Multi-omics Predictive Model for Personalized Breast Cancer Prognosis

Yu, Audrey (School: Germantown Academy)

Breast cancer is the most common cancer worldwide and a complex disease characterized by genetic heterogeneity. Despite advancements in cancer management, the risk of recurrence and death remains high among patients with stage II or III breast cancer. This study aims to develop reliable prognostic models that integrate molecular and phenotypic features to enhance personalized benefit-risk assessments and optimize treatment decisions. Clinical and multi-omics data, including mRNA expression, copy number alterations (CNAs), and somatic mutations, were extracted from the high-risk early-stage (stage II/III) breast cancer cohort in the METABRIC database. Principal component analysis (PCA) and LASSO regression were applied to reduce dimensionality of the large-scale omics data and to identify the most predictive prognostic markers. Multiple Bayesian network (BN) models were then developed to integrate these data for predicting the probability of achieving at least five-year recurrence-free survival based on individual patient profiles. Data analysis identified 16 mRNAs, 6 CNAs and 6 somatic mutations as prognostic predictors. All multi-omics BN models demonstrated high accuracy (> 0.9) and strong discriminative ability (AUC > 0.9) in prediction, attributed to molecular data integration and robust variable selection. Model validation with independent TCGA data confirmed model robustness, suggesting the improvements were not due to overfitting. Future research may explore additional data sources and biomarkers to further enhance predictive performance. The models are generalizable and offer precise, interpretable recurrence risk assessments, making them valuable tools for supporting personalized prognosis and medical decisions for clinicians and patients.

Awards Won: