DILI classification model based on *in vitro* human transcriptomics and *in vivo* rat clinical chemistry data.

Danyel Jennen, Jan Polman, Mark Bessem, Maarten Coonen, Florian Caiment, Dennie Hebels, Joost van Delft, Jos Kleinjans

Department of Toxicogenomics, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands
The Netherlands Toxicogenomics Centre, Maastricht University, Po Box 616, 6200 MD Maastricht, the Netherlands

Corresponding author: Dr. Danyel Jennen, Universiteitssingel 40, 6227 ER Maastricht, the Netherlands. Phone +31 433883983, fax +31 433884146, email danyel.jennen@maastrichtuniversity.nl

---

The past decades drug induced liver injury (DILI) is the main cause of drugs to fail during clinical trials or to be withdrawn from the market (Chen *et al.* 2011). Approximately 40% of DILI cases are not detected in preclinical studies based on conventional indicators in *in vivo* rodent studies (Zhang *et al.* 2012). Therefore, alternative methods for predicting the DILI potential in humans are needed and toxicogenomics-based approaches have been considered.
Recently, we developed an *in vitro* transcriptomics-based method in the human hepatic cell line HepG2 for predicting *in vivo* genotoxicity, which showed 89% accuracy, thereby clearly outperforming the standard *in vitro* test battery (Magkoufopoulou *et al.* 2012). For the CAMDA challenge an adapted version of this *in vitro* method was used to develop an *in vitro* classification model for predicting DILI in humans.
The development of the *in vitro* classification model for DILI in human consisted of 3 steps:

1. selecting drugs from the three DILI potential groups (i.e. “no DILI”, “less DILI” and “most DILI”) for the training and validation sets;
2. establishing gene signatures between the different DILI potential groups of the training set using a leave-one-out t-test or ANOVA;
3. using these gene signatures to train and validate the prediction model in PAM (prediction analysis for microarrays) (Tibshirani *et al.* 2002).

Selection of drugs
From each DILI potential group, i.e. “no DILI” (ND), “less DILI” (LD) and “most DILI” (MD), drugs were selected based on the *in vivo* clinical chemistry measurements of alkaline phosphatase (ALP), aspartate aminoptansferase (AST), alanine aminotransferase (ALT), lactate dehydrogenase (LDH) and Y-glutamyltranspeptidase (GTP) from rats with a daily repeated treatment. In particular, 20 MD drugs were selected that showed elevated levels for four or five of the measurements. Six ND drugs that showed decreased or unchanged levels were selected. For 35 LD drugs two or three of the measurements showed elevated levels. Dose and time were not taken into consideration in the selection.

The selected drugs were used in different settings resulting in four training sets:

- all selected drugs; MD, LD and ND or total DILI (D) and ND (61 drugs)
- drugs from MD and ND (26 drugs)
- drugs from LD and ND (41 drugs)
- drugs from MD and LD (55 drugs)

The distribution of drugs over the DILI groups for the training and validation set is summarized in Table 1.

Table 1. Distribution of drugs over the DILI groups for the training and validation set.

	training set	validation set	total
MD	20	21	41
LD	35	13	48
ND	6	2	8
total	61	36	97

Gene signatures

Microarray data from human primary hepatocytes exposed to high doses for 24 hours were used to establish gene signatures for each training set of drugs. The expression data were re-annotated to the MBNI Custom CDF-files and RMA normalized using the web tool arrayanalysis.org (Eijssen *et al.* 2013).

Genes with significantly different expression values ($p<0.01$) between the different DILI groups for each training set were selected from the expression data based on a series of statistical tests (ANOVA with three groups and t-test with two groups). For each test the two replicates of one of the drugs were removed (leave-one-out procedure). The significant genes that were present in all tests (the intersection of all lists) were selected for training the prediction model as signature. The resulting five gene signatures lists contained 31 to 141 genes as indicated in Table 2.

Training and validation of prediction models

PAM analysis (Tibshirani *et al.* 2002) was conducted for each of the signature lists for class prediction (threshold: 0). Misclassification errors (ME) were calculated for each prediction model and were highest (0.25) for ANOVA MD-LD-ND. The other four models had a ME <0.1.

Per prediction model the accuracy for each DILI group was calculated as indicated in Table 2. The accuracy within the training set is >90% for all prediction models except ANOVA MD-LD-ND (accuracy 67%-90%). This model also shows lowest accuracy for the validation (<62%). The other four models, MD-ND, LD-ND, MD-LD and D-ND, had a total accuracy for the validation of 87%, 80%, 50% and 89%, respectively.

The MD-ND and LD-ND models were further examined by testing the LD and MD drugs, respectively. This resulted for the LD drugs that 85% were predicted as MD and for the MD drugs that 95% were predicted as LD. This is also in line with the results (accuracy 89%) of the D-ND model. These findings indicate that both MD-ND and LD-ND models can be used for the prediction of DILI. In addition, the gene signature list from the MD-ND, LD-ND and D-ND models share 36 genes (Figure 1).

These genes were examined for GO processes in DAVID (Huang da *et al.* 2009) and were mainly involved in cell cycle, cell growth & proliferation and signal transduction related processes.

Table 2. Accuracy for training and validation sets for each prediction model. The number of signature genes and misclassification errors (ME) are indicated.

ANOVA MD-LD-ND (105 genes; ME 0.25)

	training	**validation**
MD	90%	33%
LD	88%	62%
ND	67%	0%
tot	87%	42%

t-test MD-ND (83 genes; ME 0.038)

	training	**validation**
MD	95%	95%
ND	100%	0%
tot	97%	87%

t-test LD-ND (79 genes; ME 0.024)

	training	**validation**
LD	97%	92%
ND	100%	0%
tot	98%	80%

t-test MD-LD (31 genes; ME 0.091**)**

	training	**validation**
MD	100%	33%
LD	91%	77%
tot	95%	50%

t-test D-ND (141 genes; ME 0.049)

	training	**validation**
D	95%	94%
ND	100%	0%
tot	95%	89%

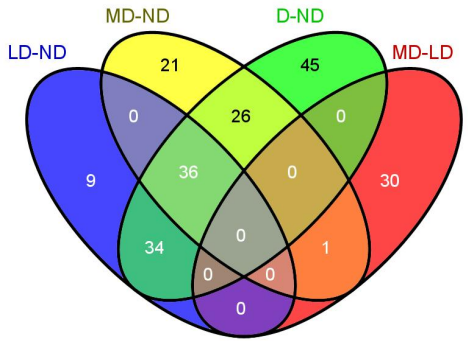

Figure 1. Comparison of the t-test based gene signatures (i.e. number of genes) for the different DILI group combinations.

Conclusions

The results of the *in vitro* human transcriptomics based models are very promising with up to 89% correct prediction for DILI potential. However, it should be noted that the two ND drugs in all validation sets are wrongly predicted and that improvement is definitely needed for distinguishing MD drugs from LD drugs.

Further analyses will be performed in which the following aspects will be considered:

- inclusion of time and dose relationships and/or additional clinical chemical measurement in the selection of drugs for the training set;
- increasing the number of ND drugs from other data repositories;
- performing analysis on transcriptomics data from other time and dose levels;
- enhancing the biological interpretation of gene signature lists.

References

Chen M, Vijay V, Shi Q, Liu Z, Fang H, Tong W (2011) FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov Today 16 (15-16):697-703

Eijssen LM, Jaillard M, Adriaens ME, Gaj S, de Groot PJ, Muller M, Evelo CT (2013) User-friendly solutions for microarray quality control and pre-processing on ArrayAnalysis.org. Nucleic Acids Res

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4 (1):44-57

Magkoufopoulou C, Claessen SM, Tsamou M, Jennen DG, Kleinjans JC, van Delft JH (2012) A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo. Carcinogenesis 33 (7):1421-1429

Tibshirani R, Hastie T, Narasimhan B, Chu G (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99 (10):6567-6572

Zhang M, Chen M, Tong W (2012) Is toxicogenomics a more reliable and sensitive biomarker than conventional indicators from rats to predict drug-induced liver injury in humans? Chem Res Toxicol 25 (1):122-129