Aka, Ida
Feasibility Study of Citalopram and Escitalopram Response in Children and Adolescents Using Electronic Health Records

Akwo, Elivs
Resistant hypertension potentiates the risk of End-Stage Renal Disease in African Americans in the Million Veteran Program

Ayala, Jennifer
The All of Us Research Program: Data and Research Center

Breeyear, Joseph
Association of Polygenic Risk Scores for Body Mass Index and Systolic Blood Pressure in a Pediatric Cohort Requiring Surgery for Congenital Heart Defects

Byrne, Daniel
Using Randomization to Assess Impact of Personalized Medicine on Health Outcomes

Chakravarthy, Rohini
Determinants of Stage at Diagnosis of HPV-related Cancer including Area Deprivation and Clinical Factors

Fox, Zachary
The Personalization of Evidence: Using Intelligent Datasets to Inform the Process

Glazer, Andrew
High-throughput patch clamping identifies dozens of novel Brugada Syndrome-associated SCN5A variants

Goleva, Slavina
Determining the clinical epidemiology and genetic architecture of psychogenic nonepileptic seizures

Huang, Hui
Mechanisms of KCNQ1 Channel Dysfunction in Long QT Syndrome Involving Voltage Sensor Domain Mutations

Jain, Neha
A prospective pragmatic trial to identify barriers to automating clinical trial Matching

Keaton, Jacob
Using electronic health records to investigate repurposing potential of drugs for hypertension implicated by genetic associations

Kroncke, Brett
Moving beyond a dichotomous view of variant pathogenicity: from common to ultra-rare

Lenert, Matthew
Are Those Patients Like My Patients? Using Python to Visualize Similarity Between Research Cohorts and US Counties

Liggin, Mariel
Breast cancer risk management implications for families of ATM and CHEK2 mutation carriers

Martucci, Victoria
A Clinical Algorithm to Identify Cases of Chronic Obstructive Pulmonary Disease
Mautz, Brian
A Phenome-Wide Association and Network Studies Approach to Identifying Novel Disease Associations Using Large-scale Electronic Health Records

Miller-Fleming, Tyne
Evaluating the Genotype-Phenotype Relationships and Imputed Gene Expression Profiles of Congenital Malformation Syndromes in BioVU

Moore, Annah
APOE ε4-specific Associations of VEGF Gene Family Expression with Cognitive Aging and Alzheimer’s Disease

Morley, Theodore
Leveraging phenotype data to identify potential carriers of deleterious genetic variants

Muhammad, Ayesha
A high-throughput cell-surface trafficking assay can identify functionally deleterious KCNE1 variants

Mukherjee, Souhrid
Helping solve medical mysteries: using machine learning and systems biology to identify digenic disease

Nakkam, Nontaya
HLA-A*32:01: Identification of a strong association with vancomycin drug reaction with eosinophilia and systemic symptoms and development of a single allele PCR assay to facilitate clinical translation

Neely, Katelyn
CYP2D6 polymorphisms and adverse events to risperidone in children and adolescents

Niarchou, Maria
Can you clap to the beat? Findings from the first large-scale genome-wide association study of a musical rhythm trait in 606,825 individuals

Piekos, Jacqueline
Uterine Leiomyomata Polygenic Risk Scores (PRS) Confer Novel Relationships in the Clinical Phenome

Sealock, Julia
LabWAS: A method to Investigate relationships between biomarkers and genetic risk scores

Sheehan, Jonathan
Personalized Structural Biology Program at Vanderbilt University

Stone, Cosby
A Personalized Care Model for Removing Unnecessary Penicillin Allergy Labels

Tezak, Ann
Breast and ovarian cancer risk management practices among female BRCA1/2 carriers: Facilitators and barriers to recommended follow-up care

White, Katie
Drug-Induced Immune-Mediated Skin Hypersensitivity Defined by Single Cell Approaches

Yang, Jing
Mela-Immu portal: an open platform for exploring immunotherapy-related multidimensional genomic profiles in Melanoma

Yeh, Vivian
Can precision medicine actually help people like me? African American and Hispanic perspectives on the benefits and barriers of precision medicine
Authors: Ida Aka and Sara L. Van Driest

Title: Feasibility Study of Citalopram and Escitalopram Response in Children and Adolescents Using Electronic Health Records

Background: There is wide variability in clinical response to citalopram and escitalopram (collectively (es)citalopram), including efficacy and adverse events (AEs). Given their common use in children to treat conditions such as depression and anxiety, electronic health records (EHR) datasets may provide the opportunity to define risk factors for non-response and AEs, if adequate data are available to discern these outcomes from clinical documentation.

Aims: To assess the feasibility of evaluating (es)citalopram response and AEs in children and adolescents using EHR data.

Methods: This pilot retrospective cohort study was performed using BioVU, a de-identified DNA biobank linked EHR data. Cohort inclusion criteria are availability of DNA specimen in BioVU; ≥3 days of (es)citalopram exposure; age ≤18 years; and sufficient data in the EHR to determine efficacy and/or AEs. AEs are defined as any untoward event or lab abnormality attributed to (es)citalopram. Responders are those with documented favorable outcome. 100 randomly selected charts of individuals with current age ≤25 and ≥1 mention of (es)citalopram were reviewed from the first mention to end of therapy, loss to follow up, or the current date. For those meeting inclusion criteria, clinical and demographic data, (es)citalopram efficacy (response, non-response, or unknown), and AE status (AE or no AE) were manually extracted from EHR data.

Results: In all, 2,046 potential exposures were identified. 40 of 100 met inclusion criteria, leading to a projected final cohort of approximately 800 children/adolescents. The median age was 13.75 (IQR 9.6-16.5) years and 21 (53%) were female; Drug efficacy was unknown in 1 (2%), 23 (58%) were responders, and 16 (40%) were non-responders. AEs were documented in 18 (45%). The most common AEs were aggression, agitation, and activation/hyperactivity. With our estimated sample size of 800, we will have 80% power to detect risk factors with effect size over 0.36 with $\alpha = 0.05$ for both outcomes.

Conclusions: This pilot study demonstrates feasibility of using EHR data to evaluate (es)citalopram response and AEs in children and adolescents. Future study will identify and phenotype an expanded cohort, followed by analyses to determine the clinical and genetic predictors of therapeutic response and AE with (es)citalopram therapy in children and adolescents. This approach may be generalized to other drugs commonly used in pediatric patients.
Resistant hypertension potentiates the risk of End-Stage Renal Disease in African Americans in the Million Veteran Program

Elvis A. Akwo², Cassianne Robinson-Cohen², Cecilia Chung², Peter W. Wilson⁵, Christopher J. O’Donnell⁴, Todd L. Edwards², Csaba P. Kovesdy³, Adriana M. Hung¹

¹VA TVHS & Vanderbilt University, ²Vanderbilt University Medical Center, ³University of Tennessee Health Science Center, ⁴Boston Veterans Administration, ⁵Emory University

Background: African Americans are four times as likely as Whites to develop end-stage renal disease (ESRD). Resistant hypertension (RH), a severe form of hypertension (HTN) is associated with increased risk of cardiovascular (CV) and renal outcomes. We investigated how ESRD risk is modified by race.

Methods: We performed a time-to-event analysis of a retrospective cohort of 240,038 hypertensive veterans enrolled in the Million Veteran Program (MVP) with a GFR >30 ml/min. The primary exposure was incident RH (time varying). The primary outcome was incident ESRD during the 13.5 year follow up (2004-2017). Secondary outcomes were myocardial infarction (MI), stroke, and death. Incident RH was defined by American Heart Association Guidelines as failure to achieve outpatient blood pressure (BP) <140/90 mmHg with 3 anti-HTN drugs, including a thiazide, or use of 4 or more drugs, excluding BPs when pain score was >5, when interfering medications were given or when there were secondary causes of HTN. Poisson regression was used to estimate incidence rates (IR) and biologic interaction with race was tested on the additive scale. Cox models with competing-risks regression as sensitivity analyses were used to identify independent effects.

Results: Median age was 60 years; 20% were African American; 5% were Hispanic and 6% were women with 23,385 Incident RH cases (9.7%). RH patients had higher IR (per 1000 PY) of ESRD (4.5 vs. 1.3), MI (6.5 vs 3.0), stroke (16.4 vs 7.6) and death (12.0 vs 6.9) than non-resistant HTN (NRH); number needed to harm = 30. In Cox models adjusted for age, sex, race, eGFR, smoking, serum lipids, body mass index, diabetes, coronary artery disease, peripheral artery disease, chronic obstructive pulmonary disease and cancer; patients with RH had a 2.0-, 1.67-, 1.9- and 1.14-fold higher risk of ESRD, MI, stroke, and death, respectively. Standardized IR differences (per 1000PY) of ESRD (for RH vs. NRH) were greater in African Americans (4.7) compared to Whites (2.5). In Poisson models, African Americans with RH had a 2.5-fold higher risk of ESRD compared to African Americans with NRH; 3-fold the risk of Whites with RH, and 9-fold the risk of Whites with NRH [p for additive interaction < 0.01; relative excess risk due to interaction: 2.5 (95%CI: 1.6, 3.5)].

Conclusion: RH was associated with a higher risk of ESRD (and CV outcomes), especially in African Americans. Interventions (behavioral, drug choices and others) that improve reaching BP targets in RH patients, could have a major impact on ESRD incidence in this high-risk population, particularly in African Americans.
The All of Us Research Program: Data and Research Center

Jennifer Ayala, PhD, Kelsey Mayo, PhD, Melissa Basford, MBA, Robert Cronin, MD, MS, Robert Carroll, PhD, Andrea Ramirez, MD, MS, Consuelo Wilkins, MD, MSCI, Joshua Denny, MD, MS

Vanderbilt University Medical Center

The All of Us Research Program is a national effort to partner with one million or more participants across the United States to create a biomedical data resource aimed at accelerating health research and medical breakthroughs. All of Us Participants may consent to provide health information, including data from health surveys, electronic health records, physical measurements, and biospecimens. The Data and Research Center (DRC) led by Vanderbilt University Medical Center sits at the confluence of these varied data, providing human and technical systems to securely intake and organize All of Us participant data. The DRC is also creating a secure, cloud-based platform capable of delivering a large, rich All of Us biomedical dataset to a broad scientific community and catalyzing a robust ecosystem to shape and support the future of biomedical research. Central to the DRC’s mission is a strong commitment to participant engagement and community involvement to ensure that feedback and preferences of individuals from many backgrounds are reflected in the design and implementation of the All of Us Research Program.
Association of Polygenic Risk Scores for Body Mass Index and Systolic Blood Pressure in a Pediatric Cohort Requiring Surgery for Congenital Heart Defects

Joseph H. Breeyear¹, Jacob M. Keaton¹,2,3, Eric S. Torstenson¹,2, Andrew H. Smith⁴,⁵, Sara L. Van Driest⁴, Jeffery G. Weiner, Prince J. Kannankeril⁵, Todd L. Edwards¹,2,3

¹Vanderbilt Genetics Institute, Vanderbilt University, Nashville, Tennessee
²Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee
³Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, Tennessee
⁴Departments of Pediatrics and Medicine, Vanderbilt University Medical Center, Nashville, Tennessee
⁵Thomas P. Graham Jr. Division of Cardiology, Department of Pediatrics, Monroe Carell Jr. Children’s Hospital at Vanderbilt, Nashville, Tennessee

Aims
Congenital malformations are the leading cause of infant mortality in the United States. Congenital heart disease (CHD) represents the most prominent congenital malformation, affecting 40,000 US births per year. Approximately half of these children require surgical intervention(s), many of which experience postoperative complications. To date, no studies have evaluated associations of polygenic risk scores (PRS) for anthropometric traits with postoperative outcomes in a pediatric CHD cohort. Further, the utility of PRS to predict outcomes such as hypertension and obesity in pediatric subjects is largely unknown.

