



Research Article

Integrating Adverse Outcome Pathways (AOPs) and High Throughput *In Vitro* Assays for Better Risk Evaluations, a Study with Drug-Induced Liver Injury (DILI)

Kapil K. Khadka, Minjun Chen, Zhichao Liu, Weida Tong and Dong Wang

National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA

Abstract

The emergence of high throughput *in vitro* assays has the potential to significantly improve toxicological evaluations and lead to more efficient, accurate, and less animal-intensive testing. However, directly using all available *in vitro* assays in a model is usually impractical and inefficient. On the other hand, mechanistic knowledge has always been critical for toxicological evaluations and should not be ignored even with the increasing availability of data. In this paper, we illustrate a promising approach to integrating mechanistic knowledge with multiple data sources for *in vitro* assays, using drug-induced liver injury (DILI) as an example. The adverse outcome pathway (AOP) framework was used as a source for mechanistic knowledge and as a selection tool for high throughput predictors. Our results confirm the value of AOPs as a knowledge source for understanding adverse events and show that the majority of drugs classified as most-DILI-concern were mapped to AOPs related to liver toxicity found in AOPwiki. AOPs were also used effectively to select a subset of assays from the Tox21 and L1000 projects as the predictors in predictive modeling of DILI risk. Together with previously published drug properties for daily dose, lipophilicity, and reactive metabolite formation, these assay endpoints were used to build a penalized logistic regression model for assessing DILI risk. This model obtained an accuracy of 0.91, thus confirming the potential power of integrating mechanistic knowledge with high throughput assays for toxicological evaluations. The results also provide a roadmap for data integration that could be used for other adverse effects.

1 Introduction

The increasing need to perform toxicological evaluations on a large collection of drug candidates and other chemicals efficiently has led to widespread awareness of the inadequacies of animal-based testing. A promising development in addressing this problem is the emergence of high throughput *in vitro* assays (Kavlock and Dix, 2010; Tice et al., 2013). Combined with *in silico* predictions, high throughput *in vitro* assays have been proposed as potentially useful for generating a detailed toxicological profile of a chemical to be used in risk assessment. This paradigm has been discussed in two National Research Council (NRC) reports (NRC, 2007, 2017). It is also compatible with the increased societal preference for reducing animal use whenever possible

using *in vitro* assays and computational modeling approaches (Locke et al., 2017).

Several large-scale projects aiming to develop high throughput *in vitro* tests for toxicological evaluations have generated data sets for substantial collections of chemicals. The Tox21 program aims to identify specific mechanisms leading to adverse effects in humans (Collins et al., 2008; Tice et al., 2013), with more than 50 assays conducted on around 9000 chemicals. The Environmental Protection Agency's (EPA) ToxCast program focuses on evaluating biological responses relevant to human adverse effects with high throughput assays (Judson et al., 2014; Richard et al., 2016). To date, it has completed evaluations of more than 2000 chemicals regarding a range of cellular response and signaling pathways. As a follow-up to the Connec-

Disclaimer: The opinion expressed in this paper is that of the authors. It does not necessarily reflect the position of the U.S. Food and Drug Administration.

Received August 15, 2019; Accepted November 8, 2019;
Epub November 8, 2019; © The Authors, 2020.

ALTEX 37(2), 187-196. doi:10.14573/altex.1908151

Correspondence: Dong Wang, PhD
Division of Bioinformatics and Biostatistics
National Center for Toxicological Research/FDA
3900 NCTR Road
Jefferson, AR 72079, USA
(dong.wang@fda.hhs.gov)

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



tivity Map (CMap; Lamb et al., 2006) project, the L1000 project has developed a high throughput transcriptomic assay using 978 “landmark” genes from human cells (Subramanian et al., 2017). Whole transcriptome profiles for multiple cell lines have been generated in response to around 20,000 small molecule perturbagens, thus providing an excellent resource for studying chemical toxicity at the transcriptome level. These are just a few prominent examples of high throughput data sets that could be used for toxicological evaluations in conjunction with bioinformatics approaches.

It is expected that by combining these data sources for high throughput assays with those for organism level phenotypes (e.g., ToxRefDB, Liver Toxicity Knowledge Base (LTKB)), we will be able to build predictive models for organism level toxicity. On the other hand, achieving this goal poses significant challenges. As is the nature of high throughput data, thousands of endpoints are potentially available for each chemical. These could be cellular, cell free, protein binding, gene expression reporter, cell viability, mitochondrial integrity, or other types. However, it is of common knowledge in data sciences that blindly increasing the dimension of predictors can lead to models with poor performance (e.g., Hastie et al., 2016). This is especially challenging in the context of toxicity testing, where an individual problem’s training data usually contains chemicals numbered in the hundreds, which is often fewer than the number of potential predictors (assays). Moreover, as a practical matter, performing all available assays for all compounds all the time is rarely desirable.

On the other end of the spectrum, mechanistic knowledge is a critical part of toxicity evaluations. With solid mechanistic knowledge of a chemical’s effect on biological systems, one can focus on specific entities (genes, proteins, organelles) that are expected to be perturbed by the chemical and screen out the irrelevant assays. This could drastically reduce the number of assays to be considered and simplify the modeling problem. One challenge here is the difficulty of comprehensively collecting the mechanistic data or knowledge. The adverse outcome pathway (AOP) framework provides a convenient way to access the mechanistic knowledge concerning toxicity (Ankley et al., 2010; Krewski et al., 2010). AOPs are represented as biological maps with a sequence of events, consisting of molecular initiating event (MIE), key events (KE), and adverse outcome (AO). These describe a toxicity pathway by linearly organizing the causal information (Allen et al., 2014; Villeneuve et al., 2014; Burden et al., 2015). Efforts have been made to map pathways from transcriptome data analysis to AOPs for biomarker discovery (van der Veen et al., 2014; Labib et al., 2016) and to accelerate AOP development by fusing transcriptomic data and literature knowledge (Bell et al., 2016; Nymark et al., 2018). However, using AOPs alone to extract relevant information from data of high throughput assays is still a challenging task.

