiom SNP genotyping platform. He is also involved in a number of projects utilizing RNA-Seq, whole genome, whole exome and target capture sequencing using the new sequencing platforms. In addition, he has developed a single-molecule analysis method to perform genome-wide mapping of structural variations and provide scaffolds for de novo genome sequence assembly. As the Faculty Director of the Genomic Core Facility at UCSF, he directs the use of state-of-the-art genotyping and sequencing platforms on campus.

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
Steven Salzberg
Diane M. Becker
Lewis Becker
Larry Brody
Josef Coresh
Derek Cummings
Gary Cutting
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
Jeffrey Leek
Justin Lessler
Brion Maher
Rasika Mathias
Shruti Mehta
Ana Navas-Acien
Elaine A. Ostrander
Elizabeth A. Platz
Stuart Ray
Debra Roter
Robert Scharpf
Alan Scott
Margaret Taub
David Thomas
Kala Visvanathan
Jeremy Walston
Xiaobin Wang
Alexander Wilson
Robert Wojciechowski
Peter Zandi

Poster No.	Presenter	Title	Page No.
1	Jean-Philippe Fortin	Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data	16
2	Ryan Longchamps	Genome-Wide Interrogation of Spouse Selection Indicates Lack of Assortative Mating	25
3	Amanda J. Price	Profiling Cell-Type Specific Epigenomic Landscapes Across Human Cortical Development and Aging	30
4	Emily Holzinger	A variable selection method for identifying complex genetic models associated with human traits	19
5	Ferdouse Begum	Identified structural variants associated with multiple phenotypes of COPDGene African American Study Cohort	9
6	Shan Andrews	Methylation quantitative trait loci enhance genome-wide association study results for autism spectrum disorder across tissue type	8
7	Fei Chen	Whole exome sequencing and linkage analysis of patients with pulmonary nontuberculous mycobacterial infection	12
8	Gianluca Ursini	GWAS derived risk profile score is associated with schizophrenia only in individuals exposed to obstetric complications	33
9	Pingwu Zhang	Genomic and RNA variants at the ARMS2/HTRA1 Locus	36
10	Norazlin Kamal Nor	Abnormal Growth in children with ASD in the SEED Study	21
11	Priyanka Nandakumar & Adrienne Tin	Differential Transcriptome Profiling of African Americans with Uncontrolled Hypertension and Chronic Kidney Disease (CKD) versus Controlled Hypertension and without CKD	28
12	Chelsea Qinjie Zhou	T cells in the necrotizing enterocolitis brain: how the gut disease leads to the brain developmental impairment	37
13	Chang Shu	Integrating Expression Quantitative Brain Loci in ASD GWAS analyses	31
14	Anthony M. Musolf	Linkage Analyses Reveals Significant Association for Myopia	27
15	Nicole Eckart	Regulatory Function of Schizophrenia-Associated Variants in CACNA1C	13
16	Samantha Bomotti	The distribution of ABCA4 variants in Stargardt disease from the ProgStar studies	10
17	Margaret Hoang	Accumulation of somatic mutations in normal and cancerous tissues with age	18
18	Jack Fu	Whole Exome Association of Rare Deletions in Multiplex Oral Cleft Families	17
19	Carrie Wright	Increased expression of histamine signaling genes in Autism Spectrum Disorder in postmortem human brain	35
20	Rebecca Eggebeen	Determining Mitochondrial DNA Copy Number from Next Generation Sequencing Data	14
21	Kai Kammers	Integrity of induced pluripotent stem cell (iPSC) derived megakaryocytes as assessed by genetic and transcriptomic analysis	22
22	Martha F Brucato	Comparison of Illumina Infinium 450K Methylation BeadChip preprocessing methods in an Epigenome Wide Association Study	11
23	Genevieve Stein-OBrien	Genome specific transcriptional signatures predict differentiation biases in Human ES/IPS cells	32
24	Julius S. Ngwa	Differential Analysis of Gene and Transcript Abundance for RNA-Seq Data using STAR and HISAT Aligners	29

25	Qing Li	Trio Random Forest: Post Analysis of Tree Structure To Reveal Interactions	24
26	Stephanie Loomis	Exome array analysis of nuclear sclerosis in the Beaver Dam Eye Study	26
27	Candelaria Vergara	Sequencing Analysis of Interferon Lambda Loci in individuals with Spontaneous Hepatitis C virus Clearance and Persistence	34
28	Deyana Lewis	Follow-Up and Replication Study of Caries in the Permanent Dentition	23
29	Zhicheng Ji	TSCAN: Pseudo-time Reconstruction and Evaluation in Single-cell RNA-seq Analysis	20
30	Kipper Fletez-Brant	Correlation Between Histone Mark and Gene Expression is Highly Dependent on Peak Calling Strategy	15

Methylation quantitative trait loci enhance genome-wide association study results for autism spectrum disorder across tissue type

Shan Andrews[1,2], Shannon E. Ellis[3], Kelly M. Bakulski[4], Andrew P. Feinberg[5,6], Dan E. Arking[2,3], Christine Ladd-Acosta[1,2,5], and M. Daniele Fallin[2,5,7]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health (JHSPH)
[2] Wendy Klag Center for Autism and Developmental Disabilities, JHSPH
[3] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine (JHMI)
[4] Department of Epidemiology, University of Michigan School of Public Health
5 Center for Epigenetics, Johns Hopkins School of Medicine, 855 N. Wolfe Street, Baltimore, MD 21205
6 Department of Medicine, Johns Hopkins School of Medicine, 855 N. Wolfe Street, Baltimore, MD, 21205
7 Department of Mental Health, JHSPH

Presented by Shan Andrews

Many of the SNPs previously associated with autism spec trum disorder (ASD) are intragenic and/or do not have a clear functional consequence. A potential function may be expression regulation via epigenetics. Examining methylation quantitative trait loci (meQTLs), or SNPs that appear to control DNA methylation (DNAm) levels at particular CpG sites, with respect to previously reported ASD-related variants may provide a functional context for their ASD associations. We defined ASD-related loci using the autism results of the Psychiatric Genomics Consortium (PGC), a large mega-analysis of ASD GWA studies. We defined meQTLs using joint genotype and peripheral blood DNAm data from the Study to Explore Early Development (SEED), a national multi-site autism case-control study of children aged 2-5 years. We found that ASD-related variants were enriched for meQTLs at a p-value of 0.029. We will present results detailing the extent to which this genome-wide enrichment is observed in meQTLs derived from post-mortem brain tissue. We have also identified novel ASD candidate genes via interrogation of the nature and extent of DNAm control at specific PGC-identified ASD loci. We will present several of these loci discovered via the SEED peripheral blood meQTLs, the post-mortem brain meQTLs, and cord blood meQTLs using samples from an enriched risk birth cohort (Early Autism Risk Longitudinal Investigation; EARLI). The utility of these analyses will be to further understand the contribution of meQTLs towards ASD etiology across tissue type.

Content Area: Computational Genetics, Genetic Epidemiology
Keywords: DNA methylation, methylation quantitative trait loci, autism, ASD, GWAS

Identified structural variants associated with multiple phenotypes of COPDGene African American Study Cohort

Ferdouse Begum [1], Ingo Ruczinski[2], Shengchao Li[3], Margaret M Parker[1], Jacqueline Hetmanski[1], Terri H. Beaty[1], Edwin K. Silverman[4], James Crapo[5], COPDGene Investigators

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
2 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
3 Cancer Genomics Research Laboratory (CGR), Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA
4 Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
5 Department of Medicine, National Jewish Health, Denver, USA

Presented by Ferdouse Begum

Chronic obstructive pulmonary disease (COPD) is the third leading cause of mortality in USA. Though COPD has a well-recognized environmental risk factor (i.e. cigarette smoking), recent well-powered genome-wide association studies (GWAS) have identified multiple genomic regions where single nucleotide polymorphic (SNP) markers are strongly and consistently associated with COPD risk. To thoroughly investigate the genetic architecture underlying COPD and related phenotypes, it is also important to explore the role of structural variants including copy number variants (CNVs), since they can alter gene expression and have been shown to be causal for some diseases.

We delineated CNVs using PennCNV on 9,076 COPDGene study subjects using genome-wide marker data generated using Illumina's Omni-Express array. COPDGene subjects include one-third African-American and two-thirds Non-Hispanic white adult smokers, with or without COPD. After employing rigorous quality control procedures to reduce the false positive CNV calls, we tested for association between CNV components (defined as disjoint intervals of copy number regions within racial groups) and several COPD-related phenotypes.

We detected hemizygous deletions that achieved genome-wide significance on chromosome 5q35.2, near the gene FAM153B, in tests of association with total lung capacity assessed by chest CT among African-Americans. We also detected hemizygous deletions on chromosome 3p26.1 associated with two smoking behavior related phenotypes.

