|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT312 | |
| **Submission Date** | 10/30/2018 | |
| **Project Title** | Assess penetrance of cancer among mutations carriers for hereditary breast cancer genes. | |
| **Tentative Lead Investigator** *(first author)* | Katherine Crew | |
| **Tentative Senior Author**  *(last author)* | Wendy Chung | |
| **All Other Authors** | Yufeng Shen, Chunhua Weng, George Hripcsak, Ning Shang, Emily Groopman, Krzysztof Kiryluk, Lynn Petukhova | |
| **Sites Participating** | Any other interested eMERGE sites | |
| **Background / Significance** | The estimate of the penetrance of cancer genes is likely biased by individuals at higher risk who enroll in research studies and have clinical genetic testing. In addition, there are several novel genes for cancer susceptibility that have recently been identified for which there are no good estimates of penetrance. eMERGE data will help to provide more accurate population based age and gender cancer risks. This proposal will focus on genes that increase the risk of breast cancer. | |
| **Outline of Project** | We will classify variants in BRCA1/2, PTEN, TP53, PALB2, CHEK2, and ATM using the ACMG guidelines and specifically using *in silico* prediction methods, developed by our team (*MVP*) and published functional data to assess missense variants*.*   1. Annotation of rare variants from the entire eMERGEseq data set using MVP and other new methods. 2. Phenotype definition. We will focus on cancer phenotypes, defined using ICD9 codes and the breast cancer phenotype we developed. 3. We will assess the age and sex specific penetrance for breast, ovarian, and other cancers with pathogenic/likely pathogenic variants in the genes above. | |
| **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 breast cancer, cancer  No | |
| **Planned Statistical Analyses** | 1. Data quality control 2. Calculating age and sex specific penetrance | |
| **Ethical Considerations** | Genomics data and phenotypic data will be de-identified to protect confidentiality. | |
| **Target Journal** | Genetics in Medicine, Genome Medicine. | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 1. Complete QC and annotation of eMERGEseq data by Dec 2018 2. Obtain and curate phenotype data of eMERGEseq subjects by Feb 2019 3. Statistical analysis by April 2019 4. Manuscript by July 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 for breast cancer
* 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