An Interpretable Machine Learning Framework to Predict Cisplatin Sensitivity

Dharani, Mythreya (School: Bergen County Academies)

One major challenge with cancer treatment is the widespread molecular differences that patient tumors can have, meaning only specific chemotherapies may work well. Machine learning models predicting patient tumor response to anti-cancer drugs are currently limited by low accuracy, due to the small dataset sizes, and limited interpretability, meaning it is difficult for clinicians to understand how a prediction was made by the model. To this end, a machine learning framework was developed to predict cisplatin sensitivity of tumors based on genomic information. Patient data from the Cancer Genome Atlas and preclinical data from the Genomics of Drug Sensitivity in Cancer project was collected. Individual genomic expression features and similarities to cell lines with known cisplatin IC50 values were integrated to extract maximal information. A multi-penalty binomial regression model was adapted to predict sensitivity using these features, with an increased performance compared to baseline models. Genes ranked highly by the model predicted sensitivity and resistance in a new cohort from the Oesophageal Cancer Clinical and Molecular Stratification study. Genes important for model prediction like RNA-binding protein tristetraprolin were investigated in vitro to confirm their pan-cancer role in cisplatin's mechanism of action. A375 melanoma cells were treated with a cytotoxic dose of cisplatin for 72 hours. A qPCR assay found a 2-fold increase in ZFP36 (tristetraprolin encoding) expression. RNA sequencing found high-fold changes in other genes with large model weights. Overall, this research advances interpretable drug sensitivity prediction, highlighting the potential of machine learning for precision medicine.

Awards Won: