CHARGE Investigator Meeting

St. Louis 2019

Working Group Abstracts

Poster Blitz and Reception

Thursday, June 27th

5:00 – 7:00PM

Vote for the best poster:

http://www.chargeconsortium.com/stlouis

Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Investigator Meeting
St. Louis, 2019

Poster 'blitz' participants are denoted with the symbol: ⚡

Travel award recipients are denoted with the symbol: 🚀

	Abstract title	**Author**
1	The Value of Diversity in Genetic Studies	Kristin Young, PhD, MS-CR
2	Evidence for novel human height loci using rare genetic variant aggregation tests in the Trans-Omics for Precision Medicine (TOPMed) program	Mariaelisa Graff
3	TOPMed Whole Genome Sequence Analysis of Waist-to-Hip Ratio and Waist Circumference	Elise Lim
4	Height as a Causal Risk Factor for Atrial Fibrillation: A Mendelian Randomization Study	Michael Levin, MD
5⚡	Genotyping On ALL Patients (GOALL); bringing GWAS to the clinic	Jeroen van Rooij
6⚡	PheWAS of genetics of serum urate and MR of genetic determinants of uric acid and CKD & Hypertension	Elvis Akwo, MD PhD
7⚡	A retrospective landscape of drug prescriptions and clinically actionable genetic variants in Penn Medicine patient population	Anurag Verma
8	Genetic studies in eMERGE network and UK Biobank offer new insights into pleiotropy across cardiovascular diseases and central nervous system disorders.	Xinyuan Zhang, Bachelor of Science
9	Multi-ancestry genome-wide association meta-analysis of the electrocardiographic PR interval	Patricia B Munroe, PhD
10⚡	Multi-ethnic genome-wide association study of decomposed cardioelectric phenotypes illustrates strategies to identify and characterize evidence of shared genetic effects for complex traits	Antoine Baldassari, MSPH
11⚡	Sex differences in age trajectories of DNA methylation age acceleration in the Rhineland Study	Dan Liu, Ph.D student
12⚡🚀	Genome-wide interaction analysis with smoking behavior identifies 25 novel blood pressure loci in UK Biobank	Ioanna Ntalla, PhD
13⚡🚀	Multi-ancestry analysis of gene-sleep interactions in 126,926 individuals identifies multiple novel blood lipid loci that contribute to our understanding of sleep-associated adverse lipid profile	Raymond Noordam, Ph.D.
14⚡	Multi-ancestry genome-wide meta-analysis accounting for gene-education interactions in up to 227,850 individuals identifies several novel lipid loci	Yun Ju Sung, Ph.D.
15	Colocalization of TOPMed Whole Genome Sequencing Analysis and Tissue-Specific eQTL Signals Detects Target Genes for Type 2 Diabetes Risk	Mindy Szeto, BS
16⚡	Whole Genome Sequence Analysis of Type 2 Diabetes Risk in the TOPMed Study	Sridharan Raghavan, MD, PhD

17⚡	Novel CpG sites of glucose and insulin homeostasis: an integrative cross-omics analysis	Jun Liu, MD, MSc
18⚡	Meta-analysis of Epigenome-wide Association Study of Serum Urate Levels	Adrienne Tin, Ph.D
19⚡🚀	Association of mitochondrial DNA copy number with cardiometabolic traits: Results from NHLBI's TOPMed program	Xue Liu, M.S., MPH
20	Mitochondrial DNA heteroplasmic mutations are under autosomal genetic control	Chunyu Liu
21⚡	Genome-wide association study of plasma total tau levels in the Framingham Study	Chloé Sarnowski, Ph.D.
22⚡	Interpretable Neural Networks for Schizophrenia Risk Prediction based on Whole Exome Sequencing Data	Arno van Hilten, ir.
23⚡	The genetic architecture of MRI-derived extremes of cerebral small vessel disease	Aniket Mishra, PhD
24⚡	Investigation of interactions between sugar-sweetened beverage consumption and CHREBP genetic variants on triglyceride and HDL-C concentrations	Danielle E. Haslam, M.A., M.S.

Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Investigator Meeting
St. Louis, 2019

(1)

The Value of Diversity in Genetic Studies

Kristin L Young (1), Genevieve L Wojcik (2), Mariaelisa Graff (1), Katherine K Nishimura (3), Ran Tao (4), Jeffrey Haessler (3), Christopher R Gignoux (1), Heather M Highland (1), Yesha M Patel (5), Stephanie Bien (3), Steven Buyske (6), Chris Haiman (5), Charles Kooperberg (3), Loic Le Marchand (7), Ruth JF Loos (8), Tara C Matise (6), Ulrike Peters (3), Eimear E Kenny (8), Christopher S Carlson (3), Kari E. North (1)

1. University of North Carolina at Chapel Hill, Chapel Hill, NC; 2. Stanford University, Stanford, CA; 3. Fred Hutchinson Cancer Research Center, Seattle, WA; 4. Vanderbilt University Medical Center, Nashville, TN; 5. Keck School of Medicine, University of Southern California, Los Angeles, CA; 6. Rutgers University, New Brunswick, NJ; 7. University of Hawaii, Honolulu, HI; 8. Icahn School of Medicine at Mount Sinai, New York, NY

Genome-wide association studies (GWAS) have laid the foundation for investigations into the biology of complex traits, drug development, and clinical guidelines. However, the majority (86%) of discovery efforts have been conducted in populations of European ancestry. Given existing differential genetic architecture between populations, bias in representation in research studies can hamper discovery efforts and exacerbate health disparities. Critical variants that are low frequency or absent in European populations will be missed, especially as the field shifts its attention towards rare variants, which are more likely to be population-specific. Additionally, effect sizes and risk prediction scores derived in one population may not be applicable to other populations with different linkage disequilibrium (LD) patterns.

To address these issues, we describe recent work by the Population Architecture using Genomics and Epidemiology (PAGE) study, demonstrating the value of diverse participants in large-scale genomic studies. PAGE conducted a GWAS of 26 clinical and behavioral phenotypes (including adiposity and other cardiometabolic traits) in 49,839 non-European individuals. Using analytic strategies designed for multi-ethnic and admixed populations, we identified 27 novel loci and 38 secondary signals at known loci, as well as replicated 1,444 GWAS catalog associations across these traits. PAGE data also revealed evidence of effect-size heterogeneity across ancestries for published GWAS associations, substantial benefits for fine-mapping using diverse cohorts, and insights into clinical implications. Given that US minority populations have a disproportionately higher burden of chronic conditions, the persistent lack of diversity in genetic research will result in inequitable access to precision medicine for those with the highest burden of disease. We strongly advocate for continued, large genome-wide efforts in diverse populations to maximize genetic discovery and reduce health disparities.

(2)

Evidence for novel human height loci using rare genetic variant aggregation tests in the Trans-Omics for Precision Medicine (TOPMed) program

Misa Graff (1), Wenjian Bi (2), Jennifer A. Brody (3), and Leslie A. Lange (4), on behalf of the TOPMed Anthropometry and Adiposity Working Group

1. University of North Carolina, Chapel Hill, NC; 2. University of Michigan, Ann Arbor, MI; 3. University of Washington, Seattle, WA; 4. University of Colorado-Denver, Aurora, CO.

Background: Genome-wide association studies for human height have identified >800 associated variants, jointly explaining approximately one-fifth of the heritability for this trait. Whole-Genome Sequencing allows us to more thoroughly study the cumulative effects of uncommon and rare coding variation on height. Methods: Using data from 40,018 multi-ethnic participants from the Trans-Omics for Precision Medicine (TOPMed) program, we assessed associations between gene-based aggregations of nonsynonymous and likely deleterious variants with minor allele frequency [MAF] <5% and height. Analyses were performed using height that was inverse-normalized by study, sex and ancestry group. Models were further adjusted for ancestry PCs, age, and age-squared. We used a linear mixed model algorithm (SAIGE-Gene), which accounts for the degree of relatedness among participants, to perform SKAT-O tests. Results: Three gene-based tests were significant (P<2.5x10-6): ACAN (Chr. 15; P=9.4x10-12), ADAMTS10 (Chr. 19; P=3.3x10-9), and MATN3 (Chr. 2; P=1.2x10-6). While common (MAF>5%) variants in ACAN have been previously associated with height, the SKAT-O test remained significant (P=6.1x10-10) after conditioning on the lead common (MAF>5%) variant in this gene. ADAMTS10 is responsible for the recessive form of Weill-Marchesani syndrome, a rare connective tissue disorder characterized by short stature. Individual ADAMTS10 variants have recently been identified for body fat distribution in the UK Biobank. An uncommon MATN3 variant was reported to be associated with height in a large-scale ExomeChip meta-analysis. MATN3 is one of several genes associated with the rare genetic syndrome Dominant Multiple Epiphyseal Dysplasia (MED). MED is characterized by skeletal malformations, and adult height in the lower range of normal. Conclusions: We observed evidence for aggregations of uncommon and rare nonsynonymous/likely deleterious variants to be associated with height. Evidence for association in ACAN remained after conditioning on the lead common variant. We plan to further investigate the genetic architecture of these loci with additional TOPMed samples.

(3)

TOPMed Whole Genome Sequence Analysis of Waist-to-Hip Ratio and Waist Circumference

Elise Lim (1), Jennifer A. Brody (2), Kari E. North (3), L. Adrienne Cupples (1,4), Ching-Ti Liu (1)

1. Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118; 2. Cardiovascular Health Research Unit, School of Medicine, University of Washington, Seattle, WA; 3. Department of Epidemiology and Carolina Center of Genome Sciences, Chapel Hill, NC 27514; 4. Framingham Heart Study, National Heart, Lung, and Blood Institute and Boston University, Framingham, MA 01702

Body mass index (BMI), a simple weight-for-height index, is typically used to determine obesity status, but it is not an accurate way to measure one's risk for obesity since body fat distribution can vary significantly from person to person with the same BMI. Body fat distribution is a strong and important risk factor for metabolic dysfunctions and obesity-related diseases. To further enhance our understanding of the genetic basis of body fat distribution, we performed whole genome sequence association analyses of waist-to-hip ratio (WHR) and waist circumference (WC) within the Trans-Omics for Precision Medicine Program (TOPMed) for both common and rare variants.

Analyses were performed on 28,303 individuals from 12 cohorts. Using linear mixed models, we adjusted for age, age2, sex, BMI, ethnicity, cohort, and the first 10 principal components and then inverse normal transformed the residuals. To reduce computational burden, we first fit null models that exclude the genotypes, while adjusting for familial relatedness with an empirical kinship matrix. All subsequent analyses were done using GENESIS in R package.

Single variant analyses were performed on variants with minor allele count of at least 25. At genome-wide significance of 5e-8, we confirmed three known loci associated with WHR (ZRANB3, p=5.3e-8; RSPO3, p=1.54e-8; KNTC1, p=4.73e-8) and three known loci for WC (JAZF1, p=5.63e-10; VEGFA, p=1.47e-8; ADAMTSL3, p=4.44e-8). SKAT was used to perform rare variant analyses, which included variants with minor allele frequency of 1% and those annotated as either loss of function, missense, or protein altering indels. We identified the same novel locus, AMMERCR1L, for WHR (p=2.59e-8) and WC (p=8.46e-6), although the association for WC was not genome-wide significant.

Our study replicated previously known genes associated with WHR and WC and discovered one novel locus that we plan to investigate further.

(4)

Height as a Causal Risk Factor for Atrial Fibrillation: A Mendelian Randomization Study

Michael G. Levin, MD (1); Renae Judy (4); Regeneron Genetics Center (6); Daniel J. Rader (1,3); Benjamin F. Voight, PhD (2,3); Scott M. Damrauer, MD (4,5)

Departments of 1. Medicine, 2. Systems Pharmacology and Translational Therapeutics, 3. Genetics, 4. Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 5. Philadelphia VA Medical Center, Philadelphia, PA; 6. Regeneron Genetics Center, Tarrytown, NY

Introduction: Observational studies have identified an association between increasing height and increased risk of atrial fibrillation (AF), although the mechanism and causal nature of this relationship have remained unclear. Recently, genome-wide association studies (GWAS) have identified genetic variants associated with height (>700 loci) and atrial fibrillation (>100 loci). Mendelian randomization (MR) enables the use of genetic variants, randomly assorted during meiosis, as instrumental variables to estimate the causal effect of increased height on risk of atrial fibrillation.

Methods: Using summary statistics from the 2018 meta-analysis of the UK Biobank and GIANT Consortium GWAS of height (N ~700,000 individuals), we constructed a genetic instrument for height consisting of 464 independent variants that were also present in the 2018 AFGen multi-ethnic GWAS of atrial fibrillation (65,446 AF cases and >522,000 referents). Two-sample MR, including inverse-variance weighted, MR-Egger, MR-PRESSO, and weighted median studies were performed. Individual-level phenome-wide association study (PheWAS) and instrumental variable analysis using a genetic risk score for height was performed in a sample of 7023 individuals of European ancestry enrolled in the Penn Medicine Biobank, adjusted for clinical and echocardiographic risk factors for AF.

Results: In inverse-variance weighted meta-analysis, a 1 standard deviation increase in height was associated with increased odds of atrial fibrillation (OR 1.33; 95% CI 1.28-1.4; p = 4.1x10-37). These findings remained robust to sensitivity analysis. Individual-level PheWAS confirmed observational associations between height and AF, and MR confirmed a strong relationship between height and AF (OR 1.13; p = 2.5x10-10), robust to adjustment for clinical and echocardiographic risk factors.

Conclusion: Using Mendelian randomization, height appears to be a strong, causal risk factor for atrial fibrillation. These results raise the possibility of investigating height/growth-related pathways as a means for gaining novel mechanistic insights, as well as the possibility of height-based targeted screening strategies for atrial fibrillation.

(5)

Genotyping On ALL Patients (GOALL); bringing GWAS to the clinic

Jeroen van Rooij(1), Annemieke Verkerk(1), Linda Broer(1), Jard de Vries(1), Gaby van Dijk(1), Erwin Bovenkamp(1), Ron van Schaik(2), Robert Hofstra(3), Joyce van Meurs(1), Andre Uitterlinden(1)

1. Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands. 2. Department of Clinical Chemistry, Erasmus MC, Rotterdam, the Netherlands. 3. Department of Clinical Genetics, Erasmus MC, Rotterdam, the Netherlands.

Genetic testing is increasingly used in clinical practice. Up to now, this is restricted to a limited set of common variants (using PCR methods) and rare variant analysis in families (using whole exome sequencing). Array-based genotyping allows for highly consistent high-throughput variant detection at a much lower cost. In addition, the many and increasingly bigger mega-GWAS have resulted in composition of so-called polygenic risk scores (PRS). We will investigate the implementation of array-genotyping on all patients at Erasmus MC. We will start by answering the following main questions: 1) which individual variants currently used in clinically testing can be replaced with a genotyping array: 2) how might PRS provide additional clinically relevant information: 3) what infrastructure and knowledge is needed to return these results to the clinic.

First, we will compare array data with diagnostic WES-data in 1,000 patients with a diagnosed rare variant. We will investigate which (rare) disease-causing mutations in these patients detected by WES can also be reliably observed through arrays. This technical validation will also be done using data from the Rotterdam Study population (n=1,000). These comparisons will determine how well the arrays perform on rare variant detection, and which classes of variants retain in the realm of WES detection.

For the second question, we will generate disease specific PRS's in three pilot clinical populations: breast cancer, age-related macular degeneration and psychiatric patients. Together with the clinical departments, we will investigate when and how these PRS contribute to medical decision-making and can be implemented in daily practice. This same will be done in the healthy parents in the patient population from the Clinical Genetics Department and retrospectively in all 12,000 participants of the Rotterdam Study cohort to validate the PRS and investigate the patient's interest in return of results.

The third question will be investigated in parallel, to determine the required infrastructure and knowledge for clinical implementation. This also includes discussion surrounding the ethical consideration in reporting polygenic scores and the clinical populations that might benefit from (selected) large-scale reporting of array-based genetic information.

At the CHARGE St. Louis Meeting we will present an overview of the GOALL project and preliminary pilot results. We believe these results might be interesting to CHARGE collaborators as personalized risk profiling and returning of genetic results are relevant in all (patient) population cohorts.

(6)

PheWAS of genetics of serum urate and MR of genetic determinants of uric acid and CKD & Hypertension

Elvis Akwo (1,2), Hua-Chang Chen (1,2), Ran Tao (1), Eric Torstenson (1), Otis Wilson (2), Ayush Giri (1), Cecilia Chung (1,2), Cassianne Robinson-Cohen (1,2), Adriana Hung (1,2)

1. Vanderbilt University Medical Center, Nashville, TN 2. Tennessee Valley Healthcare System, Nashville, TN

We performed a phenome-wide association study (PheWAS) to identify disease outcomes associated with SUA genetic risk loci. The following outcomes were identified in the PheWAS: gout and inflammatory arthropathy, osteoporosis, kidney stones, Chronic kidney disease and renal failure, peripheral vascular disease, ischemic heart disease, hypertension non specified, hypertensive renal and heart disease, metabolic disorders such as obesity, mixed hyperlipidemia, nonalcoholic liver disease and cirrhosis, splenomegaly, hereditary hemochromatosis, idiopathic peripheral neuropathy, mixed hyperlipidemia and hypercholesterolemia, diabetes with and without complications, diabetic retinopathy, hypothyroidism and testicular hypofunction. The association with diabetes was only observed with HNF1A rs1800961 and the GCKR rs1260326. We present the pheWAS to each variant and build a series of genetic risk scores for medelian randomization for the outcomes hypertension and kidney disease.

(7)

A retrospective landscape of drug prescriptions and clinically actionable genetic variants in Penn Medicine patient population

Anurag Verma Ph.D.(1), Shefali S. Verma Ph.D(1), Marjorie Risman, M.A.(1), Anastasia M. Lucas B.A (1), Scott Damrauer MD(1), Dan Rader MD(1), Abigail T. Berman MD MSCE(2), Joshua Lubin, MBA(2), David Roth MD Ph.D.(2), Teri Klein, Ph.D. (3), Sony Tuteja Pharm.D, MS BCPS(4), Marylyn D. Ritchie Ph.D(1)

1. Department of Genetics, University of Pennsylvania, Philadelphia PA 19104; 2. Penn Medicine, University of Pennsylvania, Philadelphia PA 19104;3. Stanford University,4. Department of Medicine, University of Pennsylvania, Philadelphia PA 19104

One of the main goals of personalized medicine is to identify the right patient to give a right course of treatment at the right time. Pharmacogenomic markers are proven to help in identifying the right patients who may harbor genetic polymorphisms that affect the efficacy of one or multiple drugs or may also have genetic variants that cause adverse drug reactions. Penn Medicine is one of the primary health care providers in Philadelphia and surrounding areas of Pennsylvania. It consists of EHR data on approx. 3.6 million patients and 19,131 of these patients have enrolled in the Penn Medicine biobank (PMBB). They have been genotyped for research purposes and all of their genetic data is linked with EHR. Implementing pharmacogenomics in a health system requires the knowledge of prescribing trends that might be specific to the patient population or the patients recruited into studies. For our pilot analyses, we extracted data from October 2011- December 2017. For each patient in each year, we extracted information on 48 CPIC Level A drugs to identify the prescription pattern of these drugs. Further, we used PharmCAT, a bioinformatics tool to annotate pharmacogenomic variants utilizing a set of pre-specified annotations of CPIC Level A genes. Notably, we observed that in 2016 alone over 50,000 patients are prescribed more than one CPIC level A drug. Approx. thirty-one thousand unique patients were prescribed warfarin, and 25,000 patients have prescribed clopidogrel in Penn EHR between 2011 and 2016. In the genotyped population, approximately 96.67% of samples have at least one non-referent CPIC level A variant. PharmCAT annotation also identified that 19.10% of samples are carriers of one or more PGx actionable variants. For example, we specifically looked at gene-drug interactions for warfarin (CYP2C9 and CYP4F2) and clopidogrel (CYP2C19) and found 253 samples that are poor metabolizer for warfarin and 403 samples that contained polymorphisms for adverse drug reactions such as mycoses, acute coronary syndrome, coronary artery disease and myocardial infarction due to clopidogrel. Our analyses on 19,131 genotyped individuals demonstrate that the frequency of patients containing polymorphisms in one or more clinically actionable variants is very high. Our study aimed at highlighting essential considerations for clinicians and researchers for implementing pharmacogenomics in clinical practice.

(8)

Genetic studies in eMERGE network and UK Biobank offer new insights into pleiotropy across cardiovascular diseases and central nervous system disorders.

Xinyuan Zhang(1), Yogasudha Veturi(1), Anurag Verma(1), Theodore G. Drivas(1), eMERGE consortium, Marylyn D. Ritchie(1)

1. Department of Genetics and Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA

Cardiovascular diseases and central nervous system disorders are two leading causes of death worldwide and frequently co-occur in patients. Development of effective disease treatment requires knowledge of the genetic nexus of multiple diseases in these categories. The contribution of pleiotropy remains largely unknown given that previous studies mostly focused on single diseases for genetic variant discovery. Here, we go beyond the Genome-wide association study framework to Phenome-wide association studies (PheWAS) and multi-trait joint association studies to identify pleiotropic genetic variants. A comprehensive set of phenotypes was curated for eMERGE and UK Biobank using ICD codes from the electronic health record. Following quality control, we analyzed 61 cardiovascular diseases and 28 central nervous system disorders in 43,015 European adults with 7 million SNPs from the eMERGE network dataset. We performed a GWAS-PheWAS for these 89 diseases, as well as a single multi-trait joint association test and compared the results. Both methods are able to identify previously known disease-associated SNPs, such as rs429358 with Alzheimer's disease and rs1333049 with coronary artery disease. Among 565 unique Bonferroni significant SNPs identified by PheWAS, 359 of them have shown Bonferroni significance by multi-trait joint analyses. We conducted replication studies in 294,753 European adults with 619,865 SNPs from UK Biobank. There were 82 Bonferroni significant SNPs that have been replicated by both PheWAS and MultiPhen, which were mapped to HLA, LPA, CDKN2B-AS1, NECTIN2, TOMM40, APOE and APOC1 regions. We further performed a formal test of pleiotropy using sequential multivariate analyses to pinpoint the exact associated diseases. We characterized novel variants that showed significance across cardiovascular diseases and central nervous system disorders. Our research discovered novel pleiotropy across a comprehensive list of cardiovascular diseases and central nervous system disorders, which will improve our understanding of disease etiology and may assist in drug repositioning in future research.

(9)

Multi-ancestry genome-wide association meta-analysis of the electrocardiographic PR interval

I. Ntalla1, L. Weng2,3, P.B. Munroe1,4* , S.A. Lubitz2,3,5, on behalf of the CHARGE EKG consortium

William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK 2 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 3 Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA 4 National Institute for Health Research Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK 5Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, Massachusetts, USA

The electrocardiographic PR interval reflects atrioventricular conduction, and it is associated with conduction abnormalities, pacemaker implantation, atrial fibrillation, and cardiovascular mortality. We conducted a multi-ancestry genome-wide meta-analysis for PR interval (N=293,051), discovering 210 loci, of which 149 are novel. Genetic variation at all loci nearly doubled the percentage of heritability explained, from 33.5% to 62.6%. We report enrichment for genes involved in cardiac muscle development/contraction and the cytoskeleton, and highlight key regulation processes for atrioventricular conduction. Additionally, there is variation at many genes underlying monogenic forms of heart disease. We observed genetically determined PR interval prolongation is an endophenotype for cardiovascular disease risk, including distal conduction disease, atrial fibrillation, atrioventricular pre-excitation, non-ischemic cardiomyopathy, and coronary heart disease. These findings advance our understanding of the polygenic basis of cardiac conduction, and the genetic relationship between PR interval duration and cardiac arrhythmias.

