|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* |  | |
| **Submission Date** | 12/16/2019 | |
| **Project Title** | Trans-ethnic genome wide association study of sarcoidosis in eMERGE | |
| **Tentative Lead Investigator** *(first author)* | Todd Edwards | |
| **Tentative Senior Author**  *(last author)* | Digna Velez Edwards and Giorgio Sirugo | |
| **All Other Authors** | Todd Edwards, Giorgio Sirugo, Jacob Keaton, Gabrielle Hampton, Jacklyn Hellwege | |
| **Sites Participating** | Open to all sites  Current participants:  Vanderbilt | |
| **Background / Significance** | There is currently poor understanding of the genetic causes of sarcoidosis and only a small proportion of heritable risk has been explained. Sarcoidosis is a complex inflammatory disease of unknown origin with a known racial disparity. Sarcoidosis has an incidence of five in 100,000 for those of European ancestry and 39 in 100,000 for those of African ancestry in the US. Sarcoidosis also affects African Americans at a younger age than European ancestry individuals and has a known sex disparity, affecting women more frequently than men. It primarily affects adults aged 20 to 40. Sarcoidosis affects multiple organs but primarily the lungs and lymph nodes and results in abnormal masses and nodules (granulomas). Published GWAS and candidate gene studies have pointed to several genetic risk loci for sarcoidosis, such as *BTNL2 ANXA11*, a locus on 11q13.1 and several loci in the *HLA* region on 6p21 but existing studies have been plagued by sample-size limitations and European bias. | |
| **Outline of Project** | We will conduct a genome-wide association study (GWAS) of sarcoidosis and meta-analysis across BioVU, the eMERGE Network, and the Penn Biobank. GWAS summary statistics will be used to construct race-specific polygenic risk scores (PRS) for sarcoidosis. We will perform PheWAS of the sarcoidosis PRS stratified by race the eMERGE data and BioVU, adjusted for sex and principal components of ancestry. We will summarize the results of the GWAS, with secondary analysis of Phecode groupings, as well as network analysis of results. | |
| **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)*** |  | |
| **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** | PheWAS | |
| **Ethical Considerations** | None | |
| **Target Journal** | Depends on results, likely a genetics journal such as Human Molecular Genetics or a general journal such as elife, Nature Communications, or Scientific Reports | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Gather data from coordinating center: 2/2020  Conduct statistical analyses: 2-6/2020  Write manuscript: 6-8/2020  Circulate and submit manuscript: 9/2020 | |