2023 ACG-IRG Pilot Grant

Improving the predictive power of polygenic risk scores for breast cancer in diverse and admixed populations using a novel ancestry-agnostic approach

Matteo D'Antonio, PhD

Abstract:

Background and hypothesis

Breast cancer is the most common cancer in women globally, with significant variations in incidence and mortality rates among different populations. The identification of genetic risk factors through Genome-Wide Association Studies (GWAS) has shown promise in improving breast cancer risk prediction. However, the majority of existing GWAS studies have focused on individuals of European ancestry, limiting the generalizability of Polygenic Risk Scores (PRS) to diverse and admixed populations. This grant proposal aims to *develop an ancestry-agnostic approach to GWAS and PRS for breast cancer, with a focus on enhancing patient stratification and reducing health disparities*.

Specific aims and study design

Aim 1: Develop an innovative haplotype-based approach to genotype individuals, independent of their discrete population labeling. Haplotypes, defined as combinations of Single Nucleotide Polymorphisms (SNPs) inherited together, can capture genetic variability at a finer level and do not require a discrete ancestry label.

The proposed haplotyping pipeline will convert phased SNP genotypes into haplotypes, and dimensionality reduction techniques, such as Principal Component Analysis (PCA), will be applied to reduce the complexity of haplotype data. Ancestral haplotypes will also be reconstructed, offering an advantage over SNP-based approaches by allowing us to infer ancestral sequences for GWAS.

Aim 2: Integrate the ancestry-agnostic haplotypes into GWAS and PRS calculation and compare their performance to standard SNP-based methods. GWAS will be conducted on both haplotypes and SNPs using large-scale datasets, including the UK BioBank and All of Us cohorts, which comprise hundreds of thousands of diverse and admixed individuals. The association between breast cancer and genetic variation will be assessed using logistic regression models, considering local dimensionality reduction for haplotypes. The resulting GWAS will provide insights into the genetic basis of breast cancer in diverse populations.

To evaluate the efficacy of PRS, correlations between PRS estimates obtained from haplotypes and SNPs will be computed. The study will investigate the portability of PRS between populations and test the hypothesis that haplotype-based GWAS improves PRS accuracy for diverse and admixed populations.

Cancer relevance

The proposed research has the potential to significantly impact breast cancer risk prediction and patient stratification. By developing ancestry-agnostic PRS models, the study aims to overcome the limitations of existing GWAS studies, which primarily focus on individuals of European ancestry. This approach will lead to improved breast cancer risk prediction for individuals from diverse and admixed populations, thereby enhancing the accuracy of personalized medicine and reducing health disparities.

In conclusion, this grant proposal seeks to advance breast cancer genetics research by incorporating diverse and admixed populations into GWAS and PRS studies. The development of an ancestry-agnostic

haplotype- based approach holds promise in revolutionizing the field, providing insights into the genetic basis of breast cancer in underrepresented populations, and facilitating tailored interventions for at-risk individuals. *The successful execution of this research will have far-reaching implications for breast cancer prevention, diagnosis, and treatment in diverse populations, leading to improved health outcomes and equitable healthcare access*.