

## Identifying Key Genes involved in Drug-Induced Liver Injury using Machine Learning on Human in vitro data sets

Created on: 14-02-2024 - Last modified on: 15-02-2024

### Organisation

**Name of the organisation** Vrije Universiteit Brussel (VUB)

**Department** In Vitro Toxicology and Dermato-Cosmetology

**Specific Research Group or Service** In Vitro Toxicology and Dermato-Cosmetology

**Country** Belgium

**Geographical Area** Brussels Region

### SCOPE OF THE METHOD

<b>The Method relates to</b>	Human health
<b>The Method is situated in</b>	Translational - Applied Research
<b>Type of method</b>	In silico

### DESCRIPTION

#### Method keywords

Feature selection  
Machine learning  
Supervised classification  
Wrapper feature selection  
transcriptomic signatures  
machine learning algorithms  
prediction models  
Toxicogenomics  
artificial intelligence

#### Scientific area keywords

Drug-induced cholestasis  
drug development  
preclinical drug toxicity testing  
Hepatotoxicity  
hepatic toxicity

#### Method description

Drug-induced intrahepatic cholestasis (DIC) is a main type of hepatic toxicity that is challenging to predict in early drug development stages. Preclinical animal studies often fail to detect DIC in humans. *In vitro* toxicogenomics assays using human liver cells have become a practical approach to predict human-relevant DIC. The present study was set

up to identify transcriptomic signatures of DIC by applying machine learning algorithms to the Open TG-GATEs database. A total of nine DIC compounds and nine non-DIC compounds were selected, and supervised classification algorithms were applied to develop prediction models using differentially expressed features. Feature selection techniques identified 13 genes that achieved optimal prediction performance using logistic regression combined with a sequential backward selection method. The internal validation of the best-performing model showed accuracy of 0.958, sensitivity of 0.941, specificity of 0.978, and F1-score of 0.956. Applying the model to an external validation set resulted in an average prediction accuracy of 0.71. The identified genes were mechanistically linked to the adverse outcome pathway network of DIC, providing insights into cellular and molecular processes during response to chemical toxicity. Our findings provide valuable insights into toxicological responses and enhance the predictive accuracy of DIC prediction, thereby advancing the application of transcriptome profiling in designing new approach methodologies for hazard identification.

### Lab equipment

- Open TG-GATEs database,
- GEO database,
- Affymetrix GeneChip,
- Classification algorithms.

### Method status

Internally validated  
Published in peer reviewed journal

## PROS, CONS & FUTURE POTENTIAL

### Advantages

- Such a signature could aid in *in vitro* DIC prediction, facilitating early detection of this chemical-induced toxicity.
- The identified features have biologically interpretable functions, mechanistically anchored in an AOP network, and provide new insights into molecular and cellular behavior processes during DIC development, making them valuable tools for understanding and predicting toxicological responses.

### Challenges

Differential expression (DE) analysis often generates numerous correlated candidate genes, leading to redundant information and reduced translatability for laboratory testing and lowered translatability of the DE findings for high-throughput laboratory testing. To address this, a permutation-based approach was employed to refine the results that estimate feature relevance by measuring changes in model performance upon permuting feature vectors .

## REFERENCES, ASSOCIATED DOCUMENTS AND OTHER INFORMATION

### Links

[Unraveling the mechanisms underlying drug-induced cholestatic liver injury: ide...](#)

### Other remarks

Coordinated by  This work was performed in the context of the European Science and Innovation (ESI) project cluster (ASPIS project cluster) (  [vub.be](#) )

Financed by                