Drug Vulnerabilities of the Cancer Cell Line Encyclopedia Are Revealed by Machine Learning Approaches

Elango, Mahalaxmi

Hsp90 “Heat Shock Protein 90” is a chaperone protein that helps client oncoproteins fold into their functional conformations. Since Hsp90 is upregulated in cancerous cells to possibly serve as a buffer against environmental stresses, drugs that target Hsp90, such as 17-AAG, are believed to selectively kill cancer cells. However, preliminary evidence suggests that 17-AAG performs poorly in its selectivity and specificity. In order to understand cancer’s sensitivity to Hsp90 inhibitors, we seek to identify genomic features that predict response to Hsp90 inhibitors. Using the Cancer Cell Line Encyclopedia (CCLE), a publicly available large-scale gene expression, chromosomal copy number, and tumor mutational profiling dataset across 497 human cancer cell lines for 26 anticancer drugs, we apply OLS regression, PCA, and machine learning to understand the underlying factors of and to predict cell line response to Hsp90 inhibition. To assess the relative performance of 17-AAG, we also predict the efficacy of other drugs in targeting their intended protein(s), or pathway(s). With precision and recall values at 0.57 for a linear SVM kernel, 17-AAG was very poorly predicted by our methods. Since drugs with less specific targets, particularly 17-AAG, were the most poorly predicted drugs by our machine learning methods, it suggests that either (i) the drugs are not targeting their intended purpose, (ii) the drugs are targeting correctly, but either more or less than is the most optimal target, or (iii) the drugs with several targeted proteins, or pathways, are difficult to predict. Therefore, it is imperative to improve our understanding of molecular features involved in drug response, so that we can translate our knowledge into developing more effective drugs for cancer treatment.