|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT304 | |
| **Submission Date** | 9/24/2018 | |
| **Project Title** | Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in the eMERGE Network | |
| **Tentative Lead Investigator** *(first author)* | Ozan Dikilitas | |
| **Tentative Senior Author**  *(last author)* | Iftikhar Kullo | |
| **All Other Authors** | Daniel Schaid, Mariza de Andrade, Baosheng He, Authors from eMERGE Sites | |
| **Sites Participating** | All eMERGE Sites | |
| **Background / Significance** | Coronary heart disease (CHD) is a common genetically complex disease influenced by multiple genetic variants with an estimated heritability of 40-60%. Over the past decade, genome-wide association studies (GWAS) have revealed numerous loci in the genome contributing to the genetic susceptibility for CHD. There is great interest in leveraging these findings to construct genetic risk scores (GRS) to refine risk stratification. GRSs that only utilize variants reaching genome-wide significance may miss the proportion of heritability unexplained by such variants, and may be affected by inaccurate effect size estimates of individual variants from previous studies due to limitations on statistical power. To date, the largest meta-analyses of GWAS for CHD were limited to participants of European ancestry, hence limiting generalization of these findings to individuals of different ancestries. Recently, genome-wide polygenic scores (GPS) (Khera et al., Inouye et al.) have been reported to outperform candidate loci polygenic scores (CPS). There is a need for validation and direct comparison of the predictive utility of different types of GRSs across multiple ancestries in an independent dataset. We will assess strengths of association of different GRSs with incident and prevalent CHD in patients from European and African American ancestries. | |
| **Outline of Project** | We will assess the predictive utility of GRSs by;   * Computing the GPSs (Khera et al., Inouye et al.) as well as two previously reported candidate variant based GRSs (Tikkanen et al., Tada et al.) in adult eMERGE participants * Ascertain incident and prevalent CHD status using electronic phenotyping algorithms, and extract relevant covariates such as age, sex, body-mass index, smoking status, diabetes mellitus, elevated cholesterol, and family history of CHD for the participants at baseline. * Assess GRSs by various metrics such as;   + Hazard ratios for incident CHD   + Odds ratios for prevalent CHD   among different ancestries (European vs African American) in the eMERGE Cohort (prevalent vs. incident status will be determined relative to eMERGE enrollment phase). | |
| **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)*** | * *Family history of atherosclerotic cardiovascular disease (ASCVD)* * *Smoking status* * *Blood pressure measurements* | |
| **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** | After computing individual genetic risk scores, we will proceed with the statistical analyses as follows;   * Build Cox proportional hazard regression models for each type of GRS, adjusting for age, sex and principal components, and compare the hazard ratios for incident CHD across different GRSs. * Determine the incremental predictive utility of the GRSs by likelihood ratio tests, and assessing the change in C-indices for Cox models of incident CHD that includes conventional risk factors. * Evaluate GRSs separately on individuals from European and African American ancestry as well as a joint analysis, introducing a multiplicative interaction term between GRS and PC1 to investigate any effect modification on CHD prediction by ancestry. * Plot Kaplan-Meier estimates of cumulative incidence across different quintiles of the GRSs. * Construct logistic regression models and compare odds ratios for prevalent CHD using genome-wide vs candidate variant genetic risk score approaches. * Conduct a PheWAS for the GRSs to investigate any associated pleiotropic effects. | |
| **Ethical Considerations** | None | |
| **Target Journal** | Circulation | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Approval : October 15th, 2018  Project Duration: October 15th – January 1st -2019  Draft Completion: February 1st, 2019  Submission: March 11th, 2019 | |

**\*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