An AI Approach to Drug-Induced Liver Injury Risk: Prediction of Safe Maximum Doses from Toxicogenomic Profiles

Paul Walker[1], Deniz Yuezak[2], Oliver Koch[2], Alicia Rosell-Hidalgo[1], Christopher Bruhn[2], Kirsten Headspith[1], Timur Samatov[2], Alexandra Hackmann[2], Gerald Ryan Aquino[2], Micael Fernandes dos Reis[2], Christian Vogeley[2], Nishika Sahini[2], Vladimir Galatenko[2], Gleb Goussarov[2], Ruediger Fritsch[2] & René Rex[2]

[1] *Cyprotex Discovery Ltd UK, No. 24 Mereside, Alderley Park, Macclesfield, Cheshire, SK10 4TG, United Kingdom;* [2] *Evotec International GmbH, Marie-Curie-Str. 7, Göttingen, D-37079, Germany*

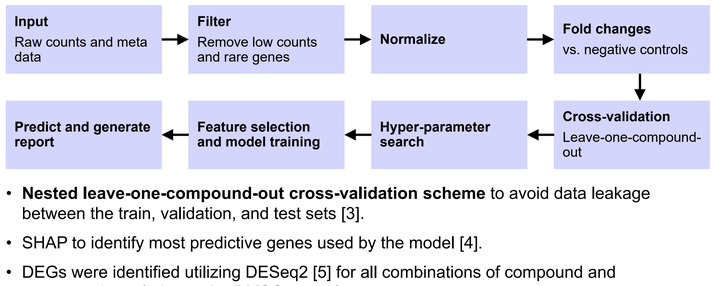
Poster number: P477

Introduction

- **DILI is a leading cause of high attrition rates** of drug candidates during development.
- There is a need to **rank drug candidates by DILI risk** and **estimate a safe dose** for clinical trials.
- We use high-throughput **transcriptomics profiles to train a supervised machine learning model** to suggest a (safe) C_{max} for novel compounds, which minimizes the risk of causing DILI in humans.
- We **predict the hepatotoxicity risk** and elucidate underlying mechanisms of **TAK-875** (fasiglifam) which was withdrawn from phase III trials due to DILI [1].

Materials and methods

- Input for model training: transcription profiles of compound-treated cells, dose range and the experimentally determined (true) C_{max}.
- **128 compounds** from the FDA Liver Toxicity Knowledge Base:
  - 68 DILI-associated compounds and 60 non-DILI-associated compounds [2].
- **Transcriptomic profiles** (using ScreenSeq™): generated from primary human hepatocytes (PHH), dosed in triplicates, eight-point dose response range, 24-hour time-point.
- The ML training and prediction code written in Python has 8 steps:



- **Nested leave-one-compound-out cross-validation scheme** to avoid data leakage between the train, validation, and test sets [3].
- SHAP to identify most predictive genes used by the model [4].
- DEGs were identified utilizing DESeq2 [5] for all combinations of compound and concentration relative to the DMSO control.
- **Dose-response analysis** was performed using BMDExpress (v 2.3 [6]) on normalized gene expression values to return BMD values.

Results and discussion

- A compound is predicted as **DILI-positive if the true C_{max} is above the predicted safe C_{max}**, and DILI-negative otherwise.
- At a 40 times C_{max} level and a 90% DILI score threshold, the model achieves the following metrics on the test set:

Metric	Mean and Standard Deviation
Accuracy	0.88 ± 0.008
Sensitivity	0.86 ± 0.032
Specificity	0.90 ± 0.020

Glossary

DILI	Drug-induced liver injury	**BMD**	Benchmark dose
C_{max}	Maximum plasma concentration	**DEG**	Differentially expressed gene
LTKB	Liver Toxicity Knowledge Database	**SHAP**	SHapley Additive exPlanations
PHH	Primary Human Hepatocytes		

DILI risk of TAK-875 and other compounds predicted by our model

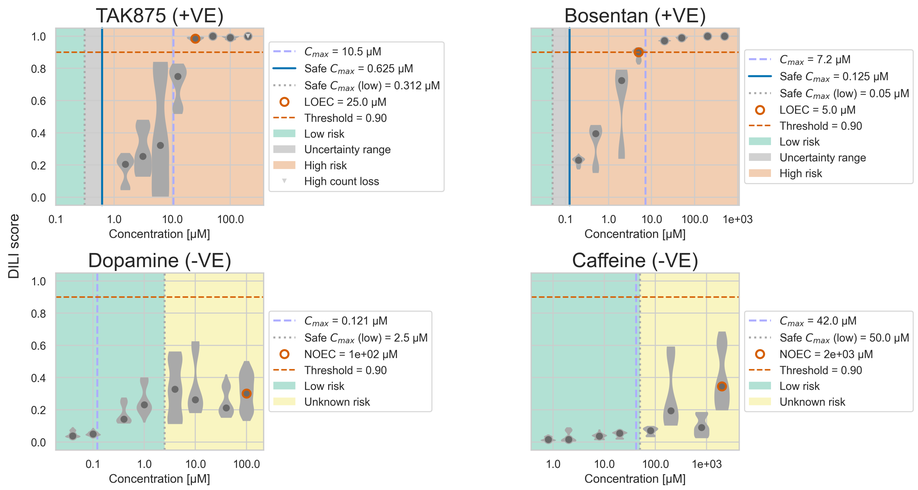


Figure 1: Dose response plots for DILI scores. Compound concentrations are shown on a log-scaled x-axis while DILI scores are shown on the y-axis. Gray violins show the distribution of the predictions for all replicates. The Lowest Observed Effect Concentration (LOEC) is shown by an orange outline on the concentration where the score exceeds the threshold of 0.9 (dashed orange line). For DILI-negative compounds, the No Observed Effect Concentration (NOEC) is marked instead. The solid blue line shows the predicted safe C_{max} and the dotted gray line the safe C_{max} (low). A down-pointing triangle indicates high count loss (likely due to cytotoxicity).

- The observed C_{max} of TAK-875 is an order of magnitude higher than the predicted safe C_{max} (figure 1).
- Our model would have **flagged TAK-875 for increased DILI risk.**
- The importance of regulated cholestasis-associated genes is captured well by the model (figures 2 and 3).
- We demonstrate the power of transcriptomics to derive signatures related to relevant mechanism of toxicity. These signatures can be used to identify specific modes of toxicity during drug discovery (figure 4).
- By using Evotec's highly multiplexed **transcriptomics pipeline** (ScreenSeq™) in combination with **AI-driven predictions of DILI**, we provide standardized, interactive reports that can **detect high-risk candidates early** in the drug discovery process.

Schematic representation of cholestasis-associated genes

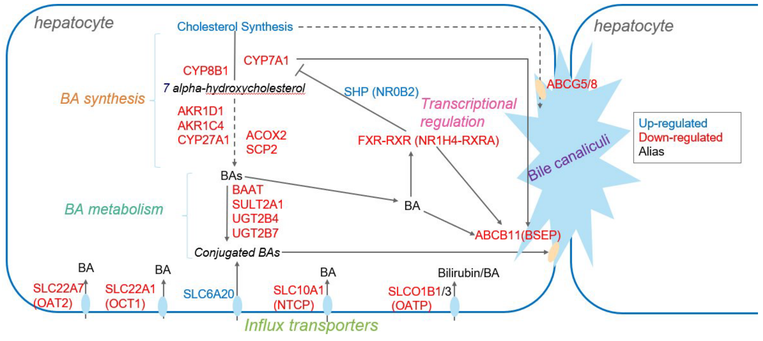


Figure 2: Schematic of cholestasis associated genes: Regulation of bile acid synthesis, metabolism, and transport. The schematic diagram illustrates gene expression changes associated with cholestasis, focusing on key processes involved in bile acid (BA) homeostasis. Upregulated genes involved in BA synthesis, metabolism, and FXR-RXR regulation are depicted in blue, while downregulated genes are shown in red. The diagram also highlights genes encoding BA influx transporters and bile canaliculi transporters, which play crucial roles in BA uptake and excretion.

A Influence (SHAP) of cholestasis genes on the prediction

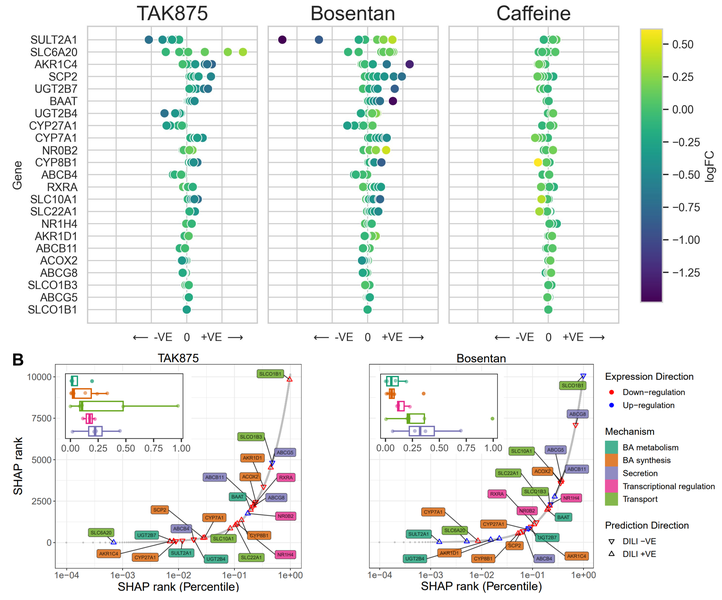


Figure 3: A) Distribution of SHAP values. Each dot represents the influence (on the x-axis) of a gene in a sample at certain concentration. Higher SHAP values drive the prediction towards +VE and vice versa. B) Accumulated SHAP rank percentiles at LOEC for all features highlighting the cholestasis-related genes. Lower percentiles are associated with higher contribution of the feature to +VE prediction. On average, genes associated with BA metabolism and synthesis contribute the most to DILI prediction.

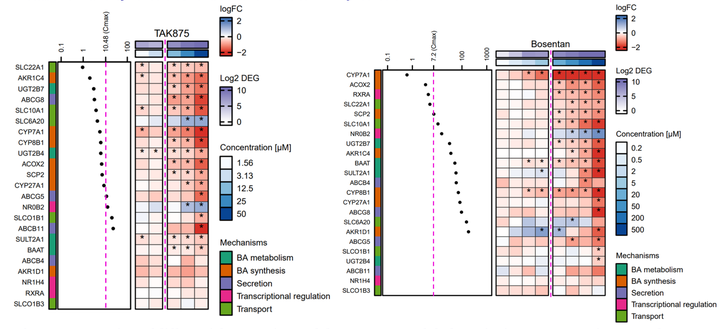


Figure 4: Overview of differential expression and dose-response of cholestasis signature genes in PHH. The direction of gene regulation is similar between TAK875 and Bosentan but the response occurs at lower concentrations in the former (within 1X Cmax).

References

1. Takeda News Release. Takeda announces termination of fasiglifam (TAK-875) development. Dec. 25 2013.
2. Chen, Minjun et al. *Drug discovery today* 21 4 (2016): 648-53.
3. Rex, Rene. *Methods in molecular biology* 2390 (2021): 421-431.
4. Lundberg, Scott M. and Su-In Lee. *Neural Information Processing Systems* (2017).
5. Love, M. I., Huber, W., and Anders, S. *Genome biology*, *15*(12), 1-21 (2014).
6. Phillips, Jason R., et al. *Bioinformatics* 35.10 (2019): 1780-1782.