|  |  |  |
| --- | --- | --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT419 | |
| **Submission Date** | 3/16/2021 | |
| **Project Title** | Exploring pleiotropy using transcriptome-based analysis approaches | |
| **Tentative Lead Investigator** *(first author)* | Binglan Li | |
| **Tentative Lead Investigator Email Address** | [binglan@stanford.edu](mailto:binglan@stanford.edu) | |
| **Tentative Senior Author**  *(last author)* | Marylyn Ritchie, marylyn@pennmedicine.upenn.edu | |
| **eMERGE Site Sponsor & Contact** | Marylyn Ritchie – University of Pennsylvania | |
| **All Other Authors** | Milton Pividori, Casey Greene, Teri Klein, Dan Rader, Scott Damrauer, Anurag Verma | |
| **Sites Participating** | We would welcome participation from all eMERGE sites. | |
| **Background / Significance** | Pleiotropy is the phenomenon where a single gene influences two or more biological traits. In many previous studies, pleiotropy has been investigated at the level of SNPs through GWAS and PheWAS approaches. Recently, the use of transcriptome-wide association studies has demonstrated the utility of gene-based TWAS approaches for the exploration of pleiotropy (Veturi et al. in press, Pividori et al. 2020 <https://advances.sciencemag.org/content/6/37/eaba2083>). Phenome-wide TWAS has been conducted in the UK Biobank and those results are publicly available. We have also conducted a Phenome-wide TWAS in the Penn Medicine BioBank, specifically in analyses stratified by ancestry (European ancestry and African ancestry). We would like to integrate the Phenome-wide TWAS in UK Biobank and Penn Medicine BioBank with a Phenome-wide TWAS in eMERGE. | |
| **Outline of Project** | We will investigate the ancestry-specific, as well as cross-ancestry, gene-disease associations by applying phenome-wide TWAS and other statistical genomic analytic tools on the eMERGE data. The gene-disease associations discovered from the eMERGE data will be validated against other biobanks and publicly available genetic association results (TWAS/GWAS/colocalization, etc.). | |
| **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)*** | *We have the data needed to complete the project* | |
| **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** | We will perform phenome-wide TWAS analyses on geographically stratified populations (African ancestry and European Ancestry), independently; and will perform meta-analyses on ancestry-specific TWAS to obtain cross-ancestry gene-disease associations. For ancestry-specific analyses, we will perform both tissue-specific and cross-tissue TWAS using ancestry-specific GWAS summary statistics to obtain as comprehensive gene-disease associations as possible. If time and computing resources permit, we will perform the same TWAS analyses using the individual-level genotype data as a validation set. We will meta-analyze ancestry-specific GWAS summary statistics so as to run cross-ancestry TWAS (tissue-specific and cross-tissue) as what will be done for the ancestry-specific analyses. To support the discoveries of gene-disease associations, we will perform colocalization analyses for the purposes of pinpointing genes and/or tissues of effect, as well as TWAS fine-mapping analyses for the purposes of identifying potential causal genes in linkage disequilibrium with statistically significant tag genes. | |
| **Ethical Considerations** | Limited to none. Only de-identified, association analyses will be performed for this study. | |
| **Available Funding or Resources** | Dr. Ritchie has an R01 and start-up funds to pay for the analyses. | |
| **Target Journal** | American Journal of Human Genetics | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | * Complete analyses – March/April 2020 * Draft manuscript prepared for network members – May/June 2020 * Submit manuscript to journal – June 2020 | |

**\*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