Development of a Novel Multi-Omics Enhanced Quantitative Structure-Activity Relationship (ME-QSAR) Model for Pancreatic Ductal Adenocarcinoma Inhibitor Discovery

Oza, Naitik (School: Freedom High School)

Traditional drug discovery exceeds 10 years and $2 billion per drug. Conventional Quantitative Structure-Activity Relationship (QSAR) models only identify single-target inhibitors and lack biological context. This project proposes a two-stage Multi-omics Enhanced QSAR (ME-QSAR) model to identify multi-target cancer inhibitors. The proposed ME-QSAR model was evaluated on Pancreatic Ductal Adenocarcinoma (PDAC), a cancer with a 7% five-year survival rate. In this model, multi-omics data (gene expression, protein expression, copy number alteration, DNA methylation, metabolite levels) were collected from the Cancer Cell Line Encyclopedia for target cell lines, each modality undergoing pre-processing. The top 50 omics features per modality were selected using minimum Redundancy Maximum Relevance and integrated into a matrix using Similarity Network Fusion. QSAR models typically require compound data tested against cancer cell lines. Therefore, molecular descriptors were generated from SMILES structures, filtered using Pearson correlation (-0.95 < r > 0.95), and reduced to 3 components via Principal Component Analysis, retaining 95% variance. The Stage 1 model used a 75/10/10 train-validation-test split, training an Elastic Net model on molecular descriptors. The Stage 2 stacking model used the same split to integrate multi-omics data for improved IC50 prediction of test compounds, enabling virtual screening of potent inhibitors with IC50 = 1 μM in PDAC. However, the Stage 2 model was limited by only 4 unique compound-target data points and had difficulty integrating chemical and biological data, so the ME-QSAR could not be completed. Still, the ME-QSAR framework shows strong potential to enhance multi-target drug discovery and treatment options.

Awards Won: