**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | 6/27/2016 |
| **Project Title** | **Genomics of structural kidney and urinary tract defects.** |
| **Tentative Lead Investigator (first author)** | Miguel Verbitsky |
| **Tentative Senior Author (last author)** | Ali Gharavi |
| **All other authors** | Simone Sanna-Cherchi, Krzysztof Kiryluk, George Hripcsak, Chunhua Weng, Ning (Sunny) Shang, Wendy Chung, Yufeng Shen, Jun Zhang, Hila Rasoully, Maddalena Marasa, and “The eMERGE Network” plus ***any additional eMERGE authors interested in participating*** |
| **Sites Involved** | A network-wide study (all sites invited to participate). |
| **Background / Significance** | Congenital anomalies of the kidney and urinary tract (CAKUT) are among the most frequent organ malformations and constitute the leading causes of end-stage renal disease in the pediatric population. They include renal hypodysplasia (RHD), obstructive uropathy, posterior urethral valves, and vesicoureteral reflux (VUR), among others. We hypothesize that CAKUT is genetically heterogeneous with contributions from common and rare variants, that may be respectively identified by genome wide association (GWA) studies and analysis rare copy number variants (CNV). In support of this hypothesis, GWA studies have been successful in revealing common risk variants for hypospadias, a common genitourinanry malformation (Geller et al. Nat Genet. 2014; van der Zanden et al. Nat Genet. 2011). There is also strong evidence implicating genomic structural variants, which include deletions, duplications and more complex rearrangements, in the genetic architecture of CAKUT. In a recent study of pediatric all-cause chronic kidney disease patients, we have demonstrated that 7.7% of children across the CAKUT spectrum had an unsuspected pathogenic genomic imbalance, suggesting genomic disorders as a risk factor for CAKUT (Verbitsky et al. *JCI* 2015). Detection of pathogenic imbalances has practical implications for personalized diagnosis and health monitoring in these cases.  We propose to perform a genome-wide analyses of SNPs and CNVs using all CAKUT cases, which based on ICD-9 (589, 593.7, 593.70, 593.71, 593.72, 593.73, 753.0, 753.2, 753, 753.3, 753.4, 753.5, 753.6, 753.7, 753.8, 753.9, 753.20) and PheWAS (599.1, 751, 751.2, 751.3, 751.22) codes and age 44 years or younger; and genetically matched controls across all EMERGE SNP datasets with the following aims:  (1) to perform a case-control GWA study to identify common variants associated with CAKUT, (2) to estimate the global prevalence of known genomic disorders among EMERGE subjects with CAKUT, (3) to assess if individuals with CAKUT and CAKUT sub-phenotypes (e.g. VUR) have higher burden of large rare genomic rearrangements, (4) to identify specific recurrent CNVs associated with CAKUT or particular CAKUT sub-phenotypes.  The EMERGE network is ideally suited to study the genetic architecture of CAKUT. 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 CAKUT phenotyping, (2) availability of high quality SNP chip data for GWA and detection of CNV using standard methods, (3) a stringent analytic pipeline for selection of high likelihood variants (already established in our laboratory), and (4) availability of a very large control population to test frequency of rare variants (a control population of >20,000 individuals is already available to us). |
| **Outline of Project** | 1. Identification of all individuals with CAKUT with available GWAS datasets network-wide. 2. CNV burden analyses and Identification of pathogenic CNVs across all CAKUT EMERGE GWAS data 3. Identification of genetically matched controls and GWA study of CAKUT in eMERGE. 4. Combined of eMERGE GWA and CNV data with independent datasets previously produced in our lab (replication and meta-analysis) 5. Manuscript preparation and submission. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | * ICD-9/10 and PheWAS codes to ascertain CAKUT phenotype\* for all genotyped EMERGE participants. * Age, sex, race/ethnicity\* * Type of GWAS cohort ascertainment (if disease-driven, what disease)\* |
| **Desired data** | * CAKUT phenotypes, e.g. ICD-9 589, 593.7, 593.70, 593.71, 593.72, 593.73, 753.0, 753.2, 753, 753.3, 753.4, 753.5, 753.6, 753.7, 753.8, 753.9, 753.20. * Genome-wide SNP intensity data for GWA and CNV analysis from all EMERGE GWAS cases with a CAKUT phenotype |
| **Planned Statistical Analyses** | 1. Ascertainment of CAKUT phenotypes 2. CNV calls across all CAKUT cases in GWAS datasets using PennCNV, 3. Identification of pathogenic and likely pathogenic rare CNVs and global burden tests 4. Case-control SNP GWA study using imputed genome-wide data available via eMERGE |
| **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\*\*** | Total Duration of the study: 2 years  Identification of CAKUT cases and genetically matched controls: 2016  CNV calling and analysis: 2016-2017  GWAS: 2017-2018  Draft of manuscript to authors: 2018  First submission: 2018 |

**\*\*** This section should include: Timeline for completion of project, including approval, project duration, first and second draft of the paper and submission.