**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | May 8, 2017 |
| **Project Title** | The Genetic Architecture of Severe and Familial Hypercholesterolemia |
| **Tentative Lead Investigator (first author)** | Xiao Fan. |
| **Tentative Senior Author (last author)** | Iftikhar Kullo (Mayo Clinic) |
| **All other authors** | Maya Safarova, Keyue Ding, Mariza de Andrade, Daniel Schaid, Marc Williams, Marylyn Ritchie, Wei-Qi Wei, Josh Denny, Beth Karlson, other investigators from eMERGE sites. |
| **Sites Involved** | All eMERGE sites |
| **Background / Significance** | Severe hypercholesterolemia (SH) defined as low density lipoprotein- cholesterol (LDL-C) levels >190 mg/dL is relatively prevalent in the general population (5-7%). A subset of individuals with SH have Familial Hypercholesterolemia (FH) due to mutations in *LDLR*, *APOB*, or *PCSK9*. The genetic basis of SH in those without an identifiable mutation in the above genes is not known. FH is clinically diagnosed based on the scoring criteria devised by Dutch Lipid Clinic Network (DLCN). FH has variable penetrance and expression possibly due to modifier genes. Also ~50% patients with FH do not develop CHD in spite of high LDL-C. Hence more rigorous analytical efforts are required to identify such modifier variants. The proposed study attempts to 1) discover genetic factors that are associated with SH and FH; 2) assess the distribution of an LDL-C genetic score in SH and FH and identify those with a polygenic basis. |
| **Outline of Project** | We will identify cases and controls (1:4) for FH and SH using an electronic phenotyping algorithm for adult FH (Mayo) on PheKB. We will identify age and sex matched controls for both phenotypes. Maximum LDL-C in the EMR will be ascertained. Secondary causes of hypercholesterolemia will be excluded. We will impute pre-treatment LDL-C levels for patients on cholesterol-lowering medications at baseline.  With the estimated SH prevalence of 5-7%, the sample size is expected to be ~4000 from the eMERGE cohort of ~75,000 adults (assuming lipid panel being available in 90%). We expect an approximate sample size of 285 for FH assuming a prevalence of 0.3%. Replication will be performed in additional local datasets including VDB at Mayo, BioVue at VU, Geisinger cohort and potentially the eMERGE Seq cohort. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | Additional Variables:  Age  Gender  Ancestry  BMI  LDL-C levels  Lipid-lowering medications  CHD cases control status, age of onset of CHD  Family history of CHD if available  Family history of hypercholesterolemia if available |
| **Desired data** | Pre-imputed phased genotypes, and imputed merged GWAS genotype data from eMERGE 1, 2, and 3 and desired additional variables. |
| **Planned Statistical Analyses** | The association between each variant and SH/FH traits will be evaluated by logistic regression with adjustment for appropriate covariates including age, sex, ancestry. |
| **Ethical considerations** | None noted |
| **Target Journal** | Circulation |
| **Milestones\*\*** | May 2017: Proposal submission  July-August 2017: Association analyses in discovery set  Sept 2017 Replication analyses  October 2017: First manuscript draft  Dec 2017: Manuscript submission |

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