|  |  |  |
| --- | --- | --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT428 | |
| **Submission Date** | August 2,2021 | |
| **Project Title** | Genetic burden of cardio-metabolic diseases on women’s health  outcomes from a study of multi-ancestry population | |
| **Tentative Lead Investigator** *(first author)* | Shefali Setia Verma | |
| **Tentative Senior Author**  *(last author)* |  | |
| **eMERGE Site Sponsor & Contact** | Shefali Setia Verma, Brenda Xiao, Marylyn D. Ritchie, University of Pennsylvania  Digna Velez Edwards, Vanderbilt University | |
| **All Other Authors** | TBD | |
| **Sites Participating** | TBD | |
| **Background / Significance** | Cardio-metabolic diseases are generally comorbid with other conditions and are associated with poor health outcomes. However, the investigation of the genetic burden of cardio-metabolic diseases with women’s health phenotypes such as ovarian cancer and many pregnancy-related complications is highly understudied in large populations. Polygenic risk scores (PRS) can be used to characterize shared genetic effects that could estimate disease risk, and the Penn Medicine Biobank (PMBB) is an electronic health record-based academic biobank consisting of genomic data on many women from multiple ancestries. To analyze the shared genetic effect of PRS from common cardio-metabolic diseases on phenotypes of women’s health, we calculated PRS for women in PMBB using PRS-CS with an external trans-ancestry LD reference panel and the largest trans-ancestry summary statistics available for seven phenotypes (body mass index (BMI), coronary artery disease, chronic kidney disease, myocardial infarction, type 2 diabetes (T2D), hypertension, and lipid measurements such as high-density lipoprotein (HDL) and triglycerides (TG)). We then tested the association of PRS with clinical lab measurements and case/control phenotypes for all participants and stratified by European and African ancestry. Our initial analysis has identified >30 significant associations reflecting shared biology among common cardio-metabolic phenotypes and diseases such as ovarian cancer, polycystic ovarian syndrome (PCOS), and pregnancy related complications such as postpartum hemorrhage. We would like to replicate and validate these findings in eMERGE dataset. | |
| **Outline of Project** | Phenotypes will be identified based on ICD and CPT codes  We will define case and control sample based on ICD codes, CPT codes and clinical laboratory measures data and then run following analyses:   1. PRS for cardiometabolic diseases 2. PRS PheWAS with women’s health phenotypes 3. Mendelian Randomization and LD score regression on significant PheWAS 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)*** | *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** | Polygenic risk scores using PRS-CS.  Medelian Randomization  LDscore regression | |
| **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 2021.  Estimated draft completion date – Oct 15, 2021  Projected manuscript submission date – November 2021 | |

**\*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