Content Area: Genetic Epidemiology
Keywords: COPD, Total Lung Capacity, smoking, CNV, deletion

The distribution of ABCA4 variants in Stargardt disease from the ProgStar studies

Samantha Bomotti [1,2], Rupert W. Strauss[2,3], Hendrik P. Scholl [2], and Robert Wojciechowski [1,2]

[1] Department of Epidemiology, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA
[2] Johns Hopkins Wilmer Eye Institute, The Johns Hopkins University, Baltimore, MD, USA
3 Department of Ophthalmology, Medical University Graz, Graz, Austria

Presented by Samantha Bomotti

Background: Type 1 Stargardt disease (STGD1; OMIM 248200) is the most common juvenile retinal dystrophy. It follows an autosomal recessive mode of inheritance at the ABCA4 locus. There is no cure and it remains difficult to diagnose due to its high allelic and phenotypic heterogeneity. We have collected genotype and clinical data from 365 STGD1 patients recruited for the ProgStar studies, which aim to characterize the natural history of STGD1 (http://progstar.org/). This is the largest cohort of STGD1 patients collected to date. We report here on the distribution of ABCA4 variants in this STGD1 population.

Methods: Data were compiled from nine clinical centers across the United States and Europe. The biological assays used to identify ABCA4 variants in genetic testing laboratories depended on the technologies available at the time of patient diagnosis (from 2004 to present). Collection of the clinical data is ongoing and will allow for in-depth phenotypic and genetic analyses once completed.

Results: The majority (89.3%) of the 365 STGD1 patients were heterozygous for ABCA4 mutations. In 23 cases (6.3%), only one ABCA4 mutation was identified, suggesting incomplete coverage of ABCA4 or locus heterogeneity. The most common variant was c.5882G>A (G1961E), found in 97 (26.6%) patients. The second most common variant, c.2588G>C (G863A), was observed in 43 (11.8%) patients. The third most common variant (c.5461-10T>C) appeared in 31 (8.5%) patients. These variants are also the most frequently found in previously published cohorts.

Conclusions: Initial variant analyses in the ProgStar studies confirm the high allelic heterogeneity of STGD1. These data will help characterize the clinical impact of variants (including rare variants) on STGD1 progression and severity.

Content Area: Genetic Epidemiology
Keywords: STGD1, ProgStar, genetics, autosomal recessive

Comparison of Illumina Infinium 450K Methylation BeadChip preprocessing methods in an Epigenome Wide Association Study

Brucato M [1], Sobreira N [2], Zhang L [2], Ladd-Acosta C [1], Ongaco C [2,3], Romm J[2,3], Baker M [2], Doheny K [2,3], Bertola D[4], Chong K[4], Perez ABA[5], Melaragno M[5], Meloni V[5], Valle D[2], Bjornsson H[6]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
[2] Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
[3] Center for Inherited Disease Research (CIDR), Institute of Genetic Medicine
[4] Unidade de Genética, Instituto da Criança, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil
[5] Genetics Division, Department of Morphology and Genetics, Universidade Federal de São Paulo, Brazil
[6] Department of Pediatrics at the Johns Hopkins University School of Medicine, Baltimore, MD

Presented by M Brucato

Kabuki Syndrome (KS; MIM 147920) is a Mendelian disorder that involves the histone methylation machinery. KS is characterized by growth retardation, intellectual disability, immunological problems, and facial dysmorphology. We collected a cohort of clinically diagnosed KS patients (n=29) and age-and-sex-matched controls (n=9) from Brazil. We compared DNA methylation patterns of KS patients with histone machinery mutations (KMT2D and KMT2A) to those of matched normal controls with the Illumina Infinium HumanMethylation450 BeadChip platform, a reliable and reproducible technology for assaying DNA methylation at 485,517 loci across the genome. We applied four different preprocessing methods to the raw data, including quantile normalization, functional normalization, noob (normal-exponential using out-of-band probes), and functional normalization plus noob. An epigenome wide association study for KS phenotype was then conducted on each preprocessed dataset. Differentially methylated positions (DMPs) and differentially methylated regions (DMRs) were identified after adjustment for patient sex, blood sample cell composition, and ancestry. KS associated DMPs and DMRs were consistently discovered regardless of the preprocessing method. The observed differences in results were minor and no preprocessing method emerged as clearly superior. Thus, we have shown that the four preprocessing methods yield comparable results in this empirical dataset. Our study conclusion is robust to the preprocessing method chosen: individuals with a genetic abnormality in histone methylation have shared changes in their DNA methylation, suggesting that there is crosstalk between histone and DNA methylation.

Content Area: Genetic Epidemiology
Keywords: epigenetics, Kabuki Syndrome, DNA methylation

Whole exome sequencing and linkage analysis of patients with pulmonary nontuberculous mycobacterial infection

Fei Chen [1], Eva P. Szymanski [2], Kenneth N. Olivier[3] , Xinyue Liu[4], Hervé Tettelin[4], Steven M. Holland[2], and Priya Duggal[1]

1 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
2 Laboratory of Clinical Infectious Diseases, NIAID, NIH
3 Cardiovascular and Pulmonary Branch, NHLBI, NIH
4 Institute for Genome Sciences, University of Maryland School of Medicine

Presented by Fei Chen

Pulmonary nontuberculous mycobacterial (PNTM) infection is a rare syndrome that often affects women over 50 years of age. These women have a distinct body morphology (tall and lean with scoliosis) and clinical features including mitral valve prolapse and pectus excavatum, suggestive of an underlying genetic mechanism. To identify genetic regions harboring PNTM associated loci, we performed whole-exome sequencing (WES) on 17 cases and 21 unaffected individuals from 10 families and 57 sporadic cases recruited at the NIH Clinical Center in 2001-2013. One family was excluded from analysis due to quality control issues. Ninety-two percent of the PNTM cases were female. We conducted a genome-wide linkage study using 8,209 independent, common genetic variants on autosomal chromosomes from the WES. Using multi-point parametric linkage analysis, we identified a significant linkage region on chromosome 6q12-q16 (HLOD = 3.324) under a recessive model with risk allele frequency of 15% and 100% penetrance. A filtering approach was applied to prioritize variants in identified linkage regions that may act as recessive homozygotes or compound heterozygotes in one or more families. This approach identified several candidate genes with a variety of cellular functions that may be involved in immune response to bacterial or viral infections, or linked to certain morphologic or clinical features of PNTM.

Content Area: Genetic Epidemiology
Keywords: WES, PNTM, family-based studies

Regulatory Function of Schizophrenia-Associated Variants in CACNA1C

Nicole Eckart[1], Ruihua Wang[1], Rebecca Yang[2], Mariela Zeledon[1], Qifeng Song[3], Heng Zhu[3], David Valle[1], Andrew McCallion[1], Dimitri Avramopoulos[1,2]

[1] Johns Hopkins University, Institute of Genetic Medicine
[2] Johns Hopkins University, Department of Psychiatry
[3] Johns Hopkins University, Department of Pharmacology and Molecular Sciences

Presented by Nicole Eckart

Schizophrenia is a chronic psychiatric disorder with 60-80% heritability. Several GWAS have repeatedly identified the SNP rs1006737, an intronic variant in the gene CACNA1C, to be strongly associated with disease risk. We and others have previously shown a correlation between genotype at this SNP and steady state levels of CACNA1C mRNA in human post mortem brains, suggesting it tags a regulatory variant at this locus. Here we report on our search among variants in high linkage disequilibrium (r2 > 0.8) for those that may be functionally relevant.

We consistently observe that the risk allele of rs4765905, a SNP tagged by rs1006737, shows significantly reduced enhancer activity (p=1.5 *10-6) in dual luciferase reporter assays on human neuroblastoma SK-N-SH cells. The two alleles of this SNP also show different affinity for proteins or complexes extracted from SK-N-SH nuclei in electrophoretic mobility shift assays (EMSA). Using protein microarrays we show allele-specific binding for rs4765905 to a number of proteins, including transcription factors such as ZKSCAN5 and HR. This evidence suggest that rs4765905 might have regulatory function.

While only rs4765905 shows differences in luciferase assays, 13 of 16 SNPs examined by EMSA show allele-specific protein binding. Based on the observed electrophoretic shift, it appears that it may be the same protein complexes that bind most of these variant sequences. Our data from the protein microarrays also show that 15 of the 4215 proteins bind 4 or more of the 9 SNPs examined thus far. These observations suggest the possibility that complex interactions involving multiple SNPs in strong LD might regulate CACNA1C expression.

CCAT is an alternative CACNA1C transcript starting from the exon 46 and encodes a transcription factor hypothesized to negatively regulate CACNA1C transcription. We applied circularized chromatin conformation capture with next-generation sequencing (4C-seq) from the CCAT promoter and found interacting fragments approximately 20kb downstream of the transcription termination site in both HEK293 and SK-N-SH cells, but interaction evidence in the area of the schizophrenia associated SNPs was weak. We are currently interrogating the CACNA1C promoter by 4C-seq.

Our data help elucidate the molecular mechanism by which one of the best-supported risk loci contributes to schizophrenia through regulation of the CACNA1C.

Content Area: Human Genetics
Keywords: Psychiatric genetics, Gene regulation, eQTL

Determining Mitochondrial DNA Copy Number from Next Generation Sequencing Data

Rebecca Eggebeen[1], Foram Ashar[1], Anna Moes[1], Megan L. Grove[2], Eric Boerwinkle[2], and Dan E. Arking[1]

1 McKusick-Nathans Institute of Genetic Medicine at Johns Hopkins University School of Medicine
2 Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston
3 Center for Research on Genomics and Global Health, National Human Genome Research Institute
4 Jules Stein Eye Institute, University of California Los Angeles

Presented by Rebecca Eggebeen

The mitochondrial genome consists of 16,569 bases, and encodes 37 genes involved in oxidative phosphorylation. Multiple copies of the mitochondrial genome exist in each mitochondrion, and the number of mitochondria per cell range from the tens to the hundreds, resulting in a variable mitochondrial DNA (mtDNA) copy number between individuals. We have demonstrated that mtDNA copy number decreases with age and is associated with frailty and overall mortality.

Currently the gold standard for determining mtDNA copy number is quantitative PCR (qPCR). We are investigating methods to estimate mtDNA copy number from other types of data, including array and next generation sequencing data. Whole genome (WGS) and whole exome (WES) sequencing data is rapidly becoming widely available in many large cohort studies. Thus, the ability to estimate multiple different types of mtDNA variation, including copy number, heteroplasmy, and sequence variants, through readily-available data is highly advantageous. One large multi-center prospective study, the Atherosclerosis Risk in Communities (ARIC) study, has over 15,000 individuals with either whole-genome sequencing, whole-exome sequencing, or both available for data analysis. Here we compare several different methods of estimating mtDNA copy number from next generation sequencing data using ARIC data. The methods include: 1) ratio of average mtDNA coverage/average single copy exon coverage; 2) ratio of the number of mtDNA reads/total number of reads in a .sam file; and 3) a program by Ding et.al called mitoCalc which estimates mtDNA copy number based on the observed ratio of sequence coverage between mtDNA and autosomal DNA.

After calculating mtDNA copy number from 29 WGS ARIC samples, all three methods were highly correlated with each other, with correlation coefficients (R) ranging from 0.75 to 1. Computing times differed between the different methods, with the ratio of number of mtDNA reads/total reads requiring the least amount of time at ~20 minutes/sample. The ratio of average mtDNA coverage/average single copy exon coverage required ~60 minutes/sample and mitoCalc required ~2 hours/sample. We chose to move forward with the simple ratio of number of mtDNA reads/total reads due to its lowest amount of computing time. WGS and WES data produced highly correlated mtDNA copy number estimates (R=0.939) for 180 ARIC samples. These estimates also correlated highly with mtDNA copy number calculated from qPCR data (R=0.52 and 0.54, respectively) for these samples. In summary, these data provide evidence that a simple ratio of number of mtDNA reads/total reads from both WGS and WES provides an alternative method to qPCR for determining mtDNA copy number.

Content Area: Human genetics, Molecular genetics
Keywords: Mitochondrial DNA, Mitochondrial DNA Copy Number, Next Generation Sequencing Data

Correlation Between Histone Mark and Gene Expression is Highly Dependent on Peak Calling Strategy

Kipper Fletez-Brant[1] and Kasper Hansen[1]

[1] IGM and Biostatistics

Presented by Kipper Fletez-Brant

Correlation analysis has driven multiple recent studies of the functional impact of the profile of histone marks at putative regulatory loci and gene expression. Specifically, strength of effect of histones marks such as histone 3 lysine 4 trimethylation (H3K4me3) on gene expression has been quantified through the use of Spearman or Pearson correlation between H3K4me3 at a locus (using ChIP-Seq) and gene expression (using RNA-seq) across multiple biological samples. However, proper peak-calling methodology, controls and tests of significance in this context are relatively unexplored. In this study we have used publicly available data to explore the relationship between H3K4me3 signature at a locus and gene expression across individuals, and describe how that relationship changes as peak-calling strategies change. Moreover, we also explore tests of significance for correlation values and report on the strengths and weaknesses of prevailing approaches.

Content Area: Human Genetics
Keywords: Genetics, genomics, statistics

Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data

Jean-Philippe Fortin[1] and Kasper D Hansen[1,2]

[1] Johns Hopkins University, Department of Biostatistics
[2] McKusick-Nathans Institute of Genetic Medicine

Presented by Jean-Philippe Fortin

Analysis of Hi-C data has shown that the genome can be divided into two compartments called A/B compartments. These compartments are cell-type specific and are associated with open and closed chromatin. We show that A/B compartments can reliably be estimated using epigenetic data from several different platforms: the Illumina 450 k DNA methylation microarray, DNase hypersensitivity sequencing, single-cell ATAC sequencing and single-cell whole-genome bisulfite sequencing. We do this by exploiting that the structure of long-range correlations differs between open and closed compartments. This work makes A/B compartment assignment readily available in a wide variety of cell types, including many human cancers.

Content Area: Computational Genetics
Keywords: Methylation, Chromatin, Epigenetics, Chromosome configuration capture, Single-cell

Whole Exome Association of Rare Deletions in Multiplex Oral Cleft Families

Jack Fu[1], Terri H. Beaty[2], Alan F. Scott[3], Jacqueline Hetmanski[2], Margaret M. Parker[3], Joan E. Bailey Wilson[5], Mary L. Marazita[6], Elisabeth Mangold[7], Hasan Albacha-Hejazi[8], Jeffrey C. Murray[9], Alexandre Bureau[10], Jacob Carey[2], Stephen Cristiano[1], Ingo Ruczinski[1], and Robert B. Scharpf[11]

[1]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
[2] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
[3] Center for Inherited Disease Research and Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore MD
[4] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital
[5] Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health
[6] Department of Oral Biology, Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh
[7] Institute of Human Genetics, University of Bonn
[8] Dr. Hejazi Clinic, Damascus, Syrian Arab Republic
[9] Department of Pediatrics, School of Medicine, University of Iowa
[10] Centre de Recherche de l'Institut Universitaire en Sante Mentale de Quebec and Departement de Medecine Sociale et Preventive, Universite Laval
[11] Department of Oncology, Johns Hopkins School of Medicine

Presented by Jack Fu

GWAS results to date show that the vast majority of the heritability of common complex disease cannot be explained by common genetic variants. It is thought that a portion of the missing heritability can be explained by rare variants. Recently, several rare single nucleotide variants (SNVs) were associated with an increased risk of non-syndromic oral cleft, highlighting the importance of rare sequence variants in oral clefts. However, the extent to which rare deletions in coding regions of the genome occur and contribute to risk of non-syndromic clefts is not well understood. By sequencing the exomes of distantly related affected individuals, disease-associated rare mutational and structural changes to coding DNA can be more readily characterized than in the sequencing of a general population. To identify putative structural variants underlying risk, we developed a pipeline to detect rare hemizygous deletions in families from whole exome sequencing and statistical inference based on rare variant sharing. Among 46 multiplex families, we identified 51 regions with one or more rare hemizygous deletions. We found 43 of the 51 regions contained rare deletions occurring in only one family member. Members of the same family shared a rare deletion in only 8 regions. Shared deletions included chr13 (53,078,416 - 53,158,768bp) with nominal statistical significance ($p = 0.004$) and the short arm of chromo- some 6p involving gene DUSP22, a gene previously reported to be involved in Duane retraction syndrome and oral clefts ($p = 0.18$). We also devised a scalable global test for enrichment of shared rare deletions.

Content Area: Statistical Genetics
Keywords: DNA copy number, rare variants, oral cleft, multiplex families

Accumulation of somatic mutations in normal and cancerous tissues with age

Margaret Hoang[1,2], Isaac Kinde[1,2], Cristian Tomasett[2,3], Thomas Rosenquist[4], Arthur P. Grollman[4], Kenneth W. Kinzler[1,2], Bert Vogelstein[1,2], Nickolas Papadopoulos[1,2]

[1] Ludwig Center for Cancer Genetics and Therapeutics and The Howard Hughes Medical Institute
[2] Department of Oncology, Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland
3 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
4 Department of Pharmacological Sciences, Stony Brook University, Stony Brook, New York

Presented by Margaret Hoang

Fundamental theories in carcinogenesis, neurodegeneration, and aging evoke the accumulation of random somatic mutations in normal tissues over time. However, absence of a simple and systematic method to characterize somatic mutations in normal tissues precludes the understanding of their functional consequences. We present Bottleneck Sequencing System (BotSeqS), a next-generation sequencing method that quantitates random somatic point mutations simultaneously across the mitochondrial and nuclear genomes of normal tissues. BotSeqS combines molecular barcoding with a simple dilution step immediately before library amplification. Using BotSeqS, we determine the mutation frequencies and spectra in normal brain, kidney, and colon from a total of 34 individuals ranging from < 1 to 98 years old. We show an age-dependent and tissue-dependent accumulation of point mutations and demonstrate that the somatic mutational burden in normal tissues can vary by orders of magnitude depending on biological and environmental factors. We further show major differences between the mutational patterns of the mitochondrial and nuclear genomes in normal tissues. Lastly, we find that the mutational spectra of normal tissues were different from each other but similar to cancers of the same tissue type, suggesting that at least a subset of mutations observed in cancers reflect tissue-specific processes. This technology can provide insights into the number and nature of mutations in normal tissues and can be used to address fundamental questions about the genomes of diseased tissues.

Content Area: Human Genetics, Molecular Genetics
Keywords: DNA sequencing, human tissues, somatic mutation, aging

A variable selection method for identifying complex genetic models associated with human traits

Emily Holzinger[1], James Malley[2], Qing Li[1], and Joan Bailey-Wilson[1]

[1] NHGRI/NIH
[2] CIT/NIH

Presented by Emily Holzinger

Purpose: Standard analysis methods for genome wide association studies (GWAS) are not robust to complex disease models (e.g. multivariable models with non-linear interaction effects), which likely contribute to the heritability of complex human traits. Machine learning methods, such as Random Forests (RF), are an alternative analysis approach that may be more optimal for identifying these effects. One caveat to RF is that there is no standardized method of selecting a set of variables with a low false positive rate (FPR) while retaining adequate power.

Methods: We have developed a variable selection method called r2VIM. This method incorporates recurrency and variance estimation into RF to guide optimal threshold selection. We assess how this method performs in simulated SNP genotype data with a variety of complex effects (multiple loci with interactions and main effects).

Results: Our findings indicate that the optimal selection threshold can identify interactions with adequate detection power while maintaining a low FPR in the selected variable set. For example, the optimal VIM threshold had an average detection power of 0.80 and an average FPR of 0.11 for a model with a two-locus interaction and no main effects (i.e. a purely epistatic model). However, the optimal threshold is highly dependent on the simulated genetic model, which is unknown in biological data. To address this, we permute the phenotype and re-run r2VIM to generate a null distribution of VIMs. The results from the permuted data are used to choose a selection threshold in the non-permuted analysis by comparing FPR estimates at different VIM thresholds, which does not require knowledge of the underlying genetic model. We tested the permutation method on an array of simulated data. Our initial results show that the best balance between FPR and detection power is produced by selecting the VIM threshold with an FPR of close to 0.05 in the permuted data. Since our method is selection based (i.e. no modeling), we also implement a novel technique in r2VIM called “entanglement maps” to guide distinction of main effects versus interaction effects. We present visualizations of these results to further aid interpretation.

Content Area: Computational Genetics
Keywords: machine learning, human genetics, random forest, epistasis, variable selection

TSCAN: Pseudo-time Reconstruction and Evaluation in Single-cell RNA-seq Analysis

Zhicheng Ji[1] and Hongkai Ji[1]

[1] Department of Biostatistics, Johns Hopkins University
[2] Department of Biostatistics, Johns Hopkins School of Public Health

Presented by Zhicheng Ji

When analyzing single-cell RNA-seq data, constructing a pseudo-temporal path to order cells based on the gradual transition of their transcriptomes is a useful way to study gene expression dynamics in a heterogeneous cell population. Currently, a limited number of computational tools are available for this task, and quantitative methods for comparing different tools are lacking. TSCAN is a software tool developed to better support in silico pseudo-Time reconstruction in Single-Cell RNA-seq ANalysis. TSCAN uses a cluster-based minimum spanning tree (MST) approach to order cells. Cells are first grouped into clusters and an MST is then constructed to connect cluster centers. Pseudo-time is obtained by projecting each cell onto the tree, and the ordered sequence of cells can be used to study dynamic changes of gene expression along the pseudo-time. Clustering cells before MST construction reduces the complexity of the tree space. This often leads to improved cell ordering. It also allows users to conveniently adjust the ordering based on prior knowledge. TSCAN has a graphical user interface (GUI) to support data visualization and user interaction. Furthermore, quantitative measures are developed to objectively evaluate and compare different pseudo-time reconstruction methods. TSCAN is available at https://github.com/zji90/TSCAN and as a Bioconductor package.

Content Area: Computational Genetics
Keywords: Single-Cell RNA-seq, Single Cell, Gene Expression, Genomics, Data Mining

Abnormal Growth in children with ASD in the SEED Study

Norazlin Kamal Nor [1], M Danielle Fallin[2], Julie Hoover-Fong[3], and Terri Beaty[1]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
[2] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health
[3] Dept of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

Presented by Norazlin Kamal Nor

Introduction: Autism Spectrum Disorders (ASD) is a brain-based neurodevelopmental disorder. Risk factors in ASD include genetic, environment and gene-environment interactions. The construct of ASD is not fully understood. Elucidation of ASD phenotypes, for example abnormal growth, could increase our understanding of the biology of ASD. The aim of this study is to estimate the association between ASD and abnormal growth in the SEED Study.

Methods: The study population was from the SEED (Study to Explore Early Development), a multi-site case-control investigation to identify risk factors for ASD. The children underwent clinical assessment for anthropometric measures as part of a clinical examination for dysmorphology classification. They had measurements of height, weight and head circumference performed. One of their parents also had assessments for height and head circumference. Grow is assessed in two ways, firstly the estimation of the three modalities of growth and secondly the estimation of the genetic potential for growth. In the first analysis, regression analysis for height, weight and head circumference will be performed. For the second analysis, binomial regression for genetic growth potential will be performed.

Conclusion: Estimation of the association between ASD and abnormal growth and growth patterns in the SEED Study is hoped to add to the body of knowledge on ASD.

Content Area: Genetic Epidemiology
Keywords: autism spectrum disorders, abnormal growth, genetic potential for growth

Integrity of induced pluripotent stem cell (iPSC) derived megakaryocytes as assessed by genetic and transcriptomic analysis

Kai Kammers[1], Jeffrey Leek[1], Ingo Ruczinski[1], Joshua Martin[2], Margaret Taub[1], Lisa Yanek[2], Dixie Hoyle[3], Nauder Faraday[2], Diane Becker[2], Linzhao Cheng[3], Zack Z. Wang[3], Lewis Becker[2], and Rasika Mathias[2]

[1] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
[2] The GeneSTAR Program, Johns Hopkins School of Medicine
[3] Division of Hematology, Johns Hopkins School of Medicine

Presented by Kai Kammers

Aggregation of platelets in the blood on ruptured or eroded atherosclerotic plaques may initiate arterial occlusions causing heart attacks, strokes, and limb ischemia. Understanding the biology of platelet aggregation is important to prevent inappropriate vascular thrombosis. GWAS studies have identified common variants associated with platelet aggregation, but because they are intronic or intergenic, it is not clear how they are linked biologically to platelet function. To examine this, we are funded to produce pluripotent stem cells (iPSCs) from people with informative genotypes, and then derive megakaryocytes (MKs), the precursor cells for anucleate platelets, from the iPSCs to determine patterns of gene transcript expression in the MKs related to specific genetic variants. To this end it is essential that the iPSC-derived MKs retain their genomic integrity during production or expansion. This was examined using three alternative measures of integrity of the MK cell lines: (1) mutation rates comparing parent cell DNA to iPSC cell DNA and onward to the differentiated MK DNA; (2) structural integrity using copy number variation (CNV) on the same; and (3) transcriptomic signatures of the derived MK cells. For the genotype and CNV data, we used the HumanOmniExpressExome-8v1 array on 14 paired donor DNA - iPSC and paired iPSC - MK lines, and for the RNASeq data we extracted non-ribosomal RNA from 14 paired iPS and MK cell lines.A comparison of genotypes between matched pairs of cell lines indicated a very low rate of discordance; estimates ranged from 0.0001%-0.01%, well below the genotyping error rate (0.37% estimated from controls). No CNVs were generated in the iPSCs that got passed on to the MKs. Finally, looking specifically for genes 'turned on' in MKs following differentiation from the iPSCs, we observed the following highly biologically relevant gene sets as being highly significant ($q < 0.001$): platelet activation, blood coagulation, megakaryocyte development, platelet formation, platelet degranulation, platelet aggregation. All three approaches strongly support the integrity of the derived MK lines.

In addition, we are currently performing extensive eQTL analysis to categorize 'functional' relevance of the GWAS-identified determinants of platelet aggregation leveraging the genotype and RNASeq data. To date, our data contains MKs from 161 people with informative genotypes. Given a high genetic and transcriptomic integrity of the iPSC-derived MKs, we found several hundred cis-eQTLs in European Americans and African Americans and see a high replication between the two groups. The majority of cis-eQTLs are unique to MKs compared to other tissue types that are reported at GTExPortal.

Content Area: Statistical Genetics
Keywords: Platelets, iPSCs, Megakaryocytes, RNA-sequencing, eQTL analysis

Follow-Up and Replication Study of Caries in the Permanent Dentition

Deyana Lewis MPH, PhD, Margaret Cooper, John Shaffer PhD,′ Eleanor Feingold PhD, Robert J. Weyant DMD DrPH, Daniel McNeil PhD, Richard W. Crout DMD, PhD, Steven E. Reis MD, Steven M. Levy DDS MPH, Alexandre R. Vieira DDS, PhD, Michael M. Vanyukov PhD, Mary L. Marazita PhD

Presented by Deyana Lewis

Recent genome-wide association studies (GWAS) of permanent dentition caries have identified novel loci (AJAP1, TGFBR1, NR4A3, LYZL2, IFT88, ISL1, CNIH, BCOR, BCORL1, and INHBA) for further study. The aim of this study is to replicate these putative genetics associations in six independent studies of non-Hispanic whites and blacks. In this study, we interrogated 158 single nucleotide polymorphisms (SNPs) in 13 race- and age stratified samples from six independent studies (n =3600). All statistical analyses were performed separately for each sample, and results were combined across samples by meta-analysis. CNIH was significantly associated with caries via meta-analysis across eight adult samples, with four SNPS showing significant associations in white adults after gene-wise adjustment for multiple testing ($p < 0.001$). These results corroborate the previous GWAS study, although the functional role of CNIH in caries etiology remains unknown. BCOR also showed significant association in four SNPs, with the strongest evidence of association was observed in white adults (p=9.11E05). Mutations in this gene results in an X-linked dominant Mendelian disorder oculofaciocardiodental (OFCD) syndrome, which is responsible for several dental abnormalities. Furthermore, in adults, genetic association was observed for IFT88 in individual white samples ($p < 0.005$). IFT88 is thought to be involved in craniofacial, salivary gland and tooth development. Overall, this study strengthens that hypothesis that IFT88 influences caries risk.

Trio Random Forest: Post Analysis of Tree Structure To Reveal Interactions

Qing Li[1] and Joan E. Bailey-Wilson[1]

[1] Computational and Statistical Genomics Branch, National Human Genome Research Institute, NIH, Baltimore, MD

Presented by Qing Li

Random forests (RF) is a machine-learning method useful to detect complex interactions among genetic markers related to a disease trait based on case-control samples. Previously, we propose a new modification of the RF algorithm for trio data analysis. RF is an ensemble method, which analyzes data and summarizes results using a large number of classification trees. During the procedure, each classification tree uses a proportion of samples and a subset of predictors. An R package, rpart, has functions implementing classification tree analysis and it can be modified to accommodate different study designs by substituting its functions of classification based on a novel criterion. For ease of implementation, our method utilizes the rpart package to conduct classification tree analysis on a subset of the samples and predictors. Then our ensemble code, also written in R, summarizes results from all trees. This ensemble method has been proven to out-perform the trio logic regression using simulation data. In this work, we investigate the potential of the tree results from CART to reveal possible gene x gene (GxG) interactions. Using simulated data, we found out that the linkage disequilibrium among markers can help we find the causal factors, but at the same time, can cause high false positive results. When the interaction effect size is small, the GxG interaction is hard to be captured by the single tree in the forest.

Content Area: Statistical Genetics
Keywords: Case-parent trio, Machine learning, Random forest

Genome-Wide Interrogation of Spouse Selection Indicates Lack of Assortative Mating

Ryan Longchamps[1,2], Josef Coresh[3] and Dan Arking[1]

[1] Mckusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore MD
[2] Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
[3] Department of Epidemiology, The Johns Hopkins University, Baltimore, Maryland

Presented by Ryan Longchamps

Epidemiological data has long supported the concept of assortative mating whereby individuals tend to choose spouses from within the same sociocultural ingroup as themselves; however, the underlying genetic architecture resulting from mating preferences is less well understood. Quantifying the effects of genetic assortative mating has major implications in statistical models calculating allele frequencies, homozygosity rates and heritability estimates which assume the presence of random mating. As a result, understanding potential genome-wide architecture driven by assortative mating within populations is necessary for accurate epidemiological studies.

Several studies attempted to elucidate the genetic contribution of assortative mating, but definitive conclusions have not yet been reached. To address this question, we evaluated Affymetrix 6.0 microarray data generated from 8,016 self-identified non-Hispanic White individuals, including 1,940 spouse pairs of the Atherosclerosis Risk in Communities (ARIC) study. Population substructure was identified through the use of Principal Components Analysis (PCA) and ingroup clusters were identified via k-means clustering. The distribution of spouse pairs was determined within and between clusters. Surprisingly, in a Chi-squared test, we demonstrate no deviation from the null hypothesis of random mating between clusters ($P = 0.263$) indicating a lack of assortative mating within sociocultural ingroups of our cohort. These findings provide compelling evidence for the absence of genome-wide assortative mating architecture.

Content Area: Human Genetics, Other
Keywords: Assortative Mating

Exome array analysis of nuclear sclerosis in the Beaver Dam Eye Study

Stephanie Loomis [1], Priya Duggal[1], Alison Klein[2], Barbara Klein[3], Ronald Klein[3], and Kristine Lee[3]

[1] Johns Hopkins Bloomberg School of Public Health
[2] Johns Hopkins Medical Institutes
[3] University of Wisconsin School of Medicine and Public Health

Presented by Stephanie Loomis

Introduction: Nuclear sclerotic cataract (NSC) is the main form of age-related cataract and is one of the leading causes of blindness worldwide. We sought to better understand the genetic factors associated with increasing nuclear lens opacity from sclerosis through interrogation of rare and low frequency coding variants using exome array data.

Methods: We analyzed exomes of 1489 participants of European descent who had not undergone cataract surgery from the Beaver Dam Eye Study using the Illumina HumanExome Array. After quality control testing, we performed both single-variant regression analysis and gene-based unified burden and nonburden sequence kernel association test (SKAT-O) for association with nuclear sclerosis grade.

Results: No variants were statistically significant after correction for multiple comparisons (p<=1.4E-06) in the single-variant analysis in either rare variants (0.003<=MAF<0.01; top SNP: rs145310495, p=8.87E-05) or low frequency to common variants (MAF>=0.01; top SNP: rs7642805, p=4.82E-05). Gene-based analysis showed similar results, with one gene, RNF149, reaching suggestive significance with nuclear sclerosis grade (p=6.54E-06, significance threshold: p=3.11E-06).

Conclusions: Our study showed exonic changes may affect risk of nuclear lens opacity.

Content Area: Genetic Epidemiology
Keywords: machine Learning, variable selection, GWAS

Linkage Analyses Reveals Significant Association for Myopia

Anthony M. Musolf[1], Claire L. Simpson[1], Federico Murgia[2], Laura Portas[2], Joan E. Bailey-Wilson[1], Dwight Stambolian[3]

[1] Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, USA
[2] Institute of Population Genetics, CNR, Li Punti, Sassari, Italy
[3] Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA, USA

Presented by Anthony M. Musolf

Myopia (nearsightedness) is a condition where overgrowth of the eye causes light to focus in front of the retina, leading to blurring of distant images. It is one of the most common causes of reduced vision worldwide, affecting 1 in 4 Americans and has reached epidemic proportions in some parts of southeast Asia.

We have genotype data (SNPs and STS) from extended pedigrees with multiple individuals affected with myopia. These families come from four discrete populations: African-Americans, Caucasians, Ashkenazi Jews from New Jersey, and Pennsylvania Amish. Two-point and multi-point linkage analyses were performed on each family.

Familywise two-point LOD scores of greater than 2 were observed at chromosomes 3, 4, 6, 10, 13, 14 and 22 in individual Ashkenazi Jewish families. Cumulative LOD scores of greater than three were observed on chromosomes 1, 6, 8, and 16; with the marker on 16 having a cumulative LOD score of 7.3.

We also observed familywise two-point LOD scores greater than 2 in individual Pennsylvania Amish on chromosomes 2, 3, 9, 15, and 16. Cumulative LOD scores of greater than three were observed on chromosomes 4 and 8.

Although no two-point LOD scores higher than 2 were observed in the African-Americans, these families were smaller and therefore less informative. Nonetheless, cumulative LOD scores across families of greater than 3 were observed at 3 SNPs on chromosome 7.

Similarly, no interesting LOD signals were observed for individual Caucasian families, though cumulative LOD scores of greater than 3 were observed on chromosomes 1, 2 and 11. The signal on 11 is quite large and contains 7 markers.

Multipoint linkage analyses are ongoing and will likely be presented. Overall, this work identifies multiple interesting linkage signals for myopia across four discrete populations, using both two-point and multi-point analyses. Many of these signals are located within genes, which may be promising candidates for further study.

Differential Transcriptome Profiling of African Americans with Uncontrolled Hypertension and Chronic Kidney Disease (CKD) versus Controlled Hypertension and without CKD

Priyanka Nandakumar*[1], Adrienne Tin*[1], Eric Boerwinkle[2], Megan L. Grove[2], Josef Coresh[1], and Aravinda Chakravarti[1]

[1] Johns Hopkins University
[2] Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

Presented by Adrienne Tin

Background: Hypertension-attributed chronic kidney disease (CKD) is highly resistant to treatment in African-Americans, and contributes to racial disparity in end-stage renal disease (ESRD). In older adults (aged 70-74), African-Americans have 4-fold higher risk of developing hypertension-attributed ESRD than European-Americans. A hypothesized mechanism linking hypertension and progressive CKD is the innate immune response and inflammation. Inflammation biomarkers have been associated with kidney function decline and incident hypertension. Gene expression in peripheral blood can provide a view of the innate immune activation profile.
Aims: To identify differentially expressed genes and pathways in peripheral blood between cases with uncontrolled hypertension and CKD versus controls with controlled hypertension and without CKD in African Americans.
Methods: Study Design: Case and control pairs (N=2x15) were selected from those without diabetes and matched by age, gender, body mass index, APOL1 risk allele count, and medication use to reduce heterogeneity. Hypertension under treatment is defined as on hypertension medication and with systolic blood pressure (SBP) ≥ 145 mmHg. CKD is defined as estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73m2. Cases were selected from those with both hypertension under treatment and CKD, and controls were selected from those with blood pressure controlled by hypertensive medications (SBP < 135 mmHg and diastolic blood pressure < 90 mm Hg) and without CKD (eGFR: 75-120 mL/min/1.73m2 and urine albumin-to-creatinine ratio < 30mg/g).
RNA Sequencing and Analysis: Sequencing of mRNA was performed on HiSeq 2000 (18.7-45.1M paired-end reads/sample) and of miRNA on HiSeq 2500 (6.2-11.5M reads/sample) from whole blood, and quality filters were applied to trim reads. For mRNA analysis, alignment was performed using STAR, read counting and differential expression analyses on expressed features were performed using featureCounts and DESeq2 at the gene level (19,200 genes), and with Stringtie and Ballgown at the transcript level (75,196 transcripts). miRNA analysis was carried out using the miRDeep2 software for mapping and quantification, and DESeq2 for differential expression analyses (1,408 precursor-mature miRNA pairs). The Weighted Gene Co-Expression Network Analysis (WGCNA) R package was used to construct highly correlated gene modules, and we tested for associations between gene modules and case status. All analyses included surrogate variables to control for artifacts.
Results: No mRNA or miRNA had significant association with case status (significance: Benjamini-Hochberg adjusted p-value<0.05) on all analyzed features; however, a focused analysis of 397 highly expressed genes in the kidney from the Human Protein Atlas identified the downregulation of SMIM24, an integral membrane protein, as having significant association with case status (p=4.4x10-5). Gene co-expression network analyses produced no significant associations for mRNA, but identified significant associations of two miRNA modules with case status (p<0.006). The miRNAs with the highest connectivity in these two associated modules were miR-17-5p and let-7a-5p respectively; both were downregulated in cases. The expression of miR17 was found to be critical for nephrogenesis, and the downregulation of let-7 family has been associated with renal fibrosis. Overall, this pilot study identified encouraging results for differential gene expression in peripheral blood in cases and controls with respect to hypertension and CKD.

Content Area: Human Genetics, Genetic Epidemiology
Keywords: African-Americans, hypertension, CKD, RNA-seq

Differential Analysis of Gene and Transcript Abundance for RNA-Seq Data using STAR and HISAT Aligners

Julius S. Ngwa[1], Melissa Liu[2], Robert Wojciechowski[2,3], Don Zack[2], Terri Beaty[3] and Ingo Ruczinski[1]

[1] Department of Biostatistics, Johns Hopkins University School of Public Health
2 Johns Hopkins Wilmer Eye Institute, Johns Hopkins University School of Medicine
3 Department of Epidemiology, Johns Hopkins University School of Public Health

Presented by Julius S. Ngwa

Single-cell RNA-Seq is becoming one of the most widely used methods for transcription profiling of individual cells. Currently there are a number of algorithms available for mapping high-throughput RNA-Seq reads against a reference genome, and for quantifying the abundance of gene transcripts. The accurate characterization of these spliced transcripts is critical in determining functionality in normal and disease cells. Here we compare gene/transcript counts obtained from Hierarchical Indexing for Spliced Alignment of Transcript (HISAT2) and Spliced Transcripts Alignment to Reference (STAR) algorithms. HISAT2 implements a large set of small graph Ferragina-Manzini (FM) indexes, spanning the whole genome to enable rapid and accurate alignment of sequencing reads. The STAR aligner consists of a seed searching step and a clustering/stitching/scoring step, and is capable of mapping full-length RNA sequences. We analyzed highly parallel genome-wide expression profiles of human and mouse cells from the publicly available Gene Expression Omnibus NCBI database (Series GSE63473). We compared the Digital Gene Expression (DGE) matrix from the aligned library as well per-cell information indicating number of genes and transcripts observed. Some large differences were found in the number of transcripts between STAR and HISAT2 aligners. In particular, the gene counts tended to be higher using HISAT2 compared to STAR. The DGE matrices obtained from these aligners showed larger differences in mouse cells compared to human cells. The STAR and HISAT2 aligners provide information on the number of reads that map to a particular genomic position, but lack information on which of the overlapping transcripts they originate from. With the presence of ambiguous reads, uncertainties in counts can result in false differential expression calls of transcripts with similar isoforms in the same gene. Thus, resolving potential fragment assignment ambiguity may be an essential issue to address in RNA-seq data. As sequencing technology evolves with large and ever increasing volumes of data, there is need for ongoing improvements on sensitivity and accuracy of these aligners.

Content Area: Statistical Genetics
Keywords: Single Cell, RNA-Seq, Sequence Aligner, Digital Expression Matrix

Profiling Cell-Type Specific Epigenomic Landscapes Across Human Cortical Development and Aging

Amanda J. Price[1,2], Nikolay A. Ivanov[1], Ran Tao[1], Wei Xia[1], Joo Heon Shin[1], Nina Rajpurohit[1], Thomas M. Hyde[1,3,4], Joel E. Kleinman[1], Andrew E. Jaffe[5,6], Daniel R. Weinberger[1,2,3,4,7]

[1] Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, Maryland, USA
[2] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
[3] Department of Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
[4] Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
5 Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
6 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
7 Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

Presented by Amanda J. Price

Human brain development is guided by highly dynamic, cell type-specific patterns of gene expression that are regulated by epigenetic factors such as DNA methylation and chromatin state. Perturbations of these finely tuned patterns are thought to contribute to psychiatric disorders with neurodevelopmental underpinnings such as schizophrenia. To understand the mechanism of their underlying aberrant brain developmental trajectories, however, one must first characterize typical developmental patterns of gene expression and epigenetic state. Identifying temporally dynamic, cell type-specific regions of the normal epigenome may offer insight into critical developmental windows during which genomic regions are particularly at risk for genetic or environmental insults.

To profile the cell type-specific epigenomic landscape of normal human brain development, we have used fluorescence-activated nuclei sorting (FANS) to isolate neuronally-enriched (NeuN+) and –depleted (NeuN-) cellular populations from 26 post-mortem brains (dorsolateral prefrontal cortex) ranging from two months to 23 years of age. Using NeuN+, NeuN-, and homogenate DNA and RNA from these donors, we have conducted whole-genome bisulfite sequencing (WGBS) to profile DNA methylation patterns; Assay for Transposase-Accessible Chromatin with high- throughput sequencing (ATAC-seq) to profile chromatin accessibility; and RNA sequencing (RNA-seq) to profile gene expression. In addition, we have conducted WGBS and RNA-seq on homogenate brain tissue from 20 fetal samples ranging from late first trimester to late gestation.

Here we present preliminary data from pilot ATAC- and RNA-seq samples from the post-natal cohort. We find that in two NeuN+ ATAC-seq samples—in line with Buenrostro et al. (Nature Methods, 2013)—peaks called using MACS2 were enriched within promoter regions (43.83-54.42%). Sequencing nuclear RNA from three NeuN+ and NeuN- samples resulted in few reads mapping to splice junctions and an abundance of intronic sequence, as expected due to pre-mRNA; however, differential expression analysis between NeuN+ and NeuN- at the gene level showed an enrichment for genes involved in determining neuronal or anti-neuronal cell fate. While correlation between RNA expression and maximum ATAC peak height in promoter sequence in paired samples was weak (R^2=0.186-0.250), promoters of genes significantly up-regulated in neurons ($p<0.05$; log2 fold change >1) contained peaks with modestly yet significantly higher summits than peaks in promoters of down-regulated genes (p-value < 2.2e-16). Expanding upon this pilot dataset with additional forthcoming ATAC-seq, RNA-seq and WGBS samples will further expand our understanding of the dynamic relationship between epigenetic regulation and gene expression in a cell-type specific context over brain development, and can complement publicly available resources like the psychENCODE Consortium.

