|  |  |  |
| --- | --- | --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT392 | |
| **Submission Date** | May 21, 2020 | |
| **Project Title** | Elucidating genetic architecture of pregnancy loss and fetal growth | |
| **Tentative Lead Investigator** *(first author)* | Shefali Setia Verma | |
| **Tentative Senior Author**  *(last author)* |  | |
| **eMERGE Site Sponsor & Contact** | Shefali Setia Verma, Marylyn D. Ritchie, University of Pennsylvania | |
| **All Other Authors** | TBD | |
| **Sites Participating** | TBD | |
| **Background / Significance** | Couples and individuals with unexplained pregnancy loss, recurrent pregnancy loss and fetal anomalies experience emotional trauma similar to a stillbirth or neonatal death, and the trauma is confounded by the fact that the etiology of their loss may not be determined, and there are only a few evidence-based diagnostic and treatment strategies available. Approximately 50% of pregnancy losses are caused by chromosomal abnormalities, such as aneuploidy. The remainder have an apparent euploid karyotype, but it is highly plausible that there are cases of pregnancy loss with other genetic aberrations that are not currently routinely detected. *Identification of genetic variations that are causative of, or predisposing to, pregnancy loss will substantially improve patient care.* | |
| **Outline of Project** | Phenotypes will be identified based on ICD and CPT codes for pregnancy loss, recurrent pregnancy loss, intrauterine fetal demise, still birth and fetal growth restriction. We will use clinical comorbidities such as preeclampsia, gestational diabetes, prolactin levels, hemoglobin A1c to dissect the genetic and exposure components for pregnancy loss.  We will define case and control sample based on ICD codes, CPT codes and clinical laboratory measures data and then run following analyses.   1. GWAS 2. PheWAS with significant GWAS hits 3. Polygenic risk scores 4. GXG analyses 5. Post -GWAS characterization of significant SNPs   We will be conducting similar analyses in Penn Medicine Biobank and UKBiobank for replications and meta-analyses. | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | X Demographics  X ICD9/10 codes  X CPT codes   * Phecodes   X BMI | * Common Variable Labs   X 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** | X 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   X No | |
| **Planned Statistical Analyses** | Generalized Regression and mixed linear models adjusting for age and PCs for GWAS and PheWAS analyses.  Polygenic risk scores using LDpred, PRSice, PLINK and PRS-CS.  GXG analyses using ridge regression approach.  LDscore regression and mt-COJO to identify correlation with other diseases. | |
| **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 April 2020 to then share with UNC. | |

**\*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