2023 – 01

Feature selection translates drug response predictors from cell lines to patients

Shinsheng Yuan1,2+, Yen-Chou Chen1+, Chi-Hsuan Tsai1, Huei-Wen Chen3 & Grace S. Shieh1,2,4,5*

May 30, 2023

Feature selection translates drug response predictors from cell lines to patients

Shinsheng Yuan[1,2+], Yen-Chou Chen[1+], Chi-Hsuan Tsai[1], Huei-Wen Chen[3] & Grace S. Shieh[1,2,4,5*]

[1]Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

[2]Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan

[3]Graduate Institute of Toxicology, College of Medicine, National Taiwan University, Taipei, Taiwan

[4]Genome and Systems Biology Degree Program and [5]Data Science Degree Program, Academia Sinica and National Taiwan University, Taipei, Taiwan

[+]These authors contributed equally

[*]Corresponding author: Grace S. Shieh, Tel: +886 227833697; E-mail: gshieh@stat.sinica.edu.tw

Abstract

Targeted and chemo-therapies are prevalent in cancer treatment. Identification of predictive markers to stratify cancer patients who will respond to these therapies remains challenging, because patient drug response data are limited. As large drug response data have been generated by cell lines, methods to efficiently translate cell-line-trained predictors to human tumors will be useful in clinical practice. Here, we propose versatile feature selection procedures, which can be combined with any classifier. For demonstration, we combined the feature selection procedures with a (linear) logit model and a (nonlinear) K-nearest neighbor, and train these by cell lines

to result in LogitDA and KNNDA, respectively. We show that LogitDA/KNNDA significantly outperform existing methods, e.g., a logistic model and a deep learning method trained by thousands of genes, in prediction AUC (0.70-1.00 for seven of the ten drugs tested) and are interpretable. This may be due to sample sizes are often limited in the area of drug response prediction. We further derive a novel adjustment on the prediction cutoff for LogitDA to yield prediction accuracy 0.70-0.93 for seven drugs, including Erlotinib and Cetuximab whose pathways relevant to anti-cancer therapies are also uncovered. These results indicate that our methods can efficiently translate cell-line-trained predictors to tumors, particularly for targeted therapies.