CWRU Institute for Computational Biology
presents

Being Precise in Precision Medicine: Measuring Exposures in Diverse Populations

and Workshop

Strategies For Enhancing The Value And Utility Of Research Through Electronic Health Records

September 28, 2017
Tinkham Veale University Center

Sponsored In Part By
National Institutes of Health/National Human Genome Research Institute – Grant R13 HG009481

INSTITUTE FOR COMPUTATIONAL BIOLOGY

SCHOOL OF MEDICINE
CASE WESTERN RESERVE UNIVERSITY

Welcome

September 28, 2017

Welcome to the third annual Case Western Reserve University (CWRU) Institute for Computational Biology (ICB) Symposium, also known as the North Coast Conference on Precision Medicine series! This year, we are discussing the challenges of measuring environmental exposures and collecting behavioral data and other social determinants of health for ***precision medicine*** research in diverse populations. Precision or personalized medicine is the incorporation of 'omic data into clinical practice to better predict, prevent, and treat disease at the individual level. The White House-led Precision Medicine Initiative (PMI) known as All of US is ramping up, and recruitment is underway to ascertain and follow 1 million Americans in a study of genetics, environment, and lifestyle to better understand their influence on human health.

This year, we have seven local and national speakers who will present on various topics related to this year's theme, including measuring physical activity, air pollution, cigarette and cigarillo use, and stress in diverse populations. We will also hear about the latest tools in data harmonization and statistical approaches for gene-environment studies. The lunch break will feature a workshop and a presentation from Dr. Sarah Pendergrass on the use of electronic health records in research.

Immediately following the symposium is a reception in the Smith Commons area of the Tinkham Veale University Center. We hope you can stop by to tell us about your thoughts on this year's as well as next year's symposium topics. Please check our website this summer for next year's symposium, including travel scholarship opportunities for trainees and junior investigators, at www.icompbio.net. You can also follow us on Instagram (smartpeoplesciencing) and Twitter (@compbio).

See you next year!

INSTITUTE FOR COMPUTATIONAL BIOLOGY

Dana C. Crawford, PhD
Chair, Organizing Committee
Associate Professor
Population and Quantitative Health Sciences
Institute for Computational Biology

Schedule

Being Precise in Precision Medicine: Measuring Exposures in Diverse Populations
September 28, 2017
Tinkham Veale University Center, Ballroom A

8:00 – 8:30 Registration and Breakfast
Tinkham Veale University Center, Foyer

8:30 – 8:40 Welcome and Introduction

Dana Crawford, PhD, Associate Professor
Department of Population and Quantitative Health Sciences
Institute for Computational Biology
Case Western Reserve University

8:45 – 9:30 Measuring air pollution and its impact on human health

Stefanie Ebelt Sarnat, ScD, MSc, Associate Professor
Environmental Health
Emory University Rollins School of Public Health

9:30 – 10:15 Using wearable devices to improve physical activity measurement

Jennifer Schrack, PhD, MS, Assistant Professor
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health

10:15 – 11:15 Break and Poster Session
Tinkham Veale University Center, Foyer

11:15 – noon Harmonization and analysis approaches for gene-environment interaction studies

Misa Graff, PhD, Research Assistant Professor
Epidemiology
University of North Carolina Gillings School of Global Public Health

Noon - 2:30pm Lunch and Workshop

Strategies for enhancing the value and utility of research through electronic health records

Sarah Pendergrass, PhD, MS, Assistant Professor
Biomedical and Translational Informatics Institute
Geisinger Health System

2:30pm – 3:15pm Community tools for data harmonization and genetic variant interpretation

Erin Ramos, PhD, MPH
Division of Genomic Medicine
National Human Genome Research Institute

3:15pm – 3:30pm Break

3:30pm – 4:15pm Stress exposure and tobacco-related disparities

Monica Webb Hooper, PhD, Professor
Director, Office of Cancer Disparities Research
Case Comprehensive Cancer Center

4:15-5:00pm Canvassing the community to measure little cigar and cigarillo use in Cleveland

Erika Trapl, PhD, Assistant Professor
Department of Population and Quantitative Health Sciences
Acting Director, Prevention Research Center for Healthy Neighborhoods
Case Western Reserve University

5pm Thank you and adjourn

Reception
Smith Commons, 1st floor Tinkham Veale University Center

Being Precise in Precision Medicine:

Measuring Exposures in Diverse Populations

INSTITUTE FOR COMPUTATIONAL BIOLOGY

Speakers

Misa Graff, PhD, MS

Dr. Misa Graff is Research Assistant Professor of Epidemiology at the University of North Carolina Gillings School of Global Public Health. Dr. Graff's research interests focus on the genetics of obesity and cardiovascular disease risk traits, with an overarching goal of integrating understanding of both genetic and environmental risk factors for disease, particularly among minority populations. Dr. Graff's current projects include the genetic epidemiology of weight gain across adolescence into young adulthood; gene-environment interactions across multiple large cohorts; the genetic architecture of adiposity in African Americans and Hispanics; and examination of genetic effects influencing cardiovascular disease risk traits and diseases burdening minority populations.

Sarah Pendergrass, PhD, MS

Dr. Sarah Pendergrass is Assistant (Investigator I) Professor in the Biomedical and Translational Informatics Program at Geisinger Health System. Dr. Pendergrass is a genetic bioinformatician who focuses on high-throughput data analysis and data-mining approaches to studying complex human diseases and traits. Dr. Pendergrass has extensive experience in using both epidemiologic and clinic-based resources to perform phenome-wide association studies (PheWAS) to identify cross-phenotype associations and pleiotropy. Dr. Pendergrass also develops software tools to visualize complex data. In recognition for her innovative work, Dr. Pendergrass was named one of Genome Technology's PIs of Tomorrow (2013).

