**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

This is project is part of the Phenotyping Workgroup’s list of Network phenotypes and the algorithm(s) is expected to be implementation-ready by 06/2016.

|  |  |
| --- | --- |
| **Submission Date** | 1/21/2016 |
| **Project Title** | **Detection of copy number variants (CNVs) and their kidney disease associations across the EMERGE network** |
| **Tentative Lead Investigator (first author)** | Miguel Verbitsky |
| **Tentative Senior Author (last author)** | Ali Gharavi, Krzysztof Kiryluk |
| **All other authors** | George Hripcsak, Chunhua Weng, Ning (Sunny) Shang, Wendy Chung, Yufeng Shen, Jun Zhang, Hila Rasoully, Maddalena Marasa, others 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** | There is strong evidence implicating genomic structural variants in the genetic architecture of chronic kidney disease (CKD). This is supported by the finding that chromosomal disorders often produce kidney developmental phenotypes and that gene deletions contribute to several known forms of kidney disease (e.g. *HNF1B* associated disorders). In a recent study of pediatric CKD patients, we have demonstrated that a substantial proportion of children with CKD (7.4% of the entire cohort)  had an unsuspected pathogenic genomic imbalance, suggesting genomic disorders as a risk factor for common forms of pediatric nephropathy (Verbitsky et al. *JCI* 2015). Detection of pathogenic imbalances has practical implications for personalized diagnosis and health monitoring in these cases. Our prior study has also demonstrated that up to 16% of individuals with congenital kidney malformations had a molecular diagnosis attributable to a copy-number disorder (Sanna-Chierchi et al. *AJHG* 2012). Strikingly, the majority of the known CNV disorders detected in this cohort had previous associations with developmental delay or neuropsychiatric diseases.  Similarly, recent studies indicate that chromosomal microarrays can identify rare genomic imbalances that can clarify the etiology of neurodevelopmental and cardiac disorders.  Structural variants include deletions, duplications, inversions, translocations, copy number variants (CNVs) or more complex rearrangements. Similar to single base variants, the vast majority of CNVs follow Mendelian inheritance and are benign, but many CNVs encompass genes and regulatory elements, and can influence gene expression and susceptibility to disease. Studies have also demonstrated that the *de novo* CNV mutation rate is 100-10,000 fold higher than the point mutation rate, suggesting that this type of variation can significantly contribute to the genesis of rare variants underlying sporadic disease. The contribution of genomic disorders to adult CKD has not been studied.    We propose to perform a genome-wide analysis of CNVs using all EMERGE SNP datasets with the following aims: (1) to estimate the global prevalence of known genomic disorders among EMERGE subjects, (2) to assess if individuals with CKD have higher burden of large rare genomic rearrangements, (3) to identify specific recurrent CNVs associated with CKD, (4) to explore pleiotropic effects of selected CNVs that are associated with CKD using network-wide PheWAS approach.  We feel that the EMERGE network is ideally suited to evaluate the potential impact of structural variants on the risk of CKD. 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 CKD phenotyping, (2) availability of high quality SNP chip data for detection of genomic structural variants 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 alleles (a control population of >50,000 individuals is already available to us). |
| **Outline of Project** | 1. Identification of CNVs genome-wide across all EMERGE GWAS datasets using PennCNV. 2. CKD phenotyping of all individuals with available GWAS datasets network-wide. 3. CNV burden and case-control association analyses. 4. Manuscript preparation and submission. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | * CKD phenotype\* for all genotyped EMERGE participants (the algorithm is being developed at Columbia as part of EMERGEIII) * Age, sex, race/ethnicity * Type of GWAS cohort ascertainment (if disease-driven, what disease)\* |
| **Desired data** | * Standardized CKD phenotypes as above * Genome-wide SNP intensity data for CNV analysis from all EMERGE GWAS cohorts * To perform PheWAS for CKD-associated CNVs we will need ICD9/ICD10 codes and all other standardized EMERGE phenotypes network-wide. |
| **Planned Statistical Analyses** | 1. Derivation of CKD phenotypes 2. CNV calls across all GWAS datasets using PennCNV 3. CNV case-control association analyses and global burden tests 4. PheWAS for CKD-associated CNVs to establish their pleiotropic effects |
| **Ethical considerations** | There are no additional risks involved. The EMR and genomic data will be stored at a secured location in the data storage system of Kiryluk and Gharavi labs. No data will be 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: 3 years  Implementation of CKD phenotyping: 2016  CNV calling and analysis: 2016-2017  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.