|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT462 | |
| **Submission Date** | 10/25/2022 | |
| **Project Title** | GWAS analysis of peripheral artery disease in eMERGE | |
| **Tentative Lead Investigator** *(first author)* | Kristjan Norland | |
| **Tentative Lead Investigator Email Address** | Norland.Kristjan@mayo.edu | |
| **Tentative Senior Author**  *(last author)* | Iftikhar J. Kullo | |
| **All Other Authors** | Derek Klarin, Scott M Damrauer | |
| **Sites Participating** | All adult eMERGE sites | |
| **Background / Significance** | Peripheral artery disease (PAD) is a heritable disease with a high prevalence in the United States (4.5% in people older than 40). Relatively few studies have evaluated PAD genetics and the number of loci reaching genome-wide significance remains low. | |
| **Outline of Project** | We aim to 1) conduct a GWAS on PAD for eMERGE participants and 2) collaborate and share the GWAS results with the PADGEN consortium. The goal of the PADGEN consortium is to discover additional genetic susceptibility variants for PAD through creation of the largest PAD dataset for genomic discovery work to date. | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI | Common Variable Labs  Common Variable Meds  Geocoding 2015 ACS variables’  Other: Case/Control status on Phase I and Phase II phenotypes |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)* | |
| **Desired Genetic Data** | eMERGE I-III Merged set (HRC imputed, GWAS)  eMERGE PGx/PGRNseq data set  eMERGEseq data set (Phase III)  eMERGE Whole Genome sequencing data set  eMERGE Exome chip data set  eMERGE Whole Exome sequencing data set  Other (not listed above): | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list  No | |
| **Planned Statistical Analyses** | This is the outline for the imputed GWAS analysis we aim to do:   1. Main Analysis:    1. PAD Case-Control GWAS with imputed variants 2. Secondary Analyses:    1. Diabetes stratified PAD case-control GWAS (separate analyses in those with and without diabetes – either type 1 or type 2)    2. Smoking stratified PAD case-control GWAS (separate analyses in those with a history of ever smoking vs. those with a history of never smoking)   We will analyze separately participants of different ancestry, as a major focus of PADGEN will be on individuals of non-European ancestry.  We will use a broad EHR-based definition of PAD:   * Presence of 2 or more diagnosis codes for PAD * Presence of 1 instance of a code for Lower Extremity Revascularization   We will use the program regenie (<https://rgcgithub.github.io/regenie/>) for GWAS analysis. | |
| **Ethical Considerations** | N/A | |
| **Target Journal** |  | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | * Run GWAS, analyze results and share with PADGEN (November-December 2022) * Draft manuscript (March-April 2022) | |

**\*Common Variables available across all datasets:**

* Demographics: sex, year of birth, decade of birth, race, ethnicity
* Codes: (repeated values & age at event): ICD, CPT, Phecodes
* BMI: (repeated value & age at event) height, weight, BMI
* Labs: (lab name, repeated lab value & age at event) Serum total cholesterol, LDL, HDL, Triglycerides, Glucose fasting/non-fasting/unknown, & White Blood Cell count
* Medications: (medication name, repeated, & age at event) Cerivastatin sodium, Rosuvastatin, Simvastatin, Fluvastatin, Pravastatin, Lovastatin, Atorvastatin, & Pitavastatin
* Other: Case/Control status on Phase I and Phase II phenotype: only on GWAS dataset participants