In this paper, we demonstrate the power of integrating mechanistic knowledge with high throughput *in vitro* assays in evaluating and predicting toxicity potential, using drug-induced liver injury (DILI) as an example. The focus is less on model development for DILI and more on the potential of integrating AOPs with *in vitro* assays. We demonstrate that, even though still under development, available AOPs (augmented with literature) could be used to interweave diverse data types from multiple sources like DrugBank, Tox21, L1000, and LTKB. This method also provides a convenient mechanism for dimension reduction for high throughput data and greatly simplifies predictive modeling. The resulting predictive model for DILI achieved excellent accuracy with a high level of simplicity. This portends well for further integration of AOPs and high throughput *in vitro* assays for better risk evaluations, with various applications.

2 Materials and methods

2.1 Data sets

DILI ranking and drug properties

DILI risk categories and drug properties were obtained from LTKB and Chen et al. (2016). The drugs were classified into three categories (most-DILI-concern, less-DILI-concern or no-DILI-concern) based on FDA labels. Table S1 in Chen et al. (2016) contains DILI risk categories, daily doses, lipophilicity (logP), and reactive metabolite (RM) formation for 354 drugs (124 most-DILI-concern, 162 less-DILI-concern, 68 no-DILI-concern). We used this data set for subsequent analysis. The daily dose (mg/day) was transformed to the log₁₀ scale in the predictive model. The data for drugs used in our analysis is reproduced in Table S1¹ of the supplementary data.

AOPs

AOPs were obtained from the AOPwiki website². Note that a proportion of these AOPs is still under development and could undergo modifications before being accepted by the Organisation for Economic Co-operation and Development (OECD). To supplement the description of AOPs related to DILI in building predictive models, we also incorporated proposed AOPs from several published papers on hepatic steatosis, liver fibrosis, cholestasis, and liver tumors (Vinken, 2013; Mellor et al., 2016; Horvat et al., 2017; Gijbels and Vinken, 2017).

Drug targets and enzymes from DrugBank

Molecular targets and metabolizing enzymes for each drug were obtained from the DrugBank website³.

Tox21 assay endpoints

Tox21 endpoints were downloaded as part of the ToxCast database (invitroDBv3⁴). The results of Tox21 assays were reported

¹ doi:10.14573/altex.1908151s

² <https://aopwiki.org/> (accessed October 2018)

³ <https://www.drugbank.ca>

⁴ <https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data>

in dose-response format, with three replicates at each dose (see Filer et al., 2016 for details). To derive a standardized score for each chemical, the maximum median response across doses was divided by the baseline median absolute deviation.

Expression levels from L1000

Data for gene expression levels from L1000 assays were obtained from Gene Expression Omnibus⁵ with the accession numbers GSE92742 and GSE70138. The replicate-collapsed z-scores (Level 5 data) were extracted with the CMapR package (Enache et al., 2019) and further processed using the R statistical software (version 3.5.0). The data represent 978 measured landmark genes and 9196 inferred genes (83% of the transcriptome). This data set includes the expression profiles for multiple cell lines with the same chemical perturbation. For our analysis, experiments using the HepG2 cell line were used whenever available. Otherwise, data from experiments using the MCF7 cell line were used. If multiple experiments were available for the same cell line, the experiment designated as “exemplar” by the L1000 team was used. Otherwise, the one with the highest transcription activity score as defined in the data set was used. The transcriptional signatures used for each drug can be found in Table S1¹.

2.2 Matching drug information in DrugBank to AOPs

DrugBank was queried with 354 FDA labeled drugs (from Chen et al., 2016); information for drug targets and metabolizing enzymes was retrieved for each drug. The drug targets and enzymes were then used to query AOPwiki to retrieve any AOPs including these terms in the description of any of its KEs or MIE. Individual AOPs with shared events were merged to form AOP networks. By covering the available possible biological contexts and mechanisms of a toxicity pathway, AOP networks likely expand the relevant mechanistic information of a chemical pathway. These networks were then visualized and analyzed with Cytoscape⁶ and the R statistical software.

2.3 Building predictive models for DILI

Combined data set

AOPs related to drug-induced liver injury were collected from AOPwiki and published literature as described above. Potential predictors of DILI (gene expression, nuclear factor binding, cellular response) were extracted from the description of MIEs and KEs for these AOPs. Data regarding these potential predictors were then extracted from the Tox21 and L1000 data sets pertinent to the drugs in Chen et al. (2016). Not all assays were performed for all drugs; the data set contained a subset of predictors and drugs, which we describe in the Results section. This set of Tox21 and L1000 data was combined with daily doses, logP, and RM (reactive metabolite formation) in Chen et al. (2016) to form the final predictor set. The corresponding data for DILI risk categories served as the response variable in our model.

Logistic regression with elastic net penalty

To build a predictive model for assessing DILI risk, we selected drugs labeled as no-DILI-concern and most-DILI-concern, which were represented as $y = 0$ (no-DILI-concern) or $y = 1$ (most-DILI-concern), respectively, in our model. Using x to denote the vector of predictors, we have the logistic regression model:

$$\log \frac{P(y=1|x=x)}{P(y=0|x=x)} = \beta_0 + \beta^T x,$$

where β_0 and β are regression coefficients. We applied an elastic net penalty (Zou and Hastie, 2005) for regularization of the model fitting as the number of predictors was still high relative to the sample size after the AOP-based selection. Specifically, the regression coefficients were estimated as:

$$(\beta_0, \beta) = \operatorname{argmin}_{(\beta_0, \beta)} -\frac{1}{N} \sum_{i=1}^N [y_i(\beta_0 + x_i^T \beta) - \log(1 + e^{\beta_0 + x_i^T \beta})] + \lambda \left[\frac{(1-\alpha)\|\beta\|_2^2}{2} + \alpha\|\beta\|_1 \right],$$

Here N represents the number of drugs, $\|\beta\|_2^2$ and $\|\beta\|_1$ represent the ridge and Lasso penalties, respectively. The parameter α , which modulates the relative strength between the ridge and Lasso penalties, was set to 0.9 in our analysis. The parameter λ specifies the magnitude of penalties, selected by leave-one-out cross-validation. The fitting of the elastic net regulated logistic regression was performed with the *glmnet* package (Friedman et al., 2010) and custom R scripts.

The elastic net method is used here as a means to prevent overfitting, which would reduce the generalizability of the model beyond the data on which the model is fitted. Once the regression coefficients have been estimated, the standard interpretation for logistic regression still applies, i.e., the log odds of a drug being most-DILI-concern is defined by the linear predictor $\beta_0 + \beta^T x$. The linear predictor is a linear combination of the measurements from *in vitro* assays. As it represents the log odds of being most-DILI-concern, it can be used as a score for DILI potential. More details about logistic regression can be found in Kleinbaum and Klein (2010) and other sources.

3 Results

To illustrate the utility of AOPwiki as a knowledge source for DILI and to demonstrate the potential to use AOPs in predictive modeling, we report the results from two different approaches. In the first approach, we used known properties of drugs (targets and enzymes retrieved from DrugBank) to query AOPwiki. With this approach, we aimed to evaluate the potential of using existing AOPs as a knowledge repository to discover adverse outcomes (DILI in particular) while having partial information about the drug. In the second approach, the AOPs are used to guide selecting potential predictors (gene expression, nuclear re-

⁵ <https://www.ncbi.nlm.nih.gov/geo/>

⁶ <https://cytoscape.org/>

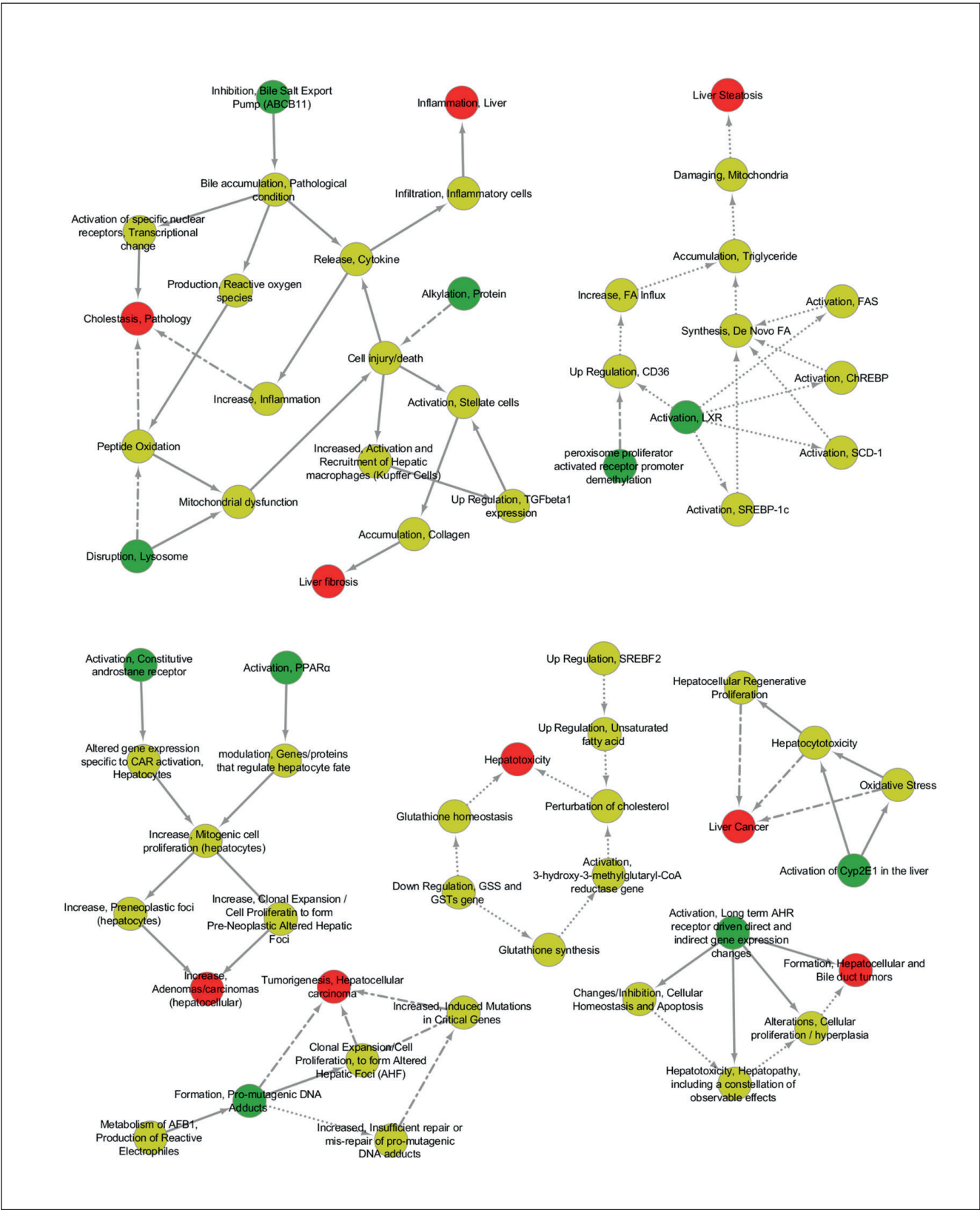


Fig. 1: AOP networks formed by AOPs with hits by most-DILI-concern drugs
The green nodes are MIEs, the red nodes are AOs, and the yellow nodes are KEs. AOPs were retrieved from AOPwiki.

