Table of Contents

Research Week 2024

Graph-based Multimodal Integration for Breast Cancer Subtype Classification

Sammantha Avaylon, Guanming Wu, Ph.D., Andrew Emili, Ph.D., Michael Mooney, Ph.D.[1], Shannon McWeeney, Ph.D.[2]

Oregon Health and Science University, Knight Cancer Institute

[1] Thesis Advisory Committee Chair
[2] Thesis mentor

Keywords

Systems biology, biological networks, machine learning, breast cancer subtypes, similarity network fusion, random walk with restart, multi-omic integration

Abstract

The significant inter- and intra-tumor heterogeneity observed in breast cancer patients presents an ongoing challenge that impedes accurate disease characterization and classification, thereby affecting precise patient stratification. Breast cancer subtyping methods, including immunohistochemistry (IHC) and PAM50 gene expression profiling, play a significant role in addressing this challenge. Specifically, the PAM50 assay delineates five intrinsic breast cancer molecular subtypes: Luminal A, Luminal B, HER2-positive, Basal-like and Normal-like (Normal), based on patient gene expression profiles. However, there remains a clinical need to further characterize the molecular heterogeneity within the PAM50 subtypes, as not all patients are cleanly classified under the existing subtypes.

Recent advancements in breast cancer diagnostics have highlighted the limitations of single-omic approaches, prompting the exploration of multi-omic integration strategies to enhance our understanding of disease pathogenesis. This study proposes the use of a graph-based multimodal model to integrate breast cancer RNA sequencing and histopathology slide data from The Cancer Genome Atas (TGCA) Breast Invasive Carcinoma (BRCA) cohort for improved breast cancer subtype classification.

To evaluate the efficacy of multi-layer network integration, we compare Similarity Network Fusion (SNF) and random walk with restart (RWR) methodologies within an intermediate fusion framework to assess the impact of local and global topology integration. Additionally, we plan to assess additional network integration approaches (early and late fusion) within the machine learning model framework to determine their impact on predictive accuracy. Our primary objective is to measure the predictive accuracy of intermediate fusion techniques utilizing SNF or RWR, while refining and characterizing existing breast cancer subtypes.

By leveraging a graph-based multimodal integration model, our study aims to deepen our understanding of breast cancer complexity, leading to more accurate disease subtype classification and prediction. This research holds the potential to provide mechanistic insights and address indeterminate subtype calls, therefore advancing precision medicine for breast cancer patients.