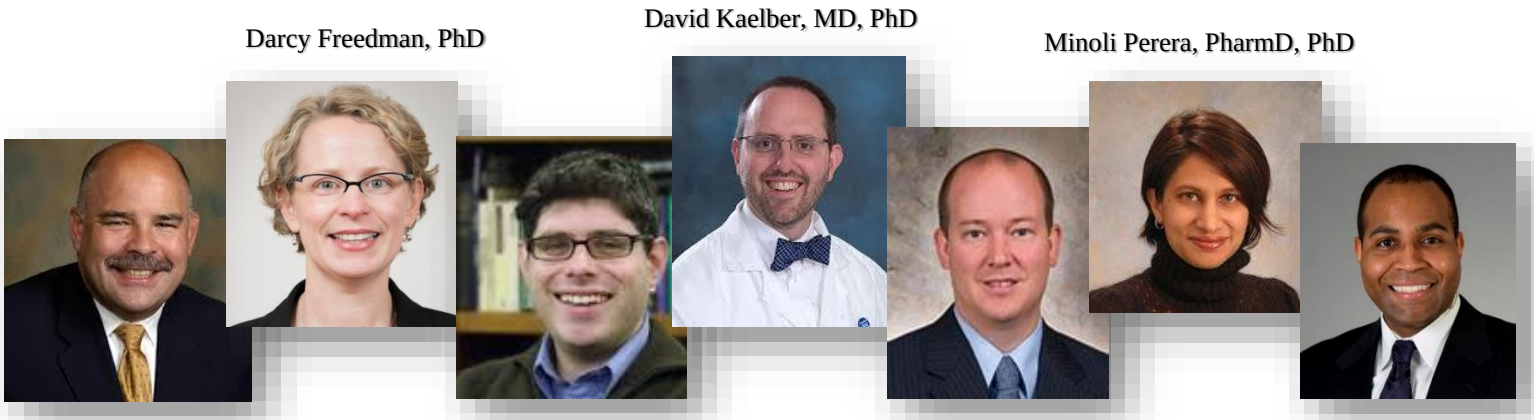
INSTITUTE FOR COMPUTATIONAL BIOLOGY

SCHOOL OF MEDICINE
CASE WESTERN RESERVE UNIVERSITY



Esteban Burchard, MD | Darcy Freedman, PhD | Aaron Goldenberg, PhD | David Kaelber, MD, PhD | Jacob McCauley, PhD | Minoli Perera, PharmD, PhD | Tim Thornton, PhD

CWRU INSTITUTE FOR COMPUTATIONAL BIOLOGY
PRESENTS

PRECISION MEDICINE FOR ALL: ENSURING DIVERSITY IN PARTICIPANTS AND PRACTICE

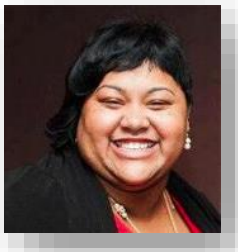
SEPTEMBER 29, 2016

TINKHAM VEALE UNIVERSITY CENTER

WITH SPECIAL GUESTS REPRESENTING THE LACKS FAMILY

Shirley Lacks

Veronica Robinson



SPONSORED IN PART BY
UCITE Nord Grant
National Institutes of Health/National Human Genome Research Institute (HG009481)
Cell and Molecular Biology Training Program (T32 GM008056-34)
CWRU Department of Bioethics
and Adaptive biotechnologies®

Welcome

September 29, 2016

Welcome to the second 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 and scientific advantages of including diverse populations in the study and implementation of ***precision medicine***. 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) is ramping up, and the first projects are being awarded to ascertain and follow 1 million Americans in a study of genetics, environment, and lifestyle to better understand their influence on human health.

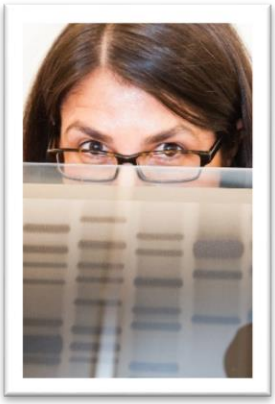
Despite the excitement surrounding these research developments, patient enrollment is a tricky business. There are many factors – personal, statistical, ethical, social, demographic, and otherwise – that may make it difficult to ensure that all populations are represented in research studies, and that all populations stand to benefit from research findings. We have seven local and national speakers scheduled to present the latest research related to these and other topics focused on diversity in precision medicine research.

The lunch break will feature a poster session and a presentation from Dr. Patrick Raber from Adaptive Biotechnologies, a biotechnology company that specializes in combining high-throughput sequencing and expert bioinformatics to profile T-cell and B-cell receptors.

We will close the symposium with special guests Veronica Robinson and Shirley Lacks, great-granddaughter and daughter-in-law of Henrietta Lacks, respectively. Henrietta Lacks was a cervical cancer patient whose biospecimens were used without her consent to create immortal cell lines, now known as HeLa cells. Since the publication of Rebecca Skloot's *The Immortal Life of Henrietta Lacks*, the Lacks family has spoken to audiences about their family's experience both as research participants and their role in informing and directing research and policy.

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!

Dana C. Crawford, PhD
Chair, Organizing Committee
Associate Professor
Epidemiology and Biostatistics
Institute for Computational Biology



INSTITUTE FOR COMPUTATIONAL BIOLOGY

Schedule

Precision Medicine for All: Ensuring Diversity in Participants and in Practice
September 29, 2016
Tinkham Veale University Center, Ballroom A

8:00 – 8:30 Registration and Breakfast

8:30 – 8:40 Welcome and Introduction

Dana Crawford, PhD, Associate Professor
Department of Epidemiology and Biostatistics
Institute for Computational Biology
Case Western Reserve University

8:45 – 9:30 The need for biomedical research in diverse populations

Esteban Burchard, MD, MPH, Professor
School of Pharmacy
University of California San Francisco

9:30 – 10:15 Public attitudes towards the use of genetic research to address health disparities

Aaron Goldenberg, PhD, Associate Professor
Department of Bioethics
Associate Director, Center for Genetic Research Ethics and Law
Case Western Reserve University

10:15 – 10:30 Break

10:30 – 11:15 Outreach and experiences in Cleveland participation and retention in research (foodNest)

Darcy Freedman, PhD, Associate Professor
Department of Epidemiology and Biostatistics
Jack, Joseph, and Morton Mandel School for Applied Social Sciences
Core faculty, Prevention Research Center for Health Neighborhoods
Case Western Reserve University

11:15 – Noon Biorepositories, data collection, and data analysis in US territories

Jacob McCauley, PhD, Associate Professor
Departments of Human Genetics and Pathology
Associate Director, Center for Genome Technology
John P. Hussman Institute for Human Genomics
University of Miami

Noon – 1:30pm Lunch and poster session

Immunosequencing: Unveiling New Molecular Biomarkers to Guide Precision Medicine

Patrick Raber, PhD
Adaptive Biotechnologies

1:30pm – 2:15pm Genetic ancestry and precision medicine

Timothy Thornton, PhD, Associate Professor
Department of Biostatistics
University of Washington

2:15pm – 3:00pm Use of electronic health records in Cleveland for diverse precision medicine research enrollment

David Kaelber, MD, PhD
Chief Medical Informatics Officer and Vice-President of Health Informatics
The MetroHealth System

3:00pm-3:15pm Break

3:15-4:00pm Pharmacogenomics (and its challenges) in diverse populations

Minoli Perera, PharmD, PhD, Associate Professor
Department of Pharmacology and Center for Pharmacogenomics
Feinberg School of Medicine
Northwestern University

4-4:45pm Biobanking and research from the participants' (and their families') perspectives

The Lacks Family
Veronica Robinson and Shirley Lacks, great-granddaughter and daughter-in-law of Henrietta Lacks.
The invitation of the Lacks family was made possible, in part, by a grant from the Nord Family Foundation through CWRU's UCITE Nord Grant, the CWRU Cell and Molecular Biology Training Program (T32 GM008056-34), and the CWRU Department of Bioethics.

5pm Thank you and adjourn

Reception
Smith Commons, 1st floor Tinkham Veale University Center

Precision Medicine for All:

Ensuring Diversity in Participants and in Practice

INSTITUTE FOR COMPUTATIONAL BIOLOGY

Speakers

Esteban Burchard, MD, MPH

Dr. Burchard is Harry Wm. and Diana V. Hind Distinguished Professor in Pharmaceutical Sciences Professor and Vice Chair, Departments of Bioengineering & Therapeutic Sciences and Medicine Director, Center for Genes, Environments & Health at the University of California, San Francisco. Dr. Burchard's laboratory investigates the genetic basis for differences in response to treatment for asthma. Dr. Burchard's laboratory is specifically interested in identifying "ethnic-specific" genetic and biologic risk factors for asthma, asthma severity, and drug responsiveness among US ethnic and racial minority groups as well as developing and applying statistical methods to admixed populations. Dr. Burchard is Principal Investigator of Gene-Environment Studies of Asthma in Hispanic/Latino Children (GALA) I and II among other NIH-funded studies. In March 2015, Dr. Burchard was named to the Precision Medicine Initiative (PMI) Working Group of the Advisory Committee to the National Institutes of Health (NIH) Director. More recently, Dr. Burchard was awarded the 2016 American Thoracic Society Innovations in Health Equality – Lifetime Achievement Award.

Darcy Freedman, PhD, MPH

Dr. Freedman is Associate Professor in the Department of Epidemiology and Biostatistics, Associate Professor in the Jack, Joseph, and Morton Mandel School for Applied Social Sciences, and Associate Director of the Prevention Research Center for Health Neighborhoods (PRC) at CWRU where she also serves as Affiliated Faculty in the Center for Reducing Health Disparities. Dr. Freedman is a population health and social scientist with expertise in food access interventions in food insecure communities. She is currently co-PI of the R01 NIDDK-funded Future of Food in Your Neighborhood (foodNEST) study, a three-year quasi-experimental pragmatic trial assessing fruit and vegetable intake among participants with access to Hub 55, a new food venue opening in the St. Clair Superior neighborhood in Cleveland, a known food desert. Additionally, Dr. Freedman is PI of the FreshLink Study, the core research of the CDC-funded Prevention Research Center at CWRU and leads the Building Capacity for Obesity Prevention study in partnership with collaborators at the Ohio Department of Health and The Ohio State University, Cooperative Extension. Dr. Freedman's latest research includes the development of evaluation technology to support nationwide implementation of healthy food incentive interventions. Dr. Freedman has extensive experience in nutrition-related community interventions and measuring outcomes through research collaborations with a range of community partners including federally qualified health centers, public housing authorities, Boys and Girls Clubs, and neighborhood associations. Dr. Freedman serves on the editorial boards of *International Journal of Food Safety, Nutrition and Public Health* and *Social Work Research,* is a merit-based member of the American Academy of Health Behavior, and is a past awardee (2011) of the NIH Loan Repayment Program in Health Disparities Research.

Aaron Goldenberg, PhD, MPH

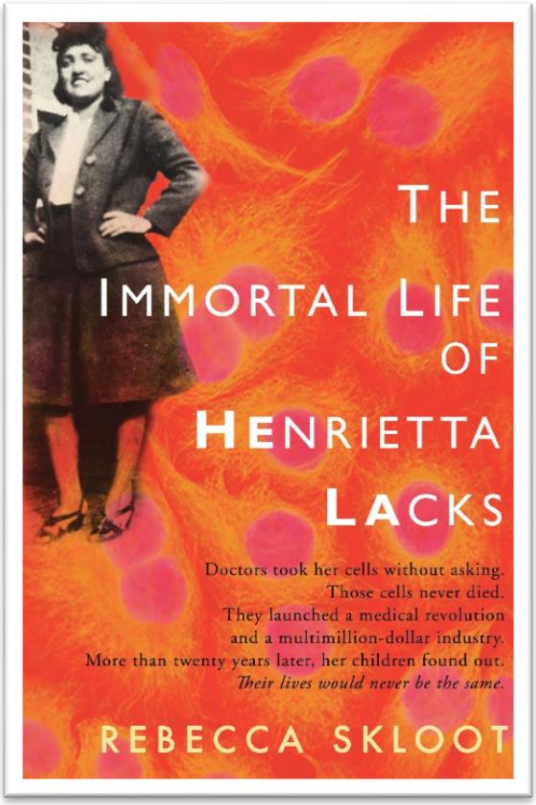
Dr. Aaron Goldenberg leads the Ethics Core for the Institute for Computational Biology and is an Associate Professor in the Department of Bioethics at CWRU School of Medicine. He is also the Director of Research for the Department. He is also the Associate Director of the Center for Genetic Research Ethics and Law, a NIH Center of Excellence in Ethical, Legal, and Social Implications Research for the National Human Genome Research Institute. He earned his PhD in Bioethics at CWRU. Since joining the faculty at CWRU, Dr. Goldenberg's work has focused on the ethical, legal, and social implications of genetics and genomics in clinical and public health settings. His research program has been grounded by a number of major project areas, including: 1) ethical implications of expanding newborn screening programs; 2) storage and use of perinatal and pediatric biological specimens for future research; 3) implications of genetics and gene-environment interactions for racial/ethnic minorities and other communities experiencing health disparities. Dr. Goldenberg is currently the Co-PI of a project funded by the Health Resources and Services Administration (HRSA) to explore the ethical and programmatic challenges of integrating genomic technology into Newborn Screening Programs, and he is the Principal Investigator on a project funded by the NIH National Human Genome Research Institute to examine parental attitudes regarding the research use of biospecimens collected from newborns. Dr. Goldenberg is also leading a project to assess how genomic advances may impact medically-underserved communities, and how clinicians and public health agencies could better assess biological and social determinants to account for gene-environment interactions. In addition to these scholarly initiatives, Dr. Goldenberg is Director for Ethics, Policy and Practice for the National Newborn Screening Clearinghouse, also known as Baby's First Test. He is a member of the Ethics and Legal Workgroup for the Newborn Screening Translational Research Network and the Legal and Legislative Workgroup for the American Public Health Laboratory Association. He is a member of the Pediatric Task Team for the Global Alliance for Genomics and Health.

David Kaelber, MD, PhD, MPH, MS, FAAP, FACP

Dr. Kaelber is a practicing internist, pediatrician, Vice-President of Health Informatics, and Chief Medical Informatics Officer at the MetroHealth System. Dr. Kaelber is also Professor in the CWRU Department of Internal Medicine with a secondary appointment in the Departments of Pediatrics and Epidemiology and Biostatistics. MetroHealth is an integrated health system and a primary teaching site for Case Western Reserve University (CWRU). Since the late 1990s, MetroHealth has been at the forefront in adopting electronic health records to advance high quality clinical care in the Cleveland area. Under Dr. Kaelber's leadership, MetroHealth implemented Care Everywhere, a health information exchange program that enables MetroHealth to share patient information with other healthcare systems in the United States. And, in 2011, Dr. Kaelber was instrumental in the development and launch of MyChart, the personal care health record complement of MetroHealth's EHR (Epic). Dr. Kaelber's research interests are rooted in pediatric populations where he has used EHR data to identify previously under-diagnosed hypertension patients and to monitor obesity trends over time. This work earned Dr. Kaelber recognition by the American Heart

Association as one of the top ten breakthroughs in cardiovascular and stroke medicine in 2007. More broadly, Dr. Kaelber's research interests span EHRs, clinical decision support, use of EHRs in chronic disease detection and management, personal health records, and medical informatics.

Shirley Lacks & Veronica Robinson



The Lacks Family are direct descendants from Henrietta Lacks, a cervical cancer patient whose biospecimens were used without her consent to create immortal cell lines now known as HeLa cells. The impoverished Lacks Family, with no knowledge of the cell lines or their use in biomedical research, had no opportunity to understand or benefit from the potential commercial applications of the biospecimens. Since the publication of Rebecca Skloot's *The Immortal Life of Henrietta Lacks*, the Lacks Family has spoken to audiences about their family's experience both as research participants and as participants informing and directing research and policy.

We are delighted that Ms. Shirley Lacks and Ms. Veronica Spencer will be speaking with us about the legacy of their family's experience in biomedical research and its potential impact on participation and ethics in this modern era of precision medicine.

