|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT 347 | |
| **Submission Date** | 5/22/2019 | |
| **Project Title** | SIM1 Rare Variation Association with Erectile Dysfunction | |
| **Tentative Lead Investigator** *(first author)* | Melody Palmer\* and Ian Byrell Stanaway\*  \*contributed equally | |
| **Tentative Senior Author**  *(last author)* | Hunter Wessells and David Russell Crosslin | |
| **All Other Authors** | Gail P. Jarvik, Elisabeth Rosenthal,  David Carrell, Eric Larson, Lester Kirchner, Sarah Pendergrass  other interested eMERGE co-authors | |
| **Sites Participating** | All eMERGE Sites | |
| **Background / Significance** | We will query SIM1 variations from the electronic MEdical Record and GEnomics (eMERGE) Network imputed data GWAS participants [1] and perform a logistic regression burden test in the SIM1 candidate gene for association with Erectile Dysfunction. Recent previous studies have found common variants adjacent to the SIM1 gene in a regulatory region associated with erectile dysfunction in two independent analyses [2,3]. It is likely that rare variants (MAF < 0.01) in coding exonic and splice sequences will also be associated with erectile dysfunction in SIM1 and would localize this putative signal due to the rare variation in the gene. Many rare variants impute well (quality > 0.3) in the eMERGE set including in SIM1. We will replicate in the Geisinger biobank.  References   1. **Stanaway**, et al., 2019 "The eMERGE Genotype Set of 83,717 Subjects Imputed to ~40 Million Variants Genome Wide and Association with the Herpes Zoster Medical Record Phenotype.", Genetic Epidemiology, 2019 2. Jorgenson E, Matharu N, **Palmer MR**, Yin J, Shan J, Hoffmann TJ, Thai KK, Zhou X, Hotaling JM, **Jarvik GP**, Ahituv N, **Wessells H**, Van Den Eeden SK. Genetic variation in the SIM1 locus is associated with erectile dysfunction. Proc Natl Acad Sci U S A. 2018 Oct 23;115(43):11018-11023. 3. Bovijn J, Jackson L, Censin J, Chen CY, Laisk T, Laber S, Ferreira T, Pulit SL, Glastonbury CA, Smoller JW, Harrison JW, Ruth KS, Beaumont RN, Jones SE, Tyrrell J, Wood AR, Weedon MN, Mägi R, Neale B, Lindgren CM, Murray A, Holmes MV. GWAS Identifies Risk Locus for Erectile Dysfunction and Implicates Hypothalamic Neurobiology and Diabetes in Etiology. Am J Hum Genet. 2019 Jan 3;104(1):157-163. doi: 10.1016/j.ajhg.2018.11.004. Epub 2018 Dec 21. | |
| **Outline of Project** | 1. Select Variants from SIM1.  2. Select Phenotype data for inclusions and exclusions.  3. Perform logistic regressions of known common variants and aggregated rare variant burden. We will adjust for site, principal components, the previously associated common variants and assess the inclusion of diabetes and BMI as covariates.  4. Replicate results in Geisinger biobank  5. Write the paper. | |
| **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)*** | *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** | 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): eMERGE 1 660k CNV calls and the other genotype chips CNV calls if completed before the finish of this project. | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list  No | |
| **Planned Statistical Analyses** | Perform logistic regressions of known common variants and aggregated rare variant burden.  Stratify by rare variants to show direction of effect of each variant.  Assess inclusion of BMI and Diabetes as covariates. | |
| **Ethical Considerations** | None | |
| **Target Journal** | Nature Genetics | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 1. Approval ~06/2019 2. Begin writing paper and refine analyses based on co-author input ~07/2019 3. Submit paper ~10/2019 | |

**\*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