(10)

Multi-ethnic genome-wide association study of decomposed cardioelectric phenotypes illustrates strategies to identify and characterize evidence of shared genetic effects for complex traits

Antoine R. Baldassari(1), Colleen M. Sitlani(2), Heather M. Highland(1), Dan E. Arking(3), Steve Buyske(4), Dawood Darbar(5), Rahul Gondalia(1), Misa Graff(1), Xiuqing Guo(6, 7), Susan R. Heckbert(8), Lucia A. Hindorff(9), Chani J. Hodonsky(1), Yii-Der Ida Chen(6, 7), Robert C. Kaplan(10), Ulrike Peters(11), Wendy Post(12), Alex P. Reiner(11), Jerome I. Rotter(6, 7), Ralph V. Shohet(13), Nona Sotoodehnia(8), Ran Tao(14), Kent D. Taylor(6, 7), Jie Yao(6, 7), Eimear E. Kenny(15, 16, 17, 18), Henry J. Lin(6, 7), Elsayed Z. Soliman(19), Eric A. Whitsel(1), Kari E. North(1, 20), Charles Kooperberg(11), Christy L. Avery(1)

1. Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill NC ; 2. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA ; 3. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland ; 4. Department of Statistics & Biostatistics, Rutgers University, New Brunswick, NJ ; 5. Department of Medicine, University of Illinois at Chicago, Chicago, Illinois ; 6. Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California ; 7. Department of Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA ; 8. Cardiovascular Health Research Unit, Division of Cardiology, Department of Medicine, University of Washington, Seattle, WA ; 9. Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD ; 10. Albert Einstein College of Medicine, Bronx, NY ; 11. Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA ; 12. Departments of Medicine and Epidemiology Johns Hopkins University, Baltimore, MD ; 13. Center for Cardiovascular Research, John A. Burns School of Medicine, Honolulu, HI ; 14. Department of Biostatistics, Vanderbilt University, Nashville, TN ; 15. Center for Genomic Health, Icahn School of Medicine at Mount Sinai, New York City, NY ; 16. Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY ; 17. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York City, NY ; 18. Department of Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY ; 19. Epidemiological Cardiology Research Center (EPICARE), Wake Forest School of Medicine, Winston-Salem, NC ; 20. Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill NC

Background: Published genome-wide association studies (GWAS) remain European-centric, examine a narrow view of phenotypic variation, and infrequently interrogate evidence of genetic effects shared across traits. We therefore examined the degree to which a multi-ethnic, combined phenotype GWAS of phenotypes that map to well-defined biology enabled the detection and characterization of complex trait loci.

Methods: Using 1000 Genomes Phase 3 imputed data in n=34,668 participants (15% African American; 3% Chinese American; 51% European American; 30% Hispanic/Latino), we performed covariate-adjusted univariate GWAS of six contiguous electrocardiogram (ECG) traits that decomposed an average heartbeat and two commonly reported composite ECG traits that summed contiguous traits. Combined phenotype testing was performed using the adaptive sum of powered scores test (aSPU).

Results: We identified six novel and 87 known ECG trait loci (paSPU<5×10-9). Lead SNP rs3211938 at novel locus CD36 was common in African Americans (minor allele frequency=10%) and near-monomorphic in European Americans, with effect sizes for composite trait QT interval among the largest reported. Only one novel locus was detected for the composite traits, reflecting varying directions of effects across contiguous traits that decreased to near-zero when summed. Combined phenotype testing did not detect novel loci unapparent in univariate testing, although this approach aided locus characterization, particularly when loci harbored multiple independent signals that differed by trait

Conclusions: In an era of mega GWAS in predominantly European ancestral populations, this study, conducted in a population one-third the size of the largest published ECG trait GWAS, underscores the merits of prioritizing diversity and phenotype measurement.

Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Investigator Meeting
St. Louis, 2019

(11)

Sex differences in age trajectories of DNA methylation age acceleration in the Rhineland Study

Dan Liu (1), Loes C. Rutten-Jacobs (1), Monique M.B. Breteler (1,2)

1. Population Health Sciences, German Centre for Neurodegenerative Diseases (DZNE), Bonn, Germany; 2. Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Bonn, Germany

Background: Biological aging as estimated through DNA methylation patterns (DNAmAge), is associated with age-related phenotypes. Previous studies showed faster epigenetic aging in men than women. How these sex differences evolve during chronological aging is not well characterized. We examined sex differences in DNAmAge acceleration across a wide age range in a population-based cohort.

Methods: This cross-sectional analysis is embedded within the ongoing Rhineland Study, a population-based prospective cohort study in Bonn, Germany, that recruits individuals aged ≥ 30 years. DNA methylation levels were measured in buffy coat using the Illumina MethylationEPIC BeadChip. DNAmAge was calculated using Horvath and Hannum algorithms. DNAmAge acceleration (AgeAccelHorvath, AgeAccelHannum) was calculated as residuals from regressing DNAmAge on chronological age. We accessed DNAmAge acceleration by sex using the t test and by sex and 10-year age categories using analysis of variance.

Results: We included 1735 participants (median age 54.0 years (range 31 - 91), 56.0% women). Overall, men (mean= 0.64, SD= 4.23) had significant higher AgeAccelHorvath than women (mean= -0.50, SD= 3.82), with similar pattern in AgeAccelHannum (men: mean= 1.14, SD= 3.93; women: mean= -0.89, SD= 3.87). Sex differences increased with increasing age only in AgeAccelHannum (sex*age (p=0.001). AgeAccelHorvath was low in the youngest agegroup (30-39: mean=-0.77, SD=3.56), increased significantly with age before 70 years old (40-49: mean=0.42, SD=3.96; 50-59: mean=0.28, SD=3.67; 60-69: mean=0.49, SD=4.41) and then declined (70-79: mean=-0.67, SD=4.23; > 80: mean=-0.60, SD=5.35). The same pattern was seen for AgeAccelHannum, but this was only statistically significant for the 50-59 year olds (30-39: mean=-0.77, SD=3.59; 40-49: mean=0.21, SD=3.69; 50-59: mean=0.61, SD=3.83; 60-69: mean=0.09, SD=4.55; 70-79: mean=-0.28, SD=4.41; >80: mean=-1.28, SD=3.98).

Conclusion: Men showed higher rates of DNAmAge acceleration than women, and this sex gap widened with increasing age. DNAmAge acceleration varied across age groups with the highest acceleration in midlife.

(12)

Genome-wide interaction analysis with smoking behavior identifies 25 novel blood pressure loci in UK Biobank

Ioanna Ntalla (1^), Karen Schwander (2^), RJ Waken (2), Stefan Van Duijvenboden (1), Mary Feitosa (3), Amy Bentley (4), Yun Ju Sung (2), and Patricia B. Munroe (1), on behalf of the Gene-Lifestyle Interactions Working Group*

Joint contributions^; *This work is supported by NHLBI grant HL 118305

1. Clinical Pharmacology, Queen Mary University of London, UK; 2. Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA; 3. Division of Statistical Genetics, Washington University School of Medicine, St Louis, MO, USA; 4. Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

Introduction: Smoking is a known risk factor for high blood pressure (BP), but it remains unclear how an individual's genetic profile interacts with smoking behavior to affect that risk. We investigated whether accounting for gene-smoking interactions can help identify novel BP loci.

Methods: We performed genome-wide interaction analyses with two smoking measures: "Current Smoker (yes/no)" and "Ever Smoker (yes/no)". Analyses were performed on systolic BP, diastolic BP, and pulse pressure in 421,264 unrelated European UK Biobank participants. Novel variants at $p<1\times10\text{-}6$ were taken forward into replication using the Gene-Lifestyle Interactions WG smoking-BP results. We used a joint two-degree of freedom test of the variant main and interaction effects. A locus was considered replicated if it was significant ($p<5\times10\text{-}8$) in the combined analysis, with evidence in the replication ($p<0.05$) and same direction of effects.

Results: We identified and replicated 25 novel BP loci; for 6 loci there was evidence ($p<0.05$) for contribution of interactions. Newly identified loci include several promising genes, including ISL1 and ANK3 (discovery driven with contribution from interactions), both involved in cardiac development, and DLC-1 (driven by main effects), a small GTPase involved in RhoA signaling, a known BP homeostasis pathway.

Conclusions: 25 novel BP loci were identified in this study accounting for gene-smoking interactions in the UK Biobank. African-ancestry analyses and investigation for independent signals driven by interactions in known loci are ongoing. Our findings may further promote our understanding of the complex relationship between genes and the environment, and its role in BP regulation.