Methods
We used high-quality imputed genotypes from pediatric participants requiring surgery for CHD (mean age at surgery ± SD = 3.38 ± 5.39 years, n = 1978 subjects). Base data for a body mass index (BMI) PRS used the Genetic Investigation of ANthropometric Traits (GIANT) consortium GWAS transethnic 2015 BMI data (n_max = 322,154 subjects). Base data for a systolic blood pressure (SBP) PRS used recently published GWAS SBP data (n_max = 760,226 subjects). Target data were pruned for linkage disequilibrium at an r² threshold of 0.1 at a maximum distance of 250 kilobases and scores were calculated in PLINK.

Results
Associations of PRS for BMI and SBP with BMI and length of hospital stay following surgery (LOS) were modeled using linear regression in R and adjusted for age, sex, and 10 PCs. The SBP model was also adjusted for BMI. BMI PRS was associated with BMI (PRS p-value threshold = 0.001, n SNPs = 1,603, β ± SE = 0.58 ± 0.15 kg/m², p = 8.5x10⁻⁵) and LOS (PRS p-value threshold = 0.05, n SNPs = 16,244, β ± SE = 1.67 ± 0.83 days, p = 0.045). The SBP PRS was associated with LOS (PRS p-value threshold = 0.0001, n SNPs = 2,985, β ± SE = -0.45 ± 0.15 days, p = 3.02 x10⁻³).

Conclusions
These results demonstrate the ability of PRS developed in adults to predict pediatric traits and outcomes. Specifically, the association of the BMI PRS with BMI provides proof of principle for the use of adult-derived PRSs in younger populations. PRS associations in pediatric participants may identify individuals at higher risk of negative cardiometabolic outcomes.
Using Randomization to Assess Impact of Personalized Medicine on Health Outcomes

Daniel Byrne, Henry Domenico.

Aims.
To explore how randomization can be used to assess impact of personalized medicine on health outcomes.

Methods.
At Vanderbilt, we have conducted a number of pragmatic trials in our Learning Health Care System that can serve as a blueprint for studies of personalized medicine. Such trials would randomize patients to usual care, or usual care plus personalized medicine. An important aspect of personalized medicine is risk stratification from real-time predictive modeling. With our Cornelius predictive modeling project we have several years of experience in building predictive models, displaying them for a random half of the patients, and incorporating them into Epic.

Results.
Big data and observational studies have value but will not be sufficient to assess whether personalized medicine improves outcomes. We have demonstrated that pragmatic trials using our Learning Health Care System model can produce rigorous randomized trials that can be used to assess personalized medicine initiatives.

Rigorous study designs will also enable comparison of different approaches to personalized medicine. Precision medicine needs to be expanded beyond genomics to environmental and behavioral factors. Many commonly available variables also provide valuable information for precision medicine. For example, body mass index predicts Type II diabetes more accurately than all of the genetic data combined. In addition, commonly available laboratory tests provide valuable low-cost information for precision medicine.

Conclusions.
The goal of personalized medicine is to improve health outcomes. Randomized controlled trials can and should be used to assess this impact. Vanderbilt has strengths in personalized medicine, clinical research, and pragmatic trials, which makes it the ideal place to conduct this work. Funding, however, must be allocated to support rigorous evaluations.

Vanderbilt University Medical Center is a national leader in personalized medicine research and its delivery at the bedside. To remain in this leadership position, we must answer the big question: "Does personalized medicine improve outcomes?".
Determinants of Stage at Diagnosis of HPV-related Cancer including Area Deprivation and Clinical Factors

Rohini Chakravarthy, BS1; Sarah Stallings, PhD2; Digna Velez-Edwards, PhD3,4; Sifang Kathy Zhao, MPH5; Douglas Conway, MS6; Aliya Gifford, PhD4; Sunil Rao, PhD7; Melinda Aldrich, PhD, MPH8; Erin Kobetz, PhD9; Consuelo H. Wilkins, MD, MSCI 2,3,10

1Vanderbilt University School of Medicine, Nashville, TN
2Meharry-Vanderbilt Alliance, Nashville, TN
3Institute for Medicine and Public Health, Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN
4Department of Biomedical Informatics, Vanderbilt University, Nashville, TN
5Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN
6Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN
7Department of Public Health Sciences, University of Miami School of Medicine, Miami, FL
8Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN
9Department of Medicine, University of Miami, Coral Gables, FL
10Department of Medicine, Division of Geriatrics, Vanderbilt University Medical Center, Nashville, TN

Aims: There are challenges to incorporating social determinants of health data into health information systems and clinical decision making on a large scale. Collecting this information can be time consuming and information systems are being modified to better accommodate collection of this type of information. The American Community Survey provides an opportunity to map individuals to sociodemographic characteristics on a large scale. The objective of this study is to identify individual and aggregate census tract variables that are associated with stage of human papillomavirus (HPV) related cancer at diagnosis.

Methods: Our study population consisted of men and women with HPV-related cancers from an urban academic medical center. We tested for the association between individual and aggregate census variables and stage of cancer presentation for HPV-related cancers using ordered logistic regression, adjusting for age, race, and sex.

Results: In a sample size of 3,247 cases of HPV-related cancers, The average age at diagnosis was 57. The average stage at diagnosis was SEER Stage 3. 43% of patients were female and 87% were white. There was no association between individual or aggregate census variables (area deprivation index) and cancer stage.

Conclusions: Area deprivation index is not associated with cancer stage at presentation. These results may reflect a lack of socio-demographic diversity in our population and they expose the lack of consensus about how to measure the effect of deprivation at the population level.
Title: The Personalization of Evidence: Using Intelligent Datasets to Inform the Process

Authors: Zachary Fox¹, Dario Giuse², Taneya Koonce¹, Marcia Epelbaum¹, Annette Williams¹, Mallory Blasingame¹, Sheila Kusnoor¹, Jing Su¹, Elizabeth Frakes¹, Patricia Biggerstaff³, Travis Osterman², Nunzia Bettinsoli Giuse¹,²

¹Center for Knowledge Management, Vanderbilt University Medical Center, Nashville, TN; ²Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; ³HealthIT, Vanderbilt University Medical Center.

Background: At the time when healthcare is focusing its effort towards exploring artificial intelligence, particularly how we best help with the diagnosis and care for patients, we reflect on the fact that one of the main uses of technology in healthcare is the ability to equip providers with data to inform their decision-making. The need for data cannot be overestimated, especially given the preponderance and rapid growth of the evidence base. Our understanding of what constitutes this evidence base has grown beyond traditionally published studies in the biomedical literature to also include social determinants of health factors that impact an individual’s overall health and well-being.

Over the past two decades, the Center for Knowledge Management (CKM) has provided clinicians with filtered synthesized evidence summaries to answer patient-specific complex clinical questions. In 2005 CKM began leveraging the message basket feature in StarPanel to scale this provision of evidence for medical center care teams. CKM’s experience is not limited to the delivery of evidence through the medical record; the team has also established a nationally recognized track record for tailoring complex health information for patients, accounting for both health literacy and learning styles, and for evaluating social determinants of health questionnaires for incorporation into the healthcare encounter.

Objectives: The primary objective of this project is to leverage existing knowledge about patients through newly developed tools and processes to personalize the provision of evidence, for clinicians and patients alike.

Description: With the institution’s transition to eStar comes an opportunity to act upon the collective wisdom we’ve gained, reimagining our approach to answering complex clinical questions. To effectively capitalize on this opportunity, CKM is taking a multifaceted approach to redefining the evidence provision process. The first step is partnering with HealthIT to establish a new message basket within Epic to serve as a seamless healthcare provider/information professional communication portal. Questions received through the portal will be triaged internally, as they have been in the past, to the CKM team member with expertise in the area of inquiry. Once questions have been received, CKM information professionals will utilize the electronic health record’s novel word frequency cloud feature, built into StarPanel by Dario Giuse, to gain insight into the patient. The Word Cloud is built automatically in real time by applying NLP techniques to extract medical concepts from all textual documents and mapping them to the Unified Medical Language System (UMLS). A high-level visual presentation gives a time-oriented view of documented concepts and their relative frequency. Incorporation of the Word Cloud will not only allow for greater personalization of evidence summaries based on a patient’s comorbidities and health history, but, the capture and highlighting of relevant social and behavioral determinants of health will enable information professionals to further tailor the information they share. The information can then be personalized based on the patient’s education and health literacy level for possible sharing directly with the patient. Additionally, the information delivered to the clinician will include suggestions for beneficial community and public health resources that can be shared with their patient.
High-throughput patch clamping identifies dozens of novel Brugada Syndrome-associated SCN5A variants
Andrew M. Glazer,1 Yuko Wada,1 Lynn Hall,1 and Dan M. Roden1,2,3
1Department of Medicine, Division of Clinical Pharmacology
2Department of Pharmacology
3Department of Biomedical Informatics

Introduction: The accurate classification of variants in Mendelian disease genes is a major challenge for the successful implementation of personalized medicine. Variants in SCN5A are a major cause of inherited arrhythmia syndromes, including partial or complete loss of function SCN5A variants which can cause Brugada Syndrome. The pathogenicity of SCN5A variants is often unknown or disputed; indeed, most observed SCN5A variants are classified as a variant of uncertain significance (VUS). The widely adopted American College of Medical Genetics and Genomics (ACMG) classification scheme includes in vitro functional data as a strong criterion for variant classification.

Aim: To determine the in vitro function of dozens of variants in SCN5A to improve variant classification.

Methods: After a literature review of 1712 reported SCN5A variants, we selected 57 variants that had not been functionally characterized (10 suspected benign and 47 suspected Brugada-associated). We also selected 12 variants that had been previously studied by patch clamping that had a range of functional defects. For each variant, we used high-throughput automated patch clamping to measure 5 important functional parameters: peak current, late current, activation and inactivation kinetics, and recovery from inactivation. Patch clamp data were integrated into the ACMG classification scheme to reclassify each variant.

Results: All 12 previously studied variants had functional properties similar to previously published manual patch clamp results. 9 of the 10 suspected benign variants had wildtype-like channel function. 34 of the 47 suspected Brugada-associated variants (72%) had partial or total loss of channel function. In total 47 of 57 previously unstudied variants were reclassified with the patch clamp data, including the reclassification of 27 variants from VUS to likely pathogenic and 9 variants from VUS to benign.

Conclusions: High-throughput in vitro functional characterization can help reclassify SCN5A variants, including identifying new disease-associated variants. This method can identify dysfunctional variants in SCN5A at scale and help “de-VUS” this important gene.
Determining the clinical epidemiology and genetic architecture of psychogenic nonepileptic seizures
Slavina B Goleva, Kevin F Haas, Lea K Davis

Introduction:
Psychogenic non-epileptic seizures (PNES) are a conversion disorder characterized by convulsions. PNES present like epileptic seizures, and 80% of PNES patients are initially misdiagnosed as such, but PNES do not arise due to aberrant brain electrical signaling. Rather, PNES are thought to arise from psychological or stress-related origins. PNES patients have disproportionately high rates of psychiatric disorders and obesity, and around 75% of PNES patients are females. PNES are understudied, and there is no International Classification of Diseases (ICD) code for PNES, which makes studying PNES using electronic health records (EHR) difficult. Furthermore, nothing is known about the genetic contribution to PNES.

Aims:
The overarching aim of this project was to characterize the clinical and genetic characteristics of PNES in a hospital population.

Methods:
We identified PNES patients in the Vanderbilt University EHR (VU-EHR) by developing an algorithm using convulsion ICD codes and natural language processing of keywords in patient charts. We performed chart review on 50 of the identified cases along with a clinical neurologist.

We characterized the pattern of comorbidities associated with PNES across the lifespan. First, we identified VU-EHR cases of 13 common psychiatric disorders using ICD codes, and tested them for comorbidity with PNES. We used the Phi correlation coefficient to determine comorbidity within males and females. We then compared the likelihood of association in females vs. males using odds ratios. We then determined which other phenotypes in the VU-EHR are associated with having a PNES diagnosis by conducting a Phenome Wide Association Study (PheWAS) of PNES case/control status.

We next determined which disorders shared underlying genetic architecture with PNES. We calculated polygenic risk score (PRS) for PNES cases (n = 205) and controls (n = 48492) using GWAS summary statistics for the commonly comorbid psychiatric phenotypes, epilepsy, and type 2 diabetes. To determine which of these variables contributed significantly to the variance in PNES case/control status, we ran a multivariable logistic regression using the PRS’s calculated for PNES cases and controls.

Results:
Using our algorithm, we identified 4,267 PNES cases in the VU-EHR out of 2,346,808 people (prevalence 0.18%), 205 of which were genotyped.

We identified strong associations with PNES in a host of psychiatric disorders confirming previous reports. We also found that several of these psychiatric disorders were more likely to be comorbid with PNES in females than in males (e.g. PTSD, OR 2.72, 95% CI 2.29 - 3.24). Our PheWAS identified a novel association with cerebrovascular disease and PNES (OR 1.08, 95% CI 1.06 – 1.1, p = 6.8 E -19).

Our multivariable logistic regression results showed that PRS for type 2 diabetes (p = 0.0033), suicide attempts (p = 0.0021), focal epilepsy (p = 0.036), and insomnia (p = 0.037) were significantly correlated with PNES case/control status.
Conclusions:
This project has increased epidemiological and genetic knowledge of PNES. Knowledge of disorders associated epidemiologically with PNES, and knowing which disorders share genetics with PNES, can help clinicians identify and treat PNES patients better, be aware of what comorbidities exist, and when they arise.
Mechanisms of KCNQ1 Channel Dysfunction in Long QT Syndrome Involving Voltage Sensor Domain Mutations

Hui Huang\textsuperscript{1,2}, Georg Kuenze\textsuperscript{2,3}, Jarrod A. Smith\textsuperscript{1,2}, Keenan C. Taylor\textsuperscript{1,2}, Amanda M. Duran\textsuperscript{2,3}, Arina Hadziselimovic\textsuperscript{1,2}, Jens Meiler\textsuperscript{2,3,4}, Carlos G. Vanoye\textsuperscript{5}, Alfred L. George Jr.\textsuperscript{5}, and Charles R. Sanders\textsuperscript{1,2,6}

\textsuperscript{1}Department of Biochemistry, Vanderbilt University, Nashville, TN, USA
\textsuperscript{2}Center for Structural Biology, Vanderbilt University, Nashville, TN, USA
\textsuperscript{3}Departments of Chemistry and Pharmacology, Vanderbilt University, Nashville, TN, USA
\textsuperscript{4}Department of Bioinformatics, Vanderbilt University Medical Center, Nashville, TN, USA
\textsuperscript{5}Department of Pharmacology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
\textsuperscript{6}Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

\textbf{Aims} Mutations that induce loss of function (LOF) or dysfunction of the human KCNQ1 channel are responsible for susceptibility to a life-threatening heart rhythm disorder, the congenital long QT syndrome (LQTS). Hundreds of KCNQ1 mutations have been identified, but the molecular mechanisms responsible for impaired function are poorly understood. Our study is aimed to investigate the impact of 51 KCNQ1 variants located within the voltage sensor domain (VSD), with an emphasis on elucidating effects on cell surface expression, protein folding and structure, and channel function.

\textbf{Methods} Nuclear magnetic resonance, flow cytometry, automated planar patch clamp, MD simulation.

\textbf{Results} Our results provided the basis for classifying each mutation into one of six mechanistic categories, highlighting heterogeneity in the mechanisms resulting in channel dysfunction or LOF from mutant to mutant. More than half of the KCNQ1 LOF mutations examined were seen to destabilize the structure of the VSD, generally accompanied by mistrafficking and degradation by the proteasome, an observation that underscores the growing appreciation that mutation-induced destabilization of membrane proteins may be a common human disease mechanism. We observed that five of the folding defective LQTS mutants are located in the VSD S0 helix, where they interact with a number of other LOF mutation sites in other segments of the VSD. These observations reveal a central role for the S0 helix as a central scaffold to help organize and stabilize the KCNQ1 VSD and, most likely, the corresponding domain of many other ion channels.

\textbf{Conclusions} LQTS associated mutations in KCNQ1 most often destabilize the protein, leading to mistrafficking and degradation.

\textit{This work was supported by NIH Grant RO1 HL122010.}
A prospective pragmatic trial to identify barriers to automating clinical trial matching

Neha Jain, Alison Culley, Christine Micheel, Travis Osterman, Mia Levy

**Aims:** Clinical trial enrollment is a laborious process and requires providers, trial investigators, and other clinical staff to actively initiate and maintain enrollment efforts. Setting up automated process triggers to perform reflex clinical trial matching can kick-start the process without requiring human intervention. As of now, there are no private or public databases that contain reliable information about cohort-level recruiting statuses. We conducted a pragmatic study to understand the obstacles to clinical trial enrollment at VICC and steps in the enrollment process that can be automated for a faster, more efficient, and more streamlined end-to-end clinical trial matching workflow.

**Methods:** Using a clinical trial matching service using software developed in collaboration with GenomOncology, we used the receipt of sequencing test results as a process trigger to perform reflex clinical trial matching on all patients with solid cancers. Providers were randomized to receive clinical trial recommendations or no recommendations. An automated data query from our clinical trial management system allowed daily updates to the recruiting statuses for all open trials enhancing accuracy of trial results. Additional refinements of these results were applied using multi-faceted filtering by treatment setting and trial phase. At this point in the process, a research nurse performed an initial manual prescreening to evaluate eligibility. The final set of “matched” trials for individual patients was sent out to providers in the intervention cohort via an EMR message with suggested next steps. Provider responses, response times, and prescreening outcomes were recorded in a REDCap database.

**Results:** Prescreening was performed on 99 patients. 88% of the trials that initially matched to the patient based on diagnosis and biomarkers were eventually found to be false matches, and only 12% (n=87) of trials were found to be possible matches. 72% of the false matches were attributed to inaccurate open/close status of trial cohorts as well as limited slot availability on trials.

Trial recommendations were marked by the intervention providers as: (i) will consider treatment, (ii) will consider treatment in future, and (iii) not clinically appropriate. 46% of the trial suggestions were acknowledged as appropriate current or next line treatments for patients, while 18% were deemed inappropriate due to a change in patient’s disease state or change in the patient’s care provider. No provider response was received on 36% of the recommendations.

**Conclusions:** It is not possible to make substantial improvements to the clinical trial enrollment workflow without building infrastructure and processes to maintain accurate status of the trial arms, cohorts, and slot availability. Such efforts can drastically reduce the burden of manual work needed to prescreen and enroll patients to trials. Uptake of such efforts by the National Cancer Institute (NCI) or the National Library of Medicine (NLM) has the potential to radically change the accuracy of clinical trial matching services and can be beneficial to the entire oncology community.
Using electronic health records to investigate repurposing potential of drugs for hypertension implicated by genetic associations

Jacob M. Keaton1,2,3,4, Jacklyn N. Hellwege2,5, Ayush Giri1,2,3,6, Qiping Feng7, Wei-Qi Wei4, Todd L. Edwards1,2,3

1Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee.

2Vanderbilt Genetics Institute, Vanderbilt University, Nashville, Tennessee.

3Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, Tennessee.

4Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee.

5Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee.

6Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, Tennessee.

7Division of Clinical Pharmacology, Vanderbilt University Medical Center, Nashville, Tennessee.

AIMS: Hypertension is a risk factor for cardiovascular disease and a leading cause of death worldwide. Current antihypertension medications can have undesirable side-effects and lack efficacy in some patients. New and inexpensive treatment options are needed.

METHODS: Previously we reported 288 drugs, not currently prescribed as antihypertensives, targeting genes whose predicted expression had a significant effect on blood pressure (BP) and with a direction of effect complementary to the drug-gene relationship (e.g. higher predicted expression associated with higher BP and targeted by a drug with an antagonistic effect). To evaluate drugs for potential repurposing, we compared (t-test) median systolic BP (SBP), diastolic BP (DBP), and pulse pressure (PP) before and after treatment with one of the drugs in up to 84,913 individuals from Vanderbilt’s electronic health record (EHR).

RESULTS: A total of 26, 12, and 47 drugs (of 191 available) significantly lowered SBP, DBP, and PP, respectively. Dibunate, a cough suppressant, had the most significant lowering effect for SBP ($\beta = -6.9$ mmHG, $P = 2.15 \times 10^{-40}$) in analysis of 558 patients. Neostigmine, a myasthenia gravis treatment, had the most significant lowering effect for DBP ($\beta = -0.57$ mmHG, $P = 2.68 \times 10^{-63}$) in analysis of 84,913 patients. Oxymetazoline, a topical decongestant, had the most significant lowering effect for PP ($\beta = -1.07$ mmHG, $P = 4.26 \times 10^{-122}$) in analysis of 59,529 patients.

CONCLUSIONS: These results demonstrate that GWAS summary statistics combined with publicly available drug databases and EHR data can be used to quickly and effectively identify opportunities for drug repurposing for treatment of complex diseases.
Title: Moving beyond a dichotomous view of variant pathogenicity: from common to ultra-rare

Aims: A major emerging challenge in genomic medicine is how to best predict disease risk from rare or novel variants found in disease-associated genes. The traditional approach classifies variants as deterministically “pathogenic” or “benign” with a third category for variants where not enough data are available (“variant of uncertain significance” or VUS). However, the expanding volume of data generated by electronic medical records and whole exome/genome sequencing offers an opportunity to reinterpret genetic liability of disease risk. We propose a variant interpretation framework where disease risk is estimated as a continuous and probabilistic parameter, penetrance (the fraction of all variant carriers that will present with disease). We demonstrate this methodology using a well-established disease-gene pair, the cardiac sodium channel gene \textit{SCN5A} and the heart arrhythmia Brugada syndrome.

Methods: We developed a pattern mixture algorithm, based on a Bayesian Beta-Binomial model, to generate variant-specific penetrance priors.

Results: The resulting priors correlate with mean Brugada syndrome penetrance posteriors with a cross validated $R^2$ of 0.41. Variant function and structural context prove most predictive of Brugada syndrome penetrance.

Conclusions: We suggest annotating variants as continuous and probabilistic is more precise than, but consistent with, the discrete pathogenicity classification approach currently used. We suggest Bayesian estimates of penetrance can efficiently integrate variant-specific data (e.g. functional, structural, and sequence) to accurately estimate disease risk attributable to individual variants.
Are Those Patients Like My Patients? Using Python to Visualize Similarity Between Research Cohorts and US Counties

Matthew C. Lenert, MS¹, Dara E. Mize, MD MS², Colin G. Walsh, MD MA²
¹Vanderbilt University; ²Vanderbilt University Medical Center, Nashville, TN

Aims: When patients and doctors collaborate to make healthcare decisions, they rely on research results – particularly results of clinical trials – to guide discussions. Clinicians seek to apply their judgment on the latest evidence to determine which therapies, diagnostic tests, and/or preventive services best fit their patients’ needs and values. Research cohorts, such as the All of US cohort, are designed to recruit diverse participants, but the question remains – how well do these results apply to me or to people who live in our area?

Methods: We developed and evaluated software for cohort visualization and comparison using one complete clinical trial dataset (SPRINT) and one published study (ACCORD). The resultant Python package, “cohort-compare”, makes use of the Center for Disease Control’s Community Health Status Indicators (CHSI) Dataset for US county features.

To compare counties to a trial cohort, we mapped the inclusion criteria for the trial to public health variables in CHSI. We then determined which public health variables could be compared to variables measured in the trial. For the SPRINT trial, the county comparison was made using the average rate of chronic heart disease deaths, stroke deaths, obesity, active smoking status, and African-American race, after conditioning county data for the presence of hypertension.

To evaluate the similarity measure, we quantified similarity of counties near trial recruitment sites compared to those counties farther away. We defined “nearness” as county centroids within a certain radius of a trial site. We then expanded the definition of “near” a trial site in one-mile increments starting at a radius of five miles. At each increment, we performed a Wilcoxon Rank Sum test to determine if a significant difference of similarity between “near” sites and “far” sites existed. We repeated this procedure until no significant difference remained, factoring in a Bonferroni-Holm correction.

We also analyzed if the site similarity of the control group compared to the control group’s trial average explained divergence in the effectiveness of the intervention in the intervention group. We used robust regression to the measure the absolute difference of the primary outcome rates between the site intervention group and the trial average intervention group using a linear model of similarity (site controls compared to trial controls), number of participants at the site, and an interaction term between the two variables.

Results: We developed and tested this package in the All of US environment. It is publically available at https://pypi.org/project/cohort-compare/.

Counties up to 495 miles to the closest SPRINT trial site and up to 712 miles to the closest ACCORD trial site had populations that were significantly more similar to the study cohort than counties farther away.
Both similarity and the number of site participants were statistically significant predictors of intervention effect differences between a site and the trial average. The model suggested that greater similarity leads to smaller differences in intervention group outcomes.

**Conclusion:** We detailed a generalizable method for both assessing recruitment gaps in large multicenter trials and for creating maps to provide clinicians intuition on trial applicability in their area.
**Title:** Breast cancer risk management implications for families of ATM and CHEK2 mutation carriers

**Authors:** Mariel Liggin, [1]; Ann Tezak, MA, MPH, [2]; Anne Weidner, MPH, [2]; Tuya Pal, MD, FACMG [2]

[1] Tennessee State University, [2] Vanderbilt University Medical Center

**Introduction/Aims:** Multi-gene panel (MGP) testing analyzes multiple inherited cancer genes. ATM and CHEK2 are moderate penetrance breast cancer (BC) genes included in many MGPs. Mutations in these genes confer a > 20% lifetime BC risk, the threshold for considering high-risk BC screening. Among a registry-based sample of ATM and CHEK2 carriers, we assessed lifetime BC risk among unaffected female relatives of these gene carriers and determine the impact of family history and gene positivity on BC screening recommendations for these at-risk relatives.

**Methods:** Clinical and family history data was collected from 56 CHEK2 and 56 ATM carriers enrolled in the Inherited Cancer Registry. Data was abstracted on living female first-degree relatives (FDR) and second-degree relatives (SDR) ≤ age 80 without a BC diagnosis. Lifetime BC risk was calculated on each relative through the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA), a public web-based program used to calculate lifetime breast and ovarian cancer risks in women based on family history. Relatives were categorized based on BC risk of < 20% versus ≥ 20%, given that national guidelines recommend high-risk breast MRI screening when patients have a BC risk of ≥ 20%. Relatives of individuals who carry the ATM 7271T>G mutation were excluded given the higher BC risks reported in literature. Relatives of CHEK2 carriers were compared with relatives of ATM carriers using Pearson’s chi-square tests.

**Results:** Among the 56 CHEK2 carriers, lifetime BC risk was calculated for 102 FDR and 143 SDR who met inclusion criteria. Of these, 77% of FDR and 88% of SDR had a lifetime BC risk < 20% based on family history, thus the presence of a CHEK2 mutation would increase their risk over the 20% threshold at which high-risk BC screening would be appropriate. Among the 56 ATM carriers, lifetime BC risk was calculated for 108 FDR and 149 SDR who met inclusion criteria. Of these, 66% of FDR and 81% of SDR had a lifetime BC risk < 20% based on family history, thus the presence of an ATM mutation would increase their risk over the 20% threshold at which high-risk BC screening would be appropriate. When comparing the two groups, there was not a significant difference in the number of FDR and SDR who had a lifetime BC risk < 20% (p=0.06 and p=0.102, respectively).

**Conclusions:** Among female FDR and SDR, the majority have a BC risk < 20%, thus identification of a CHEK2 or ATM mutation would alter BC surveillance recommendations. Our results highlight the potential for moderate penetrance genes to impact patient cancer risk management among relatives of CHEK2 and ATM carriers, beyond recommendations based solely on family history of cancer.
Aims. Chronic obstructive pulmonary disease (COPD) is a leading cause of mortality in the United States. Individuals with COPD often present with a wide range of symptoms and clinical findings, and disease progression can vary greatly, driving a need for more personalized medicine approaches. However, the heterogeneity in disease presentation is currently not well understood. Research in large populations with COPD is needed to improve personalized disease monitoring and treatment. Electronic health records provide large-scale healthcare data for clinical research, but have been underutilized in COPD research due to challenges identifying these individuals, especially in the absence of pulmonary function testing data. Our objective was to determine whether clinical features available from electronic health records can electronically phenotype individuals with COPD at a large tertiary care center.

Methods. We identified individuals over 45 years of age at last clinic visit within the Vanderbilt University Medical Center electronic health records. We tested phenotyping algorithms using combinations of variables versus both pulmonary function tests and chart review as the gold standard.

Results. The best-performing algorithm required cases to be smokers with a combination of ICD codes for COPD and mentions of oxygen use on the problem list. This algorithm had a sensitivity of 96.4%, specificity of 95.8%, positive predictive value of 90.0%, and negative predictive value of 98.6% in chart-reviewed records. We also found that a simpler algorithm requiring cases to have 3 or more COPD ICD codes had a sensitivity of 94.1%, specificity of 88.1%, positive predictive value of 71.1%, and negative
predictive value of 98.0%. To validate the performance of the best-performing algorithm, we tested
known genetic associations for COPD in our algorithm-defined cohort. In the majority of variants tested,
our results were consistent with those previously published.

**Conclusions.** Phenotyping algorithms for COPD that combine ICD codes and additional clinical data
perform well in a large EHR. Based on our genetic analyses, the population identified by our algorithm is
similar to cohorts used in previous genetic studies, which helps validate our algorithm’s performance.
The development and implementation of reliable phenotyping algorithms may expand opportunities for
clinical research in COPD. By identifying large populations of individuals with COPD, we can better
understand the disease heterogeneity and outcomes associated with different clinical presentations,
which will advance personalized medicine for individuals with COPD.

Words: 380 (Max: 500)
A Phenome-Wide Association and Network Studies Approach to Identifying Novel Disease Associations Using Large-scale Electronic Health Records.

Brian S. Mautz, Sarah H. Jones, Eric S. Torstenson, Jacklyn N. Hellwege, Todd L. Edwards, Digna R. Velez Edwards

Aims: Uterine fibroids affect up to 70% of women by menopause and result in estimated costs of up to $34 billion annually in the USA alone. Prior observational studies have investigated only a limited number of clinical factors associated with fibroid risk. Hence, little is known about the broader clinical diagnoses associated with uterine fibroids. Large scale electronic health records (EHR) provide a unique opportunity for a more comprehensive and agnostic investigation of clinical characteristics associated with women diagnosed with uterine fibroids.

Methods: Utilizing the Synthetic Derivative, the electronic health record (EHR) database at Vanderbilt University Medical Center, cases and controls in black (N = 3,568 cases/12,521 controls) and white women (N = 7,577/60,296) were identified using a previously validated method. First, we conducted a ‘phenome-wide association study’ (PheWAS) to provide a robust test for associations of all diagnoses across patient clinical records with fibroids. Second, we performed a network analysis in women with uterine fibroids to identify novel patterns of diseases.

Results: After controlling for body mass index and correcting for multiple tests, we found 289 and 426 diagnoses were significantly associated with fibroids in black and white women, respectively. As expected, across racial groups the most significant associations were previously related to menstruation, such as excessive or frequent menstruation (both p < 1.0 x 10^{-274}) and dysmenorrhea (both p < 2.58 x 10^{-152}). However, we also detected numerous novel associations with cervical cancer, endometriosis, other malignant neoplasms, and diverticulosis. There were similarities in networks between black and white women with diagnoses such as disorders of lipid metabolism, sleep, and urinary tract infections appearing as the most connected diagnoses in both. Size, complexity, and connectedness of networks varied between black and white women.

Discussion: These results provide novel insight into fibroids and suggest associative disease complexes. We detected numerous novel association between fibroids and individual diseases such as cervical cancer, malignant ovarian neoplasms, endometriosis, and polyps of female genital organs. In the broader disease networks, the most connected diseases appear to act as hubs between disease diagnoses. These hubs provide potential research and treatment points for future research. However, investigation of temporal patterns of diagnoses and disease mechanisms is needed.
Title: Evaluating the Genotype-Phenotype Relationships and Imputed Gene Expression Profiles of Congenital Malformation Syndromes in BioVU

Authors: Tyne W. Miller-Fleming¹,², Digna R. Velez-Edwards²³, Georgia Chenevix-Trench⁴, Nancy J. Cox¹²

¹Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA.
²Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA.
³Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA.
⁴QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

Aims:
Congenital malformations (CM) are physical or structural defects present from birth. CM are detected in approximately 3% of births and are the leading cause of infant mortality. Toxic environmental exposures, nutritional deficiency, and maternal age are known risk factors; however, many defects arise due to undetected genetic variation. Identifying the genetic risk factors, and the individuals at greatest risk for CM, could provide insight into the prevention or treatment of these disorders. To better understand the genetic and phenotypic relationships that underlie CM, we utilized the medical phenome and linked genotype data in Vanderbilt’s biobank, BioVU.

Methods:
De-identified electronic health records linked to genetic information allow researchers to evaluate genotype-phenotype relationships with large sample sizes among populations with strong phenotypic diversity. We investigated the medical phenome associated with CM using phenotype risk scores (PheRS). PheRS can be used to capture multiple, distinct phenotypes that contribute to disease across an entire population. We developed a PheRS for CM using all phecodes that represent congenital disease in BioVU for 23,000 individuals of European ancestry. Additionally, we calculated the PheRS for syndromic CM, including VACTERL syndrome (characterized by vertebral defects, anal atresia, cardiac defects, tracheoesophageal fistula, renal anomalies, and limb abnormalities) across the same population. Using BioVU genotype information and the GTEx reference transcriptome, we imputed gene expression and evaluated the relationship between gene expression and the PheRS for general CM and VACTERL syndrome.

Results:
After calculating the PheRS for general CM (i.e. using all phecodes that represent congenital disease in BioVU), we found no significant associations with predicted gene expression. Interestingly, when we calculated the PheRS for VACTERL, we identified significant associations with imputed gene expression of 22 genes, many of which have been previously implicated in congenital disease. One of the genes we identified encodes the mitochondrial enzyme, DHODH (dihydroorotate dehydrogenase), which has been linked to Miller syndrome (postaxial acrofacial dysostosis). Additionally, we identified the LILRB4 gene which encodes leukocyte immunoglobulin like receptor B4. Previous studies have demonstrated a role for LILRB4 in connection with Toxoplasma gondii infection and can result in miscarriage and CM.

Conclusions:
Our findings identified several genes associated with VACTERL syndrome, although the PheRS for general CM yielded no significant results. These studies suggest that implementing the PheRS and imputed gene expression strategies can identify genes associated with syndromic CM, which may be distinct for each congenital disease. Future studies will take advantage of the diverse medical phenome in BioVU to determine the genotypic and phenotypic signatures between different congenital syndromes.
**APOE ε4-specific Associations of VEGF Gene Family Expression with Cognitive Aging and Alzheimer’s Disease**

Annah M. Moore, BS\textsuperscript{a,b,c}, Emily Mahoney, BS\textsuperscript{a,b}, Logan Dumitrescu, PhD\textsuperscript{a,b}, Philip L. De Jager, PhD\textsuperscript{d,e}, Mary Ellen I. Koran, MD, PhD\textsuperscript{f}, Vladislav A. Petyuk, PhD\textsuperscript{g}, Renã AS Robinson, PhD\textsuperscript{h}, Sandeep Goyal, PhD\textsuperscript{h}, Julie A. Schneider, MD\textsuperscript{i}, David A. Bennett, MD\textsuperscript{i}, Angela L. Jefferson, PhD\textsuperscript{a}, Timothy J. Hohman, PhD\textsuperscript{a,b}

\textsuperscript{a}Vanderbilt Memory and Alzheimer’s Center, Vanderbilt University Medical Center, Nashville, TN
\textsuperscript{b}Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN
\textsuperscript{c}Pharmacology Department, Vanderbilt University Medical Center, Nashville, TN
\textsuperscript{d}Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY
\textsuperscript{e}Cell Circuits Program, Broad Institute, Cambridge MA
\textsuperscript{f}Stanford Hospital, Department of Radiology, Stanford, CA
\textsuperscript{g}Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA
\textsuperscript{h}Department of Chemistry, Vanderbilt University, Nashville, TN
\textsuperscript{i}Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago, IL

**Total Word Count:** 462

**Aims:** Several studies have suggested vascular endothelial growth factor A (VEGFA) is protective against Alzheimer’s disease (AD). High levels of cerebrospinal fluid (CSF) VEGFA are associated with slower rates of hippocampal atrophy and less cognitive decline in AD biomarker-positive subjects, suggesting VEGFA is especially protective among those at highest risk for clinical AD and cognitive decline. \textit{APOE-ε4} carriers represent a highly susceptible population for the development of AD and cognitive deficits. Thus, VEGF may confer distinct protection among those at high genetic risk for cognitive decline. Studies that analyze the entire VEGF family are required to determine which ligands and receptors along the complex VEGF signaling cascade contribute to cognitive protection. This work evaluates interactions between prefrontal cortex expression of VEGF family genes and \textit{APOE-ε4} status on clinical AD and cognition to clarify which VEGF ligands and receptors modify the association between ε4 and cognitive decline.

**Methods:** Cognitive data and autopsy measures of gene expression were obtained from the Religious Orders Study and Rush Memory and Aging Project. Binary logistic regression assessed \textit{APOE-ε4} x VEGF expression interactions on AD diagnosis. Linear and mixed-effects regression assessed \textit{APOE-ε4} x VEGF expression interactions on global cognition and cognitive change. Models covaried for age at death, postmortem interval, and sex. Secondary analyses investigated VEGF associations stratified by \textit{APOE-ε4} status. Correction for multiple comparisons adjusted for all 10 genes in the VEGF family.

**Results:** \textit{NRP1} and \textit{NRP2} interacted with \textit{APOE-ε4} on AD diagnosis (corrected p-values≤0.05). \textit{NRP1} and \textit{VEGFA} interacted with \textit{APOE-ε4} on cross-sectional cognitive performance (corrected p-values≤0.05). In all cases, higher expression of these genes
was associated with worse outcomes among ε4 carriers, but better outcomes among ε4 non-carriers.

Conclusions: Results suggest VEGFA, NRP1, and NRP2 modify the risk for clinical AD and cognitive decline based upon APOE-ε4 status. These three genes contribute to brain angiogenesis, and previous studies have suggested small vessels in APOE-ε4 brains are more prone to leaking, perhaps placing young vessels formed during angiogenesis at risk for ischemia. Our results suggest that therapeutics aimed at increasing brain angiogenesis may only be beneficial among non-carriers of the ε4 allele. Future work will seek replication and extend results to the protein level to better understand the biological context in which angiogenesis is helpful or harmful.

Figure 1. NRP1 x APOE-ε4 interaction on cross-sectional cognition
Global cognitive performance is presented on the y-axis, normalized mRNA levels of NRP1 measured in prefrontal cortex are presented along the x axis. Points and lines are colored by APOE-ε4 whereby ε4 carriers are presented in red and non-carriers are
NRP1 interacts with APOE-ε4 on global cognition (β=-0.28, p=0.007) whereby higher levels of NRP1 are associated with slower rates of cognitive decline among ε4 non-carriers (β=0.11, p=0.004), but faster rates of cognitive decline among ε4 carriers (β=-0.17, p=0.02).
Leveraging phenotype data to identify potential carriers of deleterious genetic variants

Theodore Morley, Lisa Bastarache, Nancy Cox, Douglas Ruderfer

Aims:
Heterogenous clinical symptoms caused by rare genetic variation may complicate identification of the underlying cause, prolonging diagnostic odyssey and delaying opportunity for potential medical benefits. Recent literature has shown that there exists a substantial population of individuals carrying deleterious genetic variants that cause known syndromes that go undetected in our current healthcare system. Given this, there is a clear need to maximize our ability to identify patients carrying genetic variants that could inform diagnosis, prognosis and potentially treatment. We hypothesize that unique combinations of phenotypes seen infrequently in the population are likely to correspond to increased likelihood of an underlying genetic variant. Here, we aim to use only phenotypic information from electronic health records (EHR) in order to identify individuals at increased risk for possessing deleterious genetic events with the ultimate goal of validating these predictions through genetic assays (e.g. exome-sequencing).

Methods:
To define our case population, we identified 2,388 individuals who received a form of genetic testing designed to identify structural variants from notes within the synthetic derivative. We then matched these cases in a one to four ratio to a set of 9,552 controls based on age, sex, ancestry, number of unique years in which the patient has visited VUMC, and length of medical record. Utilizing only phecode data, we considered four potential inputs including a broadly defined phenotype risk score including all phcodes (weights calculated over ~800k VUMC medical home patients), a binary matrix of presence and absence of phcodes and two dimensionally reduced versions of the binary matrix leveraging principal component analysis (PCA) and uniform manifold approximation and projection (UMAP). We also considered four different prediction models including logistic regression, naïve Bayes, random forests and support vector machines. Using four-fold cross validation with an additional held out test set consisting of twenty percent of the original sample, we selected the best performing model based on positive predictive value (PPV) for assessment on the held-out test set. We next applied this model to ~800k VUMC medical home patients defined as having visited during at least two different years, and having at least four years of medical record length

Results:
The binary phcode matrix and the random forest model provided the best performance in cross-validation and had a PPV of 73%, NPV of 91% and total accuracy of 88% in the held-out test dataset. Further, our model was well calibrated with Brier score of 0.08. Projecting the model to an age comparable subset of the medical home patients demonstrated a significant increase in predicted probabilities among individuals known to have visited the genetic clinic. Ongoing work will provide additional performance metrics on this dataset and rank order those with high probability of carrying a genetic variant for chart review and genetic validation.

Conclusions:
We conclude that phenotypic information in the EHR even at the level of phecodes can provide evidence for a patient carrying a deleterious genetic variant and that this approach has potential to inform those who should receive genetic testing.
Title: A high-throughput cell-surface trafficking assay can identify functionally deleterious \textit{KCNE1} variants

Background: With the increase of genetic testing in clinical care, an emerging challenge in precision medicine is to accurately classify variants in Mendelian disease genes. One such gene is \textit{KCNE1}, which encodes an accessory subunit for a potassium channel important for cardiac repolarization. Loss of function mutations in \textit{KCNE1} have been reported to disrupt interactions with other potassium channel subunits or reduce trafficking of channels to the cell surface, leading to type 5 long QT syndrome, associated with susceptibility to serious arrhythmias including sudden cardiac death.

Aims: The purpose of this study is to generate all possible protein coding mutations in the 127-amino acid long KCNE1 subunit and to assess their function in a high throughput multiplexed fashion.

Methods: We have developed a flow-cytometry based trafficking assay that identifies variants that disrupt KCNE1 cell surface trafficking. We have generated a comprehensive library of multiple variants and expressed it in specifically engineered HEK293 cells that harbor a single variant per cell.

Results: The library showed a bimodal distribution of cell surface KCNE1 staining, with peaks of the distribution corresponding to wild-type and known trafficking-deficient mutants respectively. Future work will quantify each variant’s presence on the cell surface via a "trafficking score", and will test the library in a cell-death assay to characterize variants that disrupt subunit interactions.

Conclusions: We have successfully developed a high-throughput functional assay to comprehensively interrogate large numbers of protein variants for an ion channel subunit. Our results will be used to supplement the American College of Medical Genetics and Genomics functional assay criterion and thereby classify and reclassify \textit{KCNE1} variant pathogenicity.
Helping solve medical mysteries: using machine learning and systems biology to identify digenic disease

Introduction

Previous research in clinical genomics has led to the identification of causative protein-coding variants in specific genes for hundreds of diseases. To address the challenges of diagnosing rare genetic diseases, the Undiagnosed Diseases Network (UDN) was formed. The UDN has had success in diagnosing many patients by sequencing their whole exomes or genomes and analyzing the effects of each individual variant in the patient. Nonetheless, more than half of enrolled UDN patients remain undiagnosed.