Erin Ramos, PhD, MPH

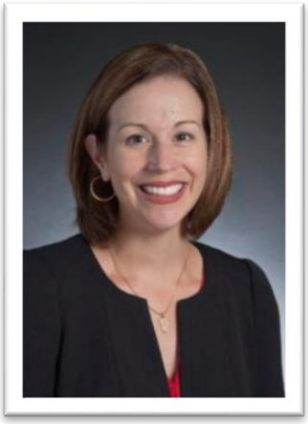
Dr. Erin M. Ramos is an epidemiologist in the Division of Genomic Medicine, National Human Genome Research Institute (NHGRI). She manages a portfolio of research in population genomics including a collaborative project to develop a set of standardized

phenotypic and exposure measures for use in genome-wide association studies and related research (PhenX). She also manages the Clinical Genome Resource (ClinGen), which aims to collect phenotypic and clinical information on variants across the genome, develop a consensus approach to identify clinically relevant genetic variants, and disseminate information about the variants to researchers and clinicians. Dr. Ramos previously served as the chair of the Data Access Committee (DAC) for the Genetic Association Information Network (GAIN) and as a member of NHGRI's DAC.

Dr. Ramos received her MPH and PhD in the multidisciplinary field of public health genetics from the University of Washington where her research focused on the genetic epidemiology of Alzheimer's disease and the ethical, legal, and social implications that surround genomics research. Her current research interests include the genetic epidemiology of dementia; genome-wide association studies and gene-environment interactions in complex disease; and Ethical, Legal and Social Implications (ELSI) research, including informed consent for large-scale genomic studies.

Stefanie Ebelt Sarnat, ScD, MSc

Dr. Stefanie Ebelt Sarnat is Associate Professor of Environmental Health at Emory University. Dr. Sarnat's research focuses on examining cardiovascular and respiratory effects of ambient air quality using population- and panel-based approaches. Dr. Sarnat leads several large-scale time-series studies of ambient air quality and acute morbidity, with specific interests in assessing the health impacts of air pollution, meteorological conditions, and weather extremes. Her research also includes field investigations, with detailed exposure and health outcome data collected in panels of susceptible individuals.



Jennifer Schrack, PhD, MS

Dr. Jennifer Schrack is Assistant Professor of Epidemiology at the Johns Hopkins Bloomberg School of Public Health (JHSPH). Dr. Schrack holds a joint position with the Krieger School of Arts and Sciences, and she is affiliated with several Centers including the Center for AIDS Research, the Johns Hopkins Center on Aging and Health, the Multicenter AIDS Cohort Study, and the Welch Center for Prevention, Epidemiology and Clinical Research.

Dr. Schrack's primary area of research is the epidemiological assessment of physical activity and its associations with health and longevity in older populations. In collaboration with the SMART group from the JHSPH Department of Biostatistics, Dr. Schrack is working to establish methods to analyze and translate heart rate and accelerometry data into accurate and reliable measures of physical activity in older populations. In recognition of these efforts, Dr. Schrack was recently awarded a 2017 Johns Hopkins University Catalyst Award and a U01 from the National Institute on Aging.

Erika Trapl, PhD, MS

Dr. Erika Trapl is Assistant Professor in the Department of Population and Quantitative Health Sciences at Case Western Reserve University and Acting Director of the Prevention Research Center for Healthy Neighborhoods. Dr. Trapl's research interests include little cigar and cigarillo use in the Cleveland area.

Monica Webb Hooper, PhD

Dr. Monica Webb Hooper is Director of the Office of Cancer Disparities Research at the Case Comprehensive Cancer Center and Professor of Oncology, Family Medicine, and Psychological Sciences at the Case Western Reserve University School of Medicine. Dr. Webb Hooper is a clinical health psychologist whose research interests are in the health behavior change of cancer risk behaviors, with an emphasis on the intersection between cancer prevention and control, and minority health and disparity elimination.

Being Precise in Precision Medicine:

Measuring Exposures in Diverse Populations

INSTITUTE FOR COMPUTATIONAL BIOLOGY

Travel Scholarship Awardees

Made possible by NIH/NHGRI R13HG009481

Basile (Okula), Anna

Graduate student
Department of Biochemistry and Molecular Biology
The Pennsylvania State University
University Park, Pennsylvania

Understanding the nature of genetic contribution to complex disease susceptibility has been at the heart of genetic research for the past decade. GWAS have largely driven the field and identified many SNPs associated with complex disease. Despite these successes, the inability of GWAS to explain more than a fraction of specific causal loci has led to the belief that many genetic and epigenetic factors likely contribute to complex diseases, including rare variants. Rare variants are believed to have larger effects than common variants, and studying their influence may add to our understanding of disease heritability. While the advent of next-generation technologies has presented an opportunity for discovering rare variants, multiple challenges exist. Because these variants are individually uncommon, there is low statistical power for detecting association with a trait. Large sample size requirements can often be prohibitive, especially when considering allele frequencies below 1%. Although multiple methods and statistical tests exist, few approaches provide an automated analysis framework. Also, the presence of genetic and phenotypic heterogeneity in complex disease presents a challenge in rare variant analysis. Heterogeneity has important implications in the discovery and replication of disease-causing genes as well as medical treatments appropriate for patients. My research aims to address challenges in rare variant analysis with 1) an automated biological approach for both collapsing and association testing of rare variants, and 2) the application of an advanced method for identifying phenotypic patterns present in clinical data for integration with sequence data to explore the genetic architecture of disease.

