**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Reference Number** | NT253 |
| **Submission Date** | 8/31/2017 |
| **Project Title** | Evaluating global and local ancestry across PheWAS phenotypes in eMERGE |
| **Tentative Lead Investigator (first author)** | Jacob Keaton |
| **Tentative Senior Authors (last author)** | Digna R Velez Edwards |
| **All other authors** | Todd L Edwards, Lea Davis, Jacklyn Hellwege, Dan Roden, Joshua Denny and “The eMERGE Network” plus ***any additional eMERGE authors interested in participating*** |
| **Sites Involved** | A network-wide study (all sites invited to participate). |
| **Background / Significance** | Current approaches to precision medicine focus on a patient’s clinical history and are often combined with known genetic risk factors. Multiple diseases have shown strong prevalence differences across racial groups and as a result race has been incorporated into clinical practice as a factor to consider in personalizing a treatment plan. However, multiple studies have shown that administratively determined race or self-reported races are imprecise estimates of an individual’s actual genetic ancestry. Furthermore, recent work by several groups has shown that for some diseases genetic ancestry (i.e., global ancestry) may directly interact with a patient’s clinical characteristics to modify disease risk and that this interaction varies at specific points in their genome (i.e., local ancestry). Within this project we leverage the rich phenotypic information available from large electronic health record (EHR) biobanks to comprehensively evaluate the relationship between disease risk and genetic ancestry and identify specific clinical characteristics of patients that interact with global and local genetic ancestry. Additionally, we will use 1000 genomes to not only map disease risk to specific genetic ancestries but use this information to better understand geographic origins of disease. We will accomplish this through the following analyses:  **1. Identify diseases associated with differences in global ancestry proportions** Using a data from a large EHR biobank, we will conduct global ancestry analyses using diseases with existing PheWAS codes. Ancestry will be determined for all subjects (regardless of reported race) with existing genome-wide association study (GWAS) using ADMIXTURE software, and proportions of ancestry specific to each reference population from 1000 Genomes will be determined. These diseases will then be evaluated for association with proportions of genetic ancestry for each1000 genomes reference population. Models will be adjusted for available disease specific candidate covariates. *We hypothesize that racial disparities in disease prevalence are driven by differences in specific genetic ancestries.*  **2. Identify loci that show evidence of global and/or local ancestry associating with disease risk interacting with patient clinical characteristics using eMERGE.** We will conduct an ancestry by clinical characteristic (AncxCC) interaction analysis evaluating interactions with body mass index (BMI), sex, and smoking status (if available) , and limiting to diseases that demonstrate ancestral differences in Aim 1. We will evaluate multivariable regression analyses testing for AncxCC interactions, and assess interaction effects with likelihood ratio tests. This will identify diseases where global ancestry interacts with clinical characteristics to modify risk for disease. For those models with evidence of an interaction we then run admixture mapping AncxCC analyses using subjects with existing GWAS data to identify genomic regions where local ancestry interacts with clinical characteristics, using RFMIX. Finally, those genomic regions with strong evidence of AncxCC will be further evaluated with focused single SNP association analyses using available GWAS data. *We hypothesize that clinical characteristics of patients interact with both global and local genetic ancestry to modify risk for disease.* |
| **Outline of Project** | 1. Estimate global ancestry among all participants in eMERGE 3. 2. Test for global ancestry associations with PheWAS phenotypes. 3. Using only those diseases with evidence of global ancestry associating with disease risk, test for interactions between global ancestry and BMI, sex, and smoking status on risk for association with PheWAS phenotypes 4. Test for interactions between BMI, sex, and smoking status and local ancestry on risk for individual PheWAS phenotypes that show from significant associations from analysis #3 above 5. Conduct local association analyses of regions where local ancestry interacts with a given clinical characteristic 6. Manuscript preparation and submission. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | * BMI, age, sex, smoking status (if available), and race/ethnicity\* * PheWAS codes |
| **Desired data** | * All eMERGE imputed GWAS data from across all racial/ancestral groups * PheWAS codes |
| **Planned Statistical Analyses** | 1. Please refer to outline of project section |
| **Ethical considerations** | There are no additional risks involved. The data will be stored in Dr. Velez Edwards servers at Vanderbilt. No data will be shared. We will also abide by the EMERGE guidelines. |
| **Target Journal** | TBD |
| **Milestones\*\*** | Total Duration of the study: 1 year |