**Prediction of Adverse Events for LINCS Drugs using L1000 Gene Expression Data, Cell Morphology Profiles and Chemical Structure Features**

Zichen Wang1,2,3, Neil R. Clark1,2,3, Avi Ma’ayan1,2,3

1Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029

2BD2K-LINCS DCIC

3KMC-IDG

Abstract

Adverse-events are unintended and undesirable therapeutic response to drugs. Adverse-events are a major public health concern as well as a significant barrier in the drug development process. Failure to detect adverse-events for drugs before they are marketed can lead to substantial patient morbidity. Many efforts have been made to predict adverse-events using chemical and biological features of drugs, whereas the integration of features from chemical and biological origin has proven to improve adverse-events predictability. However, until recently, biological features about the properties of drugs are often limited to the known therapeutic targets of the drugs, or data from specific low-content assays. In contrast, for this study, we used gene expression responses and cell-morphology profiles measured in cell lines upon drug treatment. To predict adverse event we utilized the LINCS L1000 dataset and the Molecular Libraries Probe Production Centers Network (MLPCN) dataset as features to predict side adverse events. Additionally, we extracted features from the chemical structure of the drugs and combined these features with the expression and imaging features to better predict adverse events. First, we converted the LINCS L1000 gene expression profiles of drug treatments to Characteristic Directions in gene expression space to train a k-Nearest Neighbor classifier to predict adverse events. This method alone is already highly predictive (AUROC=0.892). We then combined the matrix derived from processing the gene expression L1000 data, with cell-morphology profiling data, and features extracted from the chemical structure of the drugs to train a multi-label classifier to predict drug adverse events. We found that combining multiple features using ensemble methods is able to achieve even better predictability. The predictive model was applied to all (~20,000) LINCS small molecule compounds profiled in by both L1000 LINCS and MLPCN projects and the results are delivered in an interactive freely available web-site at: <http://www.maayanlab.net/SEP-L1000>.