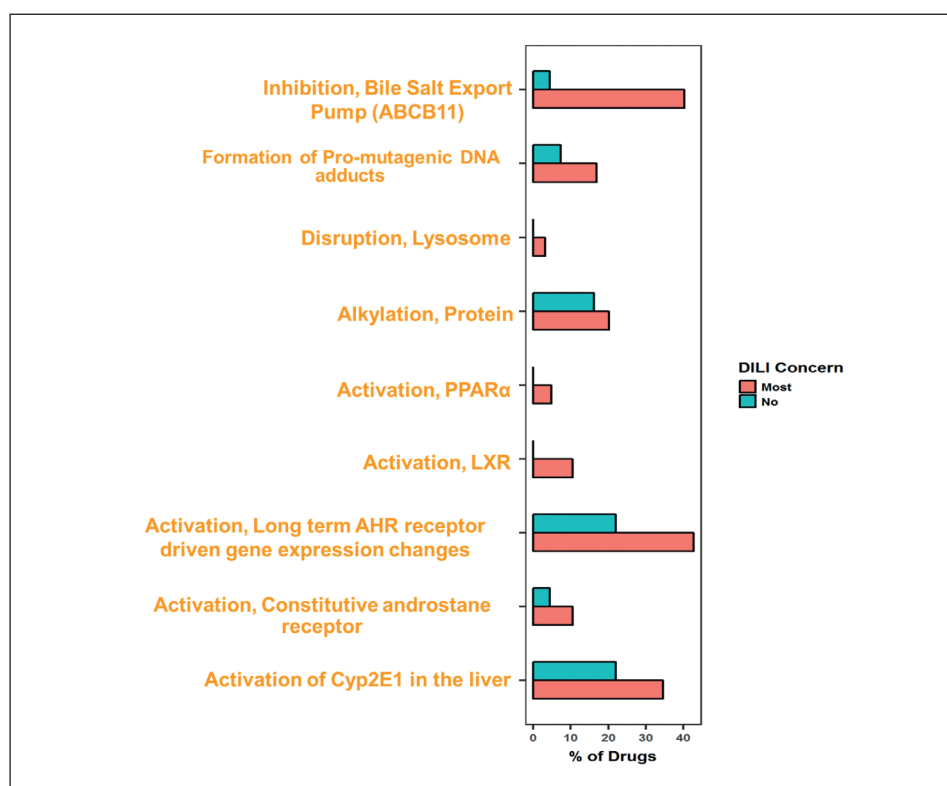


Fig. 2: Frequency of hits on some of the common AOPs with liver related AOs by most-DILI-concern and no-DILI-concern drugs

The AOPs are labeled by their MIEs. The percentage was based on 124 most-DILI-concern and 68 no-DILI-concern drugs.

ceptor binding, cellular functions) to build a predictive model using predictors from multiple data sources. This demonstrates the potential of AOPs to serve as the backbone for data integration from diverse sources as well as being a means of dimension reduction for effective modeling.

3.1 Matching DrugBank information with AOPs

Using the list of 124 drugs categorized as most-DILI-concern in Chen et al. (2016), we were able to match the drug targets and enzyme information in DrugBank to the MIEs or KEs of 23 AOPs in AOPwiki. Of these, 10 AOPs are liver related (defined as those with liver related adverse outcomes). As some of the events are shared among AOPs, these AOPs form seven distinct AOP networks. Figure 1 illustrates the AOP networks related to DILI. Detailed information regarding these AOPs is given in Table S2¹. The number of drugs that potentially perturb these seven different pathway networks varies. The pathway with the MIE “long term AhR receptor activation driven direct and indirect gene expression changes” was found to be the one with most hits by most-DILI-concern drugs (~43% of 124 drugs), followed by the pathway with the MIE “inhibition of bile salt export pump (ABCB11)” (~41%). Similarly, about 35% of the most-DILI-concern drugs investigated have a hit on the pathway “activation of Cyp2E1 in the liver” as the MIE.

In total, 64 drugs in the most-DILI-concern category have apparent associations with liver related AOPs solely based on in-

formation from DrugBank. Similarly, queries using 68 no-DILI-concern drugs resulted in hits on 17 different AOPs, of which six are related to liver and are shared with those identified by most-DILI-concern drugs. Not surprisingly, some pathways involved in a wide range of metabolisms are also hit by no-DILI-concern drugs, though at a lower frequency than that of the most-DILI-concern drugs. AOPs related to “long term AhR receptor activation driven direct and indirect gene expression changes”, and “activation of Cyp2E1 in the liver” were among the examples. In contrast, the AOP with the MIE “inhibition of bile salt export pump (ABCB11)” was found to be considerably less common (only ~4%) for no-DILI-concern drugs versus most-DILI-concern drugs (~41%). The frequencies of hits on the most common liver AOPs are illustrated in Figure 2 for both most-DILI-concern and no-DILI-concern drugs. In the supplementary data¹, we also provide code for an R-Shiny application⁷, with which users familiar with the R statistical software can explore the AOP networks related to DILI and examine drugs connected to each AOP. Figure S1¹ illustrates the hits on AOPs that do not have liver related adverse outcomes.

3.2 Develop predictive model for assessing DILI risk

To build a predictive model for DILI risk, we utilized drug property data from LTKB (daily dose, logP, and RM formation), nuclear receptor binding activities from Tox21, and gene expression data from L1000. To select a suitable subset of Tox21 and

⁷ <https://shiny.rstudio.com/reference/shiny/1.1.0/>

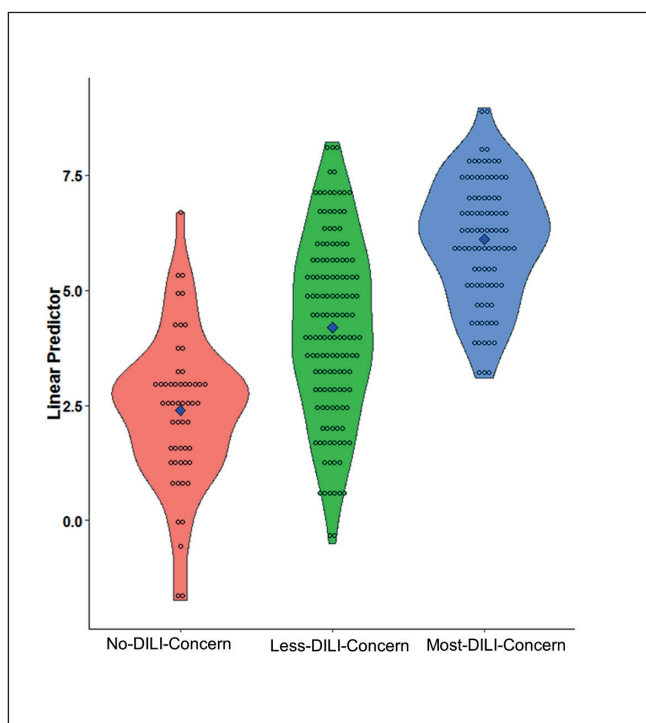


Fig. 3: The linear predictor in the penalized logistic regression model can be used as a quantitative measure of DILI risk

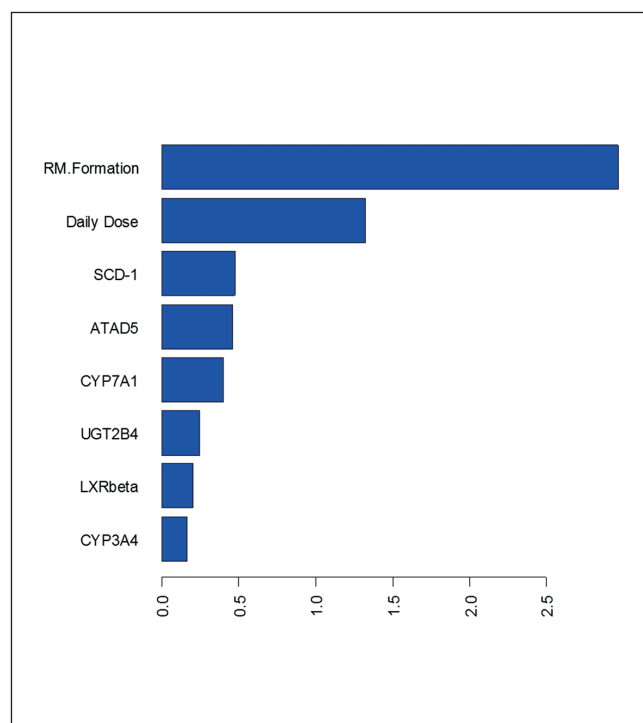


Fig. 4: Predictors with the largest coefficients in the penalized logistic regression model

L1000 data, information for relevant genes, nuclear receptors, and cellular functions was extracted from AOPwiki and literature, as described in the Methods section. These potential predictors are listed in Table 1. The genes ATAD5 and ATF6 were added as surrogates for DNA damage and endoplasmic reticulum stress as represented by KEs in the network. This resulted in a total of 50 potential predictors, of which the data were available from either Tox21 or L1000 for 41 predictors. These 41 high throughput assay endpoints were combined with the three drug properties, resulting in a total of 44 predictors forming the matrix of independent variables in the logistic regression model. After combining these predictors with the response variable for DILI risk (most-DILI-concern or no-DILI-concern), we had a data set of 92 drugs labeled most-DILI-concern and 54 drugs labeled no-DILI-concern. The values for all the predictors for these drugs are provided in Table S3¹. For predictors derived from Tox21 and L1000 data sets, the values were preprocessed, as described in the Methods section.

A logistic regression model was then fitted to the data with the elastic net penalty; the magnitude of penalty (determined by the value of λ) was selected by cross-validation. The result of model fit is reported in Table 2A; it obtained a sensitivity of 0.96 (88 of 92 most-DILI-concern drugs) and a specificity of 0.83 (45 of 54 no-DILI-concern drugs), with an accuracy of 0.91. For comparison, fitting the logistic model with three drug properties – daily dose, logP, and RM – in Chen et al. (2016) resulted in an accuracy of 0.84. Model fit with only Tox21 data or L1000 data led to models fitting only most-DILI-concern values due

to the imbalance in the data (92 most-DILI-concern, 54 no-DILI-concern), resulting in an accuracy of 0.63. Combining selected L1000 variables with the three drug properties increased the accuracy to 0.90, while adding selected Tox21 variables resulted in a slight further improvement to 0.91. Including a random set of molecular predictors did not improve the performance appreciably. To evaluate the utility of the model's predictions for drugs that were not in the model, leave-one-out cross-validation was performed using the same model fitting procedure, with the result reported in Table 2B. Under cross-validation, the model obtained a sensitivity of 0.88 and a specificity of 0.72, with an accuracy of 0.82. Without the AOPs as guides, the accuracy was reduced to 0.73. We also fitted weighted versions of the logistic model; the results suggest that the accuracy is not overly sensitive to the proportion of positives and negatives in the data. It is, however, possible that the performance might differ for target populations with very skewed ratios of positives to negatives.

Of note is that the linear predictor, $\beta_0 + \beta^T x$, in Equation (1) can serve as a quantitative indicator of DILI risk on a logistic scale. We calculated this linear predictor for 286 drugs with complete data in Chen et al. (2016), including 140 less-DILI-concern drugs in addition to the most-DILI-concern and no-DILI-concern drugs. The results are displayed in Figure 3, and showed a good separation of most-DILI-concern and no-DILI-concern drugs, while less-DILI-concern drugs occupied a range concentrated in the middle.

The regression coefficients of the top eight predictors obtained using the penalized logistic regression model with all three data

Tab. 1: Potential predictors using high throughput cellular assays

These predictors were extracted from AOP descriptions in AOPwiki and the literature. Those with corresponding measurements in the L1000 or Tox21 data (indicated in the Source column) were included in building the predictive model. NR, nuclear receptor binding; gene, changes in gene expression.

Name	Type	AOP network	Source
CD36	gene	steatosis	L1000
ChREBP	gene	steatosis	L1000
SREBP-1c	gene	steatosis	L1000
FAS	gene	steatosis	L1000
SCD-1	gene	steatosis	L1000
HSD17B10	gene	steatosis	L1000
PCK1	gene	steatosis	L1000
LDLR	gene	steatosis	L1000
CYP1A1	gene	steatosis	L1000
ACC-1	gene	steatosis	L1000
FOXA2	gene	steatosis	L1000
CPT1A	gene	steatosis	L1000
HMGCS2	gene	steatosis	L1000
HSD17B4	gene	steatosis	L1000
MTTP	gene	steatosis	L1000
ApoB100	gene	steatosis	L1000
NTCP	gene	cholestasis	L1000
OATP1B1	gene	cholestasis	L1000
CYP7A1	gene	cholestasis	L1000
OSTA	gene	cholestasis	L1000
OSTB	gene	cholestasis	L1000
MRP2	gene	cholestasis	L1000
UGT2B4	gene	cholestasis	L1000
CYP3A4	gene	cholestasis	L1000
SULT2A1	gene	cholestasis	L1000

Name	Type	AOP network	Source
MRP3	gene	cholestasis	L1000
CYP2E1	gene	tumor	L1000
apoptosis	general	fibrosis	
PPARgamma	NR	steatosis	Tox21
PPARalpha	NR	steatosis, tumor	
PPARbeta	NR	steatosis	
AHR	NR	steatosis, tumor	Tox21
NR1I3 (CAR)	NR	steatosis, cholestasis	
HNF4A	NR	steatosis	
PXR	NR	steatosis, cholestasis	
NRF2	NR	steatosis	Tox21
FXR	NR	steatosis, cholestasis	Tox21
SHP	NR	steatosis, cholestasis	
LXRalpha	NR/gene	steatosis	
LXRbeta	NR/gene	steatosis	
PPARgamma	NR/gene	steatosis	L1000
AKT2	gene	steatosis	L1000
ATAD5	gene	fibrosis	L1000
ATF6	gene	fibrosis	L1000
ESR1	NR	steatosis	Tox21
NR3C1	NR	steatosis	Tox21
MMP	mitochondrial	fibrosis	Tox21
NFkB	NR	fibrosis	Tox21
RARA	NR	steatosis	Tox21
RXRA	NR	steatosis	Tox21

Tab 2: Performance of the penalized regression model

A. Model Fit

		Actual	
		No concern	Most concern
Fitted	No concern	45	4
	Most concern	9	88

B. Leave-one-out cross-validation

		Actual	
		No concern	Most concern
Predicted	No concern	39	11
	Most concern	15	81

sources are plotted in Figure 4. The generation of reactive metabolites (RM) and the daily dose have the largest coefficients, followed by ATAD5 (an indicator of DNA damage), SCD-1 (a key enzyme in fatty acid metabolism), CYP7A1 (an important enzyme in cholesterol metabolism and bile acid synthesis), and others.

4 Discussion

The increasing availability of *in vitro* assays for toxicological evaluations (Kavlock and Dix, 2010; Tice et al., 2013) is a major 21st century development in the field of risk assessment for chemicals and drugs. Researchers can now obtain measure-



ments on hundreds of endpoints for various biological functions, which raises the prospect of fast, cheap, and less animal-intensive approaches to toxicity testing. However, incorporating diverse high dimensional data in toxicity testing, especially with the limited sample sizes used in drug development, is no easy task. Based on experiences in high dimensional modeling, it is well-known that naively including ever-increasing dimensions of data will lead to poor predictions (Hastie et al., 2016). It has also been demonstrated that, in many cases, parsimonious models based on concrete biological knowledge can outperform complex algorithms (see an example in Banerjee et al., 2017). Additionally, a biologically anchored testing scheme involving a moderate number of tests is preferable to indiscriminately carrying out hundreds of tests each time, for practical and financial reasons. In most cases, however, some mechanistic knowledge is available but is incomplete. It is thus extremely valuable to integrate the diverse sources of high dimensional *in vitro* assays with available mechanistic knowledge in toxicological modeling. Here, we present our work in this direction using DILI as an example, illustrating the excellent potential, and some challenges, of this approach.

Systematically reviewing mechanistic knowledge usually requires significant expertise and time. The AOP concept provides a convenient framework for encoding mechanistic knowledge collaboratively and communicating it in a standardized fashion. Although many AOPs are still under development, we wanted to evaluate the potential of AOPs as the source of easily accessible mechanistic knowledge for toxicological evaluations. In the first section of our study, we evaluated the utility of AOP networks that we queried with molecular properties of drugs in providing alerts for DILI risk. The molecular targets and enzyme information in DrugBank, though by no means complete, were used to perform the queries, as they reflect the typical level of knowledge on approved drugs at the molecular level. The results confirmed the value of AOPs in providing mechanistic knowledge. More than half of most-DILI-concern drugs were linked to AOPs related to DILI, with concentrations in well-known pathways such as AhR receptor activation and inhibition of bile salt export pump (Lee et al., 2010; Stieger, 2010; Morgan et al., 2010; Montanari et al., 2016). Thus, simply referencing AOPs could provide insight into potential adverse outcomes.

On the other hand, the result also points to significant challenges. Nearly half of the most-DILI-concern drugs in this exercise were not connected to AOPs related to liver toxicity. The cause was likely not an actual absence of connection, but a combination of incomplete knowledge in the form of DrugBank entries, and the difficulty of performing the mapping. Another challenge was that only some pathways, such as the inhibition of bile salt export pump, are highly specific; some AOPs were matched to both most-DILI-concern and no-DILI-concern drugs. This was not surprising, given that metabolic pathways are often shared by various biological functions, making those AOPs less useful for evaluating the DILI risk despite their frequent involvement in toxicity. The AOP networks generated from AOPwiki were not perfect and could be further rationalized manually, though

they are expected to improve with the continued refinement of AOPs. Some AOPs with adverse outcomes not related to the liver can nonetheless be informative to liver toxicity; however, we did not pursue using them in this proof-of-concept study. Despite these challenges, our results regarding DILI suggest that AOPs are a useful framework and knowledge source for discovering potential risks of adverse effects of drug candidates when certain molecular properties are present. In the future, both querying of AOPs and predictive model building could benefit from more integration of molecular pathways with AOP networks. Nymark et al. (2018) and others have already attempted this.

Having confirmed the relevance of AOPs in detecting DILI risk, we explored a second approach. To build a predictive model, we used AOPs and knowledge from the literature to identify predictors from high throughput assays. With an effective predictive model, one can systematically obtain measurements of relevant predictors and evaluate the risk provocatively. Using AOPs to narrow the range of predictors is highly effective in this setting because, although high throughput assays like those in ToxCast/Tox21 and L1000 can provide a comprehensive profile of a biological state under chemical perturbation, directly using all endpoints is not practical or efficient.

When different sources of data for *in vitro* assays need to be integrated, the problem is even more daunting. To keep the workload at a manageable level, we limited the consideration of liver toxicity related AOPs to AOPwiki and three recent papers. Focusing on gene expression, nuclear receptor binding, and cellular functions, we extracted a total of 50 potential predictors from these AOPs. To obtain data for these potential predictors, we focused on the Tox21 and L1000 projects due to their well-established presence in the research community and their wide coverage of drugs. This gave us data on 41 predictors. Also included were the three drug properties from LTKB and Chen et al. (2016): daily dose, logP, and RM, due to the well-established importance of these variables in DILI risk. The resulting 44 predictors found with all the measurements were of a manageable size for our data on 146 drugs. Still, a regularization procedure, in our case the elastic net, was needed to achieve optimal performance and prevent overfitting.

Due to its critical importance in drug development, DILI risk has been extensively studied. Chen et al. (2016) and other papers have suggested that daily dose, logP, and RM formation are valuable indicators for daily risk. Our result confirms this point, as the logistic model with these three variables is already useful in separating most-DILI-concern and no-DILI-concern drugs. Adding *in vitro* assay endpoints from AOP-selected L1000 and Tox21 assays further improved the model's performance, resulting in an accuracy of 0.91. Without applying existing knowledge from Chen et al. (2016) and AOPs, the model with only *in vitro* assays performs poorly. This demonstrates the importance of integrating existing knowledge whenever possible, even in the age of data-driven science. On the other hand, the *in vitro* assay data used for the modeling is far from ideal, as the Tox21 and L1000 assays were carried out on varying cell lines, using doses and treatment conditions that might not be a good match for our

purpose. Reliance on the HepG2 cell line may lead to underestimated endpoints due to its lack of metabolic capacity. Some of the L1000 gene expression endpoints were also inferred from the golden set rather than directly measured.

This is not meant to be a criticism of Tox21, L1000, or similar projects. As large-scale surveying projects, they provide the community with snapshots of biological responses for a huge number of chemicals with standardized protocols and cannot be expected to meet all modeling requirements optimally. But, with more dedicated and optimized measurements of the endpoints in our model, the *in vitro* assays could make an even larger contribution to the model's performance. The approach reported here provides a path to selecting a relatively small number of assays (Fig. 4) for further development and optimization that can be of practical use in toxicological evaluations.

This paper is not focused on developing the best model for DILI prediction; rather, we aimed to illustrate the potential of integrating mechanistic knowledge and high throughput assays in developing new approaches for toxicological evaluation. Our results confirmed the feasibility of AOPs as a knowledge source for understanding adverse events and showed that querying the AOP database with partial information of molecules could provide useful alert signals. AOPs were also used effectively to select a subset of assays from the Tox21 and L1000 projects as the predictors in predictive modeling of DILI potential. This led to a penalized logistic regression model that demonstrated superb performance in assessing DILI risk, thus confirming the potential power of integrating mechanistic knowledge with high throughput assays for toxicological evaluations. We also expect similar patterns to hold for the investigation of other toxicity problems and adverse events. Effective integration of mechanistic knowledge and high throughput cellular assays could point to promising approaches for reducing the use of animals for early identification of human health risk. With continued development, AOP networks can be expected to play significant roles in this process.

Supplementary Materials¹

Tab. S1: Descriptions of drugs used to build the predictive model

The table includes columns for LTKB ID, compound name, DILI concern category (Chen et al., 2016), daily dose, RM formation, logP, ToxCast code for retrieving Tox21 data, and the L1000 signature ID for retrieving differential gene expression data.

Tab. S2: Detailed information for AOPs in Fig. 1.

Tab. S3: Data matrix of predictors in the penalized logistic regression model

The LTKB ID is also included for linking with Tab. S1. The daily dose is on log10 scale, the RM formation is coded as 0 (negative) or 1 (positive). Measurements for molecular predictors from Tox21 and L1000 have been preprocessed according to procedures described in Methods and Materials.

Fig. S1: Frequency of hits on some AOPs with adverse outcomes not related to the liver by both most-DILI-concern and no-DILI-concern drugs

AOP.liver.r and **AOP.liver.RData** are the R code and data files to run an R-Shiny application for visualization of DILI related AOP networks.

References

- Allen, T. E., Goodman, J. M., Gutsell, S. et al. (2014). Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. *Chem Res Toxicol* 27, 2100-2112. doi:10.1021/tx500345j
- Ankley, G. T., Bennett, R. S., Erickson, R. J. et al. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 29, 730-741. doi:10.1002/etc.34
- Banerjee, S., Zare, R. N., Tibshirani, R. J. et al. (2017). Diagnosis of prostate cancer by desorption electrospray ionization mass spectrometric imaging of small metabolites and lipids. *Proc Natl Acad Sci* 114, 3334-3339. doi:10.1073/pnas.1700677114
- Bell, S. M., Angrish, M. M., Wood, C. E. et al. (2016). Integrating publicly available data to generate computationally predicted adverse outcome pathways for fatty liver. *Toxicol Sci* 150, 510-520. doi:10.1093/toxsci/kfw017
- Burden, N., Sewell, F., Andersen, M. E. et al. (2015). Adverse outcome pathways can drive non-animal approaches for safety assessment. *J Appl Toxicol* 35, 971-975. doi:10.1002/jat.3165
- Chen, M., Borlak, J. and Tong, W. (2016). A model to predict severity of drug-induced liver injury in humans. *Hepatology* 64, 931-940. doi:10.1002/hep.28678
- Collins, F. S., Gray, G. M. and Bucher, J. R. (2008). Transforming environmental health protection. *Science* 319, 906-907. doi:10.1126/science.1154619
- Enache, O. M., Lahr, D. L., Natoli, T. E. et al. (2019). The GCTx format and cmap {Py, R, M} packages: Resources for the optimized storage and integrated traversal of dense matrices of data and annotations. *Bioinformatics* 35, 1427-1429. doi:10.1101/227041
- Filer, D. L., Kothiya, P., Setzer R. W. et al. (2016). tcpl: The ToxCast pipeline for high-throughput screening data. *Bioinformatics* 33, 618-620. doi:10.1093/bioinformatics/btw680
- Friedman, J., Hastie T. and Tibshirani R. (2010). Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33, 1-22. <https://www.jstatsoft.org/article/view/v033i01>
- Gijbels, E. and Vinken, M. (2017). An update on adverse outcome pathways leading to liver injury. *Appl In Vitro Toxicol* 3, 283-285. doi:10.1089/aivt.2017.0027
- Hastie, T., Tibshirani, R. and Friedman, J. (2016). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. New York, USA: Springer. <https://web.stanford.edu/~hastie/ElemStatLearn/>
- Horvat, T., Landesmann, B. and Lostia, A. et al. (2017). Adverse



- outcome pathway development from protein alkylation to liver fibrosis. *Arch Toxicol* 91, 1523-1543. doi:10.1007/s00204-016-1814-8
- Judson, R., Houck, K. and Martin, M. et al. (2014). In vitro and modeling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast Program. *Basic Clin Pharmacol Toxicol* 115, 69-76. doi:10.1111/bcpt.12239
- Kavlock, R. and Dix, D. (2010). Computational toxicology as implemented by the US EPA: Providing high throughput decision support tools for screening and assessing chemical exposure, hazard, and risk. *J Toxicol Environ Health B Crit Rev* 13, 197-217. doi:10.1080/10937404.2010.483935
- Kleinbaum, D.G. and Klein, M. (2010). *Logistic Regression: A Self-Learning Text*. New York, USA: Springer. <https://www.springer.com/gp/book/9781441917416>
- Krewski, D., Acosta, D. J., Anderson, M. et al. (2010). Toxicity testing in the 21st century: A vision and a strategy. *J Toxicol Environ Health B Crit Rev* 13, 51-138. doi:10.1080/10937404.2010.483176
- Labib, S., Williams, A., Yauk, C. L. et al. (2016). Nano-risk science: Application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. *Part Fibre Toxicol* 13, 15. doi:10.1186/s12989-016-0125-9
- Lamb, J., Crawford, E. D., Peck, D. et al. (2006). The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-1935. doi:10.1126/science.1132939
- Lee, J. H., Wada, T., Febbraio, M. et al. (2010). A novel role for the dioxin receptor in fatty acid metabolism and hepatic steatosis. *Gastroenterology* 139, 653-663. doi:10.1053/j.gastro.2010.03.033
- Locke, P. A., Westphal, M., Tischler, J. et al. (2017). Implementing toxicity testing in the 21st century: Challenges and opportunities. *Int J Risk Assess Manag* 20, 198-225. <https://ssrn.com/abstract=2928986>
- Mellor, C. L., Steinmetz, F. P. and Cronin, M. T. (2016). The identification of nuclear receptors associated with hepatic steatosis to develop and extend adverse outcome pathways. *Crit Rev Toxicol* 46, 138-152. doi:10.3109/10408444.2015.1089471
- Montanari, F., Pinto, M., Khunweeraphong, N. et al. (2016). Flagging drugs that inhibit the bile salt export pump. *Mol Pharm* 13, 163-171. doi:10.1021/acs.molpharmaceut.5b00594
- Morgan, R. E., Trauner, M., van Staden, C. J. et al. (2010). Interface with bile salt export pump function is a susceptibility factor for human liver injury in drug development. *Toxicol Sci* 118, 485-500. doi:10.1093/toxsci/kfq269
- NRC – National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC, USA: The National Academies Press. <https://www.nap.edu/catalog/11970/toxicity-testing-in-the-21st-century-a-vision-and-a>
- NRC (2017). *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC, USA: The National Academies Press. <https://www.nap.edu/catalog/24635/using-21st-century-science-to-improve-risk-related-evaluations>
- Nymark, P., Rieswijk, L., Ehrhart, F. et al. (2018). A data fusion pipeline for generating and enriching adverse outcome pathway descriptions. *Toxicol Sci* 162, 264-275. doi:10.1093/toxsci/kfx252
- Richard, A., Judson, R., Houck, K. et al. (2016). The ToxCast chemical landscape – Paving the road to 21st century toxicology. *Chem Res Toxicol* 29, 1225-1251. doi:10.1021/acs.chemrestox.6b00135
- Stieger, B. (2010). Role of the bile salt export pump, BSEP, in acquired forms of cholestasis. *Drug Metab Rev* 42, 437-445. doi:10.3109/03602530903492004
- Subramanian, A., Narayan, R., Corsello, S. M. et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171, 1437-1452. doi:10.1016/j.cell.2017.10.049
- Tice, R. R., Austin, C. P., Kavlock, R. J. et al. (2013). Improving the human hazard characterization of chemicals: A Tox21 update. *Environ Health Perspect* 121, 756-765. doi:10.1289/ehp.1205784
- van der Veen, J. W., Soeteman-Hernández, L. G., Ezendam, J. et al. (2014). Anchoring molecular mechanisms to the adverse outcome pathway for skin sensitization: Analysis of existing data. *Crit Rev Toxicol* 44, 590-599. doi:10.3109/10408444.2014.925425
- Villeneuve, D. L., Crump, D., Garcia-Reyero, N. et al. (2014). Adverse outcome pathway (AOP) development I: Strategies and principles. *Toxicol Sci* 142, 312-320. doi:10.1093/toxsci/kfu199
- Vinken, M. (2013). The adverse outcome pathway concept: A pragmatic tool in toxicology. *Toxicology* 312, 158-165. doi:10.1016/j.tox.2013.08.011
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J R Stat Soc Series B* 67, 301-320. doi:10.1111/j.1467-9868.2005.00503.x

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgment

The authors want to thank Drs Rui Xiong, Qiang Gu, and two anonymous reviewers for valuable comments. The authors also thank Joanne Berger, FDA Library, for manuscript editing assistance.