My central hypothesis is that the disease phenotype in many of the unresolved UDN patients is likely to be caused by more than one genetic variant. A UDN patient has been diagnosed previously where the complex phenotype of the patient resulted from mutations in two genes. I postulate that interactions between variants synergistically lead to the manifestation of disease. A database of known digenic combinations has been published previously, and it was found that it was likely for digenic gene pairs to have defining biological characteristics. A machine learning approach was implemented to distinguish true digenic disease cases among all the digenic combinations on the digenic database.

I will be implementing a Random Forest machine learning classifier to identify digenic gene pairs among all the genes mutated in a UDN patient. The classifier would encompass biological parameters such derived from systems biology, biological networks, evolutionary biology and genomics. My results will be integrated into the UDN pipeline for prioritizing genes for experimental validation to facilitate identification of potential causal variants.

Results

The classifier implemented using only the CFs performed well for all the non-digenic sets (mean ROC AUCs > 0.88), with the performance being lowest for the matched set (mean PR AUC ~ 0.3) owing to replicating the distribution of features for the digenic pairs. EBGFs improved the performance of the classifier (data not shown; ROC AUCs > 0.90, PR AUCS > 0.39) and addition of all features yielded a classifier with high performance (mean ROC AUCs > 0.97, mean PR AUCs > 0.60). The feature importance values revealed different features being most indicative for different non-digenic set, demonstrating the robustness of the classifier to rely upon different features while identifying digenic pairs with varying characteristics.

Broader impact

This interdisciplinary integrative approach would facilitate the use of curated biological data to design a machine learning classifier that could be of direct consequence to patients. With high
quality sequencing information being more available, this approach has the potential to identify likely gene pairs related to the patient phenotype, prioritizing them for experimental validation. A digenic classifier is the need of our data intensive times which provides us with the means to rely upon previously published biological data to solve medical mysteries.
HLA-A*32:01: Identification of a strong association with vancomycin drug reaction with eosinophilia and systemic symptoms and development of a single allele PCR assay to facilitate clinical translation

Authors: Nontaya Nakkam, PhD1,2*, Katherine C Konvinse, MS1*, Francois Rwandamuriye, MS7, Rebecca Pavlos, PhD6, Ian James, PhD7, Christian M Shaffer, BS8, Jason A Trubiano, MBBS, PhD3,4,5, Cosmin A Bejan, PhD9, Mark A Pilkinton, MD, PhD10, Misha Rosenbach, MD11, Jeffrey P Zwerner, MD12, Kristina B Williams, RN, BSN, CCRP10, Jack Bourke, MBBS13, Patricia Martinez, MBBS, PhD13,14,15, Mark Watson, PhD7, Linda Choo7, Alec J. Redwood, PhD7, Katie D White, MD, PhD10, Simon A Mallal, MBBS1,7,10 Abha Chopra, PhD7 and Elizabeth J Phillips, MD1,7,10,16

Affiliations:
1. Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, USA.
2. Department of Pharmacology, Faculty of Medicine, Khon Kaen University, Khon Kaen, THAILAND
3. Department of Infectious Diseases, Austin Health, Heidelberg, AUS.
4. The National Centre for Infections in Cancer, Department of Infectious Diseases, Peter MacCallum Cancer Centre, Parkville, AUS.
5. Department of Medicine, University of Melbourne, Parkville, AUS.
6. Telethon Kids Institute, University of Western Australia, Nedlands, AUS.
7. Institute for Immunology and Infectious Diseases, Murdoch University, Murdoch, AUS.
8. Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, USA.
9. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, USA.
10. Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, USA.
11. Department of Dermatology, Hospital of the University of Pennsylvania, Philadelphia, USA.
12. Department of Dermatology, Vanderbilt University Medical Center, Nashville, USA.
13. Department of Clinical Immunology, Fiona Stanley Hospital, Murdoch, AUS.
14. Department of Clinical Immunology, Royal Perth Hospital, Perth, AUS.
15. Division of Pathology and Laboratory Medicine, School of Medicine, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia, Crawley, AUS.
16. Department of Pharmacology, Vanderbilt University Medical Center, Nashville, USA.
**Background:**
Vancomycin is a commonly used antibiotic for resistant gram-positive infections and is a prevalent global cause of a severe drug hypersensitivity syndrome called drug reaction with eosinophilia and systemic symptoms (DRESS) that occurs on average 2-8 weeks following first dosing of vancomycin. HLA class I alleles have been strongly associated with T-cell mediated adverse drug reactions which has led to pre-prescription preventive screening strategies for drugs like abacavir. Translation of screening for an HLA risk allele into clinical practice can be facilitated by the availability of cost-effective single allele PCR-based HLA assays.

**Methods:**
We identified cases that were probable or higher vancomycin DRESS based on the RegiSCAR and Naranjo causality assessments and we matched these 1:2 with tolerant controls based on sex, race and age from Vanderbilt BioVU and from three centers prospectively and also compared with entire BioVU population with imputed HLA typing available. Prospectively collected DRESS cases had interferon gamma ELISpot assay performed to vancomycin and other drugs to aid in drug causality. Associations between DRESS and carriage of HLA class I and II alleles were assessed by conditional logistic regression. We conducted a time to event analysis of those exposed to vancomycin with and without the identified risk allele. Upon finding a strong HLA class I association with vancomycin DRESS we then went onto develop a method for a real-time specific PCR assay to facilitate use in the clinical setting. Primers were designed within exon 2 of the HLA locus to create a primer set combination predicted to amplify all but 3 HLA-A*32 alleles. To validate the HLA-A*32:01 typing assay a research assistant blinded to the HLA type of the samples performed the single allele PCR assay on 458 randomly selected samples that had previously been sequenced with high resolution typing.

**Results:**
Twenty three individuals met inclusion criteria for vancomycin associated DRESS and of these 19/23 (82.6%) carried HLA-A*32:01 compared to 0/46 (0%) of the matched vancomycin tolerant controls (p=1 x 10⁻⁸) and 6.3% of the overall BioVU population (n=54,249) (p=2 x 10⁻¹⁶). A time to event survival analysis of vancomycin DRESS development during treatment amongst the HLA-A*32:01 positive group indicated that 19.2% developed DRESS within 4 weeks as opposed to 0% in a matched group not carrying HLA-A*32:01). Of 30/458 samples known to be HLA-A*32:01, 100% were accurately identified as positive for the HLA-A*32:01 allele on the single allele melting curve PCR assay. Samples were called positive or negative for HLA-A*32:01 based on the presence or absence of the HLA-A*32:01 specific melt peaks. The 30 HLA-A*32:01 positive samples showed two peaks at 88.5°C. The other 428 non-HLA-A*32 samples showed a single peak at 76.07°C.

**Conclusion:**
HLA-A*32:01 is strongly associated with vancomycin DRESS and almost 20% of those carrying this allele will develop vancomycin DRESS if on it 2 weeks of longer. A simple single allele PCR-based assay was developed to facilitate cost-effective testing in clinical practice. Based on this a model for pre-emptive screening to risk stratify patients can be proposed whereby HLA-A*32:01 testing is done following initiation of vancomycin in patients with intent to treat with vancomycin for longer than one week.
**CYP2D6 polymorphisms and adverse events to risperidone in children and adolescents**

Katelyn M. Neely¹, MD, Kazeem A. Oshikoya², MD, PhD, Ida T. Aka¹, MPH, Robert J. Carroll³, PhD, Angela C. Maxwell-Horn¹, MD, Dan M. Roden²,3,4, MD, and Sara L. Van Driest¹,2, MD, PhD

**Affiliations:** Departments of ¹Pediatrics, ²Medicine, ³Biomedical Informatics, ⁴Pharmacology Vanderbilt University School of Medicine, Nashville, Tennessee, United States

**Aim:** Risperidone is commonly prescribed for a variety of pediatric behavioral concerns and is most often used for off label indications. Approximately 1 in 3 children treated with risperidone experience an adverse event (AE). Risperidone is primarily metabolized by the enzyme cytochrome P450 2D6 (CYP2D6). There are few and conflicting data on the role of CYP2D6 polymorphisms in relation to risperidone AEs in children. This study assessed the extent of the association between CYP2D6 metabolizer status and risk for risperidone AEs in children and adolescents.

**Methods:** This retrospective cohort study used BioVU, a de-identified DNA biobank linked to Electronic Health Record (EHR) data and included children ≤18 years of age with ≥ 4 weeks of oral risperidone exposure. After DNA sequencing and copy number variant assessment, individuals were classified as CYP2D6 slow (including poor and intermediate), normal, or ultrarapid metabolizers. The primary outcome of this study was AEs, defined as any untoward event or laboratory abnormality attributed to risperidone and was assessed by manual review, blinded to CYP2D6 metabolizer status. Associations between AEs, metabolizer status, and clinical variables were determined using Fisher's exact test and multivariate regression analyses.

**Results:** The cohort included 257 individuals with a median age of 8.3 (IQR 6.3-10.5) years and 188 (73%) were male. In all, 33 (13%) were slow metabolizers, 218 (85%) were normal metabolizers, and 6 (2%) were ultra-rapid metabolizers. AEs were identified in 76 individuals (30%). The most common AEs were weight change, sedation, and abnormal movements. AEs were more common in slow metabolizers than normal metabolizers (15/33, 45% vs. 58/218, 27%, p=0.04). In multivariate analysis adjusting for concomitant CYP2D6 inhibitor use, age, race, sex, and dose, slow metabolizers had increased AE risk compared to normal metabolizers (AOR 2.5, 95% CI 1.1-5.3, p=0.02).

**Conclusion:** Children with CYP2D6 poor or intermediate metabolizer phenotypes are at greater risk for risperidone AEs compared to normal metabolizers. Pre-prescription genotyping could identify this high-risk subset of patients for alternate therapy, reduced risperidone dosing regimen, and/or increased monitoring for AEs.
Abstract

Title: Can you clap to the beat? Findings from the first large-scale genome-wide association study of a musical rhythm trait in 606,825 individuals.

Maria Niarchou, J. Fah Sathirapongsasuti, Nori Jacoby, 23andme Research Team, J. Devin McAuley, Eamonn Bell, Miriam Mosing, Peter Straub, Nicole Creanza, Fredrik Ullén, Nancy Cox, David Hinds, Lea K. Davis, Reyna L. Gordon

Aims: Across musically trained and non-musically trained individuals, there is substantial variability in the ability to perceive and produce rhythms accurately. Individual differences in musical rhythm have been linked to a subcortico-cortical network of brain regions, involving primarily auditory, motor, and subcortical/basal ganglia circuitry. Family-based studies demonstrate a moderate genetic contribution to rhythmic ability. However, understanding the molecular basis of rhythm necessitates genome-wide interrogation in a large well-powered sample.

Methods: Here we applied Genome-Wide Association Study (GWAS) methodology to identify common genetic variants associated with musical rhythm, collected from N=606,825 research participants from the personal genetics company 23andMe. Individuals responded to the question ‘Can you clap in time with a musical beat?’. To validate this single question, we also conducted a separate (behavioral) study using Mechanical Turk in N=734, and showed that individuals who answered Yes (vs. No) to this self-report question also performed better on a musical rhythm perception task (p=0.0006).

Results: In the genetic cohort, preliminary GWAS revealed 68 independent loci that surpassed the threshold for genome-wide significance. We found two loci on chromosome 4 (4q34.2 and 4q22.1), replicating prior findings of linkage to musicality in this region, as well as new loci including 16p11.2 (a known locus of neurodevelopmental disorders), 2p16.1 (a region linked to mental health and sleep phenotypes) and 17q21.31 (previously associated with cortico-basal degeneration and intracranial volume). GWAS results held after conditioning the analyses on known markers of IQ, using mtcojo, revealing independence of genetic markers of rhythm and IQ. LD-score regression showed 5% observed SNP-heritability of the rhythm phenotype.
Conclusions: Taken together, these findings provide promising evidence of genetic architecture that may be involved in rhythmic ability in humans.
Uterine Leiomyomata Polygenic Risk Scores (PRS) Confer Novel Relationships in the Clinical Phenome

Jacqueline A. Piekos1, Jacklyn N. Hellwege1,2, Yanfei Zhang8, Eric S. Torstenson1,3, Gail P. Jarvik7, Ozan Dikilitas9, Iftikhar J. Kullo9, Daniel J. Schaid10, Sarah A. Pendergrass, Dan Roden1,4, Josh C. Denny1,5, Todd L. Edwards1,3, Digna R. Velez Edwards1,5,6

1Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 3Genomic Medicine Institute, Geisinger, Danville, PA; 4Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center; 6Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA; 7Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN; 8Biomedical and Translational Informatics, Geisinger, Danville, PA; 9Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Center, Nashville, TN; 10Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 11Division of Quantitative Science, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN

Aims: Uterine leiomyomata (UL) are the most common pelvic tumor in women, with an incidence rate of 77% and poorly understood etiology. To gain understanding about the shared genetic contribution across many clinical phenotypes, we developed a UL polygenic risk score (PRS) and combined it with a Phenome-Wide Association Study (PheWAS) approach.

Method: We constructed a PRS with PRSice software using effect sizes derived from a previously published imaging-confirmed, predominately European American (EA), UL Genome-Wide Association Study (GWAS) data (N=2,651 cases and 4,326 controls) in a subset from the Electronic Medical Records and Genomics (eMERGE) network. Optimization was done in an independent set of imaging-confirmed UL cases and controls (N=5,179) from BioVU. The optimal p-value threshold for SNP inclusion in the PRS was p<0.001, resulting in 4,448 variants, and the model was significant (r² = 0.0017, p-value = 0.041).

PheWAS analyses were performed across 53,116 non-Hispanic EA and 5,583 non-Hispanic African-American (AA) men and women from eMERGE, excluding samples used for construction, using the PRS as the predictor for clinical disease phenotypes (N = 1,738) adjusted for sex, age, body mass index, and 10 principal components.

Results: In both the sex combined and female only non-Hispanic EA PheWAS, UL was the most significant phenotype (p=9.58x10⁻¹⁶⁷). In the sex combined analysis, we detected 40 (p<2.7x10⁻⁵) significant and 16 suggested significant phenotypes; in a female only stratification, 47 significant and 19 suggested significant phenotypes were detected. Top phenotype results include excessive or frequent menstruation, other benign neoplasms of the uterus, endometriosis, and ovarian cysts. Other associated phenotypes fell into categories of genitourinary, neoplasms, and sense organs. The sense organ phenotypes were comprised of various eye diseases including astigmatism, optic atrophy, and myopia. In the non-Hispanic AA PheWAS, no phenotypes reached genome wide significance.

Conclusion: The PRS we developed works very well in non-Hispanic EA, indicated by the top result in the PheWAS analysis being UL. The lack of replication in the non-Hispanic AA subset supports the claim the PRS are generally race specific. We discovered several novel associations between UL and eye diseases, suggesting that they might share some genetic architecture. Some of the diseases we detected, like osteoporosis and optic atrophy, are thought to be caused by hormone changes in an aging population. With UL now associated with these diseases, it is possible that UL may share this etiology with these diseases.
LabWAS: A method to Investigate relationships between biomarkers and genetic risk scores

Aims:
Clinical laboratory testing provides physicians with targeted biochemical measurements (i.e., biomarkers) to aid in diagnosing and treating patients for a variety of diseases. Identifying blood-based biomarkers for psychiatric traits has largely stalled due elusive biology. Combining genetics of psychiatric disorders to biomarker and diagnostic data from electronic health records could point towards mechanisms of disease, provide information on the relationship between biomarkers and diagnosis, and facilitate the development of diagnostic panels. We describe a method that pairs polygenic scores with lab data derived from EHRs to screen for associations.

Methods:
Results from 3,800 labs were extracted for 70,184 individuals in Vanderbilt University Medical Center’s biobank, BioVU. Usable labs were required to have a minimum of 1,000 observations over at least 100 individuals, and 70% of measurements to be recorded in one set of units. The remaining 475 labs were filtered to exclude observations outside 4 standard deviations from the mean. Next, the median value of each lab for each individual was calculated, inverse normalized, and adjusted for cubic splines of age at measurement. Labs exhibiting no calculable heritability using GCTA were excluded leaving 272 labs for analysis. To screen for associations between PRS and lab measurements, we created a lab-wide association study (LabWAS) pipeline. This pipeline fits a linear regression model to each lab and determines the association between the PRS after further adjusting for sex, top 10 principal components of ancestry, and genotyping batch. Polygenic scores were calculated for BioVU individuals of European genetic ancestry (n=29,768) using PRS-CS, which uses a Bayesian framework to model linkage disequilibrium from an external reference and a continuous shrinkage prior on SNP effect sizes that is robust to various genetic architectures.

Results:
In a proof-of-principle analysis, PRS for coronary artery disease reproduced known associations with canonical heart disease risk factors. The strongest association was nucleated red blood cells (OR = 0.97, p-value = 1.43e-8), followed by HDL cholesterol (OR = 0.96, p-value = 9.51e-7), and blood glucose levels (OR = 1.03, p-value = 1.12e-4). To test our method in neuropsychiatric disorders, we generated a PRS for depression and applied the LabWAS approach. The depression PRS-LabWAS showed associations with white blood cell count (OR = 1.04, p = 3.03e-8), nucleated red blood cells (OR = 0.97, p-value = 4.11e-7), and immature granulocytes (OR = 1.03, p-value = 1.84e-4). We performed a conditional analysis which include the addition of a covariate for depression diagnosis to control for potential confounding effect of diagnosis or treatment and white blood cell count remained significantly associated with MDD-PRS (OR = 1.04, p-value = 6.57e-6).

Conclusions:
Our preliminary results for depression replicate previous reports of a link between depression pathology and an elevated immune system. Results from LabWAS provide a starting point for
further sensitivity analyses to determine shared genetic architecture, relevant points in life, and effects of medication. We will share the development of the pipeline, available software, and the results of ongoing replication efforts.
Title: Personalized Structural Biology Program at Vanderbilt University

Authors: Jonathan Sheehan, Mike Sivley, Greg Sliwoski, Brett Kroncke, Jeff Mendenhall, Bian Li, Tony Capra, Jens Meiler, Chuck Sanders

Aims: Genetic sequence determines protein structure, which determines function. Combining analyses of genetic mutations and protein structures provides a powerful tool for personalized medicine. We built on Vanderbilt’s strengths in genetics, cancer biology, structural biology, biomedical informatics, and biomedical ethics, to establish a Program in Personalized Structural Biology.  

Aim 1: Develop foundational technologies using genomic and structural data in the same context. Aim 2: Demonstrate the power of structural biology in personalized medicine for cancer and undiagnosed diseases.

Methods: 1. Develop a computational pipeline for integrating all known human genetic variants into all available protein 3D structures and models. Genomic databases were used to identify locations of disease-associated variants, and locations of benign variants. These were mapped to proteins in 3D, using structures or models. This provides a tool to visualize spatial relationships between known disease-associated domains of a protein and regions that are genomically distant but reside in close proximity when the protein is folded. 2. Establish feasibility and utility of personalized structural biology. Candidate mutations from patients were identified in collaboration with investigators in the Vanderbilt Ingram Cancer Center and the Undiagnosed Disease Network. We used our newly created pipeline (Method 1) to integrate structural biology into interpretation of germline and somatic mutations.

Results: We highlight four proof-of-concept projects that illustrate the power of Personalized Structural Biology. 1. We identified 21 novel RTE1 variants that segregated with Familial Interstitial Pneumonia (FIP). We used comparative modeling to predict the tertiary structure of RTE1 and identified a spatial cluster of variants with known association to FIP in RTE1. We then developed an algorithm to classify missense variants of unknown significance (VUS) based on their spatial proximity to pathogenic vs. neutral variants. The approach out-performed other VUS classification methods in cross-validation and accurately predicted pathogenicity of disease-segregating VUS. 2. We performed computational modeling of a novel EGFR mutation identified in a lung tumor from a 33-year-old male never-smoker: a kinase domain duplication (EGFR-KDD). The structural model suggested that the mutation caused autoactivation of EGFR, providing a mechanistic explanation for in vitro findings that EGFR-KDD is constitutively active. EGFR-KDD-transformed cells proved sensitive to EGFR TKIs and the patient had a partial response to the EGFR TKI afatinib. 3. NGS of a tumor from a relapsing breast cancer patient showed development of a second HER2 mutation. Comparative modeling of this doubly-altered HER2 protein provided an explanation of the resistance mechanism and suggested the use of an alternative experimental drug. 4. Long-QT syndrome (LQTS) predisposes patients to sudden death by arrhythmia. Genetic testing to identify mutations encoding cardiac ion channels can guide treatment, but is complicated by the presence of VUS. We designed a
framework that integrates biochemical, functional and structural data to develop a machine learning algorithm capable of accurately predicting the likelihood that a VUS will cause LQTS.

**Conclusions**: We applied Personalized Structural Biology to genomic data from four patients with diverse diseases, demonstrating its power to provide actionable mechanistic interpretations of patient genetic data and improve treatment decisions.

Acknowledgements: J-N Gallant, Ariella Hanker, Jon Kropski, Christine Lovly, John Phillips, John Newman, Joy Cogan, Carlos Arteaga
A Personalized Care Model for Removing Unnecessary Penicillin Allergy Labels

Cosby A. Stone, Jr., MD, MPH
Christopher Lindsell, PhD
Joanna Stollings, PharmD
Mary Lynn Dear, PhD
Reagan Buie, MHA
Todd Rice, MD, MSCI
Elizabeth J. Phillips, MD

Disclosures:
- Dr. Stone receives funding from 5T32 GM007569-41 from NIH/NIGMS, and has received project related support specific to this research from both the VICTR Learning Healthcare System Platform under CTSA award No. UL1 TR002243 from NCATS/NIH and VICTR CTSA # VR52841. The contents presented here are solely the responsibility of the authors and do not necessarily represent official views of NCATS or the NIH.
- Dr. Phillips receives funding related to this project from: National Institutes of Health (1P50GM115305-01, 1P30AI110527-01A1, R21AI139021 and R34AI136815)
- This project is being performed under Vanderbilt IRB approvals: #181180 and #181734
- This study is registered with ClinicalTrials.gov: NCT03702283

Abstract Word Count: 416 excluding headings, max 500

Aims: 1) Develop a validated risk stratification tool to identify low risk patients labeled as penicillin allergic, 2) Study the efficacy, safety and effectiveness of direct amoxicillin challenge in penicillin allergy label removal.

Background:
Although 8-15% of the US population carries a penicillin allergy label, less than 5% of adults can be shown to be truly allergic using validated testing. A recent study suggests that 75% of penicillin allergy diagnoses are made by age three and this diagnosis is usually long-lasting. Diagnoses of penicillin allergy lead to considerable individual and public health morbidity including use of less efficacious and less safe antibiotics as well as downstream effects such as antibiotic resistance.

Methods:
A literature-based risk stratification tool was developed and validated via retrospective analysis of 318 sequential patients tested for penicillin allergy in the Vanderbilt Drug Allergy Clinic. This approach was used to determine the proportion of patients with a low risk penicillin allergy and to assess the proportion of those patients with a positive skin test or a symptomatic oral challenge. In a subsequent prospective pilot study, this tool is being applied in the ICU by a team of clinical pharmacists and physicians to determine safety and assess barriers to hospital implementation. Low risk penicillin allergy patients willing to undergo evaluation of
their allergy are directly challenged with 250mg amoxicillin without preceding skin testing. If asymptomatic, patients have their penicillin allergy label removed at the point of care.

**Results:**
Our retrospective analysis stratified 195/318 drug allergy clinic patients (61%) as low risk via the penicillin allergy risk stratification tool. Being in the low risk category carried a >99% negative predictive value, 95%CI (96.4, 99.9) for a positive penicillin skin test. Of low risk patients who agreed to undergo oral challenge to penicillin, amoxicillin or both in drug allergy clinic 184/184 (100%) were asymptomatic, demonstrating they no longer needed to be considered allergic to penicillin. In the first six weeks of the prospective ICU pilot study, 20/35 (57%) patients assessed had a low risk penicillin allergy history. Subsequently, 11/20 (55%) low risk patients agreed to undergo an observed amoxicillin oral challenge, tolerated it with no symptoms, and had their penicillin allergy label removed.

**Conclusions and Future Directions:**
Risk stratified approaches to penicillin allergy evaluation offer a safe and efficacious approach to personalized management likely to improve personal and public health outcomes. Upon the successful completion of our pilot study, we will employ a stepped wedge randomized, controlled trial to evaluate the enterprise-wide implementation of the risk stratification tool, integrated into EPIC, as clinical decision support.
Title:
Breast and ovarian cancer risk management practices among female BRCA1/2 carriers: Facilitators and barriers to recommended follow-up care

Authors:
Ann Tezak, MA, MPH, [1]; Anne Weidner, MPH, [1]; Debi Cragun, PhD, [3]; Kate Clouse, PhD, MPH, [2]; Tuya Pal, MD, [1]

[1] Vanderbilt University Medical Center; [2] Vanderbilt University; [3] University of South Florida

Abstract:

Aims:
Approximately 5-10% of breast cancers are inherited, most commonly due to mutations in the BRCA1 or BRCA2 (BRCA) genes. BRCA carriers have a 60-70% risk of developing breast cancer and up to 44% risk of developing ovarian cancer. These risks may be reduced by 90% or more through cancer risk management (CRM) practices, such as removal of breast tissue and/or ovaries. Benefits of BRCA testing are only realized through follow-up care involving CRM and family sharing of genetic test results. This mixed methods study examined the uptake of CRM practices among female BRCA carriers with and without a cancer diagnosis to better understand barriers and facilitators to appropriate follow-up care.

Methods:
Using online surveys, we collected demographic, clinical, and family sharing data from 186 female BRCA carriers recruited through prior Vanderbilt studies. A sub-group of 24 participants were purposively selected for in-depth interviews to ensure diversity in CRM and family sharing practices, with oversampling of those who self-identified as African American and/or Hispanic/Latina.

Results:
Of the 186 participants, 96% followed breast and ovarian CRM guidelines despite barriers related to insurance, financial issues, and/or difficulty establishing or scheduling care. Facilitators to following through with recommended follow-up care included: 1) support from family and friends; 2) healthcare providers empowering patients and supporting their autonomy; 3) and perceived susceptibility to and severity of a cancer diagnosis, particularly ovarian cancer. Additionally, when annual breast screening practices (i.e., mammograms and breast MRIs) were chosen over removal of breasts (i.e., risk-reducing bilateral mastectomy), often the deciding factor was due to fear of surgical procedures, including mastectomies and reconstruction, and difficult recovery.

Conclusions:
Additional interviews and analyses hope to examine non-adherence to breast and ovarian CRM guidelines among the 4% who are not adherent and gather additional insight into how most participants overcame barriers and followed through with recommended follow-up care.
Abstract for Personalized Medicine Day (500 word limit)

**Drug-Induced Immune-Mediated Skin Hypersensitivity Defined by Single Cell Approaches**


**Affiliations:**
1Vanderbilt University Medical Center, Nashville, Tennessee.
2Institute for Immunology and Infectious Diseases, Murdoch University, Perth, Australia.
3University of Washington, Seattle, Washington.
4Chang Gung Memorial Hospital, Taipei, Taiwan.
5National Yang-Ming University, Taipei Taiwan.

*Authors contributed equally to this work.

**Background:**
Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are serious cutaneous reactions that annually affect over 60,000 persons globally, are associated with mortality of up to 50%, and cause significant long-term morbidity. SJS/TEN is a class I HLA restricted, CD8+ dependent disease and we have previously identified oligoclonal T cells expressing the known cytolytic peptide granulysin at the site of tissue damage.

**Methods:**
Single-cell suspensions from blister fluid and PBMCs from two African-American women with allopurinol-associated SJS/TEN were enriched for live cells using annexin V negative selection and processed for single-cell RNA sequencing (scRNA-seq) on the 10X Chromium platform. Clusters of single cells were visualized through PCA dimensional reduction through application of clustering algorithms using UMAP for visualization. Candidate drug-reactive T cell receptors (TCRs) were replicated from previous studies of single cell T cell receptor sequencing (scTCR-seq) and whole transcriptome scRNA-seq from the same patients. A TCR reporter system then used autologous CD8+T cells transduced with the candidate TCRs, HLA class I single antigen lines and pharmacologically relevant concentrations of oxypurinol with interferon-readout to confirm HLA restriction and drug reactivity.

**Results:**
Here, we defined the transcriptomic profile of the previously identified cytotoxic, dominant T cell populations found in these patients. We confirmed that these cells and their peers in the blister fluid are predominantly CD8+ T cells expressing T cell activation and cytotoxic molecules, including granulysin, perforin, and the granzymes. In comparison to stimulated and unstimulated T cells found in the peripheral blood of these patients, we observe a marked upregulation of the activating killer-like immunoglobulin receptor KIR2DL4. T cells expressing KIR2DL4 in the blister fluid also express high levels of the chemotactic molecule CXCL13, suggesting that these CD8+ T cells bear an unusual T follicular-helper-like phenotype.

For subject 1, CD8+ T cells bearing a candidate drug-reactive TCR demonstrated IFN-g production in the presence, but not absence, of oxypurinol when tested against a panel of antigen presenting cells, thereby confirming TCR drug-reactivity. This aligns with the IFN-g
transcriptomic profile that we identified via scRNA-seq in this novel CD8+CXCL13+KIR2DL4+ single T cell population.

Conclusion:
Our studies 1) identify a novel CXCL13+ cytotoxic CD8 T cell population, 2) suggest that SJS/TEN pathology is reminiscent of a tertiary lymphoid structure, and 3) identify and confirm the antigen specificity of oxypurinol-recognizing TCRs that mediate T cell effector function within the blister fluid. These studies represent an innovative and productive mechanism to confirm novel HLA-TCR-drug pairs from the site of tissue damage in SJS/TEN.
Mela-Immu portal: an open platform for exploring immunotherapy-related multidimensional genomic profiles in Melanoma

Jing Yang1,2, Qi Liu1,2,* and Yu Shyr1,2,*

1 Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
2 Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA

* Correspondence: qi.liu@vanderbilt.edu, yu.shyr@vanderbilt.edu

Abstract

Aims: Although cancer immunotherapy using immune checkpoint blockade (ICB) has achieved remarkable benefits for the treatment of advanced melanoma, the overall response rates are low. There are increasing studies to reveal genetic basis for clinical response. Data generated by these studies, however, not easily available to the cancer immunotherapy community. We developed Mela-Immu, (http://bioinfo.vanderbilt.edu:3838/Mela-Immu/), which is an open-access resource for interactive exploration of a variety of known and novel genetic signatures predictive of immunotherapy response across multiple datasets in melanoma.

Methods: Mela-Immu collected all public available datasets which have both genomic profiling and ICB therapy outcome. Each dataset is stored at different genetic profiling, including genomics, transcriptomics and cell subpopulations and is associated with three types of clinical outcome, immunotherapy response, overall survival and progression free survival.

Results: The current version of the Mela-Immu (http://bioinfo.vanderbilt.edu:3838/Mela-Immu/) contains 889 samples derived from both bulk sequencing and single cell RNA sequencing. The association between each or integrative genetic profiling and clinical outcome can be queried, downloaded and visualized across multiple datasets to explore the consistency of signatures. Each genetic profiling provides multiple types of features, including the known or novel signatures predictive of immunotherapy response. Genomic profiling supports tumour mutational burden, and gene mutation, mutational signatures. Transcriptomic profiling includes gene/gene sets expression, gene expression relations, and immune cell components. The single cell RNA-seq not only explore specific gene expression signals but also specific cell populations between responders and nonresponders. To be noted, Mela-Immu allows users to upload their own datasets, therefore they can be analysed independently or co-analyzed with Mela-Immu’s existing datasets to validate the reliability and generality of identified signatures.

Conclusions: Mela-Immu is a high-quality portal to explore the relationship between multidimensional genetics data and ICB therapy response. This platform reduces barriers between researchers and complex genomic data, facilitating discoveries in immune-mediated therapy of melanoma.
Can precision medicine actually help people like me? African American and Hispanic perspectives on the benefits and barriers of precision medicine

Vivian M. Yeh, Erin M. Bergner, Marino A. Bruce, Sunil Kripalani, Victoria B. Mitrani, Titilola A. Ogunsola, Consuelo H. Wilkins & Derek M. Griffith

Aims: To better understand African American and Hispanic perspectives on the potential benefits of precision medicine, along with the potential barriers that may prevent precision medicine from being equally beneficial to all. We also sought to identify if there were differences between African American and Hispanic perspectives.

Methods: Six semi-structured focus groups were conducted to identify benefits and barriers to precision medicine. Three groups occurred in Nashville, TN with African American participants and three groups occurred in Miami, FL with Hispanic participants. All focus groups were conducted at community-based and university sites convenient to community partners and participants.

Results: A total of 55 individuals participated (27 in Nashville, 28 in Miami). The majority of participants were women (76.5%) and the mean age of participants was 56.2 years old. Both African Americans and Hispanics believed precision medicine has the potential to improve medicine and health outcomes by individualizing care and decreasing medical uncertainty. However, both groups were concerned that inadequacies in healthcare institutions and socioeconomic barriers would prevent their communities from receiving the full benefits of precision medicine. African Americans were also concerned that the genetic and non-genetic personal information revealed through precision medicine would make African Americans further vulnerable to provider racism and discrimination in and outside of healthcare.

Conclusions: While these groups believed precision medicine might yield benefits for health outcomes, they are also skeptical about whether African Americans and Hispanics would actually benefit from precision medicine given current structural limitations and disparities in health care access and quality.