|  |  |
| --- | --- |
| **eMERGE Network: Proposal for Analysis**  Project/Manuscript Concept Sheet | |
| **Reference Number** | NT292 |
| **Submission Date** | 6/5/2018 |
| **Project Title** | Genome-wide association study of carotid artery atherosclerosis disease in eMERGE |
| **Tentative Lead Investigator** *(first author)* | Melody R. Palmer |
| **Tentative Senior Author**  *(last author)* | Gail P. Jarvik |
| **All Other Authors** | Daniel Seung Kim, David Crosslin, Ian Stanaway, Eric Larson, Elisabeth Rosenthal, David Carrell, Iftikhar Kullo, other interested eMERGE investigators |
| **Sites Involved (Data being used)** | Geisinger  Harvard  KP/UW  Marshfield  Mayo  Mt. Sinai  Northwestern  Vanderbilt |
| **Background / Significance** | The prevalence in of symptomatic carotid artery disease (CAAD) in the United States is estimated at 2-9%, with 5-9% prevalence in patients over 65 years old and is a major risk factor for stroke. We aim to discover genetic variants associated with CAAD in adults in the electronic Health Records and Genomics (eMERGE) Network. We will perform a genome-wide association study with genotypes imputed to the Haplotype Reference Consortium.  This project was initiated in eMERGE 2, but was underpowered. This new concept sheet reflects the addition of more sites, which have already applied the phenotype algorithm, and the use of the new merged imputation data. |
| **Outline of Project** | * Application of phenotype algorithm * Merged eMERGE 1-3 imputation to HRC * Perform GWAS analysis * Write manuscript |
| **Desired Variables**  *(essential for analysis*  *indicated by* ***\*****)* | Patient-level data represented as constants or computed flag variables:   * Patient demographics * Vital status/follow up status * Smoking status * Familial hypercholesterolemia diagnoses * Type 1 diabetes mellitus (T1DM) diagnoses * Type 2 diabetes mellitus (T2DM) diagnoses * Coronary artery disease (CAD) diagnoses * Peripheral Arterial Disease (PAD) diagnoses * Endarterectomy (ENDART) procedures * Cardiac revascularization (CRV) procedures   Repeated measure data:   * BMI * Diabetes medications * Lipid lowering medications * Lipid laboratory results |
| **Desired Data** | Merged eMERGE 1-3 imputed data (available to us) |
| **Planned Statistical Analyses** | * Principal components analysis across all samples, and within the largest ancestry subsets * Logistic regressions of imputed SNPs with an additive genotype model in PLINK 1.9.   + With all samples   + Within ancestry subsets   + Adjusting for sex, age, ancestry principal components |
| **Ethical Considerations** | All data are de-identified, there are no ethical concerns. |
| **Target Journal** | American Journal of Human Genetics |
| **Milestones\*\*** | Done – Phenotype algorithm, data cleaning  Done – Genotype imputation  Done – Major analysis complete  June 2018 – Submit abstract to ASHG  August 2018 – Send manuscript to co-authors  October 2018 – Submit manuscript for publication |

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