|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT326 | |
| **Submission Date** | January, 2019 | |
| **Project Title** | Psychiatric Manifestations of Variations in ACMG59 Genes | |
| **Tentative Lead Investigator** *(first author)* | Y-C Feng | |
| **Tentative Senior Author**  *(last author)* | J.W. Smoller | |
| **All Other Authors** | E. Karlson, S Weiss, S. Murphy | |
| **Sites Participating** | We invite each site to join | |
| **Background / Significance** | Implementation of return of genetic information for medically actionable genes (ACMG59) is becoming increasingly widespread. These genes are known to be pleiotropic (Kocarnik and Fullerton, 2014), and defining their phenotypic spectrum of expression has important implications for return of genetic information to both patients and research participants. To date, there has been no systematic attempt to examine whether psychiatric phenotypes are associated with sequence variation in this set of genes. Consistent with aim 2 of our eMERGE grant application, we would like to conduct analyses to examine pleiotropic effects on psychiatric phenotypes. We will use gene burden analysis and PheWAS (focused specifically on psychiatric phenotypes) in the eMERGE-Seq dataset (N=25,000) to examine the association and penetrance of rare variants in the ACMG genes with psychiatric disorders. We will also confirm penetrance by chart review of patients enrolled in the Partners HealthCare Biobank from the Harvard site (N=2500). | |
| **Outline of Project** | We will conduct a hypothesis driven set of tests to examine the extent to which rare variants in ACMG59 genes are associated with psychiatric phenotypes. The phenotypes of interest will include those psychiatric diagnoses defined through eMERGE algorithms (ie autism spectrum disorder, ADHD, bipolar disorder, depression, schizophrenia) as well as other psychiatric disorders defined by PheCodes. We will conduct gene-based burden tests prioritizing rare variants with a predicted deleterious effect (e.g., protein-truncating, missense variants with a CADD score > 20) at different minor allele frequency bins (MAF < 1%, 0.1%, or 0.01%). In addition, we will examine a group of ultra-rare genetic variants that are absent from a population allele frequency reference (e.g., Genome Aggregation Database) and thus are more likely to be pathogenic. We will consider a variety of rare variant association testing methods including CMC, SKAT, and Fisher’s exact test. In addition to gene-based tests, we will examine individual variant association by fitting a logistic regression that uses each variant and presence or absence of the EMR-algorithm phenotype or PheWAS code as the outcome adjusting for age, sex and the first 10 principal components. We will set the p-value cut-off for a significant association using the Bonferroni correction. For PheWAS codes significantly associated with a variant, we will review a random set of 50 charts at Partners HealthCare to determine the accuracy (PPV) of the code. This will allow us to assess penetrance.  We will define penetrance as the proportion of subjects with the variant shown to be associated with a disease, for whom they have evidence in the EMR for that disease according to 1) EMR-algorithm phenotype or, 2) PheWAS code verified by chart review. | |
| **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): | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list  ADHD, Autism, Bipolar disease, depression, and schizophrenia  No | |
| **Planned Statistical Analyses** | Logistic regression  Generalized linear mixed models  PheWAS  CMC  SKAT  RCH | |
| **Ethical Considerations** | This study makes use of data that is already obtained from electronic health records. The chart reviews will be performed in the Partners HealthCare Biobank subjects, a process that is IRB approved at Partners HealthCare. | |
| **Target Journal** | American Journal of Psychiatry  American Journal of Human Genetics | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | March | Apr - Jun | July- Aug | Sept-Oct | Nov - Dec | | Data assembly and analysis design |  |  |  |  |  | | Gene/burden testing and psychiatric phewas |  |  |  |  |  | | Chart reviews |  |  |  |  |  | | Manuscript writing |  |  |  |  |  | | Manuscript completion and submission |  |  |  |  |  | | |

**\*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