MCC-C3 Padres Pedal the Cause Spring 2018

Enhanced Breast Cancer Risk Prediction from Imputed Gene Expression

Hannah Carter, PhD (MCC)
Graham McVicker, PhD (Salk)

Scientific Abstract:

Improved identification of individuals at high risk of developing cancer could significantly reduce cancerassociated morbidity and mortality and is therefore a central goal of precision cancer medicine. Genetic risk scores (GRSs) provide a strategy to identify individuals that have high risk due to heritable factors. Current risk scores rely on a small number of established risk single nucleotide variants (SNVs), which were found to be significantly correlated with disease in genome-wide association studies of cancer and healthy populations. However, the risk SNVs identified in these studies fail to account for the majority of heritable risk and provide little mechanistic insight into the underlying biology. In this proposal, we exploit new developments in the prediction of gene expression from genotypes to establish a new paradigm for breast cancer risk prediction. Using large, publicly available datasets we will train models to predict breast tissue gene expression from germline genotypes, and then look for genes that are differentially expressed between healthy individuals and individuals who developed breast cancer. To gain insight into the biological mechanisms underlying susceptibility, we will analyze the imputed gene expression values to identify transcriptional modules that are enriched for susceptibility genes. Finally, we will develop a new method to compute GRSs that leverages the relationship between gene expression and cancer susceptibility. Specifically, we will estimate cancer risk from imputed gene expression values using a network-informed regression method that is trained to separate breast cancer cases from controls and validate the resulting GRS on independent breast cancer datasets.

Lay Abstract:

Early detection remains the strongest determinant of long-term disease-free survival across cancer types, however there is a tradeoff between screening for early detection and the risk of false diagnosis and overtreatment. One strategy to optimize this tradeoff is to stratify individuals by genetic cancer risk and to screen more aggressively in high-risk populations. Genetic risk scores have shown some promise for identifying individuals at risk, however current models use a very simple combination of risk factors and do not account for the complex nature of the underlying biology. In this proposal we will develop new genetic risk scores that incorporate information about (1) the impact of genetic factors on gene expression and (2) the relationship between gene expression and disease risk. This proof-of-principle study will be implemented in breast cancer but can easily generalize to a number of cancer types and will provide a framework for future studies of cancer risk biology. The long- term goals of this research are to improve our understanding of the genetic mechanisms that drive cancer risk and to enable accurate identification of high-risk individuals that can benefit from additional screening and prevention.