|  |
| --- |
| **External Collaborator Proposal** *for* **eMERGE Network Analysis**Project/Manuscript Concept Sheet |
| **Reference Number** | NT228 |
| **Submission Date** | 4/25/2017 |
| **Tentative Lead Investigator** *(first author with contact information and affiliation)* | Yogasudha Veturi, yveturi@geisinger.eduMolly Hall, mahall2@geisinger.edu |
| **Tentative Senior Author** *(last author)* | Marylyn Ritchie, mdritchie@geisinger.edu |
| **eMERGE Site Sponsor & Contact** | Geisinger Health System, Marylyn Ritchie |
| **Project Title** | Genome-Wide Association Studies for Lipid Traits in the eMERGE network |
| **All Other Authors**  |  |
| **Other eMERGE Sites Involved** | All emerge network |
| **Background / Significance** | Blood lipids are highly heritable risk factors and clinically relevant biomarkers for atherosclerosis and cardiovascular/coronary artery disease. Previous genome-wide association studies (Willer et al., Nature Genetics 2013) have identified approximately 157 genetic loci associated with blood lipid levels using populations predominantly of European ancestry from the Global Lipids Genetics Consortium (GLGC). Likewise, studies have also been conducted to identify genetic variants associated with height (Wood et al., Nature Genetics 2014), body mass index (Locke et al., Nature 2015), and waist-hip-ratio adjusted for BMI (Heid et al., Nature Genetics 2010) from The Genetic Investigation of Anthropometric Traits (GIANT) consortium. These consortia are now conducting a large meta-analysis to identify genetic variants that contribute to lipid *as well as* anthropometric traits. |
| **Outline of Project** | This project is one arm of what is expected to be the most comprehensive GWAS to be ever conducted (approximately 1-2 million people across multiple cohorts/ancestries to study anthropometric and lipid phenotypes). In this project, we will focus on four lipid traits for individuals across eMERGE sites. We will conduct quality control, genotype imputation and association analyses within the eMERGE network and the resulting summary statistics will be used in subsequent meta-analyses conducted by the GLGC/GIANT consortia. |
| **Desired Variables** *(essential for analysis**indicated by* ***\*****)* | We seek the following variables for our analyses:* Primary phenotypes\*: fasting lipid values, i.e. high-density lipoprotein, low-density lipoprotein, total cholesterol, triglycerides
* Confounder variables\*: age at measurement, year of birth, sex, and race/ethnicity
* Related disease status\*: coronary artery disease, type II diabetes

Related phenotypes: systolic/diastolic blood pressure, body mass index, waist to hip ratio, height, weight, and fasting glucose level,  |
| **Desired Data** | eMERGE-III HRC imputed data |
| **Planned Statistical Analyses** | * Ensure that genome-wide genotypes in the eMERGE 80,000 dataset are on the correct build (37/hg19) and strand (forward)
* Impute genotypes to haplotype panel from the Haplotype Reference Consortium (being performed by David Crosslin)
* Perform quality control (separately for a major continental ancestry or samples genotyped on different arrays)

Conduct primary association analyses (using linear regression on continuous outcome) on fasting lipid values (raw and after inverse normal transformations). Phenotypic modeling will be separated by race/ethnicity, sex, disease status (if phenotype is correlated with disease status) and adult/child status. We will use software rvTests <https://github.com/zhanxw/rvtests>, BGZIP, tabix <http://samtools.sourceforge.net/tabix.shtml>, Vcftools <https://github.com/vcftools/vcftools>, and PLINK 1.9 <https://www.cog-genomics.org/plink2> to generate the necessary association summary statistics (VCF metrics, allele frequency for each variant, single variant association test statistics including effect direction, and covariance matrix for each genetic region). |
| **Ethical Considerations** | None |
| **Available Funding or Resources** | eMERGE |
| **Milestones\*\*** | 1. Complete QC by mid-April
2. Complete association analyses by April 30 (pending receipt of phenotype data and genotype imputation on HRC by David Crosslin)
3. Write relevant portions of the manuscript by end of Spring 2017.
 |

***\*\**** *This section should include the timeline for completion of project, including: approval, project duration, first and second draft of the paper and submission.*