|  |
| --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT433 |
| **Submission Date** | July 02,2020 |
| **Project Title** | Elucidating genetic architecture of varicose veins |
| **Tentative Lead Investigator** *(first author)* | Anurag Verma |
| **Tentative Senior Author** *(last author)* |  |
| **eMERGE Site Sponsor & Contact** | Anurag Verma, Scott Damrauer, Marylyn D. Ritchie, University of Pennsylvania |
| **All Other Authors**  |  |
| **Sites Participating** | TBD |
| **Background / Significance** | Varicose veins are a common condition that present as bulging of the veins of the lower extremities. Aside from cosmetic issues, they can be associated with a range of symptoms including heaviness, tiredness, and pain in the legs. The underlying pathophysiology predisposes individuals to chronic skin changes and ulceration over time. Limited genetic discovery has been published for varicose veins, largely based on UKBB analyses. |
| **Outline of Project** | We see to replicate genetic association identified through varicose veins (VV) GWAS performed in Million Veteran Program and meta-analyzed with UKBiobank. We will define case and control sample based on PheCodes and then run following analyses:1. GWAS by ancestry
2. PheWAS with significant GWAS hits
3. Polygenic risk scores
4. Post -GWAS characterization of significant SNPs
 |
| **Desired Data - Common Variables\*** *(Available from the CC)* | * Demographics
* ICD9/10 codes
* CPT codes
* Phecodes
* BMI
 | * Common Variable Labs
* Common Variable Meds
* Other: Case/Control status on Phase I and Phase II phenotypes
 |
| **Other Desired Data *(Available from participating sites)*** | *If it is possible to extract ICD code, CPT code and lab measure encounter date, it would be very useful for the analyses to identify maternal age for maternal health phenotypes.*  |
| **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** | Generalized Regression and mixed linear models adjusting for age and PCs for GWAS and PheWAS analyses. |
| **Ethical Considerations** | N/A |
| **Available Funding or Resources** |  |
| **Target Journal** | TBD |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Analysis to complete by August 2020 to then share with UNC. |

**\*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
* Other: Case/Control status on Phase I and Phase II phenotype: only on GWAS dataset participants