Multi-Omics Visible Drug Activity prediction, interpreting the biological processes underlying drug sensitivity

lunedì 28 febbraio 2022 19:05 (10)

Cancer is a genetic disease resulting from the accumulation of genomics alterations in living cells. Large scale genomics studies have been instrumental to understand the recurrent somatic genetic alterations within a cell, including chromosome translocations, single base substitutions, and copy-number alterations and for the characterization of their functional effects in transformed cells. One of the main challenging questions in this field is how to exploit all these molecular information to identify therapeutic targets and to develop personalized therapies. The understanding of the molecular features influencing sensitivity to drugs is the key element for the development of personalized therapies and to predict which patients should be treated and with which drugs and finally to evaluate eligibility criteria for oncology trials. Machine learning models are able to exploit multi-modal screening datasets such as Projects such as Genomics of Drug Sensitivity in Cancer (GDSC), Cancer Cell Line Encyclopedia (CCLE), Cancer Therapeutics Response Portal, NCI-60 and others to develop predictive algorithms useful to associate omics features with response. The basic approach is to use the data from these screenings to train a machine learning model that predicts the 50% inhibitory concentration (IC50) of a drug from the multi-omics profile of a cell line or a tissue sample. There have been several attempts at applying this approach using various machine learning frameworks such as Variational Autoencoders, Deep Networks, Convolutional Neural Networks, ensemble Neural Network models and combination of these approaches with different encodings of the features . Most of these studies use the machine learning models as "black boxes" optimized for prediction accuracy without the possibility to interpret the biological mechanisms underlying predicted outcomes.

Recently, some models were proposed to address this issue, but many of them just rely on somatic single nucleotide variations of the screened models; activity of the pathways, measured by gene expression profiling, is not taken into account, neither other important genomics alterations, such as copy number variations (CNV) that are of particular interest in cancer progression. Second, they do not take into account the unbalanced nature of the data since, in all large scale screening repositories, the values of IC50 are clustered around the value representing lack of sensitivity (for measures of sensitivity based on AUC, this value is 1) with a small minority of values representing sensitivity of a cell line to a specific drug. In order to address these limitations we propose a Multi-Omics Visible Drug Activity prediction (MOViDA) neural network model that extends the visible network approach incorporating functional information in terms of pathway activity from gene expression and copy number data into a neural network. Moreover, MOViDA is trained considering the unbalance of the dataset, we used a random sampler based on a multinomial distribution that accounts for the skewness of the dataset. We compare MOViDA with DrugCell showing that it is more accurate in predicting sensitivity to drugs, especially in the classes corresponding to lower AUC that represent those of more interest. In order to exploit the biological interpretation of network nodes we also develop an ad hoc network explanation method that scores the pathways that affect the prediction of sensitivity of a given cell line to a drug. To make this data useful for other purposes, we have identified which GOs and genes are good predictors for high sensitivity of a cell line to a drug. This explanation is the basis to hypothesize drug combinations and cell editing aimed at the identification of cell vulnerabilities.

Department

DIETI

Primary author(s) :   Dr. FERRARO, Luigi (University of Naples, Federico II, DIETI); Dr. SCALA, Giovanni (University of Naples, Federico II, Department of Biology); Prof. CECCARELLI, Michele (University of Naples, Federico II, DIETI)

Presenter(s) :   Dr. FERRARO, Luigi (University of Naples, Federico II, DIETI)