|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT360 | |
| **Submission Date** | Sept 24, 2019 | |
| **Project Title** | Multiethnic GWAS and PRS Development in Prostate Cancer | |
| **Tentative Lead Investigator** *(first author)* | NA (this will be part of the PRACTICAL Consortium) | |
| **Tentative Senior Author**  *(last author)* | Chris Haiman (USC) | |
| **All Other Authors** | Todd L. Edwards (Vanderbilt)  >300 in PRACTICAL | |
| **Sites Participating** | >150 in PRACTICAL | |
| **Background / Significance** | Prostate cancer incidence varies across racial and ethnic groups and is approximately 75% higher in African American men and 45% lower in Asian men compared with non-Hispanic White men. Age, race/ethnicity, a family history of prostate cancer and germline variation are the only established risk factors for prostate cancer, with as much as 57% of the variability in prostate cancer risk estimated to be due to genetic factors. Genome-wide association and fine-mapping studies have discovered ~270 germline risk variants for prostate cancer, with some more prevalent in specific racial and ethnic groups. To further our understanding of the genetic basis of prostate cancer and whether germline variation underlies racial and ethnic differences in prostate cancer incidence, we propose to increase the GWAS sample size through collaboration with ongoing biobanks and consortia. | |
| **Outline of Project** | The latest genome-wide association study (GWAS) of prostate cancer in the PRACTICAL Consortium included 110,406 prostate cancer cases and 126,974 controls from men of European, African, Asian, and Hispanic ancestry and identified 90 new loci. Over the next 2 years, we plan to increase the size of this study to ~200,000 cases and ~600,000 controls, through the inclusion of GWAS data from ongoing biobanks and consortia.  Genotype and imputed data will be examined for association with prostate cancer risk in eMERGE using logistic regression adjusting for relevant covariates, including age, sub-study and principal components. Results from will be combined with results in other PRACTICAL studies by a fixed-effects inverse-variance weighted meta-analysis in ancestry-specific analyses as well as across all four ancestry groups to obtain multiethnic estimates of effects. Secondary GWAS analyses will on aggressive and non-aggressive phenotypes, when available. The combined multiethnic dataset will also be used for fine-mapping of known and novel signals to identify secondary signals and better markers of risk to inform multiethnic PRS development.  Genetic risk scores will be constructed to evaluate the potential clinical utility of the combination of all common risk variants to define men at high risk of disease. Genetic risk scores will be constructed by summing the risk variants for every individual weighted by a weight constructed using the conditional effect estimates from the multiethnic meta-analysis. The GRS will be categorized by decile and odds ratios were estimated compared with men at average genetic risk in the 40-60% category using logistic regression adjusting for age, sub-study and principal components for each ancestry group. Secondary analyses will focus on PRS testing by aggressive and non-aggressive phenotypes, when available. | |
| **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)*** | Prostate cancer phenotypes, including stage, Gleason score and PSA values (for cases and controls, if available). Information on family history of cancer, if available. | |
| **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 Uterine  No | |
| **Planned Statistical Analyses** | GWAS  Fine-mapping  PRS | |
| **Ethical Considerations** | None | |
| **Target Journal** | Depends on results, likely a genetics journal such as Nature Genetics | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Gather data from coordinating center: Jan, 2020  Conduct statistical analyses: by June 2021  Write manuscript: By the end of 2021 (this depends on when we obtain similar results from other studies)  Circulate and submit manuscript: Early 2022 | |