|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT306 | |
| **Submission Date** | 09/18/18 | |
| **Project Title** | Evaluating the utility of high-throughput functional scoring of BRCA1 missense variants in eMERGE-III participants | |
| **Tentative Lead Investigator** *(first author)* | Adam Gordon | |
| **Tentative Senior Author**  *(last author)* | Lea Starita, Gail P Jarvik | |
| **All Other Authors** | Greg Findlay, Melody Palmer, Elisabeth Rosenthal, David Crosslin, Ian Stanaway, Taryn Hall, David Carrell, Kathy Leppig, Eric Larson | |
| **Sites Participating** | All adult eMERGE sites | |
| **Background / Significance** | Pathogenic variants in the cancer associated genes *BRCA1* and *2* are the most commonly identified of the 59 high penetrance actionable genes outlined by the ACMG, even in non-Ashkenazi populations. Of 3,936 germline BRCA1 SNVs currently represented in ClinVar, only 983 are classified by an expert panel as ‘benign’ or ‘pathogenic’ without conflicting interpretations. Of the rest, 1,794 are listed as Variants of Uncertain Significance (VUS) in Clinvar. An additional 218 missense variants have conflicting pathogenicity classifications from different labs. Currently, over 50% of the BRCA1 single nucleotide variants (SNVs) in ClinVar are listed as VUS by all submitters. However, women who have a VUS test result and a family history of breast cancer are reported to have high levels of distress and often undergo possibly unnecessary surgeries due to improper medical management. Historically, *in vitro* assays to determine the effect of such single nucleotide variants on protein function or splicing have been low throughput and performed in a post hoc, piecemeal fashion and therefore have not kept pace with the scaling of BRCA1 testing and the accumulation of VUS. A high throughput experimental method to measure variant function to inform variant classification would be promising, but must be linked to large repositories of patient phenotype and outcome data in order to demonstrate its accuracy and utility in variant interpretation.  In an attempt to overcome the problem of scale, the labs of Drs. Jay Shendure and Lea Starita at the University of Washington have developed a multiplexed functional assay for the breast, ovarian/fallopian, pancreatic, and prostate cancer associated gene, *BRCA1*, based on saturation genome editing to introduce and test all possible SNVs in *BRCA1* (<https://www.biorxiv.org/content/early/2018/04/05/294520>). This first report details function scores for 3,893 SNVs, encompassing 96.5% of all possible SNVs in 13 exons that encode BRCA1’s functionally critical RING (exons 2-5) and BRCT (exons 15-23) domains. These function scores were bimodal, and when dichotomized, predictions of pathogenicity were 96.7% sensitive and 98.2% specific with ClinVar entries (P/LP or B/LB only, as reported by an at least one 1-star lab). The goal of developing these high throughput functional assays is to incorporate these scores directly into the interpretation of BRCA1 missense variants. In order to do so, the concordance of these predictions with cancer risk must be established, particularly in VUS. Additionally, validation of the functional assay accuracy with eMERGE sequencing center pathogenicity classification for non-VUS would provide a second validation test of assay concordance.  We propose to evaluate the concordance the assay with sequencing center classification and with cancer diagnoses in the adult eMERGE participants. eMERGE is unique among larger databases in that we have the ability to connect genotype and phenotype data across a wide variety of traits for currently 15,000 and soon 25,000 participants, an ideal setting to test high throughput functional assays and ultimately improve variant interpretation. Specifically, within the current Data Freeze of ~15,000 eMERGEseq participants we identify 177 missense variants in BRCA1. Of these, 34 SNVs are in the domains in which the functional assay has already been completed. These are currently classified as 3 P/LP, 16 VUS, and 15 B/LB; one VUS is not found in ClinVar (p.Phe79Leu). | |
| **Outline of Project** | 1. Extract network-wide *BRCA1* variant data, compare CSG classification with functional scores 2. Identify individuals with *BRCA1* variants, confirm VUS across all eMERGE sites via orthogonal sequencing 3. Collect EHR-derived cancer phenotypes on relevant individuals and validate against chart review (only for VUS/LP/P, or B/LB with discordant functional score) 4. Collect cosegregation data on informative family members via single-site research genotyping 5. Revise variant classifications, adhering to ACMG and incorporating new data. Submit updated interpretations to ClinVar. | |
| **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)*** | Cancer diagnosis status (by chart review) among relevant *BRCA1* variant carriers | |
| **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, Ovarian Cancer  No | |
| **Planned Statistical Analyses** | * Compare CSG interpretations vs. assay predictions via simple crosstabs * Measure functional assay sensitivity and specificity using actual patient data * Calculate strength of cosegregation evidence, where applicable | |
| **Ethical Considerations** | Clearly inform family members that their cosegregation testing is not useful for clinical care if the classification remains VUS. | |
| **Target Journal** | Amer J Human Genetics | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Fall 2018 – Network data extraction & harmonization, genotype validation  Winter 2019 – EHR review, family testing,  Spring 2019 – Statistical analysis, manuscript submission | |

**\*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