(13)

Multi-ancestry analysis of gene-sleep interactions in 126,926 individuals identifies multiple novel blood lipid loci that contribute to our understanding of sleep-associated adverse lipid profile

Raymond Noordam (1), Maxime M Bos (1), Heming Wang (2), Thomas W Winkler (3), Amy R Bentley (4), Tuomas Kilpeläinen (5,6), Susan Redline (2), On behalf of the CHARGE Gene-Lifestyle Interactions Working Group

1. Leiden University Medical Center, Leiden, the Netherlands; 2. Bighams and Women's Hospital, Boston, MA, USA; 3. University of Regensburg, Regensburg, Germany; 4. National Institute of Health, Bethesda, MD, USA; 5. University of Copenhagen, Copenhagen, Denmark; 6. The Icahn School of Medicine at Mount Sinai, New York, NY, USA

Introduction: Habitual short and long sleep duration are associated with disturbances in lipid profile. We conducted multi-ancestry genome-wide sleep-SNP interaction analyses on three lipid traits (triglycerides, LDL-c, and HDL-c) to provide insights into the biology of sleep-associated adverse lipid profile.

Methods: A discovery multi-ancestry meta-analysis was conducted in 62,457 individuals (13,046 as short sleepers and 12,317 as long sleepers). Cohort-specific GWAS analyses were conducted and ancestry-specific joint meta-analyses were performed for analysis of SNP main effects and the SNP-sleep duration interaction effects on serum lipid levels together (2 df joint tests in METAL). They were subsequently combined into trans-ancestry analyses. Replication analyses were performed in 64,469 individuals (12,952 as short sleepers and 12,834 as long sleepers). Finally, discovery and replication results were combined for increased power.

Results: Using P<5E-7 in the discovery phase, P<0.05 in replication phase with similar direction of effect, and P<5E-8 in the combined analysis, we identified 49 novel lipid loci (23 for triglycerides, 12 for LDL-c, and 14 for HDL-c) when considering either long or short total sleep time in the analyses (n=126,926). Collectively, these novel lipid variants explained up to 4.25% of the total variance for triglycerides, 1.00% for LDL-c, and 0.38% for HDL-c. None of the novel loci identified with short sleep were identified with long sleep, and vice versa. An additional 10 novel lipid loci were identified in separate analyses of Europeans only (n=87,653; 3 for triglycerides, 2 for LDL-c, and 5 for HDL-c).

Conclusion: We identified different novel lipid loci with short and long sleep with zero overlap, suggesting different biological mechanisms involved in short- and long-sleep associated lipid disturbances. These novel loci have been shown to be involved in adiposity (e.g., FHIT, MAGI2), inflammation (ZNF827, NR5A2) or psychosocial traits (FHIT, SNX13), collectively contributing to our understanding of sleep-associated adverse lipid profile.

(14)

Multi-ancestry genome-wide meta-analysis accounting for gene-education interactions in up to 227,850 individuals identifies several novel lipid loci

Yun Ju Sung(1) Karen L. Schwander(1), Tuomas Kilpeläinen(2), Hugues Aschard(3), Lisa de las Fuentes(1,4)

1. Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA; 2. Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 3. Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, 75724 Paris, France; 4. Department of Medicine, Cardiovascular Division, Washington University School of Medicine, St Louis, MO, USA

Introduction: High- and low-density lipoprotein cholesterol (HDL-C and LDL-C) and triglycerides (TG) are influenced by genetic and lifestyle factors. Educational attainment is among the most widely-used indices of socioeconomic status in epidemiologic studies.

Methods: To identify novel lipid loci, we conducted a gene-education interaction study for HDL-C, LDL-C, and TG in European (N=173,885), African (N=18,925), Asian (N=18,310), Hispanic (N=13,077), and Brazilian (N=3,653) ancestry groups. We considered 2 education variables: “Some College” (yes/no) and “Graduated College” (yes/no). In Stage 1 (N=110,319), genome-wide analyses were performed at ~15 million variants imputed using the 1000 Genomes Project reference panel. All suggestive variants (P<10-6) in Stage 1 were followed up in Stage 2 (N=117,531). We used a joint two degrees of freedom test of genetic main and interaction effects.

Results: Stage 1 identified 13,851 significant and 6,835 suggestive variants. The combined analyses of Stages 1 and 2 identified 110 known and 27 novel loci (P<5x10-8). Novel loci identified for several lipid traits include genes with plausible biology. Among these is LRP1B, previously associated with insulin resistance and body mass index and overexpressed in the hypothalamus of high-fat fed mice. PTPRE plays a role in high-fat diet-induced obesity, leptin sensitivity, and glucose homeostasis. SLC1A3 and NPTX2 are linked with excitatory neurotransmitter signaling in the central nervous system. GRIN2B is suggested to play a role in mediating a crosstalk between fat and memory in the hippocampus.

Conclusions: We identified several novel lipids loci with plausible biology through a multi-ancestry study accounting for interactions between genetic variants and educational attainment. Our findings may elucidate the role and underlying mechanism of education interactions in the genetic regulation of lipid levels.

(15)

Colocalization of TOPMed Whole Genome Sequencing Analysis and Tissue-Specific eQTL Signals Detects Target Genes for Type 2 Diabetes Risk

Mindy Szeto (1), Heather Highland (2), Alisa Manning (3), on behalf of the TOPMed Diabetes Working Group

1. University of Colorado, Aurora, CO; 2. University of North Carolina, Chapel Hill, NC; 3. Broad Institute, Cambridge, MA

Type 2 diabetes (T2D) risk is highly heritable, and although many associated loci have been identified, analyses of regulatory regions in most whole-genome sequencing (WGS) studies have not been specific to relevant tissues. The availability of tissue-specific expression quantitative trait (eQTL) information can further elucidate the functional role of non-coding regulatory variants. A WGS (>38x sequencing depth) association study was performed for variants in 9,663 cases and 35,050 controls from the NHLBI's Trans-Omics for Precision Medicine (TOPMed) program. eQTL data from 48 different tissues was obtained from the Genotype-Tissue Expression (GTEx) Portal. To generate loci of interest, 200-kb windows were considered upstream and downstream of each variant found to be genome-wide significant ($p < 5x10^{-8}$) in both the TOPMed and GTEx datasets. A total of 11 regions in different tissues were significantly associated with T2D status, centered around variants such as rs6585201 in the ascending aorta (TCF7L2, TOPMed $p = 1.18x10^{-9}$, GTEx $p = 1.12x10^{-10}$, but did not localize to the strongest signal in the region), rs76895963 in the cerebellum (CCND2, TOPMed $p = 3.51x10^{-9}$, GTEx $p = 4.40x10^{-17}$), and rs4898431 in ovarian tissue (DUSP9, TOPMed $p = 1.88x10^{-8}$, GTEx $p = 4.39x10^{-8}$). We plan to implement eCAVIAR to examine causal variant colocalization, with the advantages of leveraging summary statistics and accounting for the possibility of more than one causal variant per locus. These results confirm previously published studies, but provide validation of methods we will extend to larger samples and diabetes-related outcomes.

(16)

Whole Genome Sequence Analysis of Type 2 Diabetes Risk in the TOPMed Study

Heather M. Highland (1), Timothy Majarian (2), Jennifer Wessel (3), Alisa Manning (2), Sridharan Raghavan (5) on behalf of the TOPMed Diabetes Working Group

1. University of North Carolina, Chapel Hill, NC; 2. Broad Institute, Cambridge, MA; 3. University of Indiana, Indianapolis, IN; 4. University of Colorado, Aurora, CO.

Type 2 diabetes (T2D) is heritable, with recent analyses in European ancestry populations identifying several hundred common (minor allele frequency [MAF] ≥5%) and low frequency (0.5% ≤ MAF < 5%) single nucleotide polymorphisms (SNPs) associated with disease risk. We conducted a whole genome sequence (WGS) association study of common, low frequency, and rare (MAF < 0.5%) variants in 9,663 individuals with T2D and 35,050 controls of diverse ancestry from NHLBI's Trans-Omics for Precision Medicine (TOPMed) program. We tested associations of single variants and sets of rare variants with T2D, without and with adjustment for body mass index (BMI). Rare variant sets were defined on pancreatic islet-specific regulatory annotation. In single variant analyses, 15 genomic regions were associated ($p<5x10-8$) with T2D; seven were novel (ODF2L, LMAN2, KCNV1, CBR1, VLDLR-AS1, LINC01052 and NKX2-5), and results did not differ by BMI adjustment. All novel T2D-associated single variants were non-coding and were either ancestry-specific or rare. For example, a variant near KCNV1 projected to increase T2D risk was present only in African ancestry individuals (rs11992463, MAF 5.5%, $p=1.6x10-8$, OR=1.43). Set-based testing of rare variants identified 10 regions associated with T2D, most of which were ancestry-specific. For example, a set of variants near the CBR1 gene containing predicted protein-truncating and regulatory variants was associated with T2D in individuals of African ancestry (cumulative minor allele count [cMAC]=623, $p=5.9x10-9$). Examination of the contribution of each variant to the variant-set statistic identified variants driving the T2D association. For example, a set of variants in an active enhancer on chromosome 16 near OSGIN1 was associated with T2D in Asian ancestry individuals (cMAC=48, $p=1.7x10-6$), driven by a single rare variant not meeting genome-wide significance (MAC=29, $p=6.9x10-5$, OR=2.74). In conclusion, large-scale WGS analysis in diverse samples extends our understanding of the contribution of common and rare, ancestry-specific, genetic variation to T2D risk.

