**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Reference Number** | NT254 |
| **Submission Date** | August 30, 2017 |
| **Project Title** | Association of Genes and Variants in the eMERGEseq Panel with LDL cholesterol and Triglyceride Levels |
| **Tentative Lead Investigator (first author)** | Xiao Fan |
| **Tentative Senior Author (last author)** | Iftikhar Kullo |
| **All other authors** | MS Safarova, K Ding, D Schaid, M de Andrade, Gail Jarvik, Elizabeth Karlson, AO Basile, MD Ritchie; other eMERGE investigators interested in the project. |
| **Sites Involved** | All sites |
| **Background and Significance** | * Hypercholesterolemia and hypertriglyceridemia increase risk of atherosclerosis. Understanding the genetic architecture of low-density lipoprotein cholesterol (LDL-C) and triglyceride (TG) levels will provide insights into lipoprotein metabolism and identify new drug targets. * eMERGEseq panel comprises 109 medically relevant genes and 1547 single nucleotide polymorphisms (SNPs). Several genes such as *LDLR, PCSK9, APOB*, *ANGPTL3, ANGPTL4, APOA5, APOC3, APOE, PLTP* and *PON1*, and 206 SNPs proposed by Mayo are reported to be associated with LDL-C and TG levels. * We will investigate whether rare variants and genomic regions (sets of variants) in the candidate genes listed above as well as the remaining genes and SNPs in the eMERGEseq panel influence LDL-C and TG levels. |
| **Outline of Project** | Aim I. Perform variant- and gene-level association with LDL-C and log transformed TG levels  Aim II. Replicate significant associations from Aim I in publically available datasets |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | Clinical variables:   * Median LDL-C and TG levels prior to the use of lipid-lowering medications (LLM) or corrected median LDL-C and TG levels if the patient is on LLM. * Age * Gender * Race * Ethnicity * Body mass index (BMI) closest to the date when LDL-C and TG were measured * Lipid-lowering medications (statin, niacin, fibrate) * Diabetes (>2 ICD-9 diagnosis codes of 250.xx or >2 abnormal laboratory results which is defined as fasting glucose level > 125 mg/dL or hemoglobin A1C level > 6.5%) |
| **Desired data** | All variants in 109 genes + SNPs genotyped  Phenotypes above |
| **Planned Statistical Analyses** | PCA-inferred ancestry  Sample relatedness based on identity by descent  Variant pruning based on linkage disequilibrium  Race adjusted test assuming an additive genetic model  Variant-level association using Wald Chi-Squared Test  Gene-level association using SNP-set (Sequence) Kernel Association Test (SKAT)  The association will be adjusted for age, sex, BMI, lipid-lowering medication, diabetes and principal components.  The analyses will be conducted using a combination of PLINK and R packages including SKAT R package. |
| **Ethical considerations** | None |
| **Target Journal** | * Circulation * PLOS Genetics |
| **Milestones** | September 2017: Proposal submission  October-November 2017: Association analyses in Mayo dataset  January 2018-May 2018 Association analysis in the entire eMERGEseq cohort  June- July 2018 Replication analyses  July-August 2018: Manuscript draft  August 2018: Manuscript submission |