Maastricht University

OpenRiskNet
RISK ASSESSMENT E-INFRASTRUCTURE

Toxico Genomix

Grow School for Oncology and Developmental Biology

Whole genome transcriptional profiling for genotoxicity prediction

Jumamurat R. Bayjanov, **Jos Kleinjans and Danyel Jennen**

Department of Toxicogenomics, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.
Contact: j.bayjanov@maastrichtuniversity.nl

Background

Whole genome transcriptional profiling allows global measurement of gene expression changes induced by particular experimental conditions. Toxic treatments of biological systems, such as cell models, may perturb interactions among genes and, in toxicogenomics, such perturbations assessed by transcriptional profiling are used to predict the impact of toxic compounds.

Introduction

Toxicogenomics-based approaches for predicting apical toxicities, have been dedicated to the purpose of improving predictions of genotoxicity and carcinogenicity *in vivo*. Over the past decade large amounts of transcriptional profiling data have been generated from *in vitro* study models using various chemical compounds, across multiple doses and time points as well as different organisms. We collected gene expression data from *in vitro* liver cell models of three different species exposed to different compounds and used these data to predict *in vivo* genotoxicity of these compounds. The work is part of the H2020 EU project (www.openrisknet.org).

Methods

A 5-step workflow was applied for the prediction analyses (Figure 1):
From the diXa Data Warehouse (www.dixa-fp7.eu), NCBI GEO (www.ncbi.nlm.nih.gov/pubmed/), and EBI ArrayExpress (www.ebi.ac.uk/arrayexpress/) we collected gene expression data from several human *in vitro* liver cell models exposed to 125 compounds with known genotoxicity information at different time points and dosages resulting in 822 experiments. Similarly, we also collected data for rat and mouse *in vitro* liver cell models (Table 1). We analyzed these data sets using ten different classification algorithms, thereby using 80% of the data for training and 20% for testing. We also applied 10-fold cross-validation to get more reliable results. Significant transcripts were shortlisted using iterative feature selection step and the final sets of significant transcripts for each of the three species were subjected to pathway over-representation analysis for biological interpretation.

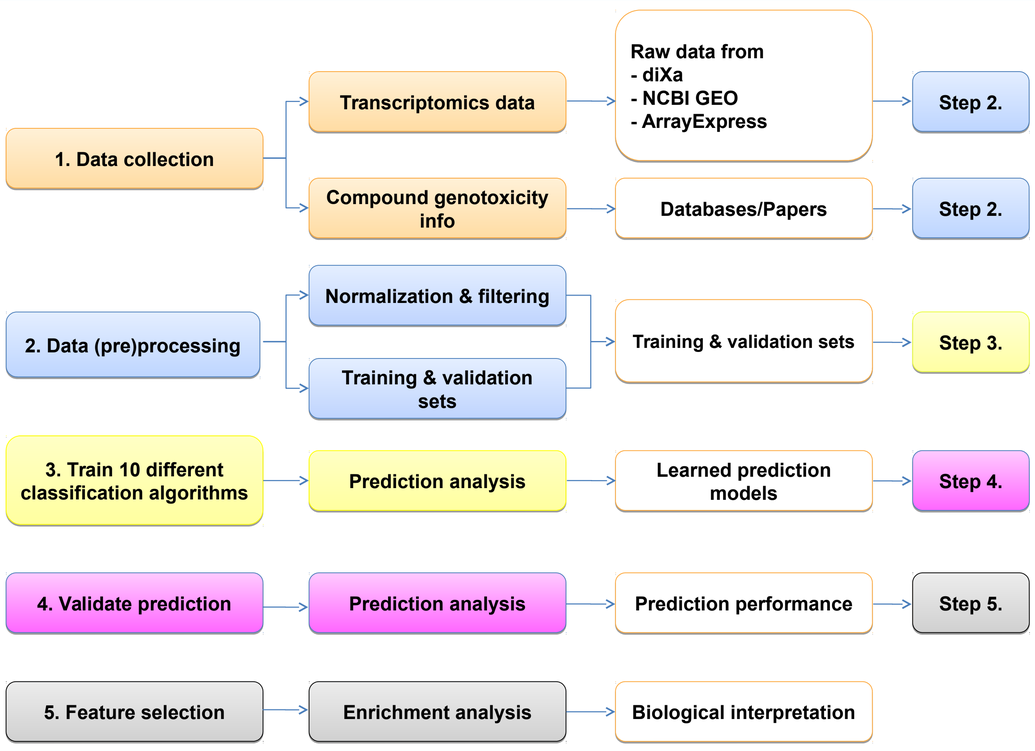

Figure 1: Workflow for genotoxicity prediction using whole genome transcriptomics data.

