|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT404 | |
| **Submission Date** | 9/28/2020 | |
| **Project Title** | **Genome wide association study of lower urinary tract dysfunction.** | |
| **Tentative Lead Investigator** *(first author)* | Miguel Verbitsky | |
| **Tentative Lead Investigator Email Address** | mmv2003@cumc.columbia.edu | |
| **Tentative Senior Author**  *(last author)* | Ali Gharavi | |
| **All Other Authors** | Krzysztof Kiryluk, Atlas Kahn, Ning (Sunny) Shang,and “The eMERGE Network” plus ***any additional eMERGE authors interested in participating*** | |
| **Sites Participating** | A network-wide study (all sites invited to participate). | |
| **Background / Significance** | Lower urinary tract dysfunction (LUTD), is associated with a range of lower urinary tract symptoms (LUTS) and pelvic floor disorders, including urinary and bowel incontinence, interstitial cystitis, urinary obstruction, abnormal voiding frequency/urgency and dysuria, genital organ prolapse among others. LUTD is often associated with genetic or acquired anatomical defects, such as genital polyps or prolapse, or congenital kidney and urinary tract anomalies (CAKUT).  Under the working hypothesis that LUTD is genetically heterogeneous with contributions from common variants we propose to conduct genome wide association studies (GWAS) for LUTD. In support of this hypothesis, GWAS have been successful in revealing common risk variants for hypospadias, a common genitourinary malformation (Geller et al. Nat Genet. 2014; van der Zanden et al. Nat Genet. 2011) and we have recently performed a GWAS on vesicoureteral reflux (VUR), identifying genome-wide significant associations (Verbitsky et al. manuscript under peer review).  We propose to perform a LUTD GWAS, based on ICD-9/10 codes and genetically matched controls across all EMERGE SNP datasets with the following aims:  (1) to perform a case-control GWAS to identify variants associated with LUTD, (2) conduct additional analyses to identify sex-specific associations and (3) assess association of common SNPs with comorbidities such as urinary tract infections and VUR in LUTD.  The EMERGE network is ideally suited to study the genetics of LUTD. The most important considerations for execution of proposed studies include: (1) the availability of large cohorts with EHR data that can be used for accurate LUTD phenotyping, (2) availability of high-quality SNP chip imputed data for GWAS using standard methods, (3) availability of a large control population for rigorously matched case-control GWAS. | |
| **Outline of Project** | 1. Identification of all individuals with LUTD with available GWAS datasets network-wide. 2. Identification of genetically matched controls in eMERGE. 3. Combined of eMERGE GWAS with independent datasets available in our lab (replication and meta-analysis) 4. Manuscript preparation and submission. | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI | Common Variable Labs  Common Variable Meds  Geocoding 2015 ACS variables’  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  No | |
| **Planned Statistical Analyses** | 1. Ascertainment of LUTD phenotypes 2. Case-control sex and genetic matching 3. LUTD Case-control SNP GWAS using imputed genome-wide data available via eMERGE. 4. Meta-analysis with external datasets 5. Sex-specific and comorbidities association analyses 6. Meta-analysis and/or replication. 7. Derivation of a polygenic risk score for LUTD | |
| **Ethical Considerations** | There are no additional risks involved. The data will be stored at a secured location in the data storage system of the Gharavi lab. No data will be shared with unauthorized third parties. Patient identity will not be compromised by the proposed analysis. We will also abide by the EMERGE guidelines in this regard. | |
| **Target Journal** | TBD, depending on the findings | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Total Duration of the study: 3 years  Identification of LUTD cases and genetically matched controls: 2020  GWAS: 2020-2022  Draft of manuscript to authors: 2022  First submission: 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