Understanding conditional associations between ToxCast *in vitro* readouts and the hepatotoxicity of compounds using rule-based methods

Samar Y. Mahmoud[a], *Fredrik Svensson*[a], *Azedine Zoufir*[a], *Dezső Módos*[a], *Avid M. Afzal*[a], *and Andreas Bender*[a,*]

[a]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

*To whom correspondence should be addressed:

Phone number: +44 (0) 1223 762 983

Email: ab454@cam.ac.uk

Keywords:

Hepatotoxicity, safety profiling, machine learning, rule models, *in vitro* toxicology, physicochemical properties

ABSTRACT

Current *in vitro* models for hepatotoxicity commonly suffer from low detection rates due to incomplete coverage of bioactivity space. Additionally, *in vivo* exposure measures such as C_{max} are used for hepatotoxicity screening which are unavailable early on. Here we propose a novel rule-based framework to extract interpretable and biologically meaningful multi-conditional associations to prioritize *in vitro* endpoints for hepatotoxicity and understand the associated physicochemical conditions. The data used in this study was derived for 673 compounds from 361 ToxCast bioactivity measurements and 29 calculated physicochemical properties against two lowest effective levels (LEL) of rodent hepatotoxicity from ToxRefDB, namely 15mg/kg/day and 500mg/kg/day. In order to achieve 80% coverage of toxic compounds, 35 rules with accuracies ranging from 96% to 73% using 39 unique ToxCast assays are needed at a threshold level of 500mg/kg/day, whereas to describe the same coverage at a threshold of 15mg/kg/day 20 rules with accuracies of between 98% and 81% were needed, comprising 24 unique assays. Despite the 33-fold difference in dose levels, we found relative consistency in the key mechanistic groups in rule clusters, namely i) activities against Cytochrome P, ii) immunological responses, and iii) nuclear receptor activities. Less specific effects, such as oxidative stress and cell cycle arrest, were used more by rules to describe toxicity at the level of 500mg/kg/day. Although the endocrine disruption through nuclear receptor activity formulated an essential cluster of rules, this bioactivity is not covered in four commercial assay setups for hepatotoxicity. Using an external set of 29 drugs with drug-induced liver injury (DILI) labels, we found that the likelihood of liver toxicity increases as compounds' promiscuity over important assays increases. *In vitro-in vivo* associations were also improved by incorporating physicochemical properties especially for the potent, 15mg/kg/day toxicity level, as well for assays describing nuclear receptor activity and phenotypic changes. The most frequently used physicochemical properties, predictive for hepatotoxicity in combination with

assay activities, are linked to bioavailability, which were the number of rotatable bonds (less than 7) at a of level of 15mg/kg/day, and the number of rings (of less than 3) at level of 500mg/kg/day. In summary, hepatotoxicity cannot very well be captured by single assay endpoints, but better by a combination of bioactivities in relevant assays, with the likelihood of hepatotoxicity increasing with assay promiscuity. Together these findings can be used to prioritize assay combinations which are appropriate to assess potential hepatotoxicity.

INTRODUCTION

The early anticipation of toxicity of compounds is a challenging yet important task for pharmaceutical and environmental stakeholders alike.[1] The traditional methods of toxicity testing, as imposed by the regulatory authorities, involve *in vitro* screening as well as acute and chronic animal testing.[2] The latter is not feasible for large compound sets either at early stages of drug development, or for the thousands industrial compounds which are untested and potentially hazardous to human and environment. Hence, *in vitro* bioactivity testing is often proposed as a potential alternative, since it provides the opportunity for detecting toxic effects of compounds in a more time and cost-efficient manner. Additionally, *in vitro* testing allows to associate biological perturbations, at molecular and cellular levels, with complex adverse effects such as hepatotoxicity.[3]

In general, for an efficient translatability of *in vitro* outcomes in to *in vivo* effects, two key factors need to be considered. Firstly, the assays and endpoints relevant to the toxicological effect in question;[4,5] and secondly, the relevant conditions such as multiple bioactivity stressors,[6] as well as exposure and bioavailability.[7,8]

Regarding the first factor, prioritizing relevant *in vitro* endpoints is not a straightforward task, primarily because hepatotoxicity is an adverse effect which involves complex pathological pathways. Based on expert knowledge and known mechanisms of hepatotoxicity, several *in vitro* models have been developed, such as cytotoxicity, bile salts pump inhibition, mitochondrial impairment[9], Cytochrome P activity,[10] and covalent binding.[11] Alternative models consider immunological activity through changes in cytokines profiles.[12] Another *in vitro* technique used to screen for hepatotoxic compounds is high content screening (HCS), which utilizes multiple cellular measurements, such as changes in nuclear size, cell count, mitochondrial mass and cell membrane integrity, as a biological response.[13–15] Considering this diversity in biological mechanisms leading to hepatotoxicity, it is not surprising that *in vitro*

assays which adopt a fraction of this bioactivity space would suffer from relatively low detection rates, with sensitivity ranging from 40 to 60%.[13–16] It has been hence recommended that *in vitro* models for hepatotoxicity should involve a broad range of bioassay endpoints that cover wider biological perturbation points and cellular phenotypes, in order to increase overall sensitivity.[14,17] However, which endpoints to consider, and how to combine their readouts, is not clear.

The second important factor for an efficient *in vitro – in vivo* translatability is exposure, *i.e.* to what extent a compound actually reaches the site where it exerts its action[8], which is related (among others) to its physicochemical properties.[18,19] Treating exposure appropriately poses considerable difficulties in predictive toxicology modelling, mostly because this information is difficult to obtain for large sets of compounds. As a proxy for exposure, parameters such as maximum plasma concentrations (C_{max}) and administered dose levels have shown improvements in the prediction of compounds *in vivo* toxicity. [9,14,15] For example, compounds had significant odds for liver injury if their C_{max} is greater than 1.1μM combined with a set of three bioactivities, namely cytotoxicity with an IC_{50} below 100μM, bile transport inhibition with an IC_{50} below 30μM, and mitochondrial impairment assays IC_{50} below 25μM.[9] It has also previously been recommended, as a rule of thumb, to have 100-fold separation between the concentration at which compounds are toxic in *in vitro* HCS assays and C_{max} value *in vivo*.[14,15] Another study has also shown that the likelihood for observing hepatotoxicity is significantly higher when the drug is administrated daily dose is higher than 100mg and the drug satisfy one of the following: i) forms active GHS adduct, ii) has 5-fold IC_{50} decrease in Cytochrome P450 metabolism-dependent inhibition, iii) binds covalently to proteins at levels higher than 200 pmol eq/mg protein.[11] These studies demonstrate how incorporating daily dose or C_{max} can be powerful in improving the predictivity of hepatotoxicity. Obtaining these measurements from

in vivo experiments, however, is generally not feasible at early stages of drug development, or for profiling large number of compounds.

In order to prioritize the important *in vitro* endpoints sufficient for detecting hepatotoxic compounds, available *in vitro* and *in vivo* data can be used to propose novel hypotheses for toxicity modes of action. To this end, the ToxCast database project[20], which was launched in 2007 and it is in its third testing phase, provides measurement data for over three thousands of compounds against more than 800 assays[21]. This enables the investigation of bioactivity endpoints and their relationship with *in vivo* toxic effects at a wider scale. Hence, several studies attempted mining the data to prioritize relevant bioactivity endpoints[5,22] as well as to generate predictive models related for liver toxicity.[23] A complementary database to ToxCast is Toxicity Reference Database[24] (ToxRefDB) of the Environment Protection Agency (EPA), which is a source for measurements from standardized animal studies with several hundred compounds overlapping with ToxCast. Therefore, with appropriate statistical methods, these databases can be mined to identify the most representative assay combinations that capture compounds inducing toxicity.

A possible approximation for exposure measures are physicochemical properties, which are associated with pharmacokinetic parameters[18,19], and which are used in estimation of exposure in form of extrapolated plasma concentrations.[25] Yet, the mechanistic understanding of how these physicochemical properties directly influence the concordance between *in vitro* measurements and *in vivo* toxic effects, and via bypassing C_{max} estimation (or other equivalent plasma/cellular concentrations), to our knowledge, has not been investigated. Therefore, understanding and prioritizing biological effects of hepatotoxic compounds, and how these effects extrapolate into *in vivo* toxicity, should involve an approach which combine expert knowledge as well as statistical methods.

The goal of this study is identify the combination of predictive *in vitro* assays from ToxCast for *in vivo* hepatotoxicity and understand how physicochemical properties influence this translatability. The current study is to the best of our knowledge the first to apply modified rule-based models on ToxCast *in vitro* measurements and ToxRefDB *in vivo* hepatotoxicity data with the incorporation of physicochemical properties. Rule based classifiers describe each class label by a combination of conditions using the input property set.[26] As such, rules can facilitate prioritizing predictive assays for the endpoint under consideration, as well as to interpret and analyze multivariate associations between *in vitro* activity and *in vivo* toxicity. However, conventional rule models are not constrained to the direction of associations between input features and outcome class, leading to associations that are either spurious or difficult to interpret.

Therefore, we modified rules according to two key assumptions: 1.) *Positive bioactivity* in an assay (and not absence of an activity) potentially contributes to hepatotoxicity; and 2.) *Multiple conditions* influence *in vitro–in vivo* associations, which means for an assay to extrapolate well into *in vivo* outcome, number of other conditions have to be met. These conditions can be a combination of bioactivities and/or physicochemical properties (related to exposure). With those two key assumptions in mind, we manually modified the rules for hepatotoxicity to enhance interpretability and biological relevance and prioritize combinations of assays with 80% overall coverage of toxic compounds. The framework we described here can also generally be used to optimize *in vitro* models for toxicity by selecting significant assay combinations, as well as identifying relevant physicochemical conditions.

MATERIALS AND METHODS

The steps followed to generate and prioritize rule for hepatotoxic compounds are summarized in Figure 1.

Data collection (Figure 1, step 1)

Hepatotoxicity endpoints. Rodent hepatotoxicity measures were extracted from the Toxicity Reference Database (ToxRefDB, version toxrefdb_v1 released on October 2014).[24] Histopathological endpoints (Supporting Information File 2, Table S1) from rat studies in liver which were observed for chronic, sub chronic, multigenerational and prenatal development were used in this analysis, recorded as the lowest effective level (LEL) in mg/kg/day for 882 compounds. These measurements were converted into a binary format by applying two toxicity thresholds of 500mg/kg/day and 15mg/kg/day which were subsequently analyzed separately. This classification was adapted from Martin *et al*[27] considering the highest and lowest quantile bins for toxic effects, corresponding to 15mg/kg/day or less, and 500mg/kg/day or less.

***In vitro* measurements.** Assay bioactivity data were extracted from the ToxCast database,[28] version December 2014, for 1,057 compounds tested in phases I and II. *In vitro* measurements in ToxCast are recorded as the concentrations at which half-maximum activity is reached (AC_{50}). These measurements are generated from dose-response curves for more than 800 assays in units of $\log_{10}$ μM concentration. Assays were annotated by their "intended target type", in the ToxCast assays summary file[28] into protein, cellular, pathway, DNA, RNA or unspecified. We used this annotation to describe the assays as target-based or phenotypic; if the intended target type was described as a protein, it was considered target-based, otherwise, for high-dimensional readouts, assays are annotated as phenotypic.

Compound set. The compound set used in this study represents the intersection between compounds in ToxRefDB and ToxCast, resulting in 673 compounds in total (see Supplementary Table S2 for full list of compounds) matched by ToxCast compound ID. Applying the toxicity threshold of 500mg/kg/day resulted in 395 toxic compounds and 278 non-toxic compounds, whereas at the threshold of 15mg/kg/day 162 compounds were annotated as toxic and 511 as non-toxic (Step 1 in Figure 1).

Dataset curation. Assay endpoints with empty fields (missing values) were considered inactive and an arbitrary (very large) AC_{50} value of 10^6 μM was assigned, adapted from Lui *et al*.[23] In order to select a data matrix that was as complete as possible only assays which had valid AC_{50} measurements for at least 5% of the compounds were selected. This step resulted in 361 assays which are listed in Supplementary Table S2. A recommended step in understanding adverse events resulted from pathway specific effects is to use assays frequently respond bellow the cytotoxicity concentrations (termed as cytotoxicity burst or CBT).[29] However, we decided to include all assays with sufficient number of AC_{50} measurements and apply rules to detect the best predictive assays for hepatotoxicity including those describing non-specific effects.

Structural preprocessing and calculation of physicochemical properties. Compounds were standardized using ChemAxon standardizer[30] (version 15.12.14.0) using the parameters cleaning 2D, mesomerisation, neutralization, tautomerization (generating the most stable tautomer) and removal of fragments of smaller sizes. Physicochemical properties were generated using RDKit[31] *via* KNIME[32] and the Calculator Plugins[33] in Instant JChem (version 15.12.14.0)[34]. A subset of 29 physicochemical properties were used in the current analysis which are listed in Supplementary Table S3.

Constructing rule-based classifiers (Figure 1, step 2)

Multivariate associations between assays and hepatotoxicity endpoints were modelled *via* rule-based machine learning classifiers (Step 2 in Figure 1) as follows. Input variables were 29 physicochemical properties and 361 ToxCast *in vitro* AC_{50} measurements for 673 compounds against two hepatotoxicity labels of 500mg/kg/day and 15mg/kg/day. The rules were generated using the C5.0 algorithm (modified from the C4.5[35] algorithms by Ross Quinlan) as implemented in the C5.0[36] and caret[37] R packages using 5-fold repeated cross-validation with 100 trials and without winnowing. Other parameters were set to default. The models with the highest correct classification rate (CCR) at each trial were retained to be used to generate the rules (accuracy distribution in Supporting Information, Figure S3). Throughout text, we refer to rules associated with hepatotoxicity as rules predictive for or describing toxicity, exchangeably.

Rule modification (Figure 1, step 3)

Each rule derived from the above procedure consists of one or more conditional statements to predict the hepatotoxicity label at a given dose, based on the input variables (ie physicochemical properties and ToxCast readouts; Step 3 in Figure 1). For example, a condition in a rule for toxic compounds may involve a bioactivity in an assay to be less than 10μM (possibly in combination with other properties), which then results (or doesn't result) in hepatotoxicity at a given dose. However, *inactive* bioassay conditions in toxicity-describing rules are not able to explain in a biologically meaningful way how a compound triggers *toxicity*, and hence such rules will be less informative (or even meaningless) as a result. Hence to extract *biologically relevant* associations between activity *in vitro* and hepatotoxicity *in vivo*, rules were manually modified to retain interpretable and biologically meaningful patterns. To

achieve this, modification was applied by removing conditions of inactivity in assays from rules describing hepatotoxic compounds. On the other hand, positive activities in assays and physicochemical properties were kept for further analysis as provided by the machine learning method. As conventional rule methods that take continuous variables as input can generate conditional splits contradicting the assumption stated above, manual curation of rules was applied to retain only splits satisfying this assumption.

Positive activity and inactivity in an assay were discriminated according to the direction of the conditional split. Since the compound potency in an assay is inversely related to the concentration at which a certain biological response was obtained (such as an IC50 value or similar), a bioactivity condition was considered active if the split represent a concentration range below a defined cutoff, and *vice versa*.

One example of how rules were modified is provided in Figure 2 and illustrated in the following. For example, in the condition Tox21_p53_BLA_p5_viability <= 0.026, the first part describes the viability of a human intestinal cell line, and the second part (<= 0.026) represents the range of AC50 (in $\log_{10}$ μM concentration) to be less than 0.026 (< ~1.06μM) for the condition to be fulfilled. As the statement describes the range below a bioactivity cutoff (split point), this example represents a bioactivity condition with positive activity. On the other hand, the two conditions, i) APR_HepG2_MitoMembPot_72_up > 2.037 and ii) Tox21_HSE_BLA_agonist_ratio > 2, represent respectively the increase of mitochondrial membrane potential and Heat Shock Protein (HSP) agonism (both relative to negative control) with AC50 values higher than 100μM. The latter conditions were considered inactive, and hence removed from rules capturing toxicity during the pruning process.

Performance assessment and rule prioritization (Figure 1, steps 4 and 5)

In the current study, the performance of rules was assessed at two levels, for individual conditions used in original rule set as well as for rules before and after modification.

Individual rule conditions were first extracted by collecting the unique set of features used by the rules. These were assessed for their association with the hepatotoxicity outcome, at both thresholds, 15mg/kg/day and 500mg/kg/day, by calculating the information gain[38] (decrease in entropy) associated with this variable. The comparison was conducted with respect to condition type, which were categorized into three groups; conditions describing positive assay activity, inactivity in an assay, and physicochemical properties.

As the applied modification may change rule performance, the next step was to assess the accuracy (confidence) and coverage of rules before and after modification (Step 4 in Figure 1).Rule confidence, represents the percentage of correctly classified compounds for a given class, i.e. here hepatotoxicity. In order to account for the imbalanced distribution of toxicity classes, the balanced accuracy of modified rules was calculated by generating 500 randomly selected balanced data subsets, each composed of 300 data points, and then averaging the accuracy. Rule coverage was calculated for the number of toxic compounds that satisfy the rule conditions (true positives).

As the modification arrived at a rule set that is overall not optimal, a set of prioritization and selection steps were conducted. The prioritized rules were subsequently assumed to capture biologically meaningful information to the maximum possible extent, given the unavoidable limitations of chemical space coverage and bias of the dataset that was available to us.

The rules predictive of toxicity were filtered based on minimum coverage and accuracy (Step 5 in Figure 1). Minimum coverage was set to 50 and 20 compounds per rule at thresholds of 500mg/kg/day and 15mg/kg/day, respectively. The cutoffs for coverage represented the median values after modification, in other words, the best 50% of rules in terms of coverage were selected (see Supporting Information for coverage distributions, Figure S4). Secondly, an

accuracy cutoff of 70% was applied for both rule sets. Another selection step was undertaken to reduce assay redundancy, that is if an assay contributes to multiple rules, then the rule with the highest accuracy that contains that particular assay was selected. Finally, the minimal rule set which covers 80% of toxic compounds at each toxicity threshold was selected (see Results and Discussion below).

To identify the key bioactivity groups captured in toxic compound in the dataset, the final rule set were clustered according to their similarity in compound membership. This was performed by generating a matrix of rules against compounds fulfilling the respective rule conditions, from which a rule similarity matrix (based on shared detected and not detected toxic compounds) was calculated based on the Jaccard index.[39] Hierarchical clustering was applied to the similarity matrix using the Agglomeration method Ward.D2[40] algorithm *via* the 'hclust' function in R.[41] The bioactivity assay conditions under each cluster were examined.

External test set

An external test set was used to examine the discriminating power of the biological space identified by rules. A set of drugs with DILI injury labels in humans[110] were used and selected by matching their InChIKey identifiers with compounds in the ToxCast chemicals. We found 29 drugs in the ToxCast chemical library with valid DILI labels but were not used in the training set (Table S10). Among these drugs, 12, 12 and 5 are of most, less and no DILI concern, respectively. The chemical similarity of these drug to the nearest neighbor in the training ranged between 0.15 and 0.89 with a median of 0.33. These drugs were tested for matching i) the prioritized rules and ii) the biological space used in rules. The biological space was determined by all ToxCast assays covering the biological targets identical to those used in rules. These assays were called here as the equivalent assay set. Positive activity calls in these assays was given for AC_{50} concentration values equal or below 100μM. Promiscuity over the

biological space was also examined which was determined by the proportion of activity calls in the equivalent assays to the overall activity calls in the all ToxCast assays.

Software

The analysis in this study was conducted in R[41] environment (version 3.3.2) . Visualizations were generated using 'ggplot2'[42] and basic R.

RESULTS AND DISCUSSION

Extracting biologically relevant rules for toxic compounds

Information content of conditions in hepatotoxicity rule-based classifiers

To have a picture of how original rules were structured, Table 1 describes the average frequency of each condition type. Overall, on average rules capturing toxicity include one positive assay activity and four inactive bioassay conditions at both toxicity levels. Close to one (0.9) physicochemical condition is on average involved in toxic rules at a threshold of 15mg/kg/day, whereas only on average 0.6 physicochemical conditions are included in rules at a threshold of 500mg/kg/day. Hence, although inactive conditions are by themselves not very information-rich, they frequently are contained in automatically derived rules. Given that inactive assay conditions cannot mechanistically meaningfully linked to toxic events this underlines the need for rule modification as our method of choice (the difficulty of deriving toxicity predictors entirely automatically from ToxCast data has also been discussed, using different methods, before).[4,43]

In order to evaluate to what extent individual assays might be able to predict hepatotoxicity we firstly examined how much information was gained from each type of conditions in the rules. Given the distribution of data points in the classes of the overall dataset, the maximum possible information gain was 0.79 and 0.98, at thresholds of 15mg/kg/day and 500mg/kg/day, respectively. Figure 3 displays the distribution of the information gain (IG), for positive bioactivities, negative bioactivities, and physicochemical conditions in toxic rules at toxicity levels of 15mg/kg/day and 500mg/kg/day, respectively, and we can observe two key trends. Firstly, the maximum observed IG obtained by any split is very low overall, with the maximum IG being only slightly higher than 0.04 and the median IG of positive bioactivity conditions being around 0.01. This indicates that single assays on their own have little predictivity for hepatotoxicity. Secondly, the median IG for physicochemical and negative bioactivity

conditions is even lower and does not exceed 0.005. This means that these conditions *on their own* are less predictive for hepatotoxicity than active assay conditions.

Although positive bioactivities provide the relatively highest IG overall, their still in absolute terms low quantitative values indicate that one condition (or assay) is certainly not sufficient to discriminate between all toxic from non-toxic compounds. This observation can be attributed to two factors: Firstly, hepatotoxicity involves diverse and complex mechanisms that cannot usually be captured by single endpoints.[6,44] Secondly, without considering exposure (or at least some proxies, such as chemical properties),[18] assay readout do not translate readily into *in vivo* outcomes,[43].[45] Our conclusion in this work is hence that, in order to improve our ability to predict *in vivo* toxicity, we need to use rules which, on the one hand, involve a combination of assay endpoints to cover wider bioactivity space, and on the second hand incorporate also physicochemical conditions.

Rule modification ***(Figure 1, steps 3-5)***

To make rules capturing toxicity more biologically meaningful, we next modified them by removing inactive assay conditions, and then keeping rules which retain the highest accuracy and coverage (as described in Methods).

The changes in error rates (accuracy) from rule modification are presented in Figure 4. modification resulted overall in a deterioration in accuracy of 10% and 20% on average for rules set at thresholds of 15mg/kg/day and 500mg/kg/day, respectively, as shown in Figure 4. The change in accuracy is broad, ranging between -20% up to +40% at both toxicity levels. However, the improvement in accuracy (negative values in Figure 4) can be seen for one fifth of rules at the level of 15mg/kg/day and less than a quarter of the rules at level of 500mg/kg/day (Figure 4).

Prioritizing predictive rule sets ***(Figure 1, step 6)***

As some modified rules have shown severe deterioration in accuracy, a selection step was next introduced to retain only highly performing modified rules, defined as those possessing at least 70% minimum accuracy, in addition to a minimum of 50 and 20 compounds being covered (corresponding to the median in coverage of all rules) at the toxicity thresholds of 500mg/kg/day and 15mg/kg/day, respectively. The resulting rules were ranked by accuracy and the rule set able to detect a minimum of 80% of all hepatotoxic compounds was selected (Step 6 in Figure 1).

Figure 6 shows the relationship between the overall compound coverage at both toxicity thresholds with i) the minimum rule accuracy and ii) the number of unique bioassays used in the rules . Firstly, if higher percentages of toxic compounds are to be detected, then more rules (some of which will have lower accuracy) are needed. For example, to obtain collectively 80% compound coverage at a threshold of 500mg/kg/day requires including 35 rules up to a lower limit of 73% accuracy. On the other hand, in order to achieve equivalent coverage at a threshold of 15mg/kg/day 20 rules with a lower accuracy limit of 81% is sufficient. Hence, potent toxicants can be captured by rules of higher confidence than compounds fall under weaker toxicity levels.

Secondly, in order to cover larger proportions of compounds, higher numbers of unique assays to be used by the rules. It can be seen that up to 50% of compounds, at both thresholds, can be described in rules using readouts from 11 assays. For 80% compound coverage, 24 and 39 assays are needed in rules, at thresholds 15mg/kg/day and 500mg/kg/day, respectively. Further for 90% coverage, 34 unique assays are needed in rules at 15mg/kg/day, whereas, more than 70 unique assays are required to cover the same proportion of toxicants at 500mg/kg/day threshold.

Hence overall, single assay endpoints, for the dataset employed here, are only sufficient to anticipate *in vivo* hepatotoxicity for rather few compounds individually. Instead, a combination of bioactivity measurements is required for enhancing the detection rate of hepatotoxic compounds, for which larger numbers of assays (with broader mechanistic coverage) are needed.

Prioritizing assay endpoints for hepatotoxicity detection

We next interpreted the bioactivity endpoints used in rules predicting hepatotoxicity, as determined from the assays used by the best-performing rules covering 80% of the toxic compounds at both LELs of 15mg/kg/day and 500mg/kg/day.

Tables 2 and 3 present the rule clusters capturing the key bioactivity classes in compounds covered by these rules. At both thresholds of 15mg/kg/day and 500mg/kg/day, the rules clustered into three bioactivity groups (according to the most common biological activity in each cluster), namely those involving primarily i) the group of Cytochrome P enzymes, ii) immune responses, and iii) nuclear receptors and transcription factor elements.

Cytochrome P

Activity against Cytochrome P enzymes is one of the key properties of hepatotoxic compounds at both toxicity thresholds (Tables 2 and 3, cluster 1). There are multiple assays in the rules describing activity against different Cytochrome P isoforms, namely against CYP3A, CYP2C18 and CYP2C19 at a dose of 500mg/kg/day, and against CYP2A, CYP2C6, CYP2C12, CYP2C13 and CYP2C19 at a level of 15mg/kg/day. All Cytochrome P enzymes were associated with information gain (IG) values between 0.021 and 0.035 that are higher than average in the selected assay set (Tables 2 and 3, cluster 1), i.e. higher than 0.02 and 0.022 at levels of 500mg/kg/day and 15mg/kg/day, respectively. The average potency in rules

covering compounds active against Cytochrome P enzymes is less than 10μM (Figure S5). Activity against Cytochrome P isoforms CYP3A4 and CYP2C19, which are responsible for the majority of drug metabolic reactions,[46] can be linked to liver injury *via* generating toxic metabolites or interfering with the metabolism of co-administered resulting in slow elimination and chemical accumulation.[47]

Immunological responses

Multiple assays utilized to anticipate hepatotoxicity in our rules were found to contribute to immunological responses at both toxicity cutoffs (Tables 2 and 3, cluster 2). Immunological response conditions in hepatotoxicity rules, in general, had shown assay potency requirements lower than Cytochrome P activities, of around 40μM (see Figure S5 for details). At a threshold of 500mg/kg/day, assays associated with the downregulation of cytokines CXCL10 (IL-10) and CD40 were used by rules, and they also had an individual information gain values above the median (0.02) at values around 0.029, namely the assays "A.15", "A.17" and "A.20". These cytokines are associated with both proinflammatory and regenerative responses depending on downstream signaling.[48,49] For example, CD40-mediated activation of IL-12 has a proinflammatory effect.[50,51] On the other hand, it can also activate IL-10 immune response, which primarily participates in regenerating and repairing hepatic cells via anti-inflammatory responses.[48,52] Similarly, the induction of CXCL10 cytokine can reduce liver injury in mice models *via* upregulation of CXCR2.[53] Yet, also blocking CXCL10 has a regenerative effect after liver damage.[49] Hence, the observed hepatotoxic effect on the organism level seems to depend on other factors than purely the direction of the regulatory effect on the gene level. Similarly, immunological endpoints detected at a threshold of 15mg/kg/day involve the downregulation of CXCL10 and CD40, as above, and in addition that of CCL2, captured by the assays "B.13", "B.15" and "B.10", respectively. At this threshold, downregulation of CCL2

had a high associated information gain of 0.037, hinting to its importance for hepatotoxicity prediction. CCL2 expression can increase tissue damage *via* IL-12 signaling[54], but simultaneously has a protective effect by activating the hepatoprotective cytokine IL-10.[54,55] A previous study has shown an overall downregulation of a variety of cytokines (CCL2, IL-10, IL-12 and IL-16) in primary and liver sinusoidal endothelial (LSEC) cell lines upon the exposure to free fatty acids in contrast to hepatocytes. The authors suggested the anti-inflammatory effect as a response to overcome liver damage.[56] Therefore, assays which detect changes in the expression of immunological cytokines, such as CXCL10 and CCL2, are biologically not entirely understood, but nonetheless they are informative and hence can be regarded as valuable in hepatotoxicity *in vitro* models.

Nuclear receptors and response elements

Activity against nuclear receptors is an important contributor to the overall bioactivity profile of hepatotoxic compounds (cluster 3 in Tables 2 and 3). Compared to the above bioactivity groups, lower potencies of approximately 50μM were used as a rule condition in this case (Figure S5). Endocrine disruption is included in predictive rules for hepatotoxicity at both thresholds, such as activity against estrogen and androgen receptors, which is captured in multiple assays ("A.3", "A24", "A.34", "B.4", "B.12", "B18", "B23"). There are established links between estrogen and glucocorticoid receptors ("A.35") with cholestasis (impairment of bile flow)[57] and steatosis (fatty liver)[58], respectively. In addition, androgen receptor antagonism is associated with a range of hepatotoxic effects with hepatitis as most commonly reported.[59] Rules obtained at both thresholds for hepatotoxicity also share activity against CAR (constitutive androstane receptor), FXR (farnesoid X receptor) and PPAR (peroxisome proliferator-activated receptor) which are described by the assays "A10", "A.11", "B.24", "A.2" and "B.17". Both CAR and FXR play a key role in preventing xenobiotic-induced

hepatotoxicity by regulating a number of genes including phase I and II metabolizing enzymes and bile acid transporters.[60,61] The activity of FXR was found to be correlated with the degree of protection from chemical-induced liver injury previously.[61] PPAR-γ agonist activity, which represented by up regulation in A.2, is known to be involved in chemical-induced liver injury.[62] Whereas, PPAR-δ has a hepatoprotective effects against toxicants[63].

Additionally, there were bioactivities contributing specifically to each toxicity cutoff under the nuclear receptor activity class. For example, at 500mg/kg/day, several transcription factor response elements were involved in cluster 3 in Table 2, such as the regulation of cyclic-AMP response element binding protein (CREB), which provided a relatively high information gain, at this dose, of 0.048 in "A.31". It has been found that the activation of CREB-binding protein/β-catenin interaction promotes liver fibrosis.[64] Another response element is the upregulation of SMAD (assay "A.28") , which mediates TGF-β-induced apoptosis and fibrosis *via* downstream immune responses.[65,66]

At a toxicity threshold of 15mg/kg/day, bioactivities under the nuclear receptors cluster included those against the vitamin D (VDR) and liver X receptors (LXR) in the assays "B.8" and "A.21". Upregulation of the VDR response element (VDRE) is associated with anti-inflammatory properties and xenobiotic metabolism,[67,68] while LXR has an anti-inflammatory effect[67] and can reduce chemical-induced toxicity.[69] Thus, the activation of these nuclear receptors can be linked to triggering protective response against xenobiotics. We observed that the cytochrome P and nuclear receptor assays clustered into separated groups. However, the NR which regulate the expression of metabolizing enzymes were embedded within the Cytochrome P group, such as CAR, FXR and VDR.

Other observed bioactivities in prioritized rules

- Phenotypic assays

Furthermore, the prioritized rules involved several phenotypic activities that cannot be attributed to a specific target. Assays describing cell morphology and cell cycle are more frequent at threshold 500mg/kg/day (Table 2), at which some selected endpoints can be attributed to a broad range of perturbations. These include as cytotoxicity, cell cycle arrest, oxidative stress and mitochondrial impairment. There are ten assays describing cell cycle or cell morphology in rules capturing hepatotoxicity at a toxicity level of 500mg/kg/day, whereas only four assays are in the rules derived at 15mg/kg/day (Table 2 and Table 3). Also, these assays had shown significant difference in the overall assay potencies, of around 50μM, between levels of 500mg/kg/day and 15/mg/kg/day. The phenotypic assays overlapping between the two toxicity levels include increase in nuclear size and mitochondrial mass, represented by "A.1", "A.25", "B.19" and "B.20". Mitochondrial dysfunction is one of the key mechanisms of chemical-induced hepatotoxicity.[44,52,70,71] Nuclear size increase, which was seen part of the cluster for nuclear receptor activity, can be a result of activating gene expression. This endpoint is also common in high content screening models due its predictivity of toxic compounds.[13,15] Additionally, at a threshold of 500mg/kg/day, two assays describe mitochondrial effects with information gain higher than the median (0.02 and 0.029), which are associated to changes in mitochondrial membrane potential, namely "A.22" and "A.23". Chemical toxicants can cause mitochondrial permeability transition (MPT) *via* the opening of permeability transition pores in the mitochondrial membrane,[52] either directly or indirectly. As a result, mitochondrial depolarization takes place which leads to ATP depletion and reactive oxygen species (ROS) release, followed by mitochondrial membrane rupture and apoptosis or necrosis.[44] Another phenotypic effect is the increase in stress kinase expression as a response to stress (assay "A.27"). This assay has the highest information gains of all assays (of 0.052) and the accuracy of its rule is 95%, which indicates that it is highly associated with toxicity at a 500mg/kg/day level. Also, rules at this level involve other cell cycle assays which screen for

cytotoxicity, oxidative stress and cell cycle arrest (see Table 2). Hence, we can conclude that less potent hepatotoxic compounds are relatively more frequently detected by phenotypic assays than more potent hepatotoxic compounds.

Despite the overall consistency in bioactivity classes in rules describing the two toxicity levels, there were also variations. More unspecific effects observed at higher doses, potent toxicants have shown narrower, more specific modes of action. Judson et al reported a greater likelihood of disrupting target specific pathways when active concentrations fall within the range of eliciting cytotoxicity.[72] What we observe here is an association between potency of toxicity *in vivo* and specificity of *in vitro* effects that are predictive for toxicity. Also, for compound to be toxic at low doses, they are required, according to the rules, to be more potent in assays (as shown in Figure S5).

- Other target-based activities

In addition to the major assay clusters described above, further rules for hepatotoxicity involved other target and phenotypic assays. For example, the transporter SLC6A3, which was described by assays "A.6" and "B.22", has relatively high information gain of 0.021 and 0.024 (see Figure 3 for distributions of information gain values) at 500mg/kg/day and 15mg/kg/day, respectively. The SLC6A3 gene encodes for the dopamine transporter (DAT). DAT-dependent neurological degeneration is linked with hepatic dysfunction related to ROS overproduction and mitochondrial impairment in rodents.[73] Another endpoint related to the toxicity threshold of 500mg/kg/day is the translocator protein (TSPO) (assay "A.9"), which is involved in the transport of cholesterol across the mitochondrial membrane. The expression of TPSO has also been found to be associated with activating macrophages in chemical-induced liver injury and hence leads to cell death.[74,75] At a toxicity level of 15mg/kg/day, upregulation of prostaglandin E2 (PGE2), in assay "B.1, was also used by rules. PGE2 is a lipid autocoid which protects

against liver damage by downregulating the expression of inflammatory cytokines.[76,77] It has also been reported that PGE2 participates in liver regeneration upon injury.[78]

Therefore, we can overall conclude that hepatotoxicity can be initiated by a variety of mechanisms (also very likely beyond the ones covered in the assays used here), hence supporting the need for broad range of endpoints when screening for potential hepatotoxic compounds.

- Combined bioactivity readouts

Our study goes beyond interpreting univariate associations into investigating rules constructed from multiple assay conditions. Some rules for hepatotoxicity involve two bioactivities at a time (linked by the symbol ⌊ in Tables 2 and 3). There are seven and six rules including multiple bioactivity features at toxicity thresholds 500mg/kg/day and 15mg/kg/day, respectively. Some assays formed diverse combinations in rules which are in some cases shared between the two toxicity thresholds. For example, androgen receptor activity was combined with mitochondrial membrane potential, at a threshold of 500mg/kg/day, Cytochrome P (CYP2A1) activity at 15mg/kg/day, and PPAR activity at both thresholds. An AR ligand, dihydrotestosterone, has shown to disrupt mitochondrial membrane potential[79], which established a link between AR modulation and disrupting mitochondrial membrane. Additionally, CYP2A1, which metabolizes 90% of the endogenous ligand for AR (testosterone) [80], is subject to modulation leading to an imbalance of serum testosterone.[81] Hence, the dual activity of AR antagonism and CYP2A1 modulation can cause compounds to be hepatotoxic. The combined activity of AR and PPAR isoforms can be explained by established bidirectional crosstalk between the two receptors by which each can influence the expression as well as the transcriptional activity of the other.[82,83] Primary hepatocytes of obese male AR-knockout mice had shown hepatic steatosis which is associated with altered PPAR-α and PPAR-γ expression.[84]

Other interactions which are seen at a level of 500mg/kg/day include multiple combinations Firstly, CYP2C19 appeared in conjunction with SLC6A3 (dopamine transporter) in rules (Table 2). It is found that the antagonists of D2 dopaminergic receptors interfere with the regulation if CYP2C enzymes.[85,86] This can support the link between interfering dopamine transportation and CYP2C19 activity highlighted in the rules. Secondly is the combination of CD40 with IL-8 cytokine. Activation of CD40 increases the secretion of IL-8 in hepatic stellate cells resulting in an amplification of proinflammatory effects.[87] Also, the hypoxia inducible factor-1 (HIF-1) was combined with CCAAT/enhancer binding protein B (C/EBPB) (Table 2). Studies have shown mutual regulation between these two transcription factors in expression and transcription.[88,89] Also, HIF-1 is one of the key transcription factors which binds to C/EBPB during liver regeneration.[90]

At 15mg/kg/day, multiple assay combinations predictive for hepatotoxicity can be seen in Table 3, including CYP2C6 with VDR and CXCL-9 with ER agonists. VDR involves in the metabolic liver damage and its expression correlates inversely with the severity of liver steatosis[91] and fibrosis.[92] In response to xenobiotics, VDR directly induces the upregulation of CYP2C6.[93] Hence, it is plausible that compounds that combine activity against CYP2C6 and upregulation of VDR are more likely to cause hepatotoxicity. Studies have shown links between ER agonists and CXCL9, at which estrogen-treated mice have shown a significant reduction in the expression of CXCL9,[94] a cytokine associated with liver fibrosis.[95]

These observations support the importance of conditional associations in studying the translatability of *in vitro* activity into *in vivo* effects, and using rules we were able to suggest which assays are most predictive for hepatotoxicity when used in combination, based on the dataset used here.

The case study of troglitazone

As a case study for the benefit of considering bioactivity combinations in screening hepatotoxic compounds, we chose troglitazone, an antidiabetic drug that was withdrawn from the market in 2000 due to incidences of hepatotoxicity.[96] It was reported that the mechanism of troglitazone liver toxicity combines mitochondrial impairment, cellular stress and the generation of reactive metabolites.[96–98] It was also argued that risk factors including genetic and environmental factors, besides biological activity of the compound itself, play a contributing role.[99]

Troglitazone is labelled in our data as toxic at the LEL level of 500mg/mg/day, but not at 15mg/kg/day in the ToxRefDB dataset. It matched ten of the toxicity rules at the level of 500mg/kg/day and two rules at the level of 15mg/kg/day, and assays comprise mitochondrial toxicity, endocrine disruption and activity of immunological responses as endpoints (see Supporting Information, Table S9). At a level of 500mg/kg/day, troglitazone fulfilled the rules describing the combinations of AR with PPAR-γ as well as IL-8 with CD40 (see "Combined bioactivity readouts"). Additionally, this compounds also matched multiple rules in which cytotoxicity was combined with specific target activities for both toxicity levels (see Supporting Information, Table S9). The average hepatotoxic compound, however, satisfied only four rules at 500mg/mg/day and two rules at 15mg/kg/day (Figure S6). Also, there is significant difference in number of satisfied rules by toxic and non-toxic compounds at both thresholds (Figure S6). Given the number of rules satisfied by troglitazone (Figur S6), it had more bioactive liabilities compared to the average toxic compound at the level 500mg/kg/day and equivalent to the average liability at the level of 15mg/kg/day. Therefore, troglitazone's promiscuity in hepatotoxic rule space predict it to be likely hepatotoxic *in vivo*.

Comparison of prioritized bioactivities with commercial hepatotoxic assay endpoints

Further, we compared the bioactivities covered by prioritized rules with the *in vitro* endpoints measured by commercial models for hepatotoxicity. Bale *et al*., reviewed a series of commercially available *in vitro* platforms for the detection of hepatotoxic compounds,[3] which we will use here. From this review, we selected four *in vitro* models that showed high diversity in screened endpoints, namely Cellciphr®,[100] 3D Insight™,[101] Hepatopac®[102] and RegeneMed[12], which are summarized in Table 4.

There is an overall large overlap between bioassay endpoints (see Table 4) screened within commercial hepatotoxicity screens, and assays prioritized by the rules generated in this study. Firstly, all models, as well as the rules derived in the current study, involve screening against mitochondrial impairment and cell stress, which are known as key signals for hepatotoxicity.[44] Secondly, phenotypic readouts associated with cell growth or morphology are used by CellCiphr®, InSphero 3D Insight™ as well as in the rules, examples of which are apoptosis, cell loss and changes in nuclear size and mitochondrial mass. Additionally, screening for Cytochrome P activity is an endpoint used in InSphero 3D Insight™, Hepatopac® and RegeneMed. 3D Insight™ and RegeneMed also screen for changes in cytokine profile, which agrees with the assays identified as important in the current study. Another endpoint is the inhibition of the bile acid transporter which is screened in 3D Insight™and Hepatopac®. While the ToxCast assay set we used in this analysis did not include the inhibition of bile acid transporters, the rules detected activity against FXR and CAR, which directly regulate the expression of bile acid transporters.[60] Inhibition of a set of ubiquitous proteins, such as albumin, urea, and fibrinogen is conducted by InSphero 3D Insight™, Hepatopac® and RegeneMed, where the (in this case rather unspecific) counterpart used in our prioritized rules would be the decrease in total protein level in the cell ("A.21").

In addition to the above endpoints, our analysis also identified assay readouts with association with hepatotoxicity, in particular nuclear receptor activity (Tables 2 and 3), that at

the current stage are less covered in commercial hepatotoxicity assays. The involvement of nuclear receptors in causing hepatotoxicity is well supported by mechanistic studies - for example, Liu *et al*. reported that estrogen and androgen antagonism are related to proliferative lesions.[23] Additionally, Hu *et al*., demonstrated a significant univariate association of two ToxCast assays for androgen receptor activity with human hepatotoxicity.[22] Hence, although it is known that estrogen and androgen disruption can cause hepatotoxicity[57,59], this endpoint is not currently used in commercial assay setups (Table 4). Therefore, while generally overlap between the features identified in our work and commercial assays for hepatotoxicity exists, we suggest that including endocrine activity among *in vitro* models can improve the coverage, and hence detection, of hepatotoxic compounds beyond the current state of the art.

Influence of physicochemical properties on improving *in vitro/in vivo* associations

We next analyzed the significance of physicochemical properties in the prioritized rule sets, to see to what extent those proxies for exposure[18,19] add value when attempting to anticipate the hepatotoxicity of compounds. For this, the deterioration in accuracy after removal of physicochemical conditions from prioritized rules was quantified with respect to the two toxicity thresholds. The absolute drop in accuracy by removing physicochemical conditions from all bioactivity rules ranges from 0% to around 15% (for details see Figure S7). The overall drop in accuracy, in average, is higher at a toxicity threshold of 15mg/kg/day, with the overall error rate typically increasing by approximately 6-11%. The distribution of error rate as a result of removing physicochemical conditions from rules at 500mg/kg/day is broad, ranging from no effect to up to 9%. Hence, incorporating physicochemical properties improve *in vitro-in vivo* associations, especially at a lower dose (and hence for the more potent toxicants).

The error rate varies not only by the toxic dose level but also by assay class, as shown in Figure 6. The error rates of physicochemical conditions are minimal in rules defined by Cytochrome P activity, particularly at 500mg/kg/day. Immunological and phenotypic bioactivities have moderate accuracy deterioration from discarding physicochemical properties, at around 5%. Yet rules describing phenotypic assays show a significant difference between toxicity thresholds, of more than 10% ($p<0.05$ Wilcoxon rank sum test). Rules related to nuclear receptor activity show the largest increase in error upon removing physicochemical conditions, especially at a threshold of 15mg/kg/day, reaching values higher than 10%. The highest accuracy drop can be observed for glucocorticoid receptor activity at almost 20%, followed by over 15% for androgen receptor activity at the toxicity level of 500mg/kg/day (Supporting Information, Table S7). Estrogen receptor activity shows an approximately 13% drop in accuracy at both toxicity thresholds (Supporting Information Table S7 and Table S8).

Direct perturbation of nuclear receptors requires compounds to penetrate not just the cell membrane but also the nucleus membrane to be toxic at low doses. Overall, for an improved association with *in vivo* effect, some assay bioactivities, such as phenotypic and nuclear related effects, are more dependent on meeting physicochemical conditions, and hence even more urgently require a proxy for exposure. This influence is overall more pronounced for more potent toxicants, than less toxic ones.

We next analyzed the most frequent physicochemical properties occurring in the rules (Table 5). At a threshold of 15mg/kg/day, these were the number of rotatable bonds (which are equal to or below 6 in rules describing hepatotoxicity), the number of hydrogen bond donors (where hydrogen bond donors are required to be absent) and the number of aliphatic rings (which needs to be equal to or smaller than two). The number of rotatable bonds has the highest frequency, occurring in over a third of the rules. This physicochemical condition is associated with an accuracy deterioration, or error rates, equivalent to 8% when it is removed from the rules. The

number of rotatable bonds is however not abundant among rules at the level of 500mg/kg/day. Instead, the number of rings (<=3) is the most abundant physicochemical property in this case and is associated with an increase in error rate of 6% and an abundance across rules at this toxicity level of 29%.

The above increase in error rates when removing physicochemical properties from rules anticipating hepatotoxicity is apparently due to a link between the physicochemical properties of compounds and *in vivo* bioavailability.[19,103] Bioavailability is governed by the extent and onset of absorption and distribution, which are linked to molecular properties such as membrane permeability and plasma protein binding. Good bioavailability *in vivo* means that, upon exposure, compounds achieve sufficient concentrations to achieve an effect at the site of action. Rotatable bond count, for example affects the magnitude of cell membrane permeability of compounds.[104] The majority of compounds that contain six or fewer rotatable bonds show oral bioavailability higher than 20%, irrespective to their molecular weight.[104] Additionally, a number of aromatic rings greater than 3 in a compound is linked to higher plasma protein binding (PPB) (>90%), irrespective of cLogP.[105] Strong PPB may slow the rate of compound distribution among body compartments which will consequently affect the concentration at the site of action (and hence hepatotoxicity).[106,107] Additionally, higher numbers for aromatic rings are linked with Cytochrome P inhibition and higher lipophilicity, both of which are associated with toxicity.[105,108] DeGoey *et al.* have introduced a simple multiparametric scoring function to describe rat oral bioavailability consisting of three properties, namely cLogD, the number of rotatable bonds and the number of rings of a compound.[109] This score sums the values of these properties with a correction for cLogD. The authors reported a negative correlation between this score and bioavailability with correlation coefficient of -0.41, in agreement with our findings above.

An additional observation, however, is that rules used different properties at different potency

levels, but consistent properties within each potency threshold. Also, different assay types have shown variable magnitudes of dependence on physicochemical properties when measurements are extrapolated into *in vivo* effects. Still, it is apparent that also simple proxies for bioavailability are able to improve the prediction of hepatotoxic compounds on the dataset employed here. Hence, given that *in vivo* exposure parameters (i.e. dose and C_{max}) are often not available at primary stages of drug development, these simple proxies such as the number of rings and rotatable bonds may act as simple alternatives to exposure measures to anticipate the hepatotoxicity of compounds at an early stage.

Applying rules to detect drugs inducing liver injury

In order to examine how the rules can be used on untested compounds, we used a set of drugs with 29 DILI injury labels in humans[110] as an external set (Table S10). Among these drugs, 12, 12 and 5 are of most, less and no DILI concern, respectively. The chemical similarity of these drug to the nearest neighbor in the training ranged between 0.15 and 0.89 with a median of 0.33. We examined how theses drug matched the rules describing hepatotoxicity. Using rules at level 500mg/kg/day, we observed that drugs with no-DILI-concern show 0-1 rule match per drug, whereas 75% and 25% of drugs with less and high DILI concern match 3 rules or more, respectively (Figure S9). Few drugs with most and less DILI concern matched rules at level 15mg/kg/day (Figure S11). This is not surprising given that the rule selection process involved removing rules describing redundant bioactivities (assays describing the same target). Therefore, we examined the biological space of drugs in the external set. This was done by identifying equivalent ToxCast assays not necessarily used by the rules but target the same gene as assays used in the rules, which resulted in a pool of 72 and 88 ToxCast assays at threshold of 15mg/kg/day and 500mg/kg/day, respectively (Tables S11 and S12). We calculated the relative promiscuity for each drug which represents the proportion of the number of activity calls (activity below 100μM) in the equivalent assays to the number of activity calls

in the all ToxCast assays. We observed that drugs with most DILI concern have shown around 20% or higher promiscuity in equivalent assays in both thresholds, in comparison to 10% or less promiscuity by drugs with no DILI concern (See Figure S8 andS9). Using assays predictive for hepatotoxicity at 15mg/kg/day, there were four times higher odds that drugs with promiscuity higher than 20% to be of high DILI concern. Also, the majority of drugs with high DILI concern had less than 6 rotatable bonds. The odds of observing high DILI concern is 5.6 when the drug combines 20% or higher promiscuity (over assays selected at level 15mg/kg/day) in addition to number of rotatable bonds less than 6. Although these observations align with findings from the training set, a larger external set is required to support a statistically significant outcome.

Overall, we can conclude that promiscuity over the relevant biological space can indicate the potential of liver toxicity. This knowledge can also be used to better leverage *in vitro* models by prioritizing the relevant biological space and physicochemical properties.

Conclusion

We proposed in this work a novel framework for generating interpretable rules for the hepatotoxicity of compounds, which use both *in vitro* bioactivity measurements and physicochemical properties. Rules generated from machine learning algorithms were pruned to remove biologically not meaningful inactive assay conditions from rules describing toxicity. The resulting interpretable rules were used first to prioritize hepatotoxicity *in vitro* endpoints, considering accuracy as well as more than 80% toxic compound coverage. The resulting rules were compared with four commercial *in vitro* models for hepatotoxicity. Finally, the influence of physicochemical properties on the derived *in vitro- in vivo* associations were investigated separately for each assay class.

Our results suggest that a set of multiple ToxCast assays are needed for a sufficiently high coverage of hepatotoxic compounds, as no single assay can discriminate toxic from non-toxic compounds. This was also apparent from the information gain derived for single ToxCast assays alone. At two toxicity threshold levels of 15mg/kg/day and 500mg/kg/day, the best-performing modified rules, which cover 80% of toxic compounds, cluster into three major bioactivity classes, namely Cytochrome P activities, immunological responses, and nuclear receptor activities. While overall assays selected for predicting hepatotoxicity overlapped with endpoints used in *in vitro* models from commercial sources, nuclear receptor activity, which represented an independent mechanistic cluster, is not currently covered in this way.

Specific bioactivity combinations were seen in the rules, such as disruption of androgen receptors combined with activities against PPARs, Cytochrome P and increase in mitochondrial membrane potential. These describe perturbation in multiple biological pathways resulting together in a greater likelihood of observing toxicity *in vivo*. Incorporating physicochemical properties, in general, also improved the accuracy of rules describing toxicity especially for potent toxicants, i.e. those toxic at the toxicity level of 15mg/kg/day. The likely explanation is that, for those compounds, bioavailability plays an important role for toxicity to be observed, which to some extent can be anticipated by physicochemical properties. The most frequent physicochemical properties used in rules, namely the number of rotatable bonds and the number of rings, are linked to bioavailability parameters, such as membrane permeability and plasma protein binding, respectively. Hence, the likelihood of a compound to be hepatotoxic *in vivo* increases both if it is active in relevant bioassays as well as showing the necessary bioavailability.

There has been an increasing interest in understanding the molecular mechanisms responsible for initiating toxic side effects, in the form of adverse outcome pathways (AOP) frameworks. The assay endpoints screened in the ToxCast project can in principle describe key

events in an AOP,[111] since they provide insights to the perturbation in biological processes in the cells by the screened compounds. Our proposed rule-based method can be used as a tool to generate molecular hypotheses to guide the identification of key events of the AOPs. In order to be practically successful in this direction, assay coverage in biological space, compound coverage in chemical space and a complete data matrix linking both domains are crucial.

Another application to this rule method framework is to optimize *in vitro* models for toxicity screening. For example, in order to improve the compound coverage of hepatotoxicity *in vitro* models, we recommend to incorporate assays from three major bioactivity classes when testing for hepatotoxicity, which are Cytochrome P activity, immune responses and endocrine disruption. This is in addition to phenotypic readouts such as cell viability, cell stress, mitochondrial impairment and changes in cellular organelles. The combination of assays from all areas will then allow for the more likely detection of hepatotoxic compounds. We also recommend considering physiochemical properties as simple proxies for *in vivo* exposure measures, such as C_{max}, when attempting to anticipate potential hepatotoxicity. While simple, those properties are fast to calculate and they are able to improve predictivity of *in vivo* hepatotoxicity, at least on the data based in this study.

The workflow presented here can finally be generalized to other types of toxicity, considering any type of chemical and biological input data, provided coverage in the chemical and biological domain for the toxic endpoint of interest is given.

FIGURES

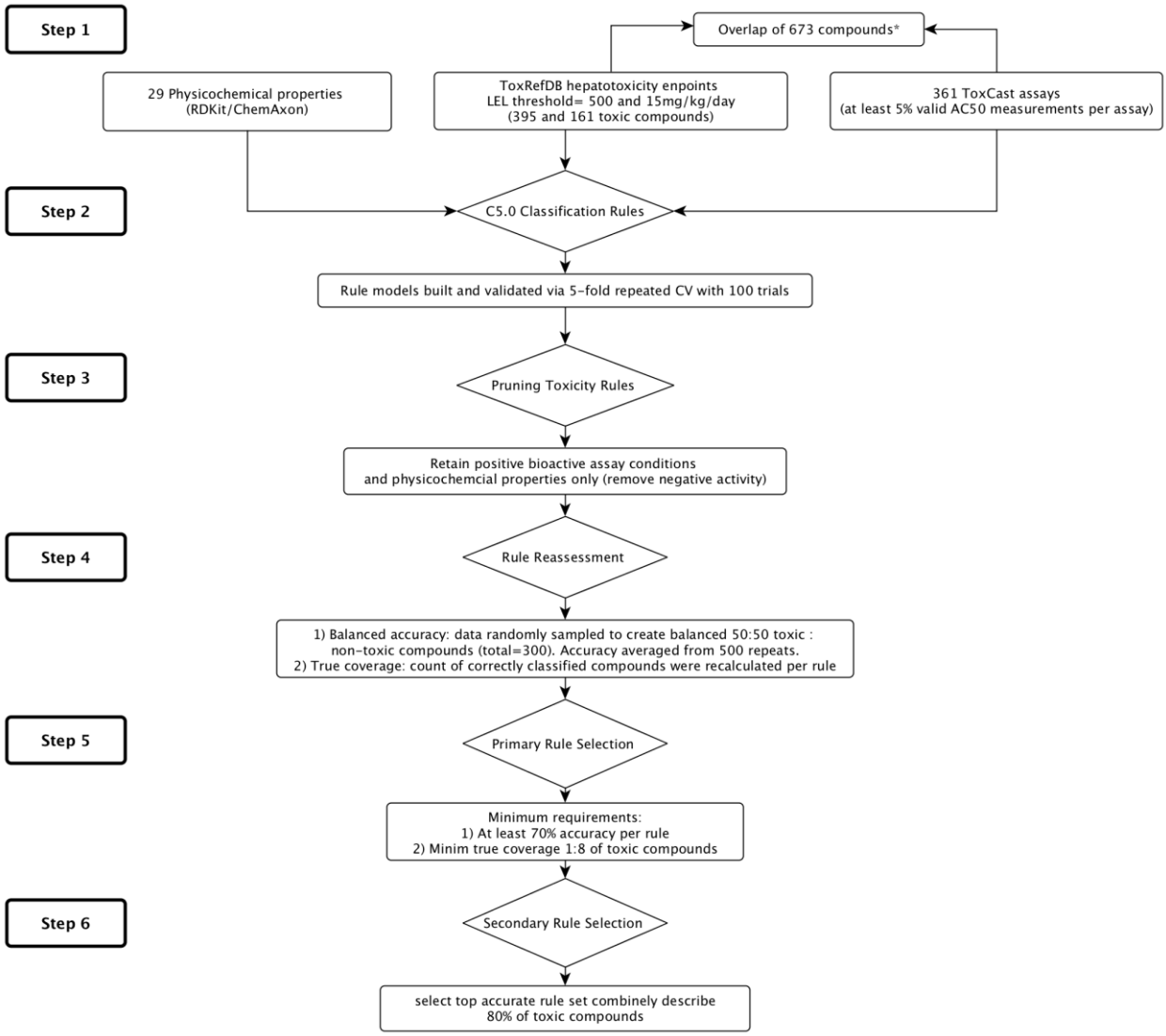


Figure 1. Dataset and workflow for extracting biologically relevant rules for hepatotoxic compounds. Firstly, ToxCast bioactivity measurements and an interpretable set of physicochemical properties were used as descriptors to generate rule-based classifiers *via* the C5.0 algorithm. Rat hepatotoxicity endpoints, from the ToxRefDB, were converted into binary labels by setting two maximum exposure thresholds for lowest effective level (LEL), 500mg/kg/day and 15mg/kg/day. At each threshold level, rules capturing toxicity were pruned by removing inactive assay statements. Next, the modified rules were reassessed in terms of balanced accuracy and number of correctly classified compounds (true positives). Prioritization of biologically relevant rules was conducted in multiple steps. The primary selection involved

performance measures, keeping rules that exerted at least 70% accuracy and median value of coverage (50 at 500mg/kg/day and 20 at 15mg/kg/day). The secondary selection was performed based on overall compound coverage, by prioritizing the combination of most accurate rules that describe 80% of toxic compound set. The final set of prioritized rules were analyzed in terms of contributing bioactivity and physicochemical conditions.

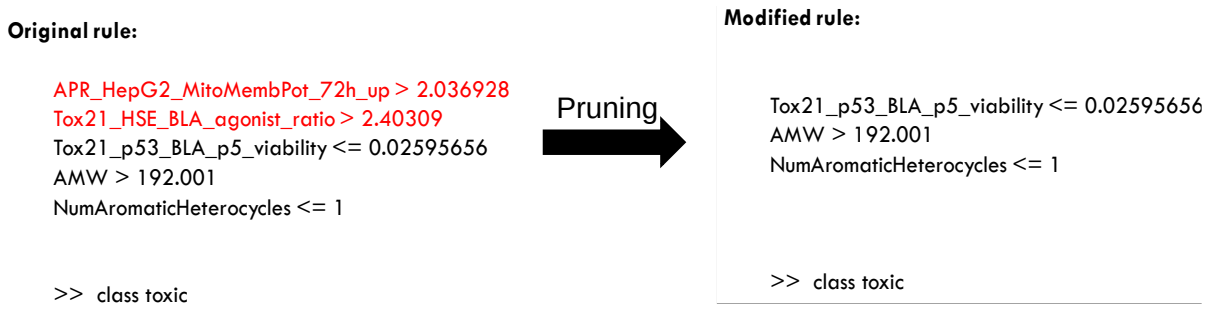


Figure 2. An example of rule pruning to retain biologically relevant rules. Left: unmodified rule, describing toxic compounds, as generated by the C5.0 algorithm. This rule consists of five conditions, namely two inactive bioactivities (shown in red), one positive bioactivity and two physicochemical properties: he increase in mitochondrial membrane potential (inactive), heat shock protein agonist (inactive), cytotoxicity assay (active), average molecular weight, and number of aromatic heterocycles. The rule was then modified by removing the inactive bioactivities to retain only positive bioactivity readout (cytotoxicity) and physicochemical properties (right), which are statistically more meaningful, and biologically more plausible to be associated with toxicity.

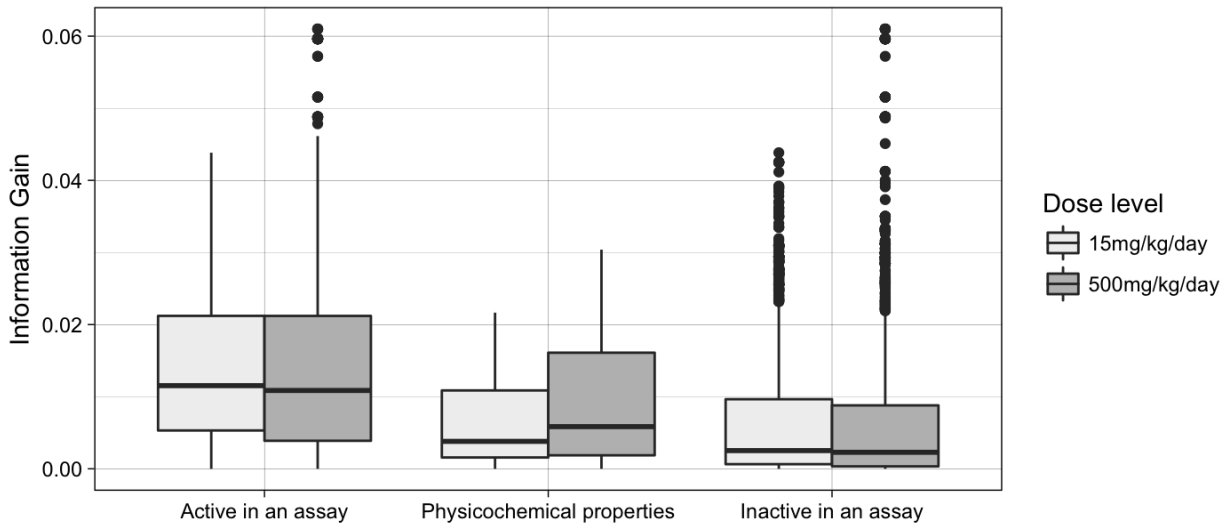


Figure 3. The distributions of the information gain (IG) for each condition type (positive bioactivity, assay inactivity and physicochemical properties), in rules capturing toxicity, at two toxicity thresholds, LEL= 500mg/kg/day and 15 mg/kg/day. IG is a metric used to describe to which extent particular condition (split) improves the homogeneity (purity) of the partitioned data, and which features are hence associated with the classes under consideration. The Overall values of IG are low compared to the maximum possible values to fully discriminate between all toxic from non-toxic compounds, which is corresponding to 0.79 and 0.98, for levels of 15mg/kg/day and 500mg/kg/day, respectively. Positive bioactivity conditions had the greatest average IG in comparison to other condition types, whereas, negative bioactivity conditions had the lowest ($p<0.05$ Wilcoxon rank sum test for IG of activity in assays and both inactivity and physiochemical properties). This means that the predictive power of a single positive activity in classifying toxic compounds is generally larger than single bioassay inactivity and physicochemical property conditions. Yet, single assays on their own are not sufficient to fully predict compounds in the dataset, which aligns with findings in previous studies.[43,112,113]

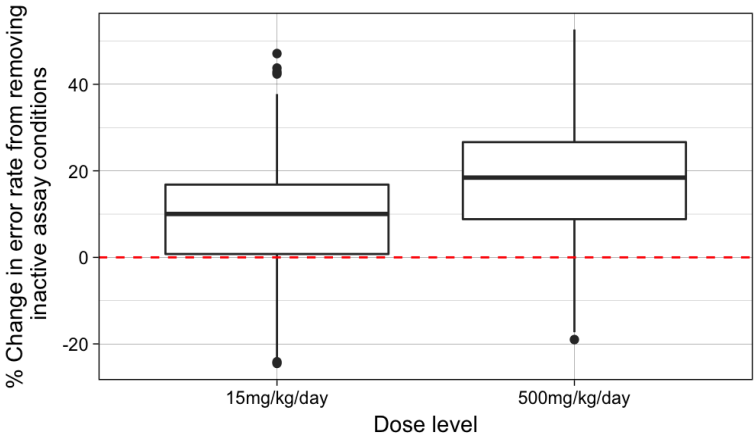


Figure 4. The change error rate (accuracy deterioration) in percentage from removing inactive bioactivity conditions from rules. The deterioration of rule accuracy after modification is, in general, greater at threshold 500mg/kg/day, at an average of 20%, whereas, the overall drop in accuracy at 15mg/kg/day after modification is 10%.

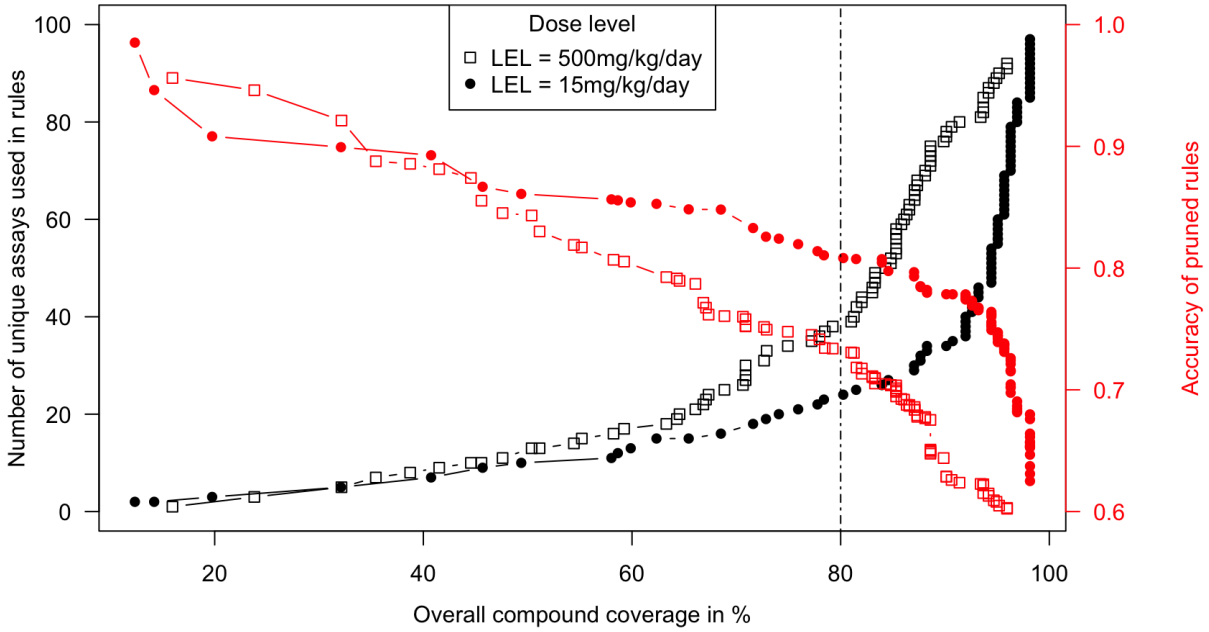


Figure 5?. Percentage of overall toxic compounds covered by the rules. The x-axis represents the percentage overall toxic compound coverage as a function of minimum rule accuracy (Y-axis, red) and number of unique bioactivity assays used in rule combination (Y-axis, black). The most accurate rule sets enough to cover 80% of toxic compounds, at each threshold level, were selected as the minimal rule set to describe hepatotoxicity. Fewer numbers of unique assays were required at a threshold of 15mg/kg/day in comparison to level of 500mg/kg/day to cover 80% of all compounds, namely 24 and 39 assays, respectively.

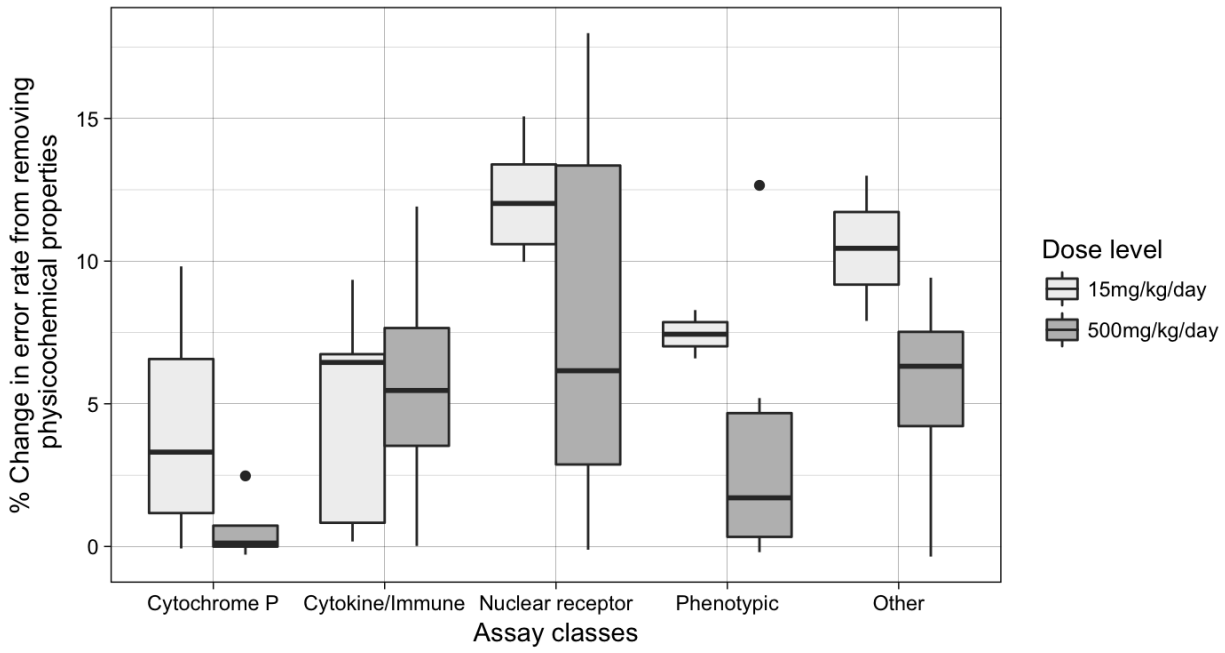


Figure 6. Change in error rates in percentage as a result of removing physicochemical property conditions from best performing rules as a function of assay class. The overall deterioration of accuracy by removing physicochemical conditions varies with the assay class. The accuracy drop is minimal among rule of Cytochrome P activity in particular in case of the higher dose level of 500mg/kg/day, while very significant accuracy deterioration is seen in rules described by nuclear receptor activity, especially at a threshold of 15mg/kg/day.

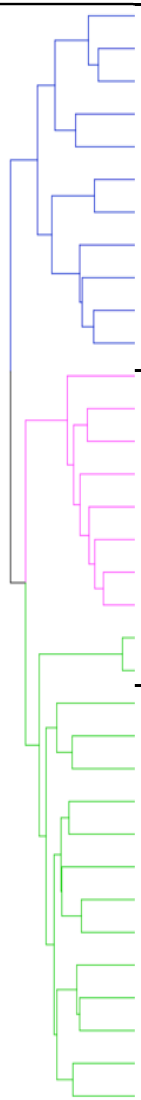
TABLES.

Table 1. Average number of conditions per toxic rule in the original set.

	Condition type in toxic rules		
Toxicity threshold	*Active in an assay*	*Inactive in an assay*	*physicochemical*
15mg/kg/day	1	3.8	0.9
500mg/kg/day	1	3.6	0.6

Overall per rule, there is one positive bioactivity, four negative bioactivities and one physicochemical property. The abundance of inactive assay conditions and physicochemical conditions is slightly lower at toxicity threshold 500mg/kg/day.

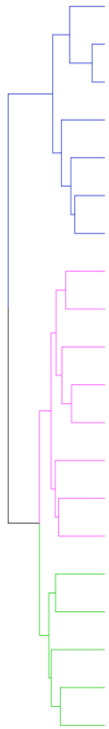
Table 2. Clustering of unique bioactivity assays in prioritized rules capturing toxicity which cover combined 80% of all toxic compounds at a threshold of 500mg/kg/day. Rules were clustered according to similarity in compound coverage, which resulted in three major clusters, namely of mainly Cytochrome P activity, immunological responses and nuclear receptor activity. Rules using two assay conditions at a time are linked by the symbol ⌊.



Bioactivity class	Index	Associated assay	Information gain (split)	Accuracy (rule)	Coverage (rule)	Gene symbol	Function	Use of Physchem*
Activity against Cytochrome P	A.1	APR_HepG2_MitoMass_24h_up	0.019	0.921	66	-	Cell morphology	✓
	A.2	ATG_PPARg_TRANS_up	**0.045**	0.874	78	PPARG	Nuclear receptor	✓
	A.3	└ OT_AR_ARSRC1_0480	**0.033**	0.874	78	AR	Nuclear receptor	✓
	A.4	NVS_ADME_hCYP2C18	**0.028**	0.817	64	CYP2C18	Cyp	✓
	A.5	NVS_ADME_hCYP2C19	**0.025**	0.752	57	CYP2C19	Cyp	-
	A.6	└ NVS_TR_hDAT	**0.021**	0.752	57	SLC6A3	Transporter	-
	A.7	NVS_ADME_rCYP3A1	**0.035**	0.881	52	Cyp3a23/3a1	Cyp	-
	A.8	NVS_ADME_rCYP3A2	**0.021**	0.767	56	Cyp3a2	Cyp	-
	A.9	NVS_MP_hPBR	0.011	0.734	62	TSPO	Transporter	✓
	A.10	NVS_NR_hCAR_Antagonist	0.017	0.760	54	NR1I3	Nuclear receptor	✓
	A.11	OT_FXR_FXRSRC1_0480	0.014	0.748	113	NR1H4	Nuclear receptor	✓
Immunological activity	A.12	APR_HepG2_CellCycleArrest_72h_dn	**0.022**	0.749	96	-	Cell cycle	-
	A.13	└ Tox21_FXR_BLA_antagonist_ratio	0.013	0.749	96	NR1H4	Nuclear receptor	-
	A.14	BSK_BE3C_uPA_down	0.016	0.791	53	PLAU	Protease	✓
	A.15	BSK_KF3CT_IP10_down	**0.029**	0.745	104	CXCL10	Cytokine	✓
	A.16	BSK_KF3CT_MMP9_down	**0.022**	0.762	77	MMP9	Protease	✓
	A.17	BSK_LPS_CD40_down	0.017	0.752	56	CD40	Cytokine	-
	A.18	└ BSK_3C_IL8_down	0.014	0.752	56	CXCL8	Cytokine	-
	A.19	BSK_LPS_MCP1_down	0.019	0.805	75	CCL2	Cytokine	✓
	A.20	BSK_SAg_CD40_down	**0.026**	0.772	90	CD40	Cytokine	✓
	A.21	BSK_SAg_SRB_down	**0.030**	0.807	100	-	Cell cycle	✓
Nuclear receptor activity/ phenotypic readouts	A.22	APR_HepG2_MitoMembPot_72h_up	**0.020**	0.819	52	-	Cell morphology	✓
	A.23	APR_HepG2_MitoMembPot_1h_dn	**0.029**	0.888	65	-	Cell morphology	✓
	A.24	└ Tox21_AR_BLA_Antagonist_ratio	**0.020**	0.888	65	AR	Nuclear receptor	✓
	A.25	APR_HepG2_NuclearSize_24h_up	0.018	0.761	63	-	Cell morphology	✓
	A.26	APR_HepG2_OxidativeStress_1h_up	**0.025**	0.789	59	-	Cell cycle	-
	A.27	APR_HepG2_StressKinase_1h_up	**0.052**	0.956	63	-	Cell cycle	✓
	A.28	ATG_BRE_CIS_up	**0.021**	0.734	75	SMAD1	DNA binding	✓
	A.29	ATG_C_EBP_CIS_up	**0.038**	0.843	54	CEBPB	DNA binding	✓
	A.30	└ ATG_HIF1a_CIS_up	0.013	0.843	54	HIF1A	DNA binding	✓
	A.31	ATG_CRE_CIS_up	**0.048**	0.731	151	CREB3	DNA binding	✓
	A.32	ATG_FoxA2_CIS_up	0.018	0.741	57	FOXA2	DNA binding	-
	A.33	BSK_SAg_PBMCCytotoxicity_up	0.019	0.758	54	-	Cell cycle	-
	A.34	Tox21_ERa_LUC_BG1_Agonist	0.009	0.793	54	ESR1	Nuclear receptor	✓
	A.35	Tox21_GR_BLA_Antagonist_ratio	0.009	0.787	61	NR3C1	Nuclear receptor	✓
	A.36	Tox21_MitochondrialToxicity_viability	0.018	0.946	51	-	Cell cycle	✓
	A.37	└ ATG_p53_CIS_up	0.014	0.946	51	TP53	DNA binding	✓

The information gain is highlighted for values higher than median (0.02)

Table 3. Description of unique bioactivity assays in best performing rules capturing toxicity, which cover, combined, 80% of toxic compounds, at threshold of 15mg/kg/day. Rules were clustered according to similarity in compound coverage, which resulted in three major clusters of rules. Similar to findings in Table 2, the predominant assay class used by each rule cluster are Cytochrome P activity, immunological responses and nuclear receptor activity. Some rules used two assay conditions at a time, and these assays are linked in the table by the symbol; ⌊.



Rule cluster/key mechanism	Index	Associated assay	Information gain (split)	Accuracy (rule)	Coverage (rule)	Gene symbol	function	Use of Physchem*
Activity against Cytochrome P	B.1	BSK_LPS_PGE2_up	0.0215	0.861	23	PTGER2	gpcr	✓
	B.2	NVS_ADME_hCYP2C19	**0.0287**	0.856	22	CYP2C19	cyp	✓
	B.3	NVS_ADME_rCYP2A1	**0.0278**	0.853	24	Cyp2a1	cyp	-
	B.4	⌊ OT_AR_ARSRC1_0480	**0.0290**	0.853	24	AR	nuclear receptor	-
	B.5	NVS_ADME_rCYP2C12	**0.0294**	0.908	20	Cyp2c12	cyp	✓
	B.6	NVS_ADME_rCYP2C13	**0.0271**	0.854	21	Cyp2c13	cyp	✓
	B.7	NVS_ADME_rCYP2C6	**0.0294**	0.985	20	Cyp2c6v1	cyp	✓
	B.8	⌊ ATG_VDRE_CIS_up	0.0203	0.985	20	VDR	nuclear receptor	✓
Immunological activity/Endocrine disruption	B.9	BSK_3C_ICAM1_down	0.0182	0.810	25	ICAM1	cell adhesion molecules	✓
	B.10	BSK_4H_MCP1_down	**0.0370**	0.814	40	CCL2	cytokine	✓
	B.11	BSK_BE3C_MIG_down	**0.0287**	0.867	25	CXCL9	cytokine	-
	B.12	⌊ Tox21_ERa_BLA_Agonist_ratio	**0.0236**	0.867	25	ESR1	nuclear receptor	-
	B.13	BSK_hDFCGF_IP10_down	**0.0402**	0.899	25	CXCL10	cytokine	-
	B.14	⌊ Tox21_MitochondrialToxicity_viability	0.0208	0.899	25	NA	cell cycle	-
	B.15	BSK_SAg_CD40_down	0.0190	0.808	20	CD40	cytokine	✓
	B.16	Tox21_AR_BLA_Antagonist_viability	**0.0251**	0.893	25	NA	cell cycle	✓
	B.17	⌊ Tox21_PPARd_BLA_antagonist_ratio	0.0130	0.893	25	PPARD	nuclear receptor	✓
	B.18	Tox21_ERa_BLA_Antagonist_ratio	0.0191	0.856	30	ESR1	nuclear receptor	✓
Nuclear receptor activity	B.19	APR_HepG2_MitoMass_72h_up	0.0162	0.833	23	NA	cell morphology	✓
	B.20	⌊ APR_HepG2_NuclearSize_24h_up	0.0181	0.833	23	NA	cell morphology	✓
	B.21	ATG_LXRb_TRANS_up	0.0141	0.824	23	NR1H2	nuclear receptor	✓
	B.22	NVS_TR_hDAT	**0.0243**	0.848	31	SLC6A3	transporter	✓
	B.23	OT_ER_ERbERb_0480	0.0098	0.826	22	ESR2	nuclear receptor	✓
	B.24	Tox21_FXR_BLA_agonist_ratio	0.0153	0.820	23	NR1H4	nuclear receptor	✓

The information gain is highlighted for values higher than median (0.022)

Table 4. Comparison of bioactivities used in prioritized rules and assay endpoints adopted by commercial *in vitro* models for hepatotoxicity. While the assays identified as important in rules in this work and a large number of endpoints used in hepatotoxicity models agree, nuclear hormone receptor activities may constitute an additional relevant endpoint to use for this purpose, as derived from ToxCast data in the current study.

	Endpoint classes							
In vitro systems	Metabolism	Viability/phenotypic changes	Cell stress	Immune response	Bile transport	Inhibition of protein synthesis	Mitochondrial impairment	Endocrine activity
Cyprotex CellCiphr®		Apoptosis Cell cycle arrest Cell loss Cytoskeletal disruption DNA fragmentation and damage response Mitosis marker Nuclear size Phospholipidosis Steatosis	Glutathione depletion Oxidative stress Stress kinase activation Reactive oxygen species				Mitochondrial function	
InShero 3D Insight™	Cytochrome activity	Apoptosis	Glutathione depletion	IL-6 release	BSEP	Albumin	Intra-tissue ATP	
Hepregen Hepatopac	Metabolites Clearance		Glutathione levels		MRP2 CMFDA	Albumin, urea	ATP MTT	
RegeneMed	Cytochrome activity Clearance		Glutathione	Cytokine profile		Albumin, urea, fibrinogen, transferrin	ATP	
Modified rules	Cytochrome activity	Cell cycle arrest Cytotoxicity Nuclear size Mitochondrial mass	Oxidative stress Stress kinase	IL-9, IL-10,CCL2 and CD40	FXR	Cellular protein content	Mitochondrial membrane potential/toxicity	Estrogen and androgen receptors activity

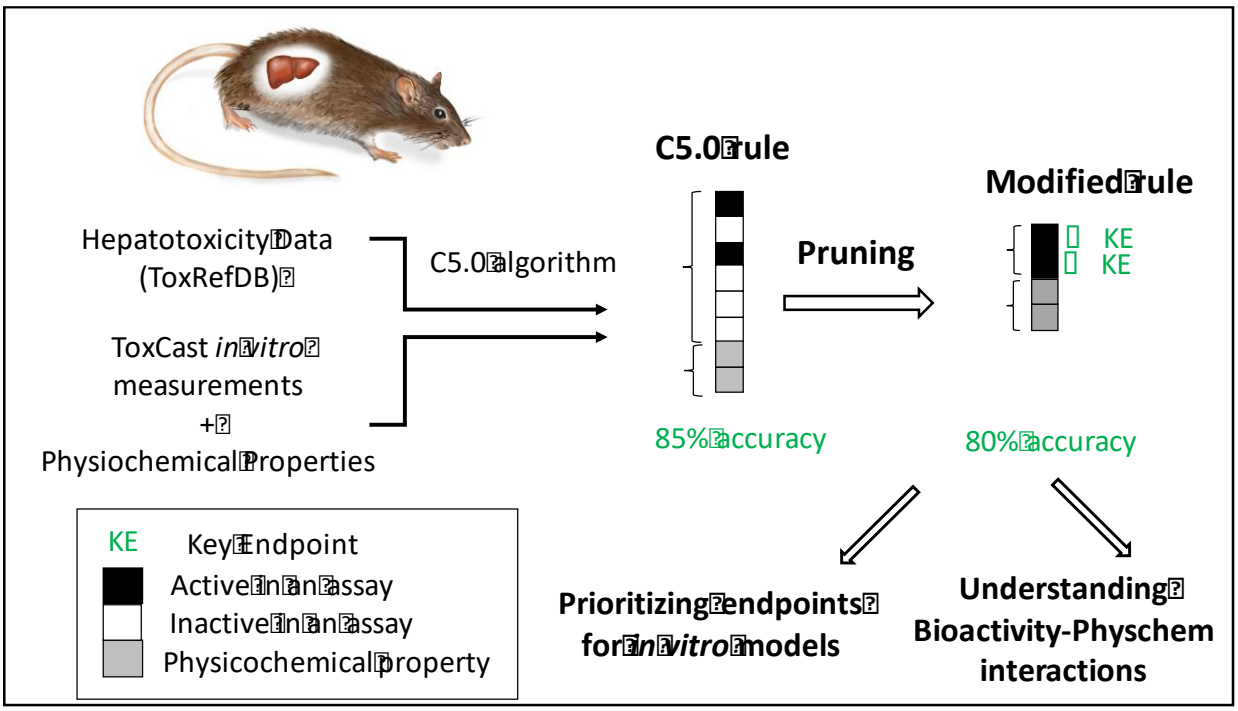
Table 5. Physicochemical properties frequently present in in prioritized hepatotoxicity rules. The most frequent physicochemical property in rules for a toxicity threshold of 15mg/kg/day is the number of rotatable bonds, while the number of rings is the most frequent physicochemical property in rules predicting toxicity at a 500mg/kg/day level.

	15mg/kg/day	
Physicochemical condition	Error rate%*	Frequency %**
NumRotatableBonds <= 6	7.8 ± 3.2	35
NumHBD <= 0	9.2 ± 3.7	10
NumAliphaticRings <= 2	2.7 ± 0.3	10

	500mg/kg/day	
Physicochemical condition	Error rate %*	Frequency %**
NumRings <= 3	5.7±3.6	29
NumHeavyAtoms <= 33	3.9±0.5	11
NumAromaticCarbocycles > 0	11.5±1.9	9

*Error rate % represents the deterioration in rule accuracy as a result of removing each of the physicochemical properties presented in the table (mean ± standard deviation).
** Frequency % is the percentage of rules containing the physicochemical property out of all prioritized rule set.

TABLE OF CONTENT FIGURE



ASSOCIATED CONTENT

Supporting Information

Supporting Information includes (1) The metric for calculating univariate associations Figure S1 and the distributions of predictive values Figure S2, (2) The distribution of original models accuracy generated by cross validation Figure S3, (3) Distribution for rule accuracy and coverage before and after modification Figure S4, (4) Plot for the values of assay conditions split points (bioactivity cutoffs) Figure S5, (5) Number of rules matched by toxic and non-toxic compounds at levels of 15mg/kg/day and 500mg/kg/day Figure S6, (6) Error rate from removing physicochemical property conditions at each toxicity level Figure S7, (7) Number of rules matched

by each toxic and non-toxic compounds in training data at levels of 500mg/kg/day Figure S8, (8) Number of rules matched by drugs of different DILI concern levels using rules describing toxicity at 500mg/kg/day Figure S9, (9) Number of rules matched by each toxic and non-toxic compounds in training data at levels of 15mg/kg/day Figure S10, (10) Number of rules matched by drugs of different DILI concern levels using rules describing toxicity at 15mg/kg/day Figure S11, (11) Percentage of activity calls in equivalent assays (determined at level of 15mg/kg/day) to overall activity calls (at 100μM) for 29 drugs with Drug Induced Liver Injury (DILI) labels Figures S12 and S13, (12) Number of rotatable bonds in drugs with different levels of DILI concern. Figure S14 ,(13) list of hepatotoxicity endpoints from ToxRefDB Table S1, (14) list of 361 ToxCast assays used in analysis Table S2, (15) list of physicochemical properties generated by RDKit Table S3, (16) List of compounds ToxCast ID, names and CAS number Table S4, (17) list the prioritized rules and corresponding performance for toxicity levels of 500mg/kg/day and 15mg/kg/day, respectively, Table S5 and Table S6, (18) the accuracy of individual rules after removing physicochemical properties for toxicity levels of 500mg/kg/day and 15mg/kg/day, respectively, Table S7 and Table S8, (19) the rules describing troglitazone at levels 500mg/kg/day and 15mg/kg/day, Table S9, (20) List of 29 drugs with DILI concern Table S10, (21) Lists of assays equivalent to biological space identified by the rules Tables S11 and S12. This material is available free of charge via the Internet at http:// pubs.acs.org.

Funding Sources

This research was supported by the Islamic Development Bank and Cambridge Trust Fund.

Acknowledgement

We would like to thank Lilia Fisk, Jonathan Vessey for their valuable feedback and Lewis Mervin for proofreading.

REFERENCES

(1) Gleeson, M. P., Modi, S., Bender, A., Robinson, R. L. M., Kirchmair, J., Promkatkaew, M., Hannongbua, S., and Glen, R. C. (2012) The challenges involved in modeling toxicity data in silico: a review. *Curr. Pharm. Des. 18*, 1266–91.

(2) Kramer, J. A., Sagartz, J. E., and Morris, D. L. (2007) The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. *Nat. Rev. Drug Discov. 6*, 636–649.

(3) Bale, S., Vernetti, L., Senutovitch, N., Jindal, R., Hegde, M., Gough, A., Mccarty, W. J., Bakan, A., Bhushan, A., Shun, T. Y., Golberg, I., Debiasio, R., Osman Usta, B., Taylor, D. L., and Yarmush, M. L. (2014) In Vitro Platforms for Evaluating Liver Toxicity. *Exp Biol Med 239*, 1180–1191.

(4) Dix, D. J., Houck, K. A., Judson, R. S., Kleinstreuer, N. C., Knudsen, T. B., Martin, M. T., Reif, D. M., Richard, A. M., Shah, I., Sipes, N. S., and Kavlock, R. J. (2012) Incorporating biological, chemical, and toxicological knowledge into predictive models of toxicity. *Toxicol. Sci. 130*, 440–441.

(5) Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., Reif, D. M., Rotroff, D. M., Shah, I., Richard, A. M., and Dix, D. J. (2010) In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environ. Health Perspect. 118*, 485–492.

(6) Kim, M. T., Huang, R., Sedykh, A., Wang, W., Xia, M., and Zhu, H. (2016) Mechanism Profiling of Hepatotoxicity Caused by Oxidative Stress Using Antioxidant Response Element Reporter Gene Assay Models and Big Data. *Environ. Health Perspect. 124*, 634–41.

(7) Wetmore, B. A., Wambaugh, J. F., Allen, B., Ferguson, S. S., Sochaski, M. A., Setzer, R. W., Houck, K. A., Strope, C. L., Cantwell, K., Judson, R. S., LeCluyse, E., Clewell, H. J., Thomas, R. S., and Andersen, M. E. (2015) Incorporating High-Throughput Exposure Predictions with Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing. *Toxicol. Sci. 148*, 121–136.

(8) Wambaugh, J. F., Hughes, M. F., Ring, C. L., MacMillan, D. K., Ford, J., Fennell, T. R., Black, S. R., Snyder, R. W., Sipes, N. S., Wetmore, B. A., Westerhout, J., Setzer, R. W., Pearce, R. G., Simmons, J. E., and Thomas, R. S. (2018) Evaluating in vitro-in vivo extrapolation of toxicokinetics. *Toxicol. Sci. 163*, 152–169.

(9) Shah, F., Leung, L., Barton, H. A., Will, Y., Rodrigues, A. D., Greene, N., and Aleo, M. D. (2015) