|  |
| --- |
| **eMERGE Network: Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT299 |
| **Submission Date** | 7/31/2018 |
| **Project Title** | Genetic Loci for Polycystic Ovary Syndrome (PCOS) |
| **Tentative Lead Investigator** *(first author)* | Yanfei Zhang, Geisinger (yzhang1@geisinger.edu) Kevin Ho, Geisinger (kho@geisinger.edu) |
| **Tentative Senior Author** *(last author)* | Ming Ta Michael Lee (mlee2@geisinger.edu) Sarah Pendergrass (spendergrass@geisinger.edu) |
| **All Other Authors**  | Dustin Hartzel Nicole Restrepo Mariusz Butkiewicz |
| **Sites Participating** | All eMERGE sites |
| **Background / Significance** | • Polycystic ovary syndrome (PCOS), the most common endocrine & metabolic disorder affecting reproductive age women and the primary cause of infertility, has a prevalence of 4% to 15% based on 2 of 3 defining criteria: hyperandrogenism, oligo-ammenorrhea, polycystic ovaries. In spite of these criteria up to 50-70% of women with PCOS are undiagnosed.• Aside from its reproductive phenotypes, PCOS is clinically significant given its adverse metabolic abnormalities in lean and obese women compared with age and weight-matched non-PCOS women beginning in adolescence. Foremost are high prevalences of impaired insulin-mediated glucose disposal and an increased and premature risk of Type 2 diabetes in 50%-70% women with PCOS. In addition to other metabolic abnormalities, women with PCOS exhibit increased cardiovascular risk of coronary artery disease and stroke, altered cardiac structure, endothelial dysfunction, dyslipidemia, and metabolic syndrome. Obesity, which affects approximately half of women with PCOS, amplifies the degree of these cardiometabolic abnormalities. Animal models reproducing PCOS phenotypes include DHT-induced hyperandrogenemic rat models.• There is a strong genetic basis for PCOS as evidenced by the identification of SNPs in genes associated with phenotypic traits in candidate gene and family-based association studies. GWAS since 2011 have identified at minimum 16 reproducible candidate susceptibility loci involving pathways for insulin signal transduction, carbohydrate metabolism, steroidogenesis, gonadotropin action, and transcriptional regulation including DENND1A, THADA, RAB5B, LHCGR, FSHR. A recent large meta-analysis of 10,074 cases identified additional loci (PLGRKT, ZBTB16, MAPRE1) and replicated 11 reported loci.• We performed GWAS of women with PCOS in the Geisinger EHR, identified by our phenotypic algorithm, using Geisinger MyCode Community Health Initiative genetic data and both matched and unmatched controls. We plan to validate our results in eMERGE dataset using similar GWAS approach. |
| **Outline of Project** | 1. We have created a validated algorithm based on the Geisinger EHR for identifying PCOS cases and controls. We would like to validate our PCOS algorithm by analyzing the eMERGE data set.2. Following the identification of PCOS cases and controls, we would like to compare and potentially validate our GWAS results from our Geisinger cohort using the eMERGE data set.3. We will perform metaanalysis of PCOS GWAS results from Geisinger and eMERGE data sets. |
| **Desired Data - Common Variables\*** *(Available from the CC)* | [x] Demographics [ ] ICD9/10 codes[ ] CPT codes[ ] Phecodes[x] BMI | [ ] Common Variable Labs[ ] Common Variable Meds[x] Other: Case/Control status on Phase I and Phase II phenotypes |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)*  |
| **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[x] Other (not listed above): Genetically informed ancestry from principle components |
| **Does project pertain to an existing eMERGE Phenotype?** | [ ] Yes, if so please list [x] No |
| **Planned Statistical Analyses** | Quality control for the imputed data, including imputation score cutoff and minor allele frequency (MAF) cutoff > 1%. We will drop out from eMERGE any shared samples in the current MyCode Community Health Initiative array based data we are using at Geisinger.Logistic regression and analyses will be adjusted for age and the number of principle components to account for ancestry.Analyses will also be adjusted for age, BMI, and principle components to identify the shift in associations when incorporating BMI as a covariate.We will perform meta-analysis of GWAS results in Geisinger with GWAS results from eMERGE data. |
| **Ethical Considerations** | All data will be de-identified, and only summary data will be shared in resultant manuscripts |
| **Target Journal** | PLOS Medicine, PLOS Genetics, American (or European) Journal of Human Genetics |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 10/18 completion of validation analyses using eMERGE data. 11/18 Submission and internal review of data. 11/18 manuscript draft completion. 12/18-1/19 Final draft and submission of manuscript. |

**\*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