|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT358 | |
| **Submission Date** | 09/11/2019 | |
| **Project Title** | A systematic evaluation of the transferability of polygenic risk score for breast cancer across diverse population and different geographic areas | |
| **Tentative Lead Investigator** *(first author)* | Cong Liu, Nur Zeinomar | |
| **Tentative Senior Author**  *(last author)* | Chunhua Weng, Mary Beth Terry | |
| **All Other Authors** | Wendy Chung, Krzysztof Kiryluk, Atlas Khan, George Hripcsak, Ning Shang, Ali Gharavi, David Fasel, Kathy Crew, and others who join the study | |
| **Sites Participating** | We propose a network-wide study (all sites are invited to participate). The analyses will be led by Columbia University. | |
| **Background / Significance** | Most recently, a breast cancer (BC) polygenic risk score (PRS) with the highest discriminatory ability to date was developed and prospectively validated in the largest GWAS datasets currently available, (79 studies in the Breast Cancer Association Consortium (BCAC), and over 190,000 women in the UK Biobank)1. Although this score is well calibrated and validated for European ancestry individuals, it is unknown how this score performs in individuals of non-European ancestry and how generalizable the BC PRS is in North Americans across different geographic areas. Moreover, although there is evidence of the heritability of breast cancer prognosis2-5, the relationship between genetic variation and breast cancer recurrence and survival has not been well characterized. eMERGE provides a potentially valuable data source to evaluate the transferability of PRS to other ancestries, settings, and outcomes. Additionally, eMERGE has rich clinical and pathological data that can be leveraged to further our understanding of the outcomes following a BC diagnosis, in addition to BC risk. By testing previously developed PRS in eMERGE dataset, we will be able to examine (1) whether a PRS well optimized for European ancestry can be generalized to other populations; (2) whether a PRS optimized in one geographic area can be applied directly to other geographic areas; (3) whether a PRS that has been developed for BC risk prediction is useful for BC prognosis; and (4) how much additional information is gained from the PRS beyond routinely assessed clinical factors, including family history. | |
| **Outline of Project** | 1. Calculate and refine breast cancer polygenic risk scores (PRS) for overall eMERGE cohort based on two previously developed PRSes (Khera et al. Nat Gen 20186 and Mavaddat et al. AJHG 20191) and compare specificity/sensitivity among eMERGE individuals and non-eMERGE individuals. 2. Compare the PRS performance in European ancestry population and non-European ancestry populations based on self-reported demographics and PCAs. 3. Optimize the PRS for one eMERGE site and test its performance across different sites and evaluate the transferability of PRS. 4. In BC cases, examine the performance of the PRS developed for BC risk on predicting outcomes following a BC diagnosis, including a second breast event, disease free survival, and overall survival. 5. Examine if the PRS for outcomes after breast cancer adds to the discriminatory ability of existing BC prognostic models such as the Nottingham prognostic index (NPI)7. 6. Develop a hybrid system to incorporate PRS with clinical routine for better breast cancer risk and prognostic prediction. | |
| **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)*   * self-reported demographic * family history when available. * Tumor size, histological grade, nodal status when available. * Estrogen receptor status of tumor when available. * Survival status when 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: breast cancer  No | |
| **Planned Statistical Analyses** | 1. Examine ROC AUC for the previously developed PRS for eMERGE subjects. 2. Compare ROC AUC for the PRS between different ancestral populations. 3. Compare ROC AUC for the PRS between different geographic areas. 4. Examine the performance of PRS in predicting outcomes after diagnosis, including second breast cancer event, disease free survival, and overall survival 5. Compare ROC AUC between PRS and NPI for breast cancer prediction and prognosis. 6. Quantify the gain in predictive performance from adding PRS to routine clinical risk assessment. | |
| **Ethical Considerations** | NONE | |
| **Target Journal** | AJHG, PLoS GENETICS, Breast cancer research (BCR), International journal of cancer, JNCI or similar | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | * 07/2019: Develop initial breast cancer PRS calculation pipe line for eMERGE; * 10/2019: Evaluate the transferability of breast cancer PRS across different geographic areas; * 1/2019: Compare the performance between PRS and other existing BC prognostic models such as the Nottingham prognostic index (NPI); * 3/2020: Manuscript draft completion; * 4/2020: 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
* 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

Reference

1. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet.* 2019;104(1):21-34.

2. Lindstrom LS, Li J, Lee M, et al. Prognostic information of a previously diagnosed sister is an independent prognosticator for a newly diagnosed sister with breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2014;25(10):1966-1972.

3. Hartman M, Lindström L, Dickman PW, Adami H-O, Hall P, Czene K. Is breast cancer prognosis inherited? *Breast Cancer Research.* 2007;9(3):R39.

4. Hemminki K, Ji J, Försti A, Sundquist J, Lenner P. Survival in breast cancer is familial. *Breast cancer research and treatment.* 2008;110(1):177-182.

5. Verkooijen HM, Hartman M, Usel M, et al. Breast cancer prognosis is inherited independently of patient, tumor and treatment characteristics. *Int J Cancer.* 2012;130(9):2103-2110.

6. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50(9):1219-1224.

7. Lee AHS, Ellis IO. The Nottingham Prognostic Index for Invasive Carcinoma of the Breast. *Pathology & Oncology Research.* 2008;14(2):113-115.