Content Area: Human Genetics, Computational Genetics
Keywords: Epigenomics, Brain development, Whole genome bisulfite sequencing, ATAC-seq, RNA-seq

Integrating Expression Quantitative Brain Loci in ASD GWAS analyses

Chang Shu[1], Christine Ladd-Acosta[2], Andrew Jaffe[3], Julie Daniels[4], Craig Newschaffer[5], Ann Reynolds[6], Diana Schendel[7], Laura Schieve[8], and M. Daniele Fallin[1]

[1] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health
[2] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
[3] Lieber Institute for Brain Development
[4] University of North Carolina, Chapel Hill, NC
[5] A.J. Drexel Autism Institute, Philadelphia, PA
[6] University of Colorado - Denver, Aurora, CO
[7] Aarhus University, Aarhus, Denmark
[8] National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta

Presented by Chang Shu

Background: Autism Spectrum Disorder (ASD) is highly heritable and there is evidence that common genetic variation plays a major role in this variability. However, genome-wide association studies (GWAS) thus far have had limited success. Genetics studies of other psychiatric disorders have shown enrichment for genetic variants that control brain expression, i.e. brain expression quantitative trait loci (eQTLs). Limiting genome-wide single nucleotide polymorphism (SNP) analyses to subsets known to be brain eQTLs (denoted "eSNPs"), and/or located in genes known to show developmental brain expression patterns, can reduce the search space allowing important association signals to separate from signals simply due to millions of tests performed.
Objectives: To perform genome-scale SNP association analyses for ASD, limited to SNPs known to be brain eSNPs or known to be located in genes expressed in early neural development. Further, to compare patterns of genome-wide association among SNP subsets defined by expression in specific brain regions.
Methods: Brain eSNPs and their proxy SNPs, based on linkage disequilibrium (LD) in 1000 genomes, were obtained from 6 published brain eQTLs studies, with annotation for 11 different brain tissue types. GWAS was performed on all SNPs, subsets of brain eSNPs, and brain tissue specific eSNPs after LD-based SNP pruning, using SNPs data from the Study to Explore Early Development (SEED) from 584 ASD cases and 725 non-ASD controls drawn from the general population. Similar annotation-based subsetting of Psychiatric Genomics Consortium (PGC) ASD SNP results are planned. Comparisons were made by examining the patterns of QQ plots.
Results: A total of 288,675 brain eSNPs were obtained after LD pruning, along with brain tissue specific eSNPs in cerebellum(3,027), frontal cortex(1,450), hippocampus(764), inferior olivary nucleus(659), occipital cortex(695) , pons(440), putamen(425), substantia nigra(359), temporal cortex(1,420), thalamus(636), and intralobular white matter(1,034). GWAS based on brain eQTLs revealed SNPs (rs7625872, rs73861956) that separated from expectation in QQ plots, while no separation was observed in the overall GWAS analysis of SEED data. The QQ patterns were also differential by subset in analyses based on eSNPs for specific brain tissues, where only QQ plots of cerebellum, frontal cortex and temporal cortex eSNP subsets showed positive separation from expectation.
Conclusion: The findings reported here are consistent with literature on the key brain regions involved in ASD etiology, namely cerebellum, frontal cortex and temporal cortex. Subsetting GWAS analysis to brain eSNPs can provide further insight on the ASD common variant signals.Content Area: Genetic Epidemiology
Keywords: Autism Spectrum Disorder, brain Expression Quantitative Loci, genome-wide association studies

Content Area: Genetic Epidemiology
Keywords: Autism Spectrum Disorder, brain Expression Quantitative Loci, genome-wide association studies

Genome specific transcriptional signatures predict differentiation biases in Human ES/IPS cells

Genevieve Stein-OBrien[1,2], Amritha Jaishankar[1], Suel-Kee Kim[1], Seungmae Seo[1], Joo Heon Shin[1], Daniel Hoeppner[1], Josh Chenoweth[1], Thomas Hyde[1,3,4], Joel Kleinman[1,3], Daniel Weinberger[1,3,4,5], Elana Fertig[6], Carlo Colantuoni[1,3,5], and Ronald McKay[1]

[1] Lieber Inst. For Brain Develop., Baltimore, MD
[2] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Univ., Baltimore, MD
[3] Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD 21205
4 Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD 21205
5 Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland
6 Oncology Biostatistics and Bioinformatics, Johns Hopkins University, Baltimore, MD

Presented by Genevieve Stein-OBrien

Predicting the effect of an individual's genetic background is key to advancing personalized medicine. To capture the mechanisms by which these effects emerge, time-course data of the transcriptome in human ES/IPS cells during pluripotency and differentiation conditions from numerous backgrounds was collected. Novel whole-genome Coordinated Gene Activity in Pattern Sets (CoGAPS) analysis of this RNA-seq data clearly separated shared developmental trajectories from unique transcriptional signatures for each individual's genome. These signatures were able to identify their respective donors in data from multiple tissues and across technical platforms, including RNA-seq of port-mortem brain, micro arrayed embryoid bodies, and publicly available datasets. Further analysis of these signatures found they were predictive of lineage biases during neuronal differentiation. Individuals whose signatures had high rankings of OTX2 and SOX21 also had enhanced induction of markers of telecephallic precursors (i.e. PAX6, WNT1). Conversely, signatures with enrichment of retinoic acid responsive genes ($p<1E-5$) corresponded to enhanced induction of hindbrain markers including HOXB1 and HOXB4 as well as OLIG2 positive cells. Further, lineage biases were consistent with early differences in morphogenetic phenotypes within monolayer culture, thus, linking transcriptional genomic signatures to stable quantifiable cellular phenotypes. Interestingly, matched single nucleotide polymorphisms (SNP) and RNA-seq from the Lieber Post-mortem Brain Collection indicate that relatedness of any two individuals transcriptional signature is independent of the race of the individuals. As many complex diseases are often attributed to the 70-90 percent of human variation found within race, the identification of signatures that define the functional rather than physical background of an individual's genome has the potential to profoundly influence understanding of human disease. For example, the relationship between pathway enrichment in these transcriptional signatures and drug response differences between individuals IPS cells may prove powerful for optimal patient stratification in clinical trials.

Content Area: Human genetics
Keywords: human ESC/iPSCs, transcriptional signatures, personalized medicine, Post-mortem brain tissue, complex disease

GWAS derived risk profile score is associated with schizophrenia only in individuals exposed to obstetric complications

Gianluca Ursini[1,2], Qiang Chen[1], Giovanna Punzi[1], Stefano Marenco[3], Richard Straub[1], Carlo Colantuoni[1], Ryota Hashimoto[4], Alessandro Bertolino[2], Daniel R. Weinberger [1,5]

[1] Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA
[2] Group of Psychiatric Neuroscience, Department of Basic Medical Science, Neuroscience and Sense Organs, Aldo Moro University, Bari, Italy
[3] Clinical Brain Disorders Branch, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA
[4] Molecular Research Center for Children's Mental Development, United Graduate School of Child Development and Department of Psychiatry, Osaka University, Osaka, Japan
[5] Departments of Psychiatry, Neurology, Neuroscience, and the McKusick Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA

Presented by Gianluca Ursini

BACKGROUND: Schizophrenia GWASs suggest that genetic risk is conferred by a large number of small effect alleles across the genome (1). Environmental factors also have a role in the pathophysiology of schizophrenia, and obstetric complications and intrauterine adversity (OCs) slightly but significantly increase risk for adult emergence of this disorder (2,3). Preliminary evidence of interactions of genes and OCs has been reported (4). Here, we test whether risk profile scores (RPSs) constructed from alleles showing association with schizophrenia (1) interact with OCs exposure in predicting case-control status.
METHODS: We analyzed the interaction between RPSs and OCs exposure in a sample of 272 healthy subjects and 228 patients with schizophrenia from the CBDB/Lieber GWAS study (all adults, white) on whom we had both GWAS genotypes and obstetrical histories. RPSs were generated as described elsewhere, using odds ratios derived from the PGC 2 datasets excluding the CBDB/LIBD dataset (1). We used different GWAS p value thresholds for selecting risk alleles (from P<0.05 to 5E-8). OCs questionnaires were completed by mothers of affected individuals and of control subjects, and were scored using the McNeil-Sjostrom Scale. Pregnancy, delivery and neonatal complications were included.
RESULTS: We first analyzed whether RPSs predict case-control status without taking into account exposure to OCs, and as expected, we found that all the RPSs, generated using different threshold for selecting risk alleles, predict case-control status (all p< 8.36e-06). OCs exposure alone did not predict case-control status (p>0.7). Strikingly, however, analysis of the interaction between OC exposure and the RPS obtained with the set of SNPs showing GWAS significant association with schizophrenia (p<5E-08) shows that OC exposure predicts case-control status (p=0.03), while RPS does not (p>0.5); moreover OCs and RPSs significantly interact in predict case-control status (p<0.01), so that only in presence of OCs exposure is the RPS associated with schizophrenia. We did not find significant interactions (all p>0.08) between OCs and RPS generated using less restrictive thresholds.
DISCUSSION: Our data suggest that the RPS obtained from SNPs showing GWAS significant association with schizophrenia interacts with OCs exposure in affecting risk for schizophrenia. More specifically, the RPS obtained from these SNPs predicts case-control status in our sample only in the presence of serious OCs exposure. Our data raise the inconvenient possibility that the weak effect sizes of these SNPs, even at the GWAS level of significance, is because they only increase risk in the context of other developmental risk factors, which are not universal among patients in these large GWAS studies.

1. Schizophrenia Working Group PGC. 2014. Nature 511(7510): 421-427.
2. Cannon M, et al. Am J Psych 159(7): 1080-1092.
3. Schmidt-Kastner R, et al. Mol Psych 17(12): 1194-1205.
4. Nicodemus KK, et al. Mol Psych 13(9): 873-877.

Content Area: Genetic Epidemiology
Keywords: risk profile score, obstetric complications and intrauterine adversity, schizophrenia, gene-environment interactions

Sequencing Analysis of Interferon Lambda Loci in individuals with Spontaneous Hepatitis C virus Clearance and Persistence

Candelaria Vergara[1], Chloe Thio[1], Margaret Taub[2], Rachel Latanich[1], Greg D. Kirk[1], Shruti Mehta[2], Andrea L Cox[1], David Thomas[1], Priya Duggal[2]

[1] Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
[2] Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA

Presented by Candelaria Vergara

Hepatitis C virus (HCV) infections affect ~170 million people and is one of the leading causes of cirrhosis and liver cancer. The infection either spontaneously is cleared or persists; persistence is more common in persons of African descent. Polymorphisms near the genes for interferon lambda (IFNL) are significantly associated with HCV persistence and enriched in persons of African descent but the causal DNA sequence and biological mechanisms are unknown. We aimed to determine the distribution of functionally annotated variants (SNVs, insertions, deletions) in and around the IFNL genes in African Americans (AA) and European Americans (EA) with HCV clearance and persistence. Whole genome sequencing was performed in 19 AA (11 clearance) and 18 EA (7 clearance) at 30X coverage using the Illumina HiSeq platform. Variants were annotated and predicted using Snpeff. We interrogated a 1Mb region on chromosome 19 (chr19:39002893-40006156) that includes the IFNL genes and tabulated the distribution of annotated variants between clearance and persistence groups.142 variants in 31 genes were observed in AA and 122 variants in EA. Only missense variants were annotated in the IFNL genes. The total number of variants was higher in the persistence group compared with the clearance group for EA (71 vs 51) and AA (74 vs 68). In EA, the Non-synonymous coding variants were more frequent in the persistence group compared with the clearance group (70 vs. 40) although this was not distinct in the AA (69 vs. 67) Most of these variants were predicted to have moderate impact. Two high impact variants were observed in the EA clearance group compared to 1 in the EA persistence group. In AA, 1 variant of high impact was identified in the clearance group, and 4 variants in the persistence group. A higher number of annotated variants (mainly non synonymous coding) were present in individuals with HCV persistence independent of ethnic group. Overall, we observed a diversity of variants with functional impact in individuals with both HCV clearance and persistence. Further analysis will help to elucidate the role of these variants in the resolution of HCV infection.

Content Area: Human Genetics
Keywords: Genetics, Interferon Lamda genes, Hepatitis C virus, Host genetics, Sequencing

Increased expression of histamine signaling genes in Autism Spectrum Disorder in postmortem human brain

Carrie Wright[1], Joo Heon Shin[1], Andrew Jaffe[1], Anindita Rajpurohit[1], Nick Brandon[2], Alan Cross[2], Thomas M. Hyde[1], Joel E. Kleinman[1] and Daniel R. Weinberger[1]

[1] Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA
[2] AstraZeneca Neuroscience iMED, Cambridge, MA, USA

Presented by Carrie Wright

BACKGROUND: The histaminergic neurotransmitter pathway plays an important role in modulating various neurotransmitter systems and has been shown to be involved in diverse brain functions. Evidence for alterations of this system has been identified in many neurodegenerative and neurodevelopmental diseases. A recent study of H3R antagonism showed promise in rescuing social deficits in a mouse model of ASD. To explore whether this system is altered in ASD we characterized the expression of histaminergic genes utilizing RNA sequencing of post-mortem brain samples in ASD subjects and matched controls.

METHODS: We evaluated RNA sequencing data of post-mortem dorsolateral prefrontal cortex (DLPFC) samples from 52 subjects (13 ASD subjects and 39 matched control subjects) using Ribosomal RNA depletion (RiboZero) with stranded-specific library preparation. Three controls were selected for each ASD subject matching for age, gender, and ethnicity. Paired-end sequencing was run on the Illumina HiSeq 2000. Tophat (v2.0.4) was used to align reads to that of the known transcripts of the Ensembl Build GRCh37.67 (and to enforce strandness for RiboZero). FeatureCounts (v1.4.4) was used to determine gene estimates. The influence of diagnosis status on all gene abundance estimates with an average mean log2 (Reads Per Kilobase of transcript per Million mapped reads (RPKM) + 1) expression value greater than or equal to 0.1 was then evaluated using a linear model regression covarying for principal components from the transcriptome-wide RNA expression values to correct for known and unknown confounders. Multiple testing correction was performed using the Benjamini Hochberg method. The expression levels of *HDC*, *HNMT*, *HRH1*, *HRH2*, *HRH3*, and *HRH4*, were then evaluated among the genome wide data. This was further evaluated using several gene set enrichment analyses.

RESULTS: We did not find a significant diagnostic effect on expression of these genes individually, but a significant elevation in the expression of *HNMT*, *HRH1*, *HRH2*, and *HRH3* as a gene set was observed. This finding was replicated using public data from 22 subjects (10 ASD cases and 12 controls) available on the Gene Expression Omnibus.

DISCUSSION: Our results suggest that histaminergic signaling may be altered in ASD. This system should be considered in future ASD studies and in therapeutic target development.

Genomic and RNA variants at the ARMS2/HTRA1 Locus

Pingwu Zhang, Julia VanBuskirk, Shannath Merbs, Don Zack

Presented by Pingwu Zhang

Age-related macular degeneration (AMD), characterized by neurodegeneration in the central part of the retina (Macular), is the principal cause of blindness among those aged over 65 in developed countries. Genetic studies have identified more than 19 gene/loci which are related to AMD. The strongest regions associated to AMD are ARMS2/ HTRA1 l (the region contains two genes of interest, ARMS2 and HTRA1) and CFH. Because both ARMS2 and HTRA1 genes are so close together, it is difficult to tell which gene is associated with AMD risk. While HTRA1 are highly conserved proteins, ARMS2 is only present in primates and there is no solid evidence for the existence of the encoded protein. The biological function of the ARMS2 gene remains unclear though about ten years have passed since it was first reported in 2005.

In order to characterize novel genomic and RNA variants in this locus and provide new hints on how to solve the puzzle. We examined the ARMS2 mRNA expression in all eye tissues and found ARMS2 mRNA has lower expression in retina. A 5A>C polymorphism in ARMS2 promoter region is associated with AMD and binds to C-Abl (Oncogene 1, Receptor Tyrosine Kinase). We found a predicted non-coding RNA (LOC105378525) overlaps with ARMS2 gene and RT-PCR and sequencing results confirmed that the RNA is expressed in the human choroid retinal pigment epithelium (RPE). These indicate that the genetic structure of the region is more complicated than we expected and may give new explanations as to how genomic and RNA variants at this locus contribute to AMD.

Content Area: Human Genetics, Molecular Genetics
Keywords: ARMS2, age-related macular degeneration (AMD), polymorphism, C-Abl

T cells in the necrotizing enterocolitis brain: how the gut disease leads to the brain developmental impairment

Qinjie Zhou[1], Chhinder Sodhi[1], and David J. Hackam[1]

[1] Department of Surgery, School of Medicine, Johns Hopkins University

Presented by Chelsea Qinjie Zhou

Necrotizing enterocolitis (NEC) is a devastating gastrointestinal disease that occurs to premature infants with high morbidity and mortality. Previously, we discovered a primary factor for the disease occurrence was a major Th17 lymphocytes influx into the NEC intestine, correlated with reduction in T regulatory (Treg) cells. Further, neonates with NEC have significantly higher tendency for neurodevelopmental impairment. We hypothesized that such damage was due to the abnormal inflammatory response, similar to the lymphocytes induced gut damage. To test this hypothesis, we examined the cytokine profile of NEC brain by real-time qPCR. We found a significant increase of Th17 related pro-inflammatory cytokines such as IL-22, IL-23 and IL-6, elevated pro-apoptosis markers such as PUMA and Bax, and a decrease of anti-inflammatory cytokine such as Foxp3 in the NEC brain. Correspondingly, we examine the presence of Th17 lymphocytes by flow cytometry gating on CD4+IL-17+ cells, and indeed we identified such cell population in the NEC brain. We conclude that similar to Th17 lymphocyte influx into the gut tissue in NEC, there was a subsets of Th17 cells present in the NEC brain. Furthermore, there were elevated pro-inflammatory and pro-apoptosis markers expressions, with a significant reduction of anti-inflammatory cytokines in the NEC brain.

Content Area: Pathogen Genetics, Molecular Genetics
Keywords: Necrotizing enterocolitis, Th17 lymphocytes, Neurodevelopment, cytokines