|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT339 | |
| **Submission Date** | 04/25/2019 | |
| **Project Title** | Phenotype risk scores identify patients at a high risk of hereditary cancer syndromes and improve variant interpretation (*tentative title as an evolution of our “Cancer PheWAS” phenotype on the phenotype tracker)* | |
| **Tentative Lead Investigator** *(first author)* | Chenjie Zeng and Lisa Bastarache | |
| **Tentative Senior Author**  *(last author)* | Josh Denny | |
| **All Other Authors** | Georgia Wiesner, Dan Roden, and investigators from any/all other eMERGE III sites interested in participating. | |
| **Sites Participating** | All eMERGE III sites | |
| **Background / Significance** | Advances in next generation sequencing have led to a drastic increase in the use of clinical genetic testing in the management of hereditary cancers. Results of genetic testing could inform screening in patients and their relatives as well as treatment decisions. Accurate interpretation of these variants has become a pressing challenge. Our previous work in phenotype risk scores (PheRS) showed that aggregating relevant clinical features could improve phenotyping and augment variant interpretation for mendelian diseases.  Many pieces of evidence support that hereditary cancer syndromes could manifest as a broad spectrum of clinical features. For example, Pilarski *et al.* found that PTEN hamartoma tumor syndrome patients could also have autism spectrum disorders, colon cancer, esophageal glycogenic acanthosis, penile macules, renal cell carcinoma, testicular lipomatosis, and vascular anomalies, all of which were rarely included in previous diagnosis criteria. However, currently known clinical features for many hereditary cancer syndromes are, in general, limited to family history, age of onset and/or histologic characteristics of tumors. Other potential related clinical features remain to be identified.  We have developed the phenome-wide association study (PheWAS) method to evaluate associations of genetic variants with a large number of clinical features documented in electronic medical records (EMRs). By applying this method to eMERGEseq data, we will identify clinical features associated with pathogenic variants in hereditary cancer genes systematically. We will validate the observed associations in additional datasets. We will aggregate the validated clinical features as well as relevant features curated by OMIM for each cancer syndrome gene and derive the corresponding PheRS for each individual in the eMERGEseq cohort. We will determine the phenotypic risk of hereditary cancer syndromes of each individual and estimate association of variants previously determined as of unknown significance with PheRS for each cancer syndrome. Because of the increased power achieved by aggregating clinical features, we anticipate detecting additional pathogenic variants and variants that have partial penetrance, which will facilitate accurate variant interpretation.  The genes we are focusing on, as part of our Cancer PheWAS, are: APC, ATM, BMPR1A, SMAD4, BRCA1, BRCA2, PALB2, CHEK2, TP53, MLH1, MSH2, MSH6, PMS2, POLD1, POLE, MUTYH, PTEN, STK11, NF2, RB1, MEN1, RET, BLM, SDHAF2, SDHB, SDHC, SDHD, TSC1, TSC2, VHL, JAK2 and WT1, as well as FANCC. | |
| **Outline of Project** | 1. Identify a range of phenotypes associated with carrying pathogenic variants in hereditary cancer genes through phenome-wide association studies (PheWAS) and validate observed association in additional datasets including the UK biobank.  2. Create PheRS for each participant based on associated phenotypes and phenotypes curated by OMIM.  3. Use linear regression to detect association between VUS and PheRS in the eMERGEseq data. Validate PheRS by testing against known P/LP and benign variants. | |
| **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)*** | *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** | 1. Phenome-wide association studies of hereditary cancer genes: logistic regression for common phenotypes with sufficient pathogenic variant carriers; fisher exact test for rare phenotypes or a small number of pathogenic variant carriers. 2. Linear regression to evaluate associations between PheRS and variants in hereditary cancer genes. | |
| **Ethical Considerations** | None | |
| **Target Journal** | TBA | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 1 year for completion | |

**\*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