Funding

OpenRiskNet is a 3 year project funded by the European Commission within Horizon2020 EINFRA-22-2016 Programme (Grant Agreement 731075)

Results

Prediction analyses:
For the human gene expression data of the training sets, Support Vector Machines algorithm showed the best genotoxicity *in vivo* prediction (Figure 2). Logistic Regression performed best for rat and mouse data sets. The validation sets had accuracies at 91-93% with 95-100% specificity and 75-87% sensitivity (Table 2).

Table 1: Number of tested compounds and experiments, where various doses of a compound were tested at different time points.

Organism	Compounds	Experiments
Human	125	822
Rat	69	619
Mouse	45	100

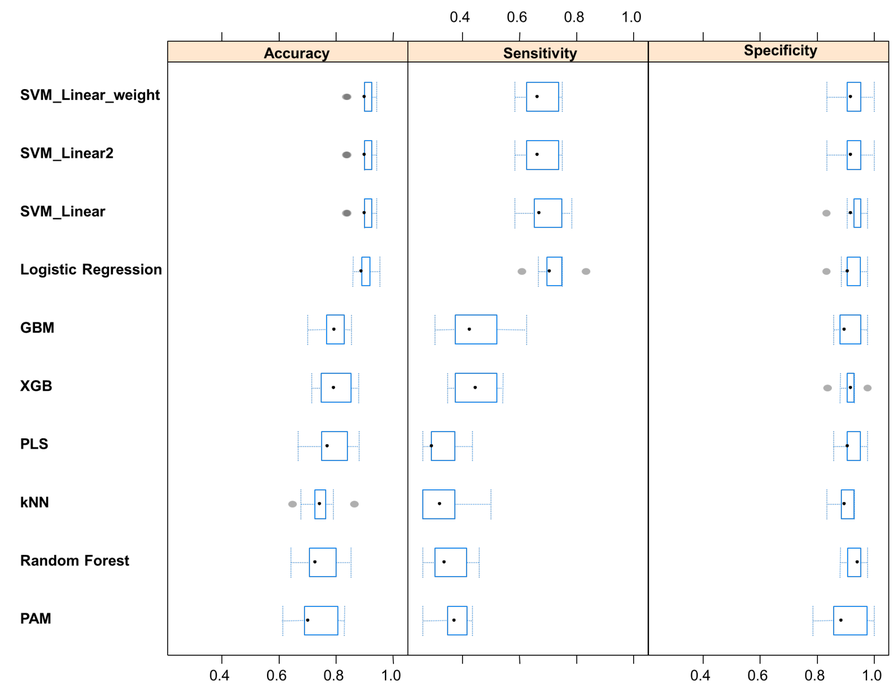

Figure 2: Performance of 10 selected algorithms for genotoxicity prediction using gene expression data of training sets for human in vitro cell models.

Table 2: Validation sets prediction performance of best classification algorithm per species.

Organism	Accuracy	Sensitivity	Specificity	Best algorithm
Human	92.5%	87%	95.3%	SVM Linear weighted
Rat	93.4%	87.2%	96.4%	Logistic Regression
Mouse	91.6%	75%	100%	Logistic Regression

Biological interpretation:
Upon identifying deregulated gene-gene interaction networks by applying ConsensusPathDB, the top 5 of affected pathways are related to p53-centered pathways.

Conclusion

The results from our meta-analysis demonstrate both high accuracy and robustness of transcriptomic profiling of genotoxicity hazards across a large set of genotoxicants and across multiple human, rat or mouse liver cell models. The resulting predictive assays can potentially be used for regulatory purposes, certainly when applied in combination with the traditional genotoxicity *in vitro* test battery. Next, we want to identify core orthologous genes among the three different species that are potential predictive targets for assessing genotoxicity and carcinogenicity across different biological systems.

GROW School for Oncology & Developmental Biology