NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

POSTGRADUATE PROGRAM
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"

MASTER THESIS

A method for parsing clinical outcomes and combining them with phopshoproteomic and genomic data for predicting drug efficacy: application in hepatocellular carcinoma

Orfeas S. Aidonopoulos

Supervisor: Stavros Perantonis, PhD, Scientific Director, N.C.S.R. Demokritos

ATHENS

NOVEMBER 2014

ABSTRACT

Systems biology has become an essential component of drug discovery, attempting to combine experimental data with computational modeling to capture different levels of cellular function (such as signaling, transcription, regulation) and integrate them in predictive models. These models are then used to best understand the drug mode of action (MOA), identify new targets and predict clinical drug efficacy and toxicity. In this study, we tried to identify signaling pathways related to drug efficacy in one of the most lethal malignancies worldwide, hepatocellular carcinoma (HCC). Particularly, gene expression data were collected for various HCC cell lines treated with anticancer compounds of known clinical efficacy. Each compound had been categorized with a 'pass' or 'fail' label according to their success or failure in human clinical trials. For labeling each drug we constructed a graphical user interface that parses clinical trials databases for clinical outcomes containing the respective compound. Thus, having available a dataset consisting of labeled drugs as observations and genes as features a supervised learning method was applied (feature selection) to identify genes predictive of the drugs' clinical efficacy. Finally, using the extracted data as an input to a pathway construction algorithm, we were able to infer signaling networks on the proteomic level that best fit the measured gene expression signatures. We identified reactions and pathways playing an important role as accurate predictors for the efficacy of nine drugs in HCC.

SUBJECT AREA: Systems Biology, Bioinformatics

KEYWORDS: Machine learning, signaling pathways, pathway construction, gene expression, data analysis