From SMILES to Synergy: Biologically-Enriched Drug Screening for Synergistic Combination Therapy in Tamoxifen Resistant ER+ Breast Cancer

Prokofev, Daniil (School: Thomas Jefferson High School for Science and Technology)

Worldwide, breast cancer kills 700,000 women annually. Although 5 years of tamoxifen (Tam) therapy cuts the risk of progression in half, 40% of patients develop resistance (TamR), driven by quiescent breast cancer stem cells (bCSCs) that survive therapy and awaken to metastasize. The objective of this study was therefore to find a novel combination drug, synergistic with tamoxifen, that kills bCSCs. Computational synergy screening is promising, but SOTA models (SynergyX, MARSY) fail to predict the synergy of unseen drugs, essential for screening. Replacing drug data with unique identifiers retains identical out-of-distribution (OOD) performance, suggesting shortcut learning. To overcome this, I developed a novel drug embedding model that learns from molecular structure (SMILES) alone, enriched during training via contrastive alignment to transcriptomic, cell viability, and knowledge graph data. This biologically grounded representation enables OOD generalization, improving downstream synergy prediction by +0.34 AUPRC over SOTA on held-out drugs. I also trained a transcription factor (TF) perturbation predictor using VIPER enrichment of drug transcriptomics data. Together, I screened 12,209 FDA approved and investigational compounds for 1) synergy with tamoxifen and 2) reversal of bCSC TF programs, defined by TamR patient single cell RNA sequencing data. All three top novel combination candidates demonstrated strong synergy (Bliss > 10) with Tam; the lead compound surpassed Vorinostat in mammosphere inhibition and bCSC oncoprotein downregulation. Bioinformatics analyses confirmed targeted suppression of bCSC maintenance networks. My generalizable framework identifies novel, synergistic, mechanistically rational combinations to overcome canonical therapy limitations.

Awards Won: