|  |  |  |
| --- | --- | --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT429 | |
| **Submission Date** | August 2,2021 | |
| **Project Title** | Endometriosis genetic analyses | |
| **Tentative Lead Investigator** *(first author)* | Shefali Setia Verma | |
| **Tentative Senior Author**  *(last author)* | Shefali Setia Verma | |
| **eMERGE Site Sponsor & Contact** | Shefali Setia Verma, Yuki Bradford, Marylyn D. Ritchie, University of Pennsylvania  Digna Velez Edwards, Vanderbilt University | |
| **All Other Authors** | TBD | |
| **Sites Participating** | TBD | |
| **Background / Significance** | Endometriosis is a poorly understood, under diagnosed and extremely deliberating gynecological problem that affects more than 200,000 women in the US every year. Women with endometriosis are at elevated risk for serious and important adverse maternal and neonatal outcomes during reproductive years and at higher risk of developing cardiovascular diseases such as myocardial infarction later in life.  Genome-wide association studies (GWAS) studies have identified common variants in certain genes in European populations, but they explain a very small proportion of variance. Variants identified from GWAS lie in intergenic regions, so functional fine mapping of these variants is essential for understanding the influence of these genes on endometriosis. In these analyses, we plan to perform GWAS of endometriosis stratified by ancestry. We will investigate the GWAS hots by fine mapping approaches. We will also validate PRS for endometriosis and run downstream analyses from PRS to identify comorbidities influenced by endometriosis. | |
| **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. GWAS of endometriosis 2. Fine Mapping of GWAS hits 3. Polygenic Risk score 4. PRS PheWAS with phenotypes obtained form EHR 5. 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 October 2021.  Estimated draft completion date – Jan 1, 2022  Projected manuscript submission date - Jan 31, 2022 | |

**\*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