(17)

Novel CpG sites of glucose and insulin homeostasis: an integrative cross-omics analysis

Jun Liu (1), Elena Carnero-Montoro (2), Jenny van Dongen (3), Ayse Demirkan (4), Cornelia M. van Duijn (1,5)

1. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands; 2. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 3. Department of Biological Psychology, Amsterdam Public Health (APH) research institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands; 5. Nuffield Department of Population Health, Oxford University, Oxford, UK

Background: Despite the growing evidence that differential DNA methylation is implicated in type 2 diabetes (T2D) and obesity, our understanding of the functional relevance of the phenomenon remains limited. Methylation of DNA in the circulation is one of the key feature of obesity. A key question to answer is whether there are functional effects of the differential methylation relevant for the pathogenesis of T2D and whether these are dependent or independent of obesity.

Purpose: We used a cross omics integrative analysis to understand the effect of DNA methylation in the early phases of T2D pathology while accounting for body mass index (BMI).

Methods: We performed a blood-based epigenome-wide association study (EWAS) of fasting glucose and insulin among 4,808 non-diabetic European individuals and replicated the findings among 11 cohorts summing up to 11,750 trans-ethnic non-diabetic individuals, mainly (58 %) from European ancestry. We integrated blood-based in silico cross-omics databases comprising genomics, epigenomics and transcriptomics collected public resource.

Results: In the discovery phase, we identified and replicated nine novel differentially methylated sites in whole blood (P-value < 1.27 × 10-7): sites in LETM1, RBM20, IRS2, MAN2A2 genes and 1q25.3 region were associated with fasting insulin; sites in FCRL6, SLAMF1, APOBEC3H genes and 15q26.1 region were associated with fasting glucose. Follow-up in silico cross-omics analyses reveals that the cis-acting methylation quantitative trait loci (meQTLs) near SLAMF1 and its expression in blood are involved in glucose level regulation in the circulation. Moreover, we find that differential methylation in FCRL6 may affect glucose level and the risk of T2D by regulating FCLR6 expression in the liver. In silico cross-omics analyses highlight that differential methylation plays a key role in the crosstalk between the adaptive immune system and glucose homeostasis. When we adjusted all the insulin-related CpG sites in the association analysis of BMI and insulin, the beta for BMI reduces by 16.9% (beta: 0.065 vs beta: 0.054).

Conclusions: We identify nine novel DNA methylation sites associated with glucose homeostasis and provide new insights into the genetics, epigenetics and transcriptomics of T2D by the integration of cross-omics data in silico. The study shows that the interplay between obesity, differential methylation and insulin metabolism is complex and may explain at least 16.9% of the association between obesity and insulin.

(18)

Meta-analysis of Epigenome-wide Association Study of Serum Urate Levels

Adrienne Tin (1), Pascal Schlosser (2), Franziska Grundner-Culemann (2), Pamela Matias (3), Antoine Weihs (4), Laura Smyth (5), Anselm Hoppmann (2), Jan Bressler (6), Megan Grove (6), Allan Gelber (1), Alexander Teumer (4), Anna Kottgen (1, 2)

CKDGen EWAS studies

1. Johns Hopkins University; 2. University of Freiburg; 3. Helmholtz Zentrum München (the German Research Center for Health and Environment); 4. University of Greifswald; 5. Queen's University Belfast; 6. University of Texas Health Science Center at Houston

Background. Uric acid has adverse and protective associations with human disease. Notably, high serum urate levels cause gout, the most common form of inflammatory arthritis in adults. Yet, urate might also have antioxidant function as evidenced by its protective association with Parkinson disease. DNA methylation is a major mechanism of gene regulation and can be influenced by the environment. Understanding the relation between DNA methylation and serum urate levels in blood cells may reveal regulatory pathways of potential urate functions.

Methods. We conducted an epigenome-wide association study of serum urate and DNA methylation in whole blood or monocytes quantified by the Illumina Infinium HumanMethylation27, 450K, or EPIC BeadChips. Each study conducted race-specific analysis of two models: model 1 (M1) adjusting for age, sex, genetic principal components, DNA methylation batch effects, and cell type composition. Model 2 (M2) additionally included known correlates of serum urate levels. Study-specific results were adjusted for p-value inflation using the Bayesian BACON method followed by inverse variance weighted fixed effect meta-analysis of all studies and European-ancestry (EA) studies. Epigenome-wide significance was set at 1.1x10-7(=0.05/440185).

Results. Data from 12,478 participants (16 studies) were combined: 6,428 EA, 2,645 African-American, 2,720 South-Asian, and 685 African individuals, and 440,185 CpG sites common to the 450K and EPIC arrays were analyzed. P-value inflation was modest (BACON estimates, trans-ethnic, M1: 1.06, M2: 1.05; EA, M1: 1.08, M2; 1.04). The trans-ethnic meta-analysis detected 477 significant CpG sites in M1 and 140 in M2. The EA meta-analysis detected 143 significant sites in M1 and 31 in M2. Significant CpG sites contain plausible biological candidates and will now be replicated in independent samples.

Conclusion. Trans-ethnic epigenome-wide study of serum urate identified significant DNA methylation sites. With replication and functional follow-up, these findings may reveal novel insights into the relationship between serum urate and transcriptional regulation.

(19)

Association of mitochondrial DNA copy number with cardiometabolic traits: Results from NHLBI's TOPMed program

Xue Liu(1), Thomas Blackwell(2), Jessica L. Fetterman(3), Dan E. Arking(4), Ryan Longchamps(4), Nathan Pankratz(5), Jerome I. Rotter(6), Jun Ding(7), Goncalo Abecasis(2), Aldolfo Correa(8), Ramachandran S Vasan(9), Daniel Levy(9), James Wilson(8), Chunyu Liu(10)

1. Department of Anatomy & Neurobiology, Boston University, Boston, MA; 2. Informatics Research Center, University of Michigan, Ann Arbor, MI; 3. Department of Medicine, Boston University, Boston, MA; 4. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 5. Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 6. Departments of Pediatrics and Medicine, Harbor-UCLA Medical Center, Torrance, CA; 7. Laboratory of Genetics and Genomics, NIA/NIH, Baltimore, MD; 8. Jackson Heart Study, University of Mississippi, Jackson, MS; 9.Framingham Heart Study, Framingham, MA; 10.Department of Biostatistics, Boston University, Boston, MA

Mitochondria are the powerhouses of eukaryotic cells and they generate energy through oxidative phosphorylation (OXPHOS). Each cell contains many copies of mitochondrial DNA (mtDNA), which encodes several key OXPHOS proteins. Abnormal mtDNA copy number (CN) may cause impaired energy production and thus result in disease. Previous studies reported inconsistent relations between mtDNA CN and several cardiometabolic disease (CMD) risk factors. The goal of this study was to investigate the association of mtDNA CN with age, sex, and several CMD risk factors using whole genome sequencing (WGS) through the NHLBI's Trans-Omics for Precision Medicine (TOPMed) program. About 4100 Framingham Heart Study (FHS) and 3400 Jackson Heart Study (JHS) participants obtained WGS through TOPMed using whole-blood derived DNA. mtDNA CN was calculated as twice the average coverage of mtDNA divided by the average coverage of nuclear DNA. In both cohorts, mtDNA CN slightly increased before middle age (51-60 years), and decreased after middle age. Women had slightly higher mtDNA CN compared to men. mtDNA CN was associated with white blood cell counts and was subject to substantial batch effects. mtDNA CN residuals were obtained after adjusting for age, age-squared, sex, and batch effects. Association analyses were performed between mtDNA CN residuals and CMD traits. Covariates included sex, age, age-squared (blood pressure traits only) and BMI (not for BMI/obesity as outcomes). In meta-analysis of both cohorts, a decrease in mtDNA CN was associated with higher BMI (Pmeta-analysis=0.003), higher systolic blood pressure (Pmeta-analysis=0.02), and higher prevalence of hypertension (Pmeta-analysis=0.0005) and diabetes (p=0.02). In contrast, a decrease in mtDNA CN was associated with lower total cholesterol (Pmeta-analysis=0.002) and lower low-density lipoprotein cholesterol (Pmeta-analysis=0.0002). Replication of our results in additional cohorts is warranted. Understanding the role of mtDNA CN in CMD may provide insights into the etiology of a range of age-related diseases.

(20)

Mitochondrial DNA heteroplasmic mutations are under autosomal genetic control

Chunyu Liu(1), Jessica L. Fetterman(2), Yong Qian(3), Thomas Blackwell(4), Ryan Longchamps(5), Dan E. Arking(5), Jerome I. Rotter(6), Aldolfo Correa(7), Ramachandran S Vasan(8), Goncalo Abecasis(4), Daniel Levy(8), James Wilson(7), Jun Ding(3)

1.Department of Biostatistics, Boston University, Boston, MA; 2.Department of Medicine, Boston University, Boston, MA; 3.Laboratory of Genetics and Genomics, NIA/NIH, Baltimore, MD; 4.Informatics Research Center, University of Michigan, Ann Arbor, MI; 5.McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 6.Departments of Pediatrics and Medicine, Harbor-UCLA Medical Center, Torrance, CA; 7.Jackson Heart Study, University of Mississippi, Jackson, MS; 8.Framingham Heart Study, Framingham, MA

The maternally inherited mitochondrial DNA (mtDNA) encodes for key bioenergetic genes and displays a high mutation rate. mtDNA is present in thousands of copies per cell. mtDNA mutations frequently coexist alongside inherited alleles at the same mtDNA loci within the cell, a phenomenon called heteroplasmy. The mechanisms by which heteroplasmic mutations are propagated in somatic tissues are poorly understood. We aimed to investigate whether the burden of heteroplasmy was associated with nuclear genetic variants using whole genome sequencing through the NHLBI's Trans-Omics for Precision Medicine (TOPMed) program. In sequencing data, an alternative allele of a site refers to a different allele when compared to the reference allele of the revised Cambridge Reference Sequence. We defined a site to be a heteroplasmic mutation if the proportion of alternative allele (PAA) was between 0.03 and 0.97 in an individual. We quantified heteroplasmy burden using two scores. The first score (Si1) sums the number of heteroplasmic mutations in individual i. The second was a weighted score (Si2), in which Si1 was weighted by the PAA of each heteroplasmic variant. Genome-wide association testing of the two scores in the Framingham Heart Study (FHS) and Jackson Heart Study (JHS) yielded the same significant genetic region at 11p11.12. Meta-analysis of the FHS and JHS identified common variants at $p<5e-8$ using both Si1 and Si2. This genetic locus encodes for several long non-coding RNAs and contains strong cis quantitative genetic loci for DNA methylation (mQTLs) and gene expression (eQTLs) based on FHS data and QTL databases. Our results indicate that mtDNA heteroplasmic mutations are under autosomal genetic control. Future studies will focus on functional inference for this region. Understanding how the nuclear genome regulates mitochondrial heteroplasmic mutations may provide insights into a range of age-related diseases.

(21)

Genome-wide association study of plasma total tau levels in the Framingham Study

Chloé Sarnowski(1), Claudia L. Satizabal (2,3,4), Vincent Chouraki(5), Matthew P. Pase(6,7), Charles DeCarli(8), L. Adrienne Cupples(1,2), Ramachandran S. Vasan(2,9), Alexa S. Beiser(1,2,3), Anita L. DeStefano(1,2,3), Sudha Seshadri(2,3,4)

1. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 2. Boston University and the NHLBI's Framingham Heart Study, Boston, MA, USA; 3. Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 4. Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, TX, USA; 5. Lille University, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related diseases; Lille, France; Labex Distalz; 6. Centre for Human Psychopharmacology, Swinburne University of Technology, Hawthorn, Australia; 7. Melbourne Dementia Research Centre, The Florey Institute for Neuroscience and Mental Health and the University of Melbourne, Australia; 8. Department of Neurology and Center for Neuroscience, University of California at Davis, CA, USA; 9. Preventive Medicine & Epidemiology, Boston University School of Medicine, Boston, MA, USA

Introduction: Few GWAS were conducted on Amyloid beta (Aβ) and tau levels, two biomarkers of Alzheimer Disease (AD), and they were limited by their sample sizes and inability to explore less common genetic variants.

Methods: We performed a GWAS of plasma total tau (t-tau) levels, measured using an ultrasensitive assay, in 6,018 Framingham Study participants using Haplotype Reference Consortium imputations. We used a linear mixed-effects model adjusted for age, sex and principal components and accounting for familial relatedness. We explored the potential pleiotropy between t-tau and three measures of plasma Aβ levels (Aβ42 or Aβ40 dosages and Aβ42/Aβ40) using bivariate analyses.

Results: We detected at the genome-wide level ($P<5x10^{-8}$) a new locus on 1p36 (rs34683021, MAF=0.26, $P=2.8x10^{-8}$) in the intron of PANK4, a gene overexpressed in the brain that belongs to a family of proteins associated with neurodegeneration. We confirmed the association of the 17q21 MAPT locus (rs242557, MAF=0.36, $P=9.9x10^{-93}$) and identified an additional distinct signal in ARL17A using conditional analyses (rs4630592, MAF=0.35, $P=1.3x10^{-50}$). We found associations ($P\leq0.05$) in loci previously related to tau (CSF p-tau and t-tau, plasma t-tau), AD, and AD-related traits (AD age of onset, cognitive decline, total ventricular and whole-brain volume, hippocampal atrophy, and small vessel stroke). Finally, we uncovered pleiotropic effects for common and low-frequency genetic variants located in C10orf11 (AD), CREB5 (cognition), LINGO2 (Parkinson Disease), ELAVL1 (AD), GOSR2 (myoclonus atrophy, epilepsy and ataxia), GLIS1 (CSF tau and AD) and PKD1L2 (calcium channel) on t-tau and Aβ levels.

Conclusions: Our GWAS reveals new variants and genes associated with plasma t-tau levels and identifies potential pleiotropic effects with Aβ in brain-related genes. Studying relevant AD biomarkers in genetic analyses can help to uncover new genes and pathways that may contribute to AD-related processes and be promising targets for developing preventive and therapeutic interventions for AD.

(22)

Interpretable Neural Networks for Schizophrenia Risk Prediction based on Whole Exome Sequencing Data

A. van Hilten (1,2), S.A. Kushner (3), W.J. Niessen (1,2,4), G.V. Roshchupkin (1,2,5)

1. Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands; 2. Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands; 3. Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands; 4. Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands; 5. Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands

Objective: Schizophrenia is a highly heritable mental disorder with a complex etiology. Recent studies have associated more than 100 different loci with schizophrenia, demonstrating the complex polygenic nature of this psychiatric disorder. Modern methods, such as neural networks, allow exploration of complex data with higher-level dependency. Applying such an approach might yield novel insights into the etiology and genetic risk of schizophrenia.

Methods: The prediction model was built by training a neural network using a Swedish case control study with 4969 cases and 6245 controls, involving ~1,2 million exome variants. Variant data was reduced to 21,390 genes using a special network layer developed for omics data analysis.

Results: The model reached an area under the curve of 0.65 and an accuracy of ~60% in the test set, with a theoretical maximum ~74% upper limit of performance calculated on the basis of the prevalence of schizophrenia and the concordance rate in monozygotic twins. The interpretability of the network allowed us to identify the genes most crucial for the prediction of schizophrenia versus control. The five most discriminant genes were: HLA-C, TTN, TRY2P, HLA-A and LINC00226.

Conclusion: We developed an interpretable neural network architecture for exome data, which incorporates prior biological knowledge (i.e., gene annotations), and can be easily extended to include genomic, tissue, cell type, or other functional annotations. Given that every node in the network is interpretable, we anticipate this approach as having the potential for uncovering novel insights into the genetic architecture of complex diseases, such as schizophrenia.

(23)

The genetic architecture of MRI-derived extremes of cerebral small vessel disease

Aniket Mishra(1), Dina Vojinovic(2), Hideaki Suzuki(3), Josh Bis(4), Nuno R Zilhao(5), Xueqiu Jian(6,7), Qiong Yang(8,9), Wei Zhao(10), Sarah Harris(11), Yasaman Saba(12), Shuo Li(8), Murali Sargurupremraj(1), Traci M Bartz(4), Edith Hofer(13,14), Ganesh Chauhan(1,15), Fen Peng(6), Thomas H Mosley(16,17), Bruce M Psaty(18,19,20,21), Lenore J Launer(22), Eric Boerwinkle(6), Yicheng Zhu(23), Bernard Mazoyer(24), Mark Lathrop(25), Anne Joutel(26), Christophe Tzourio(1,27), Jennifer A Smith(28), Michelle Luciano(11), Reinhold Schmidt(13), M Arfan Ikram(2,29), Helena Schmidt(12), Vilmundur Gudnason(5), Paul Matthews(3), W T Longstreth Jr(30), Myriam Fornage(6), Cornelia Van Duijn(2,31), Sudha Seshadri(7,32), Stephanie Debette(1,32,33)

1.Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team VINTAGE, UMR 1219, F-33000 Bordeaux, France; 2.Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands; 3.Department of Medicine, Imperial College London, UK and UK Dementia Research Institute; 4.Deparment of Biostatistics, University of Washington, Seattle, WA, USA; 5.Icelandic Heart Association, Icelandic Heart Association, Kopavogur 201, Iceland; 6.The University of Texas Health Science Center at Houston, Houston, TX, United States; 7.Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, Texas, USA; 8.Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 9.Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 10.Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, 48109; 11.Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK; 12.Gottfried Schatz Research Center, Department of Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria; 13.Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Austria; 14.Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria; 15.Centre for Brain Research, Indian Institute of Science, Bangalore, India; 16.Division of Geriatrics, School of Medicine, University of Mississippi Medical Center, Jackson, MS, USA; 17.Memory Impairment and Neurodegenerative Dementia Center, University of Mississippi Medical Center, Jackson, MS, USA; 18.Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; 19.Department of Epidemiology, University of Washington, Seattle, WA; 20.Department of Health Services, University of Washington, Seattle, WA, USA; 21.Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA; 22.Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 23.Department of Neurology, Peking Union Medical College Hospital, Beijing, China; 24.University of Bordeaux, Institut des Maladies Neurodégénératives, CNRS-CEA UMR 5293, France; 25.University of McGill Genome Center, Montreal, Canada; 26.Genetics and Pathogenesis of Cerebrovascular Diseases, INSERM, Université Paris Diderot-Paris 7, DHU NeuroVasc, Sorbonne Paris Cité, Paris, France; 27.CHU de Bordeaux, Pole de santé publique, Service d'information médicale, F-33000 Bordeaux, France; 28.Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109; Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104; 29.Department of Radiology and Nuclear Medicine, Erasmus MC, University Medical Center, Rotterdam, the Netherlands; 30.Department of Neurology and Department of Epidemiology, University of Washington, Seattle, WA, USA; 31.Nuffield Department of Population Health, University of Oxford, Oxford, UK; 32.Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 33.CHU de Bordeaux, Department of Neurology, F-33000 Bordeaux, France

Background: To perform a gene-mapping study of MRI-markers of cerebral small vessel disease (SVD) using a composite extreme phenotype design combining extreme distributions of white matter hyperintensity (WMH) burden and presence or absence of lacunes (extreme-SVD). The genome-wide association studies (GWAS) reported common variants in five loci to be associated with WMH burden in the general population and in four additional loci in stroke patients.

Methods: 17 population-based cohorts of middle-aged to older persons with available MRI measurements and genome-wide genotyping (N=19,056), whole exome sequencing (WES, N=5,098) or exome chip (EC, N=12,666) data contributed to this study. We performed meta-analyses of GWAS, gene-based tests (using VEGAS2) and multi-trait association studies (jointly with WMH burden, MRI-defined brain infarcts and stroke using MTAG) for common variant gene-mapping of extreme-SVD. We also used WES/EC data to test association of exonic variants with extreme-SVD. We performed partitioned heritability analyses and SMR/HEIDI tests for functional characterization of extreme-SVD genetic associations.

Results: We identified 12 genome-wide significant associations at 1q21.2, 2p16.1, 2p21, 2q32.1, 2q33.2, 6q25.1, 7q36.1, 10q24.33, 12q24.11, 14q32.2, 16q24.1, and 17q25.1. The VEGAS2 gene-based test identified one additional association at 17q21.3. Gene-based burden tests of rare and low frequency variants identified PHACTR4 ($p=1.96\times10\text{-}5$) as the top gene. The candidate study of five familial SVD genes identified significant association of common variants in COL4A1/2 (rs9515201, $p=9.43\times10\text{-}4$) and HTRA1 (rs2293871, $p= 2.22\times10\text{-}5$), and burden of rare and low frequency variants in NOTCH3 ($p=0.017$) with extreme-SVD. The partitioned heritability analyses showed extreme-SVD heritability enrichment in highly expressed genes in astrocytes and in caudate nucleus. The SMR/HEIDI tests showed non-pleiotropic association of extreme-SVD common risk variants at the 17q25.1 locus with TRIM47, RP11-552F3.10 and WBP2 expression in brain.

Conclusion: We report a multi-cohort gene-mapping study on extreme-SVD that expands our understanding of the genetic risk architecture of cerebral SVD.

(24)

Investigation of interactions between sugar-sweetened beverage consumption and CHREBP genetic variants on triglyceride and HDL-C concentrations

Danielle E Haslam (1), Gina M. Peloso (2), Caren E Smith (1), Josée Dupuis (2,3), Hassan S Dashti (4,5), Alice H. Lichtenstein (1), Mark A Herman (6), Nicola M. McKeown (1)

1. Jean Mayer U.S. Department of Agriculture Human Nutrition Research Center on Aging at Tufts University, Boston, MA; 2. Boston University School of Public Health, Boston, MA; 3. National Heart, Lung, and Blood Institute's Framingham Heart Study and Population Sciences Branch, Framingham, MA; 4. Massachusetts General Hospital, Boston, MA; 5. Broad Institute, Cambridge, MA; 6. Duke University School of Medicine, Durham, NC

Background: Genetic determinants of triglyceride (TG) and high-density lipoprotein cholesterol (HDL-C) concentrations have been replicated in genome-wide association studies (GWAS), but the loci discovered only account for a small fraction of the estimated total heritability. Consumption of sugar-sweetened beverages (SSB) and genetic variants at the carbohydrate responsive element binding protein (CHREBP) (also known as MLXIPL) locus have separately been linked to lipid concentrations, and ChREBP is a transcription factor that responds to SSB consumption. We hypothesized that SSB consumption may modify the associations between CHREBP variants on HDL-C and TG concentrations.

Methods: We conducted a cross-sectional analysis of data from 11 Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium cohorts (N=61,795). A total of 1,606 single-nucleotide polymorphisms (SNPs) were identified within or near the ChREBP locus. SSB consumption (sodas, fruit punches, lemonades, or other fruit drinks) was estimated from food-frequency questionnaires. The associations between SNPs and HDL-C and TG concentrations were quantified in the full sample and by category of SSB consumption. Summary statistics were combined through fixed effect meta-analyses. We screened for SNPs that differed between participants in the lowest (<1 serving SSB/month) and highest (>1 servings SSB/day) categories of SSB consumption to assess SNPxSSB interactions on TG and HDL-C concentrations.

Results: We first replicated previously observed GWAS associations between one SNP on HDL-C and two distinct SNPs on TG concentrations (Bonferroni-corrected $p<0.0001$). Additionally, we identified two distinct novel SNP associations with TG concentrations (FZD9-rs42124 and VPS37D-rs10245965). Two distinct SNPs displayed statistically significant difference in effect by category of SSB consumption on HDL-C (TBL2-rs35709627 and FZD9-rs34821369), and one SNP on TG concentrations (CHREBP-rs13240662) (all $p<0.0001$).

Conclusions: Our results indicate that high SSB consumption may modify the association between SNPs associated with lipids concentration within or near the CHREBP locus on TG and HDL-C concentrations, and that SNP analyses stratified by dietary factors may be a promising tool for uncovering gene-diet interactions.