**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | 4/25/2017 |
| **Project Title** | PheWAS of Polycystic Ovary Syndrome (PCOS) GWAS loci in the electronic health records of 38,000 adult subjects in the eMERGE Consortium. |
| **Tentative Lead Investigator (first author)** | Yoonjung (Yoonie) Joo |
| **Tentative Senior Author (last author)** | M. Geoffrey Hayes |
| **All other authors** | Jennifer A. Pacheco, Rex L Chisholm, Theresa L. Walunas, Margrit Urbanek, Andrea E Dunaif |
| **Sites Involved** | Northwestern University |
| **Background / Significance** | Polycystic ovary syndrome (PCOS) is the most common reproductive disorder among women with unknown etiology. To date, four PCOS GWAS studies have been conducted in European (Hayes *et al*., *Nat Commun* 2015; Day *et al*., *Nat Commun* 2015) and Han Chinese ancestry cohorts (Shi *et al*.,*Nat Genet* 2012; Chen *et al*., *Nat Genet* 2011) in which 25 loci were found to be associated with PCOS. Using existing phenotype and genotype data of the eMERGE Network, we will perform a phenome-wide association study (PheWAS) to identify unique risk factors with high pleiotropy and/or comorbidity with PCOS. We expect that individual risk susceptible loci could have several pleiotropic effects on distinct disease processes and/or pathways, aiming to better understand the etiology of PCOS. |
| **Outline of Project** | 1. Obtain imputed genotype data of the eMERGE phases 1, 2  2. Collect phenotype (ICD9/10 codes), covariate data of these eMERGE subjects (already in-hand at the eMERGE CC).  3. Perform PheWAS with PCOS risk variants using the PheWAS R package (e.g. Denny *et al*., *Bioinformatics*, 2010)  5. Look for replicating association results between the PCOS GWAS and ICD-9 (and potentially ICD-10) diagnosis codes for PCOS.  6. Look for evidence of comorbidity and pleiotropy with other phenotypes. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | - ICD9/ICD10 diagnosis codes to generate PheWAS codes\* (already in-hand at the eMERGE CC)  - Principal Components for covariate\* |
| **Desired data** | - Phenotype (ICD9/ICD10 codes) data for all adult subjects from the eMERGE 1,2 sites  - Imputed genotype data for 25 variants of interest |
| **Planned Statistical Analyses** | PheWAS of 25 SNPs associated with PCOS |
| **Ethical considerations** | None to report |
| **Target Journal** | *JCEM* |
| **Milestones\*\*** | Data analysis completed by June 15, 2017  Paper drafted by July 31, 2017 |