|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* |  | |
| **Submission Date** | 10/01/2020 | |
| **Project Title** | Improving Prediction for Coronary Heart Disease Risk Across Diverse Populations Using Transethnic Polygenic Risk Scores | |
| **Tentative Lead Investigator** *(first author)* | Ozan Dikilitas | |
| **Tentative Lead Investigator Email Address** | Dikilitas.ozan@mayo.edu | |
| **Tentative Senior Author**  *(last author)* | Iftikhar J. Kullo | |
| **All Other Authors** | Satoshi Koyama, Kazuo Miyazawa, Daniel J. Schaid, David Crosslin, Gail Jarvik, Harsh Patel, other eMERGE and Biobank Japan investigators, Kaoru Ito, Issei Komuro | |
| **Sites Participating** | All adult eMERGE sites | |
| **Background / Significance** | Because polygenic risk scores (PRS) are generated using cohorts of European-ancestry (EA) individuals, they may not be comparably associated with disease risk in non-EA individuals. In the eMERGE Network phase III cohort, we demonstrated that risk estimates for the association of PRSs with coronary heart disease (CHD) were similar in EA and Hispanic ethnicity (HE) individuals but lower in African-ancestry (AA) individuals. Additionally, genome-wide PRS with millions of variants showed a stronger association with CHD across all ancestry groups (Dikilitas et al, *AJHG* 2020). In a recent report by Koyama et al (*Nature Genetics* 2020), a PRS derived using transethnic genome-wide meta-analysis of CHD in EA and Japanese participants outperformed those derived from either Japanese or EA GWAS for CHD in an independent case-control cohort of Japanese participants (Top vs. bottom decile; Transethnic PRS OR 8.3 AUC 0.664, EA PRS OR 3.7 AUC 0.628, Japanese PRS OR 4.2 AUC 0.634). Leveraging large transethnic meta-analysis summary statistics may similarly improve CHD risk prediction in both EA and other non-EA populations. | |
| **Outline of Project** | In the proposed study, we will derive CHD PRSs using summary statistics from a transethnic genome-wide meta-analysis conducted by the Biobank Japan investigators (Dr. Kaoru Ito, Prof. Issei Komuro et al.) and explore improvement of risk prediction for CHD in the eMERGE network phase III/IV genotyped cohort. We will define CHD cases and controls using previously validated eMERGE phenotyping algorithms based on ICD/CPT codes among adult EA, AA, and HE participants. We will quantify the strength of association between each derived PRS and CHD in all three ancestry groups. | |
| **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** | In eMERGE Phase III v3 array dataset (n=97,555 adults), PRS will be derived using Pruning & Thresholding method. We will also explore the use of EA vs. ancestry-specific LD reference panel when we derive these PRSs.  Each score will be standardized to mean=0 standard deviation=1 within each ancestry-group prior to analysis to facilitate easier interpretation of the results.  Association of each PRS and CHD will be tested using multivariable logistic regression adjusting for age, sex, eMERGE site, EHR-depth, and first 5 ancestry-specific principal components.  Within each ancestry group, transethnic PRS will be compared to its EA-derived counterpart to quantify the change in AUC and pseudo-R2 from the base model without PRS. Estimates of these indices will subsequently undergo internal validation along with evaluation of model calibration using bootstrapping (10,000 iterations).  In order to explore whether EA- or transethnic-derived PRS exhibit a dose-response relationship with CHD vs. have an increased risk only limited to the extremes of the PRS distribution, using the best performing PRS in each ancestry group, we will model PRS with natural splines and plot estimated ORs across the entire PRS range.  We will further explore incorporation of rare variants into the risk prediction models with PRS to determine improvement in predictive performance. | |
| **Ethical Considerations** | N/A | |
| **Target Journal** | Journal of the American College of Cardiology (JACC), Circulation | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Implementation of phenotyping algorithms and data analysis (October 2020)  Draft and circulate the manuscript among co-authors (November 2020)  Initial journal submission (December 2020) | |

**\*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