|  |  |
| --- | --- |
| **External Collaborator Proposal** *for* **eMERGE Network Analysis**  Project/Manuscript Concept Sheet | |
| **Reference Number** | NT282 |
| **Submission Date** | 4/23/2018 |
| **Tentative Lead Investigator** *(first author with contact information and affiliation)* | Yogasudha Veturi, post-doctoral fellow, Department of Genetics, University of Pennsylvania [yveturi@pennmedicine.upenn.edu](mailto:yveturi@pennmedicine.upenn.edu) |
| **Tentative Senior Author**  *(last author)* | Marylyn Ritchie, [marylyn@pennmedicine.upenn.edu](mailto:marylyn@pennmedicine.upenn.edu) |
| **eMERGE Site Sponsor & Contact** | University of Pennsylvania, Marylyn Ritchie |
| **Project Title** | Identifying pleiotropic effects in cardiovascular and neurological diseases using EHR data |
| **All Other Authors** |  |
| **Other eMERGE Sites Involved** | All emerge network |
| **Background / Significance** | There is rich evidence for a strong relationship between cardiovascular and neuropathological diseases. However, the genetic basis of any such association has not been investigated yet, particularly with reference to identifying any instances of pleiotropy between these disease categories using EHR-derived phenotypes; i.e. both ICD9 codes and clinical lab measures. |
| **Outline of Project** | * In this study, we propose to: * (1) identify pleiotropic genes as well as new biomarkers associated with cardiovascular and neurological disease risk and   (2) study interactions of genetics and risk factors/environment (e.g. lifestyle, sex and age) on the considered EHR-derived phenotypes |
| **Desired Variables**  *(essential for analysis*  *indicated by* ***\*****)* | Primary ID9 codes:   * Inflammatory diseases of the central nervous system (320-327) * Hereditary and degenerative diseases of the central nervous system (330-337) * Pain (338-338) * Headache syndromes (339-339) * Disorders of the central nervous system (340-349) * Disorders of the peripheral nervous system (350-359) * Acute Rheumatic Fever (390-392) * Chronic rheumatic heart disease (393-398) * Hypertensive disease (401-405) * Ischemic heart disease (410-414) * Diseases of pulmonary circulation (415-417) * Other forms of heart disease (420-429) * Cerebrovascular disease (430-438) * Diseases of arteries, arterioles and capillaries (440-449) * Other diseases of circulatory system (451-459) * Confounding variables: age, sex, and race/ethnicity, adult/child status * Related phenotypes: systolic/diastolic blood pressure, blood lipid levels, (serum cholesterol levels), serum urate levels, body-mass-index, smoking status, drinking status, medications taken |
| **Desired Data** | eMERGE-III HRC imputed data |
| **Planned Statistical Analyses** | Conduct primary association analyses (using logistic regression) on neurological and cardiovascular ICD-9 codes. Phenotypic modeling will be separated by race/ethnicity, sex, and adult/child status. We will also apply Bayesian tests for colocalization between traits and Network Mendelian Randomization to understand the underlying causal mechanisms between pairs of diseases and genetic markers. Finally, we will apply network-based algorithms such as homogeneous conditional Gaussian regression (HGCR) to determine *network level* associations between diseases and genetic variants. |
| **Ethical Considerations** | None |
| **Available Funding or Resources** | eMERGE |
| **Milestones\*\*** | 1. Complete QC by May 2. Complete network-based analyses by June 2018. |

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