|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT417 | |
| **Submission Date** | 1/18/2021 | |
| **Project Title** | GWAS analyses of common infections | |
| **Tentative Lead Investigator** *(first author)* | Lan Jiang | |
| **Lead Investigator Email** | lan.jiang@vumc.org | |
| **Tentative Senior Author**  *(last author)* | QiPing Feng | |
| **All Other Authors** | C Michael Stein, Cecilia Chung, Alyson Dickson, Wei-Qi Wei, Jonathan Mosley, Mingjian Shi | |
| **Sites Participating** | Current participants: Vanderbilt  Open to all sites | |
| **Background / Significance** | Infectious diseases are one of the most common causes of morbidity and mortality worldwide. Susceptibility to infection is highly heritable; however, there has been a limited attempt to identify the genetic determinants underlying common infectious disease. Previous association studies have been sought in small candidate gene studies, and there is few GWAS studies relevant to US patients. | |
| **Outline of Project** | We will leverage large EHR-based biobank and physician-confirmed diagnosis codes to identify infectious disease cases. Specifically, we identified cases for 12 infectious diseases using ICD9CM and ICD10CM diagnosis codes. The diseases included UTI, pneumonia, chronic sinus infection, childhood ear infection, yeast infection, hepatitis C, strep throat, shingles, cold sores, hepatitis B, mononucleosis, positive TB. We will select controls from individuals who didn’t have any ICD code for the candidate disease by matching year of birth, gender, year at 1st VUMC visit, and year at most recent VUMC visit. We will conduct GWAS to identify genetic variants associated with altered risk of infectious disease. We will also test genetically predicted gene expression and imputed HLA alleles. | |
| **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** | GWAS, gene expression analyses, | |
| **Ethical Considerations** | None | |
| **Target Journal** | Depends on results | |
| **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: 2-4/2021  Conduct statistical analyses: 5-8/2021  Write manuscript: 9-10/2021  Circulate and submit manuscript: 11-12/2021 | |