Shirley Lacks is Henrietta's daughter-in-law and was close friends with Henrietta's daughter Deborah. Veronica Spencer Robinson is Henrietta Lacks' great granddaughter. Ms. Robinson is a Lacks Family representative on a panel for the National Institutes of Health (NIH) responsible for reviewing applications for the use of the HeLa cells in research.

Jacob McCauley, PhD

Dr. McCauley is Associate Professor of Pathology and Dr. John T. Macdonald Foundation Department of Human Genetics at the University of Miami. Dr. McCauley is also Associate Director at the Center for Genome Technology, John P. Hussman Institute for Human Genomics and Director of the Biorepository Core Facility in the Center for Genome Technology at the John P Hussman Institute for Human Genomics. Dr. McCauley is formally trained in human genetics, and his primary interest is to improve the understanding of human disease through disease gene discovery, genomics, and in-depth examination of environmental factors that influence disease outcome. His research focuses on the use of molecular techniques, bioinformatics, and statistical methods to identify genetic variation and to characterize its role in disease susceptibility and outcomes within a variety of human diseases. He has significant experience overseeing biological sample collection, tracking, quality control, genotyping, sequencing and analysis involved in large-scale human genetics projects. Dr. McCauley is a member of several multidisciplinary collaborations with colleagues both nationally and internationally. He has been involved in studying a variety of complex human diseases including autism, Alzheimer disease, stroke, multiple sclerosis, and inflammatory bowel disease.

Minoli Perera, PharmD, PhD

Dr. Perera is an Associate Professor in the Department of Pharmacology and Center for Pharmacogenomics at Northwestern University's Feinberg School of Medicine. With expertise in pharmacokinetics, clinical pharmacology and human genetics, Dr. Perera focuses her research interests on pharmacogenomics in minority populations. Examples of her work include identifying common genetic variants associated with warfarin dose and identifying genomic risk factors for venous thromboembolism, both in African Americans. Dr. Perera anticipates that results from these and other multidisciplinary studies will allow the development of an economically sustainable program in clinical pharmacogenetics, including a consult service. Overall, Dr. Perera's goal is to bring meaningful translation of pharmacogenetics research findings into clinical care, a goal very much consistent with precision medicine.

Patrick Raber, PhD

Dr. Patrick Raber is a scientific liaison with Adaptive Biotechnologies, a biotechnology company based in Seattle, Washington that specializes in combining high-throughput sequencing and expert bioinformatics to profile T-cell and B-cell receptors. Previous to his position at Adaptive, Patrick earned his PhD in the Department of Microbiology, Immunology, at Parasitology at LSU Health Sciences Center New Orleans. Based in Atlanta, Georgia, Patrick serves as scientific expert for applications of Adaptive's immunosequencing platform including oncology, immunotherapy, transplantation, infectious diseases, and autoimmunity. Patrick also assists investigators interested in incorporating Adaptive's technology into clinical trials and research protocols.

Tim Thornton, PhD

Dr. Thornton is Associate Professor of Biostatistics at the University of Washington (UW) and an Affiliate Investigator at the Fred Hutchinson Cancer Research Center. Dr. Thornton's research involves the development and application of statistical methods for the analysis of large-scale genomic data for the identification of genetic variants that influence complex diseases and quantitative traits. Currently, Dr. Thornton's collaborative research is largely focused on using statistical methods to identify novel genetic variants for complex traits in ancestrally diverse populations, including underrepresented minority populations in the U.S. (African Americans, Latinos, and American Indian and Alaska Native (AI/AN) populations). Dr. Thornton currently serves as a lead statistician and co-investigator for Hispanic Community Health Study (HCHS)/Study of Latinos (SOL), and this cohort was recently included in the multi-cohort Population Architecture using Genomics and Epidemiology (PAGE) II study.

Precision Medicine for All:

Ensuring Diversity in Participants and in Practice

INSTITUTE FOR COMPUTATIONAL BIOLOGY

Travel Scholarship Awardees

Made Possible by NIH/NHGRI R13HG009481

J Preston Campbell

Post-doctoral fellow
Department of Cancer Biology
Vanderbilt University
Nashville, TN

I am a practical translational biologist and occasional toolmaker, focused on complex systems and metastasis research. Broadly, I am interested in applying rigorous quantitative analysis to model biological systems across scales, from single-cell behavior to patient outcomes. During my graduate training pharmacology elucidated a mechanism at the intersection of psychosocial stress, adrenergic pharmacology, bone biology, cancer metastasis, and heart disease. We found a pragmatic solution to this complex problem: beta blockers to prevent bone metastasis. However, I was dissatisfied with the in vivo methods for quantifying early metastasis and developed a method to detect a single metastatic cell in a mouse bone.

During my postdoctoral work I focused on drug discovery and development and published the first in vivo validation of a RSK inhibitor in the context of breast cancer metastasis. During this time I became less enamored with the traditional reductionist biologist approach to scientist and made the transition to a more quantitative approach.

Currently I am in the lab of Vito Quaranta, exploring lung cancer heterogeneity and generating novel quantitative approaches to address chemotherapeutic resistance. Specifically I am trying to identify epigenetic changes in lung cancer that predict patient response to targeted therapy, and developing math modeling and experimental approaches to validate potential therapies. The goal of my research is to develop metrics that allow us to quantify a specific patient signature prior to treatment thus allowing the clinician to tailor treatment to both the tumor signature and the individual patient profile.

Leslie Ann Caromile

Assistant Professor
Center for Vascular Biology
University of Connecticut Health Center
Farmington, Connecticut

I earned an undergraduate degree in neurobiology from the University of Connecticut, a MS in molecular biology from California State University, Los Angles and a PhD in pathology from the University of Washington School of Medicine. I am currently an assistant professor in the Center for Vascular Biology at the University of Connecticut Health Center. My current studies, supported by my NIH/NCI K01 Mentored Research Scientist Award to Promote Diversity, focus on the transmembrane peptidase prostate specific membrane antigen (PSMA). Expression of PSMA is markedly increased in primary and metastatic prostate carcinomas where it correlates negatively with prostate cancer patient prognosis. However, the functional role of PSMA remains unclear. It has been my goal to identify the

molecular mechanisms responsible for PSMA function in this disease. Currently, I am focusing on building a mathematical model of PSMA signaling/regulation in advanced prostate cancer with the hope of using this model to predict new biology. In addition to research, I am also involved in many award winning activities related to undergraduate and graduate research training, mentoring and outreach.

My overall career goal is to become an independent researcher in the field of cancer biology within an academic setting. As a successful practicing scientist and mentor, I plan to use my own successes and familiarity with the practice and administration of science to help other Native Americans and minorities enter this profession.

Jonathan Deane

Senior Investigator I
Cancer Immunotherapeutics
Genomics Institute of the Novartis Research Foundation
San Diego, California

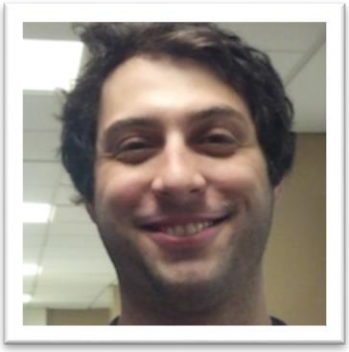
The research in my group is focused immune regulation which has implications for both autoimmunity and driving immune responses for infectious diseases and cancers. In the case of autoimmunity, we have explored projects that either inhibit pro-inflammatory pathways or activate anti-inflammatory pathways. For oncologic indications, we have focused on activating pro-inflammatory pathways in a safe and selective manner. This has led us to identify and develop low molecular weight therapeutics through phenotypic screening, and we have also worked with protein science colleagues to develop biologics. We are currently working on 7 projects that are at various stages of development, from target validation to preparation for clinical trials. As we look towards the clinic, we have to generate data showing that our compounds are safe and effective, and determine whether the safety and efficacy is generalizable or specific to patient subsets.

Alexandra Fish

PhD candidate
Program in Human Genetics
Vanderbilt University
Nashville, Tennessee

I use both statistical and machine-learning based approaches to better understand the genetic regulation of gene expression levels and how this relates to the development of complex disease. I have taken several approaches to address this, including: demonstrating that apparent epistasis influencing gene expression in humans is consistent with statistical artifacts; identifying sequence properties predictive of putative enhancers which are conserved across mammalian evolution; and an ongoing project examining the influence of differential ancestry on EHR-derived phenotypes in admixed populations.

Benjamin Glicksberg



PhD candidate
Graduate School of Biological Sciences
Icahn School of Medicine at Mount Sinai
New York, New York

My current work revolves around incorporating genetic, clinical, and epidemiological frameworks into predictive models for personalized medicine applications. I am involved in many projects that incorporate Mount Sinai's Electronic Medical Records (EMR) and genetic Bio*Me* BioBank repository. An early work of mine focused on identifying pairs of diseases that were significantly connected across 3 modalities, including clinical (EMR), genetic, and within the literature space. Additionally, I assisted on a paper using topological data analysis to identify 3 different subgroups within the highly complex Type 2 Diabetes, utilizing both EMR and BioBank datasets. This work has direct application to precision medicine, as a primary goal would be to tailor treatments to patients falling into different subgroups, based on their clinical background and genetic architecture. Additionally, we have recently completed a study identifying loss-of-function mutations within our BioBank population and determining if these mutations are associated with modulation of cardiovascular traits and/or disease risk. By classifying certain loss-of-function mutations in genes that are significantly associated with protective properties of these traits (i.e. loss-of-function in gene X being associated with lower cholesterol levels) we might be able to identify novel targets for therapeutic treatments. In the context of personalized medicine, the classification of both these deleterious (i.e. associated with undesired trait change) and protective mutations might facilitate future genetic counseling recommendations to individuals carrying these variants.

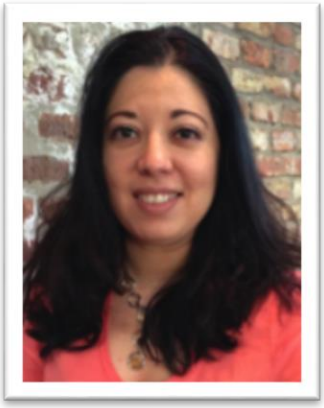
Jacob Hall

Post-doctoral fellow
Translational and Biomedical Informatics
Icahn School of Medicine at Mount Sinai
New York, New York

Acne is a common skin condition that primarily affects teenagers and young adults. Though not life-threatening, severe acne can have both physical and emotional consequences. A wide range of treatment and prevention methods exist, corresponding to the severity of acne. Drugs used to treat severe acne, such as Isotretinoin, can lead to severe birth defects; Other side effects include depression and suicidal thoughts. In addition to possible side effects, Isotretinoin only clears up severe acne in 80% of patients. Better treatment for acne is justified, and precision medicine may play a role.

In our lab, we are studying the facial microbiome to elucidate possible microbial signatures or host-pathogen interactions. As a preliminary technical study, we compared multiple collection and sequencing methods to determine the best overall approach for capturing the facial microbiome. We found that whole genome sequencing, while more expensive, is justified because it provides a much greater diversity of bacterial and non-bacterial genomic information; additionally, available WGS analysis pipelines provide better species identification and quantification. Preliminary results have shown that certain combinations of bacteria are more likely to co-occur in patients with acne. One bacterial species associated with acne is *Propionibacterium acnes*. We seek to determine whether subspecies of *P. acnes* are associated with acne severity. A better understanding of the microbial landscape for both people with normal skin and people with acne may help to develop better treatment methods that seek to establish a "good" facial microbiome and treat or prevent acne.

Wenndy Hernandez



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

I am committed to deciphering the molecular mechanisms underlying complex human diseases, particularly those that disproportionally afflict under-represented minorities. My goal is to reduce, and ultimately eliminate, disparities in stroke incidence, morbidity, and mortality in African Americans through research by identifying molecular determinants of ischemic and hemorrhagic stroke in order to enhance patient treatment. I have a broad interest in cardiovascular diseases and my overall goal is to be a leader in research and education on cardiovascular disease disparities by leading a team of investigators with appropriate expertise and collaborations to elucidate critical molecular pathways in cardiovascular diseases. In addition, I am committed to train the next generation of biological and clinician scientists, particularly from under-represented populations, with the necessary core skills needed to conduct genomics studies. My experience in patient-oriented research and statistical genetic analysis puts me in a unique position to integrate these disciplines in a meaningful manner to conduct genomic research with broad translational impact.

Rachel Hodos

PhD candidate
Computational Biology
New York University
New York, New York

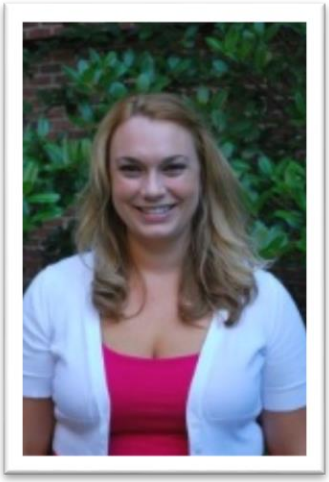
My current work is focused on omics-based drug repurposing, that is finding new uses for existing pharmacotherapies, primarily based on gene expression. The basic idea is to look for gene expression patterns that are reversed between drug and disease, hypothesizing that reversal of expression levels might indicate reversal of the higher-level disease phenotype (see Figure 8 from http://onlinelibrary.wiley.com/doi/10.1002/wsbm.1337/epdf) . I am working both on developing novel statistical methodologies, as well as applying these techniques to discover new therapies for cystic fibrosis. I am also interested in integrating a variety of data sources to better understand and predict drug mechanisms and side effects.

Joshua Hoffman

Post-doctoral fellow
Epidemiology and Biostatistics
University of California San Francisco
San Francisco, California

I am currently a postdoctoral scholar in the department of epidemiology and biostatistics at the University of California, San Francisco. My major research focus is in the elucidation of breast cancer associated risk variants in Latina and African American sample populations through GWAS, as well as admixture mapping of prostate and breast cancer risk loci. In addition to studying cancer risk in underrepresented populations, I am also involved in identifying the shared genetic basis of a range of varying cancers within Caucasians. Another avenue of my work is the use of EHR based cohorts for genome-wide association studies of asthma, allergy, and drug response.

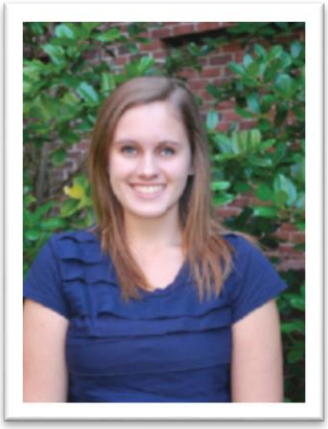
Brittany Hollister



PhD candidate
Program in Human Genetics
Vanderbilt University
Nashville, Tennessee

My research focuses on how the interaction between socioeconomic and genetic factors affects complex diseases with racial health disparities. Thus far, I have focused on extracting socioeconomic status (SES) information from de-identified electronic health records so that SES data can be more easily incorporated into genetic studies which utilize these records. The next portion of my project examines the relationship between the extracted SES data and genetic factors which affect blood pressure in African Americans. To date, very few genetic studies of blood pressure in underrepresented populations incorporate SES data, despite prior associations between SES factors and blood pressure. This could lead to misleading conclusions regarding the genetic variants contributing to blood pressure. My work will provide a more complete picture of the biology of blood pressure because it will include both SES and genetic data.

Carissa Jones



PhD candidate
Program in Human Genetics
Vanderbilt University
Nashville, Tennessee

My dissertation project has two distinct areas of research. The first aim of my dissertation research is to identify genetic variants associated with lung cancer incidence and mortality in African Americans. Despite higher lung cancer incidence and mortality rates compared to all other racial/ethnic populations, African Americans have been poorly represented in biomedical research. In my research, I have identified a variant associated with survival in African Americans that was previously identified in a European American population; however, the variant appears to have a population-specific effect. To our knowledge, this is the first genome-wide association of lung cancer survival in African Americans. Additionally, my research examines cross-cancer pleiotropic associations of lung cancer risk in African Americans.

The second aim of my dissertation is focused on characterizing the African origins of African Americans. African Americans are an admixed group of individuals with approximately 80% African ancestry and 20% European ancestry. Previous studies have shown that this African ancestry is predominantly of West African origin. Our preliminary data show that African ancestry may vary across the United States, and particularly in the South, where the majority of slaves were imported. This aim seeks to uncover fine-scale resolution of African genetic ancestry in African Americans, and overlay it with known historical data surrounding the salve trade movement.

Mike Sivley

PhD candidate
Program in Biomedical Informatics
Vanderbilt University
Nashville, Tennessee

My research focuses on the analysis of protein-coding variation and how it influences human health. Specifically, I am interested in analyzing coding variation within its functional context – protein structure – and have developed resources that efficiently map protein-coding variants into all solved and predicted structures of human proteins. We find significant differences between the spatial distributions of synonymous and missense variation, indicating spatial constraint on amino acid substitution. I have expanded on this to develop a novel, predictive feature using the relative spatial proximity of an amino acid to known pathogenic and neutral missense variation and find that it performs comparably with other methods in the presence of pathogenic variant clusters. Through our collaboration with the Undiagnosed Disease Network (UDN), we've shown that we can use spatial distributions to predict the pathogenicity of variants of unknown significance.

Precision Medicine for All:

Ensuring Diversity in Participants and in Practice

INSTITUTE FOR COMPUTATIONAL BIOLOGY

Poster Abstracts

William S. Bush
Epidemiology and Biostatistics
Case Western Reserve University

Nuclear-Mitochondrial Interactions and Neurocognitive Impairment in HIV+ Adults

Sandra Smieszek[1], Todd Hulgan[2], David C. Samuel[2], Donald R. Franklin[3], Robert K. Heaton[3], Scott L. Letendre[3], Ron J. Ellis[3], Asha R. Kallianpur[4], and William S. Bush[1,4]

1)Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, USA. 2)Vanderbilt University, Nashville, Tennessee. 3)University of California-San Diego, California. 4)Cleveland Clinic Foundation/Lerner Research Institute and Cleveland Clinic Lerner College of Medicine, Ohio.

HIV-Associated Neurocognitive Disorder (HAND) is a term that captures a wide spectrum of neurocognitive deficits ranging from mild to severe, in HIV-infected persons. The genetic underpinnings of this complex phenotype are incompletely understood. Mitochondrial function has long been thought to play a role in neurodegeneration, along with iron metabolism and transport. Abnormalities of mitochondrial function and iron metabolism have long been implicated in neurodegeneration. In this analysis, we aimed to characterize the mitochondrial DNA (mtDNA) haplogroup interactions with nuclear genes found to be associated with HAND phenotypes in the CHARTER cohort, encompassing 1025 individuals of African, admixed Hispanic and European ancestry. We first employed a polygenic modeling approach to investigate the global effect of previously associated nuclear SNPs, and to examine how the polygenic effect of these SNPs is influenced by MT haplogroups. We found evidence of interactions between nuclear genomic SNPs *en masse* and MT haplogroups within European and African ancestry individuals, as evidenced by Table 1. Subsequently, we performed an analysis of each SNP by mtDNA haplogroup combination, and detected significant interactions between two nuclear SNPs (rs17160128 and rs12460243) and European mtDNA haplogroups, with the SNPs showing a more dominant association in H and J haplogroups versus a more additive association in T and UK haplogroups. These associations highlight the role of *FBN3*, a gene that belongs to the fibrillin gene family. Fibrillins are extracellular matrix molecules, which assemble into microfibrils in many connective tissues and are important in regulating pathways of the immune response, inflammation and tissue homeostasis and involved in maintenance of blood-brain-barrier integrity. These findings utilized a novel analytic approach and indicate a new potential genetic mechanism in the pathogenesis of HAND and may shed light on the pathophysiology of this debilitating neurocognitive disease.

Ancestry Group	Genetic Variance	Residual Variance	Phenotypic Variance	Ratio	Full Model Log-Likelihood	Reduced Model Log-Likelihood	LRT	P-value	n
European Descent	0.042	0.216	0.283	0.236	64.284	62.975	2.618	0.05283	440
Admixed Hispanic	0.002	0.112	0.207	0.459	13.714	13.747	0	0.5	101
African Descent	0.011	0.109	0.126	0.133	241.253	239.336	3.835	0.0251	484

Table 1. Mixed-model analysis of GWAS-associated SNPs on continuous GDS with terms for mitochondrial interaction

Mariusz Butkiewicz
Epidemiology and Biostatistics
Case Western Reserve University

Whole Exome Sequencing of Amish AMD cases reveals suitable variants to populate novel Amish Exome Chip

Mariusz Butkiewicz[1], Dwight Stambolian[3], Margaret A. Pericak-Vance[2], William K. Scott[2], Jonathan L. Haines[1]
[1]Case Western Reserve University, Institute for Computational Biology, Cleveland, OH; [2]University of Miami, Hussman Institute for Human Genomics, Miami, FL; [3]University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania, PA.

Purpose: The genetic architecture within the Amish shows significant differences compared to the general population. Such variants that differ from the general population have the potential to reveal novel markers for complex diseases. Here, as an effort to populate a novel Exome Chip we identified rare and novel coding variants that could potentially be significant for complex diseases such as age-related Macular Degeneration (AMD). We conducted Whole Exome Sequencing on 89 Amish individuals affected by AMD. After a quality control (QC) protocol was applied to determine high quality SNPs, the filtered variants were analyzed with respect to availability in the general population accessed through the Exome Aggregation Consortium (ExAC). The value of genetic studies in relatively isolated populations such as the Amish derives from the recent expansion of the population from a small number of founders and continued cultural isolation that restricts the introduction of additional genetic variation leading to large and stable pedigrees.

Methods: Whole Exome Sequening of 89 Amish individuals with AMD provided a total of 116,811,848 called nucleotide base pairs before QC. After applying QC filters, a set of 292,162 variants was determined. These variants were analyzed with respect to two criteria: 1) variants are present and rare in the general population available through ExAC and 2) variants are not present in ExAC and only seen among variants specific to the 89 Amish individuals.

Results: Based on the outlined filtering criteria, we identified 19,371 variants that are present and rare in the general population (minor allele frequency < 0.01). Further, we found 115,668 variants that were novel and did not appear among variants in the general population (ExAC). Annotating these variants for pathogenic phenotypes through ClinVar showed no association.

Conclusions: The goal here was to identify rare and novel coding variants that could be added to a new exome chip to genotype new Amish samples to discover AMD cases and controls

Jessica Cooke Bailey
Epidemiology and Biostatistics
Case Western Reserve University

Attitudes towards centralized biorepositories among patients in Cleveland, OH: Implications for the Precision Medicine Cohort Program

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

The Precision Medicine Initiative, announced by the White House in early 2015, promises to revolutionize patient care by incorporating clinical, environmental, personal, and 'omics data in electronic health records (EHRs) for decision support at the point of clinical care. To accelerate this vision, the National of Institutes of Health in consultation with academic, healthcare, mHealth, and patient advocacy leaders, is launching the PMI Cohort Program (PMI-CP), an ambitious national effort to ascertain one million Americans for precision medicine research. A requirement of this cohort is the submission of electronic health records and biospecimens to a central biorepository facility funded by the government for storage and further processing for data generation and analysis. While this centralized biorepository model provides quality control and is cost-effective compared with a federated model, it is unclear what impact this requirement will have on ascertainment, which is expected to be socioeconomically, geographically, and racially diverse. To understand the potential impact this model has on ascertainment, we are surveying patients participating in a precision medicine research project at MetroHealth in collaboration with the Institute for Computational Biology at Case Western Reserve University in Cleveland, Ohio, a diverse metropolitan area with 20% of residents self-described as African American in the 2010 U.S. Census. MetroHealth is an academic, public provider, integrated tertiary care system serving northeast Ohio with a vendor-based EHR. The payer mix includes 21% uninsured, 24% Medicaid, 18% Medicare. Results currently include responses from 37 patients; 72% were African American, 55% were female, and average age was 61. Regarding willingness to participate in the PMI, 61% of patients responded favorably to submitting health records and genetic data to a national biorepository coordinated by the government. Only ~39% of patients indicated that return of results was very important whereas ~23% did not want results at all. When asked if they would be willing to install an application on their mobile device to track their health and to send that data to a national center, ~52% responded favorably to both. These responses, while preliminary, indicate that the attitudes of patients in a diverse health care environment towards the PMI-CP are varied.

Dana Crawford
Epidemiology and Biostatistics
Case Western Reserve University

Chronic kidney disease (CKD) is the gradual loss of kidney function over time. Patients with CKD are characterized by kidney function (estimated by glomerulal filtration rate or GFR) ranging from normal or high (GFR > 90mL/min; stage 1), mild (GFR = 60-89 mL/min; stage 2), moderate (GFR = 30-45 mL/min; stage 3), severe (GFR = 15-29 mL/min; stage 4), and end stage (GFR <15 mL/min; stage 5). Development of mild and moderate CKD (stages 2 and 3) is often associated with a variety of preventable health conditions, including hypertension and diabetes. Progression to moderate and severe CKD is irreversible leading to the eventual requirement of renal dialysis or renal transplant. Progression to moderate and severe CKD is also associated with increased risk of cardiovascular death.

CKD progression to end-stage renal disease (ESRD) is non-linear and highly variable on a patient-by-patient basis. Several factors are associated with progression including male sex, increased age, estimated GFR (eGFR), and albuminuria. Race/ethnicity is also a recognized risk factor for progression as African Americans progress faster to ESRD compared with European Americans. This major health disparity has a substantial genetic component, with a single locus (*APOL1*) accounting for as much as 70% of the observed disparity. These known factors identify many high risk patients through emerging predictive modeling strategies, but not all.

It is likely that there are additional or novel factors yet-to-be associated with CKD progression. Indeed, recent observational studies have suggested that immune status may be a marker for CKD status and progression to ESRD. Patients with ESRD have impaired cellular immune function, much of which is attributable to impairment of T cell function. It is less established how T cell function influences patient progress from early to late stages of CKD and whether or not T cell profiles early in CKD can predict progression to ESRD. Given these observations, we hypothesize that markers of immunity independently predict CKD progression to ESRD.

In collaboration with The MetroHealth System, we are ascertaining patients with electronic health records for precision medicine research in a pilot study known as MIPs (MetroHealth/Institute for Computational Biology Pilot study). We have characterized the TCR repertoire of the first five MIPs participants using Adaptive Biotechnologies' ImmunoSeq. The ImmunoSeq assay coupled with the ANALYZER bioinformatics pipeline provide individual sample-level measures of TCR diversity including number of rearrangements, number of productive rearrangements, and productive clonality (ranging from 0 to 1 representing the range from polyclonal to monoclonal). In this preliminary set, productive clonality ranged from 0.0151 to 0.1632, and the distribution of the most common clones differed across the five samples. Further TCR repertoire characterization of MIPs participants, including associations with clinical variables, is ongoing.

Alexandra Fish
Vanderbilt Genetics Institute
Vanderbilt University

The importance of epistasis - or statistical interactions between genetic variants - to the development of complex disease in humans has been controversial. Genome-wide association studies of statistical interactions influencing human traits have recently become computationally feasible and have identified many putative interactions. However, statistical models used to detect interactions can be confounded, which makes it difficult to be certain that observed statistical interactions are evidence for true molecular epistasis. In this study, we investigate whether there is evidence for epistatic interactions between genetic variants within the cis-regulatory region that influence gene expression after accounting for technical, statistical, and biological confounding factors. We identified 1,119 (FDR=5%) interactions that appear to regulate gene expression in human lymphoblastoid cell lines, a tightly controlled, largely genetically determined phenotype. Many of these interactions replicated in an independent dataset (90 of 803 tested, Bonferroni threshold). We then performed an exhaustive analysis of both known and novel confounders, including ceiling/floor effects, missing genotype combinations, haplotype effects, single variants tagged through linkage disequilibrium, and population stratification. Every interaction could be explained by at least one of these confounders, and replication in independent datasets did not protect against some confounders. Assuming the confounding factors provide a more parsimonious explanation for each interaction, we find it unlikely that cis-regulatory interactions contribute strongly to human gene expression, which calls into question the relevance of cis-regulatory interactions for other human phenotypes. We additionally propose several best practices for epistasis testing to protect future studies from confounding.

Brittany Hollister
Vanderbilt Genetics Institute
Vanderbilt University

Development of a text mining algorithm to extract socioeconomic data from electronic health records for precision medicine research

Socioeconomic status (SES) is a fundamental contributor to health, particularly when analyzing racial disparities in disease. SES data are rarely included in genetic studies and translational precision medicine research, due in part to the difficultly of collecting these data when studies were not originally designed for that purpose. The emergence of large clinic-based biobanks linked to electronic health records (EHRs) provides research access to large patient populations with longitudinal phenotypic data captured in structured fields as billing codes, procedure codes, and prescriptions. SES data however, are often not explicitly recorded in structured fields, but rather recorded in the free text of clinical notes and communications. The content and completeness of these data vary widely by practitioner. To enable gene-environment studies that consider SES as an exposure, we sought to extract SES variables from racial/ethnic minority patients (n=11,197) in BioVU, the Vanderbilt University Medical Center biorepository linked to de-identified EHRs. We developed several measures of SES using information available within the de-identified EHR, including broad categories of occupation, education, insurance status, and homelessness. 200 patients were randomly selected for manual review to develop an algorithm for extracting SES information from de-identified medical records. The algorithm consists of 15 categories of information, with 830 unique search terms. An additional 50 records were then randomly selected for algorithm evaluation. These 50 records were reviewed by two independent reviewers, with Kappa statistics ranging from 0.47 for occupation to 0.92 for retirement, and percent positive agreements ranging from 80% (Medicaid) to 93.3% (retirement). The manual review results from the 50 records were then compared to the results from the algorithm, resulting in positive predictive values of 35% (education), 63.6% (retirement), 87.5% (unemployment), 81.8% (Medicaid), and 33.3% (homelessness), suggesting some categories of SES data are easier to extract than others. The SES data extraction approach developed here will enable future EHR-based genetic studies to integrate SES information into statistical analyses. Ultimately, incorporation of measures of SES into genetic studies will help elucidate the impact of the social environment on disease risk and outcomes.

Carissa Jones
Vanderbilt Genetics Institute
Vanderbilt University

Lung cancer (LC) is one of the most commonly diagnosed cancers in the United States. Genome-wide association studies (GWAS) have successfully identified several loci associated with LC risk. However, despite higher incidence and mortality rates in African Americans, few LC GWAS have been conducted in African Americans. Given the high burden of multiple testing intrinsic to GWAS, studies have shifted toward examining variants selected based on biological plausibility. One such method of selecting variants involves the investigation of pleiotropy, which occurs when a single variant or locus affects more than one phenotype. Here, we sought to examine the pleiotropic effects of SNPs previously associated with other cancers with lung cancer risk. African American lung cancer cases (N=1,410) and controls (N=2,843) from 5 independent studies were genotyped on the Illumina 1M-Duo array and standard quality control filters were applied; imputation was performed using IMPUTE2 and the 1000G cosmopolitan reference panel (June 2014). SNPs were extracted from the National Human Genome Research Institute - European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog based on the search term “neoplasm”. Manual review filtered out any study in which the outcome was not cancer risk, resulting in 984 SNPs from 220 studies. We examined 916 SNPs, of which 629 were directly genotyped and 287 were imputed (info score >0.8, minor allele frequency >0.01). For each SNP, we performed logistic regression using SNPTest, adjusting for age at diagnosis, sex, smoking status, global African ancestry, and study site. Two SNPs, rs2853677 (in TERT on chromosome 5p15) and rs1051730 (in CHRNA3 on chromosome 15q25), were significantly associated with LC risk (odds ratio (OR) = 1.27, 95% confidence interval (CI): 1.13-1.41 and OR = 1.37, 95% CI: 1.18-1.60, respectively) applying a false discovery rate of 5%. Both SNPs have been previously associated with LC risk in African Americans. The current analysis examined only the variants directly reported in the NHGRI-EBI GWAS catalog. However, because linkage disequilibrium (LD) patterns are known to differ between racial/ethnic populations, it is possible that SNPs discovered in European or Asian populations do not tag the causative SNP in African Americans. Consideration of LD differences between populations will allow for the discovery of associations not captured by examining the reported SNP alone.

Harpreet Kaur
Lerner Research Institute
Cleveland Clinic

Associations of CSF iron transport with neurocognitive function over time in HIV+ adults

H Kaur[1], W Bush[2], S Letendre[3], R Ellis[3], R Heaton[3], S Patton[4], J Connor[4], D. Samuels[5], T. Hulgan[5], A. Kallianpur[1,6]
[1]Cleveland Clinic/Lerner Research Institute, Genome Medicine Institute, Cleveland, USA
[2]Case Western Reserve University, Cleveland, USA
[3]University of California-San Diego, California, USA
[4]Pennsylvania State University College of Medicine, Hershey, PA, USA
[5]Vanderbilt University School of Medicine, Nashville, USA
[6]Cleveland Clinic Lerner College of Medicine

Neurocognitive (NC) decline is a growing concern among HIV-infected (HIV+) individuals as they age and may involve dysregulated cerebrospinal fluid (CSF) iron transport. This longitudinal study examined relationships between CSF iron biomarkers, *APOE*-ε4 genotype, and NC decline in HIV+ enrollees in a large observational cohort study who underwent comprehensive NC testing and global deficit score (GDS). GDS was assigned at baseline and 6-month intervals during follow-up. NC change status (improved/stable/declined) was also defined at specific visits or last follow-up compared to baseline. CSF iron, heavy-chain ferritin (H-ferritin) and transferrin were measured at baseline. Biomarker associations with the GDS and with NC change status at baseline, 30, 36, and 42 months (mos.) of follow-up were evaluated by linear and repeated-measures analysis of variance, regression analyses, adjusting for potential confounders and *APOE*-ε4 carrier status. Of 403 HIV+ adults with CSF biomarker data, 157 completed follow-up at 30 mos. [26% aged≥50, 12% women, 74% on ART, 36% AIDS, mean $CD4^+$ T-cell count 468/mm^3, 65% with undetectable CSF virus, 31% *APOE*-ε4 carriers], 131 at 36 mos. and 110 at 42 mos. Higher H-ferritin at baseline was associated with NC improvement at last follow-up in HIV+ adults aged<50, adjusting for age, comorbidity, ZDV use (which affects iron transport), and *APOE*-ε4 status, with relative risk 1.17 for improved vs. stable status [p=0.01, 95% CI=1.03-1.33]. CSF transferrin and H-ferritin were significantly higher in participants aged≥50 but were unrelated to *APOE*-ε4 status. Both biomarkers were associated with GDS at 30 and 42 mos., adjusting for comorbidity [p<0.05 for H-ferritin and transferrin at both time-points (1-3% of GDS variance over time explained)]. With further adjustment for ZDV use and *APOE*-ε4 status, associations remained significant for H-ferritin at 30, 36, and 42 mos., but only in persons aged<50 (4.7% of GDS variance explained). In 10 *APOE*-ε4 carriers aged≥50, higher CSF transferrin was associated with better GDS at 30 mos. (7.2% GDS variance explained). CSF iron also showed a nominal association with GDS at 30 mos. in older *APOE*-ε4 carriers (p=0.05). CSF iron biomarkers are independently associated with GDS differences over time in HIV+ adults. Further studies are needed to understand potential interactions between age, *APOE*-ε4 genotype, comorbidity, and CSF iron transport in promoting HIV-associated NC decline.

Patrick Raber
Adaptive Biotechnologies

Adaptive Biotechnologies is at the forefront of immune-based discoveries, combining high-throughput sequencing and expert bioinformatics to profile T-cell and B-cell receptors. We bring the accuracy and sensitivity of our immunosequencing platform into laboratories around the world to drive groundbreaking research in cancer and other immune-mediated diseases. Adaptive also translates immunosequencing discoveries into clinical diagnostics and therapeutic development to improve patient care. Adaptive's immunoSEQ Assays give researchers a high-definition view of the adaptive immune system. It’s now possible to explore broad characteristics of T-and B-cell repertoires as well as the unique details of individual clones. Whether you are investigating the general trends of the immune system or specific clones, you can be confident in your data with the quantitative and reproducible results of immunoSEQ Assays behind your research.

Nicole Restrepo
Epidemiology and Biostatistics
Case Western Reserve University

Exploration of a potential FOXA1 suppressing program in Luminal A breast cancer patients from METABRIC

Nicole A. Restrepo[1], Mariusz Butkiewicz[1], Ruth A. Keri[2]
[1]Department of Epidemiology & Biostatistics, Case Western Reserve University
[2]Department of Pharmacology, Case Western Reserve University

Breast cancer is the most common malignancy in women in the United States regardless of race or ethnicity. A clinically and molecularly heterogeneous group of cancers, breast cancer subtypes (i.e., luminal A, luminal B, caludin-low, and basal) share a core gene expression signature with cellular subpopulations of the mammary gland. Currently, master regulators have been identified for each subtype which has provided partial insight into molecular pathways involved in subtype-specific tumorigenesis. One example is *FOXA1* which modulates transcriptional activity of nuclear hormone receptors and is upregulated in luminal A tumors. Expression of ERα is a strong driver of proliferation in some HER2 subtypes. While FOXA1's ability to transcriptionally activate other genes has been well studied, less known are the genes and molecular mechanisms that FOXA1 suppresses. Understanding of these suppressed genes may provide more insight into the molecular targets and pathways detrimental to tumorigenesis, proliferation, and metastasis. Here we have identified genes negatively correlated with *FOXA1* expression in luminal A tumor samples taken as part of the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC). Microarray expression data from a sample (n=10) of luminal A tumors was assessed for gene expression signatures of *FOXA1* in comparison to all other gene signatures. In general, gene expression signals were normally distributed. We found 62 genes were strongly, negatively correlated (r^2= -0.80) with *FOXA1* expression and additionally that 166 genes were strongly, positively correlated (r^2= +0.80). As expected, *FOXM1* was positively correlated while *SUMO1* and *BCL11a* were negatively correlated. Of the 62 genes negatively correlated with *FOXA1*, 27 genes were known transcriptional targets of *FOXA1* based on ENCODE Transcription Factor Targets dataset. These negatively correlated *FOXA1* target genes were enriched in pathways for insulin/IGF pathway mitogen activated protein kinase/MAP kinase cascade, cadherin signaling, and apoptosis signaling. Future directions will explore whether these genes share a protein binding motif for a yet undetermined component in the *FOXA1* transcriptional regulatory machinery.

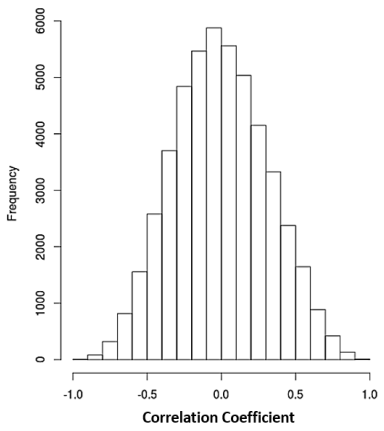


Figure 1: Histogram of the distribution of gene expression in comparison to *FOXA1* in luminal A tumors.

Satya Sahoo
Epidemiology and Biostatistics
Case Western Reserve University

Large biomedical data repositories are enabling researchers to move away from small-sized studies to “big data” with sufficient statistical power, greater time scales, and wider spectrum of attributes. However, the complexity of data and variety of methods used in different research studies requires provenance metadata to play a key role in secondary analysis of biomedical big data. Provenance describes the source or history of data, for example the study protocol, methods used to assign values to research study variables, and procedure used to curate or process data. Our project aims to develop a provenance analysis engine using the World Wide Web Consortium (W3C) PROV specifications (W3C is the Web technology standards organization) that can scale with computational requirements of big data applications. We propose to use the PROV model to represent provenance of research studies and develop a set of provenance analysis operations that can support data reproducibility, data quality view, and derive trust value during cohort identification involving multiple research studies. The provenance analysis engine builds on our earlier work in provenance modeling and efficient query execution using a “sub-graph materialization” strategy. The proposed platform will support: (a) computing provenance trails to retrieve history of research study variables, (b) provenance signature with user defined constraints to retrieve only relevant subset of study data, and (c) merge or comparison operations over provenance graphs. We propose to use the National Sleep Research Resource (NSRR), the Center for Sudden and Unexpected Death in Epilepsy (SUDEP) Research (CSR), and data from a lung cancer registry at UCLA as representative biomedical big data repositories to develop and evaluate the provenance analysis engine. The NSRR consists of 50,000 sleep studies from more than 36,000 subjects representing a wide variety of biomedical data, including clinical, biochemical, and physiological covariate data. The CSR is collecting multi-modal epilepsy data to study SUDEP and consists of 14 epilepsy centers across the U.S. and the U.K. The UCLA registry captures clinical reports, labs, imaging, and health-related questionnaire results on a growing number of patients undergoing lung cancer screening. We propose to disseminate the provenance analysis engine to the wider biomedical big data research community for use as a standalone platform and also for integration.

Lindsay Sausville
Genetics/Epidemiology and Biostatistics
Dartmouth College/Case Western Reserve University

Globally lung cancer results in more deaths worldwide than any other cancer, indicating a need for better treatments. Members of the eicosanoid metabolism pathway represent promising therapeutic targets, as several enzymes involved in the generation of these lipids are dysregulated in many cancers and their inhibition reduces lung cancer growth in mouse models. However, genetic variations of enzymes involved in eicosanoid metabolism have not been adequately examined for association with lung cancer. The goal of this study was to determine whether germline genetic variations altering eicosanoid producing enzyme function and/or expression are associated with differences in lung cancer survival.

We examined the association of genetic variation with mortality within eicosanoid metabolism genes in 395 non-small-cell lung cancer (NSCLC) cases from the Southern Community Cohort Study (SCCS). A total of 108 SNPs, both common and rare, in 19 genes, were examined for association. No common or rare variants were associated with lung cancer survival across the entire study population. However, rare variants in ALOX15B (arachidonate 15-lipoxygenase, type B) and the common variant rs12529 in AKR1C3 (prostaglandin F synthase) were associated with NSCLC mortality in women and African Americans, respectively. Rare variants in ALOX15B were associated with greater mortality in women (HR = 2.10, 95% CI = 1.25 - 3.54, p-value = 0.005). The major allele of rs12529 in AKCR1C3 associated with improved survival in African Americans (HR = 0.74, 95% CI = 0.59 - 0.92, p-value = 0.008). The lack of genetic associations among all NSCLC cases and the association among women only for rare variants in ALOX15B may, in part, explain the better NSCLC survival observed among women. These results raise the possibility that some subgroups within the NSCLC population may benefit from drugs targeting eicosanoid metabolism.

Mike Sivley
Biomedical Informatics
Vanderbilt University

The abundance of whole-genome and whole-exome sequencing data, coupled with the rapid increase in experimentally derived protein structures in the Protein Data Bank (PDB), provides a unique opportunity to study missense variation in its structural and functional context. Recent studies in cancer have identified spatial clusters of recurrent somatic mutations. These clusters are presumed to highlight functionally important regions that, when disrupted, facilitate the progression of cancer. Using spatial analytics from ecology and epidemiology, we quantify and compare the non-random spatial patterns of population-derived, pathogenic, and somatic single-nucleotide variants (SNVs) — across distance scales — to identify patterns predictive of pathogenicity.

Sandra Smieszek
Epidemiology and Biostatistics
Case Western Reserve University

Genetically predicted gene expression and HIV-associated neurocognitive disorder

Sandra Smieszek[1], Todd Hulgan[2], David C. Samuel[2], Donald R. Franklin[3], Robert K. Heaton[3], Scott L. Letendre[3], Ron J. Ellis[3], Asha R. Kallianpur[4], and William S. Bush[1,4]
1)Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, USA. 2)Vanderbilt University, Nashville, Tennessee. 3)University of California-San Diego, California. 4)Cleveland Clinic Foundation/Lerner Research Institute and Cleveland Clinic Lerner College of Medicine, Ohio.

HIV-Associated Neurocognitive Disorder (HAND) is a term established to capture a spectrum of neurocognitive (NC) deficits associated with HIV infection. The genetic underpinnings of HAND are poorly understood. CHARTER is a U.S.-based observational study of neuro-HIV outcomes in ambulatory, HIV+ adults who underwent standardized, comprehensive NC assessment (2003-7) and were assigned a a Global Deficit Score (GDS) [*normal* (GDS<0.5) or *impaired* (GDS≥0.5)]. In this study, we investigated the impact of genetic variants known to alter the expression of genes in whole blood on GDS outcomes. We used CHARTER genome-wide association study (GWAS) data (imputed to the 1000 Genomes reference) to predict gene expression using an approach called PrediXcan. It estimates the genetically regulated component of gene expression using reference panels from studies of expression quantitative trait loci (eQTL). In this study, Gene/Tissue Expression data and CHARTER genome-wide genotype data were used to model the expression profile of 11,000 genes. We then evaluated associations of these "imputed" gene expression traits with CHARTER phenotypes to identify genes of interest associated to NC impairment defined. We performed regression analyses to identify predicted gene expression values that associate to CHARTER phenotypes, adjusting for population effects and known clinical covariates (comorbidity status, current CART use, plasma viral load, nadir CD4+ T-cell count). Among the top genes were Ankyrin Repeat Domain 44 (ANKRD44), insulin receptor substrate 2 (IRS2), and Activating Transcription Factor 3 (ATF3). Using a set of 222 genes associated to CHARTER phenotypes at $p < 0.01$, we performed gene pathway enrichment analysis. Initial analysis suggests overrepresentation of iron ion binding, immune defense response, regulation of inflammatory process and mitochondrial membrane pathways. Additionally, among the most significant associations ($p < 0.01$), we found 17 with known HIV protein interactions. We therefore hypothesize that individuals with altered regulation of these HIV-interacting genes may be predisposed to HAND in the presence of HIV infection. This study provides further support for a role of iron transport in HAND pathogenesis and evaluates gene expression effects by modeling the mechanisms through which genetic variants influence NC impairment.

Andrea Waksmunski
Genetics and Genome Sciences
Case Western Reserve University

Elucidating rare variation for age-related macular degeneration using the Ohio Amish

Andrea R. Waksmunski[1], Rob Igo[2], Heather A. Voss-Hoynes[2], Jessica Cooke Bailey[2], and Jonathan Haines[2,3]
[1]*Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH*
[2]*Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH*
[3]*Institute for Computational Biology, Case Western Reserve University, Cleveland, OH*

Age-related macular degeneration (AMD) is the third leading cause of vision loss worldwide and affects about 10 million people in the United States. Although previous studies have identified 34 susceptibility loci contributing to AMD, it is suspected that about one-third of total AMD heritability is still unknown. The Amish offer an opportunity to study rare variants associated with AMD because they constitute a relatively isolated population that may be enriched in rare variation and is genetically and environmentally homogeneous. In this study, we analyzed Illumina exome chip data from general and Amish populations to identify and assess rare variants for AMD with an aim to better understand its missing heritability. The exome genotyping data were clustered and called with GenomeStudio developed by Illumina. Because GenomeStudio often misclusters SNPs from the exome chip and is more tailored to identify common variants, additional strategies were implemented to correct misclustered rare SNPs, including reclustering and several quality control filters in GenomeStudio and the implementation of zCall, a separate rare variant caller. In this study, we used 242,901 markers targeting over 200,000 nonsynonymous SNPs for 2,472 samples to identify rare variants. Of these samples, 180 were obtained from Amish individuals in Ohio, 87 of whom have confirmed AMD. Following extensive quality control and manual review, 146,571 SNPs across 2,466 samples were retained. Analysis of these variants is in progress and will seek to identify rare variants that arise specifically in Amish individuals with AMD.