UNIVERSITY of MARYLAND
MARLENE AND STEWART GREENEBAUM
COMPREHENSIVE CANCER CENTER

Population Science Seminar Series

DiAGMI: Disentangling Outcome-Predictive and Subtype-Associated Signals in Multi-omics Cancer Data

---

Wednesday, May 27, 12:00 PM EST via Zoom Link

ABSTRACT: Multi-omics cancer cohorts often contain far more molecular features than patient samples, making subtype discovery and outcome prediction difficult to separate. Existing approaches frequently combine subtype-related and outcome-predictive signals into a single representation, limiting interpretability and introducing confounding. We present DiAGMI, a two-module framework designed to disentangle these signals while improving performance in small multi-omics datasets. DiAGMI-Gen separates outcome-predictive and subtype-associated latent representations across multiple omic layers and generates high-fidelity synthetic patient samples to augment training data. We identify key modifications that prevent clustering collapse and introduce a fidelity gate to detect synthetic data drift. DiAGMI-Factor then integrates real and accepted synthetic samples through interpretable non-negative multi-view factorization to support outcome prediction, subtype-aware clustering, and cross-omics feature ranking. Across five benchmark scenarios spanning varying signal strengths, class imbalance, and multiple outcome types, DiAGMI generated realistic synthetic samples and improved prediction stability while preserving biologically meaningful feature associations. Applied to a real breast cancer multi-omics cohort, DiAGMI disentangled subtype-associated and survival-predictive variation, identifying candidate cross-omics biomarkers linked to outcome independently of subtype for future biological validation.

Speaker: Huang Lin, PhD

Dr. Lin received his PhD in Biostatistics from the University of Pittsburgh and completed postdoctoral training at NIH. He joined UMD in 2023 as a faculty member in the Department of Epidemiology and Biostatistics. His research develops statistical and machine learning methods for high-dimensional multi-omics data, particularly metagenomics and metabolomics, with applications in cancer, infectious disease, aging, and public health. Dr. Lin leads and collaborates on NIH- and NSF-funded projects focused on predictive modeling, causal inference, multi-omics integration, and biomarker discovery.

UNIVERSITY of MARYLAND
SCHOOL OF MEDICINE

Contact: Susan Holt, Office Manager – 410-706-8505 - sholt@som.umaryland.edu

NCI
Designated Comprehensive Cancer Center