Identification of Predictive Biomarkers Against Cancer with Sparse Neural Networks

Liang, Minnie (School: West Lafayette Junior/Senior High School)

A predictive gene biomarker helps determine which patients are most likely to benefit from a specific treatment option. The probability of survival from a treatment can greatly depend on if the patient is positive for the specific biomarker. The purpose of this project was to identify these predictive gene biomarkers against various anticancer drugs using neural networks. First, the neural network model predicted and identified a set of genes as potential biomarkers. Then, this initial selection was further narrowed down through RCoT, a randomized conditional correlation test. Through performing RCoT to obtain p-values for every gene in the set of potential biomarkers, only the genes with p-values less than 0.05 were chosen as the predictive biomarkers. Then the model's predictions were compared with past publications and research to show the validity. For example, the drug Paclitaxel's top selected predictive biomarker was BCL2L1 with a 0.00031 p-value. Over 1300 PubMed articles have been written about this gene and drug. Many publications, such as Lee et al. (2016) and Dorman et al. (2016), report that BCL2L1 is predictive of treatment response for Paclitaxel. For the drug Topotecan, the top selected gene biomarker identified by the neural network was SLFN11. There are over 110 articles on SLFN11 and Topotecan in PubMed that report SLFN11 as predictive of treatment response for Topotecan. These results show high promise towards the future of medicine and predictive biomarker identification. A few of the model-identified biomarkers have no articles written in literature about the drug and gene together. These new-found combinations of a drug and gene can be experimented and researched as well, potentially leading to discoveries of new predictive biomarkers.