Butkiewicz, Mariusz

Staff scientist, bioinformatics
Department of Biomedical and Translational Informatics
Geisinger Health Systems
Rockville, Maryland

Dr. Mariusz Butkiewicz is currently a Bioinformatics staff scientist in the Department of Biomedical and Translational Informatics at Geisinger Health Systems. His current research aims to bridge knowledge from WES/WGS

sequencing data with phenotype data from Geisinger's electronic medical records system. Previously, Dr. Butkiewicz earned a Diploma in Computer Science (Master's equivalent) from Leipzig University in Germany, and later earned his PhD in computational Chemistry from Vanderbilt University working with Dr. Jens Meiler (www.meilerlab.org) on developing new cheminformatics approaches to applying machine learning approaches to aid drug discovery of novel small molecule therapeutics for Schizophrenia and Malaria. During his post-doctoral training with Dr. Jonathan Haines at Case Western Reserve University, his research centered on quantitative genetic and genomic studies with an emphasis on Alzheimer's disease, and other neurological disorders.

Clement, Tasmine

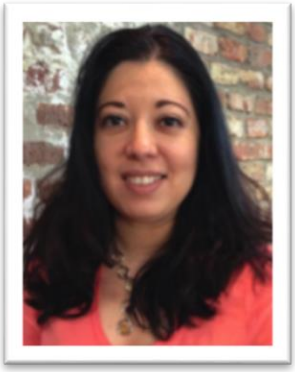
Post-baccalaureate student (PREP scholar)
Department of Biochemistry
Virginia Tech
Blacksburg, Virginia

I have previously performed clinical-translational research at Notre Dame's Center for Rare and Neglected Diseases. The goal was to develop therapies and outreach efforts for people suffering from various orphan diseases, like Niemann-Pick type C (NPC). NPC is a lysosomal storage disorder with a wide clinical spectrum that results in nervous system deterioration. NPC has a higher incidence in French Canadian (Nova Scotian) populations and Spanish-American (NM & CO) populations. As a researcher, I worked with health care practitioners and families to collect and summarize patient data electronically so that it would be easily shared between current and future parties; including researchers for clinical trials. I combed through a lifetime's worth of medical records for patients with rare diseases in order to compile information on phenotypes, treatment plans and familial history. The information was then used to diagnose patients with rare diseases through clinical and biochemical biomarkers. With my work, patients were also able to enter clinical trials for potential drug treatments like cyclodextrin. I was also responsible for developing a disability scale for NPC to better evaluate clinical therapies and to find additional phenotypic correlations. In all, I spent two years identifying markers of disease progression within individuals and quality of life for each individual based on symptomology and individual treatment plans.

Hall, Jacob

Post-doctoral fellow
Institute for Next Generation Healthcare
Icahn School of Medicine at Mount Sinai
New York, New York

Acne is a common skin condition that can have physical as well as psychological/emotional consequences. Some factors, such as age, sex, and diet, help predict risk of developing acne, yet pathogenesis is far from fully understood. Many treatment options exist to treat a range of acne severity, though response to treatment is variable. Additionally, the most aggressive treatment options don't always eliminate acne and can lead to severe side effects, so better treatment options are needed. Recently, I have used multi-omics data (genome, transcriptome, proteome, microbiome) to explore differences between acne and non-acne patients as well as differences between healthy and acne skin of acne patients. Both comparisons shed light on different aspects of acne pathogenesis. Comparing lesion and non-lesion skin of acne patients has generally highlighted the inflammatory component of acne. Comparing non-lesion (healthy) skin of acne and non-acne patients has yielded differentially expressed genes, with gene set enrichment pointing to factors such as epithelial-to-mesenchymal transition, coagulation, and UV response. Unsurprisingly, the gene expression signatures had similarities between other skin diseases, such as psoriasis, eczema, lupus, and contact dermatitis. We are currently performing PacBio sequencing of the microbiome, which will yield high-resolution, long genomic reads, to better characterize influencing factors of acne.



Hernandez, Wenndy

Post-doctoral fellow
Department of Medicine
University of Chicago
Chicago, Illinois

My graduate training provided me with a solid foundation in multiple aspects of cancer biology including molecular mechanisms, genomics, drug resistance, and translational research approaches. I have acquired analytical skills in methods that measure and apply population genetic ancestry to genetic association studies, an essential step towards personalized medicine. I have successfully integrated molecular and analytical skills to identify population-specific genetic associations to complex diseases, particularly in the context of prostate cancer, warfarin pharmacogenomics, and venous thromboembolism.

Currently, my training entails quantitative and computational methods and tools that utilize large-scale data in aggregate which coupled with my molecular background will allow me to formulate new questions through the vast amount of data generated. My ultimate goals are to utilize novel and more complex methods to analyze and integrate a wide variety of genetic data types to understand the basis of cardiovascular diseases; design novel prediction models in order to provide patients with optimal treatment; and to establish a resource database collected exclusively on African American ischemic stroke patients from which to study the relationship between genetic variation and gene expression in order to improve our understanding of the molecular signatures underlying ischemic stroke and its subtypes.

I began graduate school knowing that I would pursue a career conducting research that may one day help reduce the health disparity gap. My graduate training provided me with a solid foundation in multiple aspects of cancer biology including molecular mechanisms, genomics, drug resistance, and translational research approaches. While I had several options for a post-doctoral position, I was committed to conducting genetic research among under-represented populations. I joined Dr. Minoli Perera's lab in September 2012 and began research in pharmacogenomics that led to an algorithm for predicting therapeutic warfarin dose for African Americans. My novel warfarin dosing algorithm outperformed the most widely used algorithm and standard clinical practice for dosing African Americans, and it's currently being implemented in a pilot clinical study through the 1200 Patients project (Dr. Odonnell) at the University of Chicago. I observed that nearly 60% of the African American warfarin patients were placed on warfarin due to venous thromboembolism (VTE) which provided an opportunity to conduct a case-control study on VTE risk susceptibility -- in fact, this resulted in the genome-wide association study on VTE susceptibility among African Americans. Translational research that integrates population-specific genomic variation and complexity has tremendous potential to reduce health disparities and improve clinical practice. My future research plans will do exactly that by integrating polygenic risk models and gene expression regulation into analyses aimed to improve our understanding on the etiology of stroke and discover targets for intervention.

El Rouby, Nihal

Post-doctoral fellow
Department of Pharmacotherapy and Translational Research
University of Florida
Gainesville, Florida

I'm a Post doctoral fellow with a concentration on Pharmacogenomics in the Department of Pharmacotherapy and Translational Research at UF. My previous research as a PhD student focused on studying the genetic determinants of Resistant Hypertension. The ultimate goal of this research is to identify genetic predictors, which can be used along with other clinical predictors to identify high-risk patients and individualize their drug treatment. I realize and appreciate the value of Computational Genomics and Bioinformatics during my PhD training, and have been seizing any and every opportunity to learn different methodologies to analyze big data, with the goal of using this knowledge to learn about important genetic, racial, environmental determinants that will help in facilitating precision medicine approaches. My long term career goal is to be a successful academic researcher with a focus on Pharmacogenomics.

Martucci, Victoria

MD/PhD student
Medical Scientist Training Program (Human Genetics)
Vanderbilt University
Nashville, Tennessee

I am an MD/PhD student at Vanderbilt entering my second year of my PhD in human genetics. As a PhD student, my thesis will focus on unravelling the complexities of chronic obstructive pulmonary disease (COPD) genetics using information in Vanderbilt's electronic health record (EHR) system. COPD is the third leading cause of mortality worldwide, so understanding genetic risk factors for COPD development could allow early intervention to reduce the devastating effects of this disease. COPD is highly heterogeneous in its clinical presentation, which makes studies of its genetics and development challenging. COPD research is also complicated by difficulties identifying COPD cases. The gold standard for COPD diagnosis is pulmonary function testing via spirometry, but this test is not routinely performed as a screening tool. Therefore, individuals who are asymptomatic or have mild disease may be undiagnosed, causing misclassification biases in performing case-control studies. One of the major goals of my project will be the development of an algorithm to identify individuals with COPD without using spirometry data based on information in the EHR. Once I have identified cases of COPD, I plan to leverage the wide array of clinical information available in the EHR to identify subphenotypes of COPD, with the goal of identifying unique genetic variants associated with each subphenotype. My research using EHR and existing genetic data will help discover some of the underlying mechanisms of COPD development, which can be used to better understand, treat, and ultimately prevent this highly prevalent disease.

McDonough (Rowe), Caitrin

Research Assistant Professor
Department of Pharmacotherapy and Translational Research
Center for Pharmacogenomics
University of Florida
Gainesville, Florida

I have a strong foundational background in human genetics and pharmacogenomics. As I have always been interested in genetics, and research that impacts human disease, my studies took me from working with fruit flies as an undergraduate to studying the genetics of diabetic nephropathy in African Americans as a doctoral student at Wake Forest University to investigating cardiovascular pharmacogenomics as a postdoctoral fellow at the University of Florida (UF). Upon completing my training, I moved into a faculty position as a Research Assistant Professor in the College of Pharmacy at UF in 2013. Over the past four years, I have continued to expand my knowledge in the areas of cardiovascular pharmacogenomics, clinical trials, and precision medicine. In April 2016, I was appointed as a KL2 scholar through the UF CTSA. My KL2 project focuses on developing a resistant hypertension (RHTN) computable phenotype, utilizing electronic health record (EHR)-based data from the OneFlorida DataTrust - a database that contains longitudinal EHR-based data from providers throughout the state of Florida. Additionally, as part of the training component of the KL2 award, I am completing a Master's degree in Biomedical Informatics. I also remain active in genomic and pharmacogenomics research through collaborations in the Stroke Genetics Network, and the International Consortium for Antihypertensive Pharmacogenomics Studies, and I am involved with the Precision Medicine Program at UF. All of these activities allow me to maintain my expertise in human genetics and pharmacogenomics, while gaining additional knowledge and skills in biomedical informatics and "Big Data".

Restrepo, Nicole

Staff scientist, bioinformatics
Department of Biomedical and Translational Informatics
Geisinger Health Systems
Rockville, Maryland

I am a bioinformatics staff scientist in the Department of Biomedical and Translational Informatics at Geisinger Health Systems. Although formally trained as a genetic epidemiologist, my work in genetics and Electronic Health Records phenotyping intersect bioinformatics and precision medicine. My previous research and training involved the identification of common and rare genetic variation contributing to ocular disease (i.e., age-related macular degeneration, diabetic retinopathy, and glaucoma) in diverse populations of African American and Hispanic descent. My studies also look to dissect the interactions of genetic and environmental modifiers of ocular disease risk in unique Founder populations with a focus on precision medicine outcomes.

Singh, Abanish

Assistant Professor
Psychiatry and Behavioral Sciences
Duke University School of Medicine
Durham, North Carolina

My research training was focused on computational biology, high-throughput genomics, and big data analytics, which resulted in some of prominent findings on human genome. These findings included that the overrepresentation of short DNA elements in the human genome was a result of ancient duplication followed by degeneration activities in human DNA (Singh et al, 2007; Singh et al. 2010) and that the RNA-Seq can identify well human coding variants just using transcriptome as compared to the whole genome (PMID:20598109). With a unique skillsets as resulted from an outstanding training, my sole aim was to help improve the human health through the cutting edge translational research that may lead to the precision medicine. I became interested in understanding the measurement of biobehavioral risk factors and environmental stressors and their interactions with genes that may influence CVD risk factors and endophenotypes. My relatively recent work (Singh et al., EJHG, 2015) identified a novel CVD risk gene EBF1, where a common variant contributed to inter-individual differences in central obesity (hip, waist, and BMI) in the presence of chronic psychosocial stress. I also developed an algorithm to create a synthetic measure of stress using the proxy indicators of its components (PMID:26202568). More work that is recent (Singh et al. ASHG 2017) has elucidated the race, sex, and age related differences in the EBF1 x stress interaction and demonstrates the need for careful evaluation of environmental measures in different ethnicities in cross-ethnic gene-by-stress interaction studies.

Thomas, Kia

Post-baccalaureate student (PREP scholar)
College of Arts & Sciences
Emory University
Atlanta, Georgia

My past and current research experience is in Dr. Gregory Melikian's lab at Emory School of Medicine department of pediatric infectious disease where I do work-study. Here we are researching the method of fusion employed by HIV and Lassa virus, and my role is to do data analysis using ImageJ/Fiji. My past research also includes working in Dr. Sam Speck's lab as a part of IMSD, where I worked with a post-doc to investigate the role of antigen M2 in the murine gammaherpes virus MHV68. Here I learned microbiology lab techniques such as PCR, western blotting, gel electrophoresis, DNA purification, cell culture, and flow cytometry. My current research is at the University of Chicago department of anesthesia and critical care in Dr. Daniel McGehee's lab. Here I am working with a post-doc on how to manipulate the dopamine reward pathway to find a new treatment for Parkinson's disease, using behavioral analysis (in-vivo), imaging, microscopy, immunohistochemistry, and electrophysiology. My role in this project is to see how Parkinson's influences the synaptic plasticity of affected neurons in the mouse model. In the fall I will be doing research in Dr. Bernardo Mainou's lab in Emory School of Medicine's department of pediatric infectious disease. My research will focus on using reovirus as a mechanism to treat triple-negative breast cancer.

My current research is related to precision medicine in that HIV and its many variances affect everyone differently. It is important to specify the types of treatments to the individual and the exposure they are privy to following their diagnosis. I am aware that HIV is sometimes treated with a "cocktail" of medications, different depending on the individual's conditions. I hope to improve this method of treatment and to make it more accurate and efficient, using precision medicine. With more of my research experience having been in infectious disease, I hope to go into a doctoral program and have a following career focused on the challenge that is the treatment of HIV and other communicable diseases.

Being Precise in Precision Medicine:

Measuring Exposures in Diverse Populations

INSTITUTE FOR COMPUTATIONAL BIOLOGY

Poster Abstracts

In-silico Functional Annotation of Genomic Variants for Neurodegenerative Disease

W. Bush[1], M. Butkiewicz[1], M. Sivley, T. Capra, and J. Haines[1]

1) Institute for Computational Biology, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.; 2) Vanderbilt Genomics Institute, Department of Biological Sciences, Vanderbilt University

The Alzheimer's Disease Sequencing Project Discovery Phase is comprised of a Family-Based study containing WGS data on 584 subjects from 113 families, and pedigree data for more than 4000 subjects, and a Case-Control study comprised of WES data on 5096 cases 4965 controls. Whole exome sequence data on an additional 853 (682 Cases [510 Non-Hispanic, 172 Hispanic]), and 171 Hispanic Control subjects from families that are multiply affected with AD and in the process of being sequenced. In this work, we use bioinformatics methods to functionally characterize and prioritize variants identified in next-generation sequencing studies of neurodegenerative disease. Our annotation process illustrates the impact of including tissue-specific transcript sets and sources of gene regulatory information, and assess the potential impact of changing genomic builds on the annotation process. While these factors only impact a small proportion of total variant annotations (~5%), they influence the potential analysis of a large fraction of genes (~25%). Leveraging in silico annotation data in conjunction with brain and other tissue-based RNA-sequencing studies, we have constructed annotations of ADSP variants that enable new hypothesis-driven analyses, and ultimately provide new insights into the pathogenesis of neurodegenerative diseases.

Survey of attitudes towards biosample collection and genetic testing in a racially diverse CKD population in Cleveland, OH

Jessica N. Cooke Bailey[1], Dana C. Crawford[1], Aaron Goldenberg[2], Anne Slaven[4], Julie Pencak[4], Marleen Schachere[4], William S. Bush[1], John R. Sedor[3,4], John F. O'Toole[3]
[1]Department of Population and Quantitative Health Sciences, Institute for Computational Biology, Case Western Reserve University, Cleveland, OH
[2]Department of Bioethics, Case Western Reserve University, Cleveland, OH
[3]Departments of Medicine and Physiology and Biophysics, Case Western Reserve University, Cleveland, OH
[4]Division of Nephrology, Department of Medicine, MetroHealth Medical Center, Cleveland, OH

Background: The NIH All of Us Research Program (AURP) is an ambitious national effort to longitudinally collect health data and biospecimens from a million Americans for storage, processing, and analysis in a government-funded data repository. Previous studies have suggested minority populations would be reluctant to share personal data and samples with the Federal government. We tested this premise in a population of chronic kidney disease (CKD) patients in Cleveland, OH.
Methods: Patients in a CKD clinic were approached prior to a standard of care visit and asked to complete a structured 5 question bioethics and IRB-approved survey and to provide a blood sample for future genetic analysis.
Results: Most (86%) participants responded to our survey; 50% African American, 54% female, with average age 61.5. Responses from 111 individuals indicate the majority would participate in the AURP and were willing to send biosamples to a national repository and share de-identified data. However, fewer than half of respondents were willing to install a phone app for personal data tracking. A minority of individuals (10%) did not want any results returned. Most respondents considered results return very important - 96% of those who wanted results wanted personal (health or genetic) data returned; 41% wanted *at least* summary data about the PMI-CP cohort, 4% *only* summary data about the overall group, 76% *at least* personal health data, and 4% *only* personal health data. Genetic data was priority; 89% wanted *at least* personal genetic data while 19% wanted *only* personal genetic data. We found no significant difference between responses when comparing African American and White individuals.
Conclusions: Attitudes of CKD patients in a diverse health care environment towards the AURP are varied but, in contrast to published data, did not differ across self-reported race (African Americans and whites) in this sample. Willingness to participate in some aspect of a PMI project was high. Of those agreeing to take the survey, almost all wanted return of genetic results. Given this demand, efficient processes should be developed to provide subjects with appropriate education and context for results return. Other chronic disease populations and healthy subjects need to be studied to determine if health status, race or ethnicity modifies willingness to join the AURP.

Markers of the adaptive immune response are associated with progressively worse chronic kidney disease status

Dana C. Crawford[1], William S. Bush[1], Jessica N. Cooke Bailey[1], Kristy Miskimen[1], Penelope Miron[1], John O'Toole[2], and John R. Sedor[2,3]

[1]Institute for Computational Biology, Departments of Population and Quantitative Health Sciences, [2]Medicine, and [3]Physiology & Biophysics, Case Western Reserve University, Cleveland, OH.

Germline and somatic genomic variation represent the bulk of 'omics data available for precision medicine research. These data, however, may fail to capture the dynamic biological processes that underlie disease development, particularly for chronic diseases of aging such as chronic kidney disease (CKD). To demonstrate the value of additional dynamic precision medicine data, we sequenced somatic T-cell receptor rearrangements, markers of the adaptive immune response, from genomic DNA collected during a clinical encounter from 15 participants with CKD and associated co-morbidities. Participants were consented as part of a larger precision medicine research project at the MetroHealth System, a large urban public hospital in Cleveland, Ohio. Genomic DNA was extracted from whole blood, and T-cell receptors were sequenced with six replicates per sample using Adaptive Biotechnologies' immunoSEQ assays coupled with the Illumina NextSeq PE. All sequences were assembled using Adaptive Biotechnologies' ANALYZER bioinformatics pipeline. Demographic and clinical data closest to the time of blood draw were extracted from the electronic health record. Overall, the average age of patients was 61.73 years, more than half (60%) were female, and the majority were African American (80%). T-cell receptor diversity was estimated using productive clonality, a measure ranging from 0 (polyclonal samples) to 1 (monoclonal or oligoclonal samples). Productive clonality in this sample ranged from 0.0151 to 0.2565 with a mean of 0.1030 (standard deviation or SD=0.0669). Average productive clonality did not statistically differ by sex: females = 0.0811 (SD=0.0533) and males = 0.1358 (SD=0.0764). We then tested for correlations between T-cell receptor diversity and biomarkers of CKD, including disease status calculated using the CKD-EPI equation. Reduced T-cell diversity was associated with increased creatinine (R^2=0.0995), BUN (R^2=0.0258), and eGFR (R^2=0.066), but not with white blood cell count (R^2=0.0004). Reduced T-cell diversity was also associated with worsening CKD status (R^2=0.2362), with a higher on average productive clonality (0.0488) among stage 4 patients (n=5) compared with stage 3 (0.0330; n=8) and stage 2 patients (0.0149; n=2). These data suggest an association between advanced CKD and premature aging of the adaptive immune system and highlight the potential of dynamic 'omic data to generate novel hypotheses about disease mechanisms.

Characterization of Resistant Hypertension in a Statewide Electronic Health Record-Based Database (OneFlorida)

Caitrin W. McDonough, William R Hogan, University of Florida; Elizabeth A Shenkman, University of Florida; Rhonda M Cooper-DeHoff, University of Florida

Our objective is to create a Resistant Hypertension (RHTN) computable phenotype from Electronic Health Record (EHR) based data, and to determine the characteristics associated with RHTN within a large, diverse, EHR-based database. The OneFlorida Clinical Research Consortium includes 10 unique health care systems providing care for approximately half of the state (48%, ~10 million). OneFlorida houses a Data Trust which contains longitudinal EHR data and claims data from these providers in a common format, the PCORnet common data model v3.0. For the current project, data from five health care systems were considered. All of the adult hypertension (HTN) patients with a HTN diagnosis from an outpatient encounter were extracted from the OneFlorida Data Trust. Additional data such as demographics, prescribing, and vitals information were also extracted. The RHTN computable phenotype was created by constructing a drug exposure variable that took into consideration the number of antihypertensive medications an individual was prescribed at any point in time over the course of the OneFlorida dataset. All data extraction, computation phenotype coding, and statistical analyses were conducted using SQL or SAS. Our preliminary results show that there were n=342,026 adults with a HTN diagnosis from an outpatient visit in the dataset. After the RHTN computable phenotype was constructed, n=11,670 RHTN cases were identified from the n=130,901 HTN individuals with all of the required variables in the dataset (8.9% RHTN prevalence). Fifty-five percent of RHTN cases were Black or African American, compared to the total HTN population (25% Black/African American). RHTN cases also had more prescriptions for loop diuretics, centrally acting agents, alpha-blockers, and vasodilators compared to the total HTN population. Not surprisingly, the RHTN cases had 26% of the antihypertensive prescriptions in the dataset, and the RHTN cases had fewer blood pressure readings that were in control (only 49.4% of readings <140/90). Overall, our preliminary data shows that it is possible to create the very complicated computable phenotype of RHTN within the OneFlorida Data Trust. We found that the RHTN prevalence in OneFlorida is 8.9%, which is consistent with previous studies from NHANES. Although promising, these results require further validation of the computable phenotype and replication in other similar datasets in order to ascertain their true meaning. Once validated, the experience gained from this computable phenotype can be applied to many other phenotypes.

Mixed-model adjustments for tests of epistasis reduce confounding by other loci

N. Patel [1]; W. Bush [1]; A. Fish [2]; J. Capra [2]

1) Department of Population and Quantitative Health Sciences, Institute of Computational Biology, Case Western Reserve University, Cleveland, OH.; 2) Department of Biological Sciences, Vanderbilt Genomics Institute, Vanderbilt University, 22525 West End Avenue, Nashville, TN.

Recent studies have identified numerous potential examples of epistasis among cis-regulatory variants influencing gene expression traits. However, these tests are subject to different forms of confounding than single-variant association tests - most notably, that multi-locus genotype combinations can tag the effect of nearby cis-eQTLs. As a result, many statistically significant interaction models for eQTL SNPs are more parsimoniously explained by the presence of another nearby SNP, and post-hoc conditional analyses are necessary to identify and characterize these findings. In this study, we develop and evaluate a mixed linear model approach, to adjust for genetic relatedness, in order to correct for this potential confounding. In a prior study, we identified 1,119 cis-regulatory interactions in a discovery set of 210 individuals with array-based gene expression and dense genotyping from lymphoblastoid cell lines, following rigorous false discovery corrections. Even though many of these interactions replicated in whole blood samples from the Genotype Tissue Expression (GTEx) project, all were found to be likely false positives in post-hoc tests for confounding. We re-analyzed these potential interactions using models including a random effect adjustment for the other SNP within the cis-regulatory region. This adjustment dramatically reduced the amount of confounding; likelihood ratio tests for the effect of the interaction terms were non-significant for all but six models. These results illustrate that mixed-model adjustments are a powerful approach for addressing confounding in SNP-SNP interaction analyses.

Modifiers of severity in Autism Spectrum Disorder

Sandra Smieszek, Rob Igo, Jonathan Haines

Autism Spectrum Disorder (ASD) comprises a complex of neurodevelopmental disorders primarily characterized by deficits in verbal communication, impaired social interaction and repetitive behaviors. The genetic architecture has proved to be complex and encompasses profound clinical heterogeneity, which poses challenges in understanding its pathophysiology. We conducted a large scale association analysis of the MSSNG whole genome sequencing data to elucidate potential modifiers of ASD severity. Using the additive linear model method (PLINK) we have directly tested association between 6,198,166 SNPs (Quality Control: MAF > 0.05, HWE $P < 1 \times 10^{-6}$, Mendelian errors, removal of samples with dis-concordant sex status, twins, samples with unreported relatedness) and Vineland Adaptive Behavior Scores.

Interestingly, the top variants direct us to an 850kb region containing 21 variants within 3 genes on Chromosome 2: *LYPD1* a member of the Lynx family of neurotransmitter receptor-binding proteins implicated in anxiety, *NCKAP5* previously implicated in autism (CNV, 2 cases) and *GPR39*, a product of which has been implicated in depression. Some other interesting loci include *CACNA2D2* ($P < 1 \times 10^{-7}$, 16 markers) encoding a subunit of the voltage-dependent calcium channel complex and axon guidance receptor gene, *DCC* ($P < 1 \times 10^{-7}$). Furthermore, to leverage the size of the data we conducted a pathway enrichment analysis of the set of highly significant results ($P < 1 \times 10^{-6}$) using PARIS and DAVID software. The most significant category is tobacco use disorder, with $P < 1 \times 10^{-5}$, with 28% of genes contributing to the significance, followed by the brain development and structural component of myelin sheath pathways. Genes categorized a neurological, developmental and immune related constitute 65% of all the genes contributing in these pathways. We took variants from contributing genes from significantly overrepresented categories to test how much variability in the VABS scores can be explained by the variants. The cumulative effect of the single top pathway enrichment alone on affection status is 2% ($P = 6.34 \times 10^{-6}$).

We detect a region that may be a hallmark of severity in ASD. As genetic predisposition may be different for almost every ASD individual, understanding the common mechanisms for endo-phenotypes may help elucidate ASD causal mechanisms.

Carissa Smock, PhD, MPH Postdoctoral Fellow, Case Western Reserve University

Sonia Alemagno, PhD; Lynn Falletta, PhD; Jeffrey Hallam, PhD, Jonathan VanGeest, PhD

Introduction: Physical inactivity is associated with excess mortality and morbidity and contributes to $131 billion in health care expenditures annually. Up to 29.5% of U.S. adults are physically inactive. Place-based exercise prescriptions and referrals are an economical solution to physical inactivity. Health care provider associations and proposed government policy initiatives support adoption of exercise counsel and prescriptions. However, as few as 14% of primary care providers prescribe exercise or make exercise referrals. The purpose of this pilot study is to identify predictors of place-based exercise prescription and referral to exercise either at a specific location, or to an exercise professional.

Methods: An 88-item questionnaire was developed and administered electronically and in paper format to two hospital systems' networks of primary care physicians and nurse practitioners. Factor analysis was conducted to identify underlying constructs. Bivariate analyses were utilized to identify variables significantly related to exercise prescription and referral. A binomial logistic regression was performed to determine factors related to whether or not providers refer patients to place-based exercise.

Results: Responses were received from 158 providers (70%); 61.5% were female, 86.9% were white. Respondents' specialties were: family medicine (20.6%), internal medicine (27.8%), nurse practitioner (29.4%), and other (22.2%). Overall, 54.3% refer patients (mostly with chronic conditions) to specific exercise locations or professionals. Logistic regression of factors related to exercise referral yielded five significant variables: (1) Asking about exercise, (2) documenting exercise, (3) prescribing exercise, (4) knowledge of Physical Activity Guidelines for Americans, and (5) providers' perception that patients' need increased time to exercise.

Conclusion: Despite the known health risks of physical inactivity, referral for exercise is infrequent, about half of the providers refer a fraction of their patients. Provider's believe that patients lack time to exercise and may need guidance on how to work within patients' specific schedules and assess convenient locations to refer and prescribe exercise. Improving provider knowledge of recommended activity guidelines may result in improved adoption of physical activity prescription or referral.

Kia Thomas, Emory University

Parkinson's disease (PD) is a neurodegenerative condition that is lethal in humans. In this disease, dopaminergic neurons in the substantia nigra of the basal ganglia die, eventually leading to loss of motivation, motor control, and motor learning. There is no cure for PD, and the most common treatment is levodopa, a drug that enters the brain and is chemically converted to dopamine. Long-term treatment decreases the therapeutic effects, and most patients develop levodopa-induced dyskinesia, or uncontrolled movements. Therefore, it is important to find more effective treatments, and that effort depends upon testing in animal models of the disease. The goal of this study is to validate a mouse PD model involving striatal injection of the neurotoxin 6-OHDA, which kills dopamine neurons. We examined 6-OHDA-induced changes in behavior and dopaminergic cell loss. The results of this experiment contribute to a larger study aimed at manipulation of synaptic plasticity in striatal medium spiny neurons to relieve the symptoms of dopamine cell loss. Subjects were C57/Bl6 mice, 3-5 months old. Saline (sham control), low (1ug/ul) and high (4ug/ul) dose 6-OHDA injections were given into the left dorsal lateral striatum. Behavioral effect of the lesion was assessed using rotation tests. We did 2 tests per subject, 1) after a saline injection and 2) after an apomorphine injection. We predicted to see more ipsilateral rotation in lesioned mice with saline injection and more contralateral rotation after apomorphine. Comparison of rotation behavior between these treatments showed no difference between saline and apomorphine for sham control animals. We saw a significant difference in turning behavior for high ($p < 0.03$) and low ($p <. 0.05$) dose 6-OHDA animals1-2 weeks post-injection, but this effect was not seen in animals tested 3 weeks post-injection. Immunocytochemistry on slices of each mouse's striatum was performed to assess the level of staining for the dopamine neuron marker, tyrosine hydroxylase. The fluorescence intensity in the striatum was determined for each hemisphere and a ratio was calculated. Comparison of the immunostaining results with the behavioral data revealed an inverse correlation of the difference between rotation directions (#contralateral – #ipsilateral) to the ratio of TH staining comparing hemispheres (injected /control). Overall, our data shows that more dopaminergic cell death and more intense behavioral affects of the neurotoxin 1-2 weeks post-injection and at a dose of 4ug/ul. Future studies will use the biocytin filled neurons in these slices to assess their morphological changes induced by the 6-OHDA lesion. These results will contribute to the larger study that will hopefully lead to a new treatment for Parkinson's.

New evidence for enrichment of metabolic, signaling, and inflammatory pathways in age-related macular degeneration

Andrea R. Waksmunski[1], Jessica N. Cooke Bailey[2], Margaret A. Pericak-Vance[3], William K. Scott[3], Jonathan L. Haines[2,4], and the International Age-Related Macular Degeneration Genomics Consortium
[1]Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; [2]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; [3]Hussman Institute for Human Genomics, University of Miami, Miami, FL; [4]Institute for Computational Biology, Case Western Reserve University, Cleveland, OH

The disease etiology of age-related macular degeneration (AMD) is largely unknown. We aimed to identify biological pathways, molecular interactions, protein families, and regulatory regions enriched for genetic variations associated with AMD. We performed pathway analysis using the Pathway Analysis by Randomization Incorporating Structure (PARIS, V2.4) software tool on 445,115 common and rare variants genotyped as a part of the International AMD Genomics Consortium (IAMDGC). The samples used to generate these data were from 16,144 advanced AMD cases and 17,832 controls. To acquire a more comprehensive understanding of which curated biological and genomic entities contribute to AMD risk, we performed PARIS using multiple biological databases including KEGG, Reactome, GO, NetPath, BioGRID, MINT, Pfam, and ORegAnno. Entities with p-value<0.0001 were prioritized for further investigation. Preliminary results confirmed several pathways previously implicated in AMD, including the complement cascade, Wnt signaling, and inflammation pathways. Genes previously implicated in the pathophysiology of AMD, such as *SYN3*, were consistently enriched for significant interactions and regulatory regions identified by PARIS. Preliminary results also showed that 2 genes, *PLCG2* and *CYP1A1,* had significant signals across multiple immune, metabolic, and signaling pathways among the NetPath, KEGG, GO, and Reactome databases. These anchor genes nominate multiple biological pathways and processes in AMD etiology. Further examination is underway to identify the "driver" genes among these phenomena and determine how they may be contributing to AMD risk. This analysis highlights how the combination of genome-wide genotyping and statistical pathway analysis incorporating *in silico* functional data can be used to elucidate the genetic architecture of complex eye diseases like AMD.