|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT469 | |
| **Submission Date** | 1/6/2023 | |
| **Project Title** | **GWAS for kidney function and related traits for global meta-analyses by the CKDGen Consortium** | |
| **Tentative Lead Investigator** | Atlas Khan | |
| **Tentative Lead Email Address** | ak4046@cumc.columbia.edu | |
| **Tentative Senior Author** | TBD, Krzysztof Kiryluk ([kk473@columbia.edu](mailto:kk473@columbia.edu)) will lead emerge analyses and contributions | |
| **All Other Authors** | eMERGE CKD workgroup members  CKDGen Consortium members  + any other interested eMERGE investigators | |
| **Sites Participating** | All other eMERGE sites/investigators are invited to participate | |
| **Background / Significance** | Chronic kidney disease (CKD) affects over 10% of the population and represents a significant expense to healthcare systems in the US and globally. This study will contribute eMERGE results to the global meta-analyses of CKD and kidney function-related traits across multiple GWAS cohorts led by the CKDGen consortium. By including summary statistics based on eMERGE datasets in these meta-analyses, we hope to increase the power for new discovery and ancestral diversity of GWAS studies for kidney-related traits. | |
| **Outline of Project** | Here we proposed to perform genome-wide association analysis (GWAS) for CKD, estimated glomerular filtration rate (eGFR), albuminuria (UACR), blood urea nitrogen, and uric acid levels in the eMERGE-III dataset. We will also plan to expand this study to include additional eMERGE-IV GWAS dataset as they become available. We will perform standardized GWAS analyses by continental ancestry with adjustments for age, sex, diabetes, imputation batch, and ancestry PCs. We will generate GWAS summary statistics and contribute the results to the next phase (5th round) of global meta-analyses as part of our collaboration with the CKDGen consortium. This will include multiple CKDGen cohorts, All-of-Us, UKBB, MVP, Biobank Japan, Columbia Biobank, BioMe, PMBB, and other biobanks/consortia that are willing to contribute their results. | |
| **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-III 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): future eMERGE-IV GWAS dataset | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list: CKD Phenotype (we plan a minor modification to include the CKD subtype information)  No | |
| **Planned Statistical Analyses** | 1. We will prepare a standardized phenotype and covariates for GWAS analysis using agreed upon definitions and CKDGen consortium-wide analysis plan. 2. We will perform GWAS analysis for the eMERGE-III dataset, including genotype quality control, ancestry analysis, and association analysis adjusting each trait for age, sex, diabetes, imputation batch, and significant PCs of ancestry. 3. We will perform meta-analysis of all participating studies followed by functional annotations for the significant loci. 4. We will perform meta-PheWAS analyses of significant loci across eMERGE and other biobanks to explore potential pleiotropic effects of genome-wide significant loci | |
| **Ethical Considerations** | N/A | |
| **Target Journal** | TBD based on the results of the final meta-analysis | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Implementation of phenotyping algorithms: March 2023  Completion of GWAS in the existing datasets: May 2023  Completion of consortium-wide meta-analyses: Sept 2023  Study completion and journal submission” End of 2023 | |

**\*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