**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | February 2, 2017 |
| **Reference Number** | NT212 |
| **Project Title** | A phenome-wide association study to discover pleiotropic effects of lipid metabolism genes (*LDLR, APOB, PCSK9*, and *LPA*). |
| **Tentative Lead Investigator (first author)** | Maya Safarova |
| **Tentative Senior Author (last author)** | Iftikhar Kullo |
| **All other authors** | M de Andrade, D Schaid, JC Denny, SJ Hebbring, MS Williams, GP Jarvik, MD Ritchie, KM Borthwick, EB Larson, A Scrol, L Bastarache, DR Crosslin, TA Manolio  *Representation from other eMERGE sites* |
| **Sites Involved** | All sites |
| **Background and Significance** | * Variants in several genes implicated in the lipid and lipoprotein metabolism influence susceptibility to atherosclerotic cardiovascular disease. Increasingly, however, it is being recognized that these genes may also have pleiotropic effects. Recent studies demonstrate that *LDLR, APOB, PCSK9*, and *LPA* influence glucose homeostasis, blood pressure, triglyceride-rich lipoprotein metabolism, suggesting that genetic variants in these genes are likely to have pleiotropic effects. An emerging drug class of medications targeting these atherogenic particles LDL-C and Lp(a) is either approved or about to be approved for clinical use, motivating an investigation of associations of *LDLR, APOB, PCSK9*, and *LPA* variations with diverse phenotypes in the electronic health record (EHR). * We propose an agnostic comprehensive scan of the phenome, to test the associations between *PCSK9-LDLR-APOB-LPA* and the entire array of phenotypes in the EHR. |
| **Outline of Project** | Aim I. Perform an agnostic scan of the phenome to identify variant and gene level pleiotropic effects of *PCSK9, LDLR, APOB* and *LPA.* (eI-III merged imputed genotype data)  Aim II. Replicate phenome-wide significant associations from Aim 1 (eIII sequenced data and additional genotyped datasets from Mayo, VU, Marshfield and other eMERGE sites) |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | * Genotyped/imputed variants in *LDLR, APOB, PCSK9*, and *LPA*; Genotyping platform, quality control metrics * Total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, lipoprotein(a) / data on lipid-lowering treatment * Baseline clinical characteristics (age, gender, race, ethnicity, current smoking, hypertension, diabetes, BMI, ASCVD subtypes) * Phenocodes |
| **Desired data** | * All genotyped and imputed data for *LDLR, APOB, PCSK9*, and *LPA* variants * Case-control status via application of PheWAS algorithm * Case-control status for FH using an eAlgorithm for familial hypercholesterolemia |
| **Planned Statistical Analyses** | All variants (common, low-frequency and rare) will be analyzed for associations using both single point and agglomerative tests.  Each phenotype meeting inclusion/exclusion criteria will be tested for association at the gene and variant levels using weighted adaptive sum of powered score test and logistic regression assuming an additive genetic model adjusted for age, sex, study site and principal components, respectively.  The quality control and data analyses will be conducted using a combination of PLINK and the R statistical package, including PheWAS R package. |
| **Ethical considerations** | None |
| **Target Journal** | * Circulation * PLOS Genetics |
| **Milestones** | |  |  |  |  |  | | --- | --- | --- | --- | --- | | **2017** | | | | | | **02-05** | **05-09** | **09-10** | **10-11** | **12** | | Obtain information from all sites | Data review and analyses | First draft of manuscript circulated | Second draft of manuscript circulated | Manuscript submission | |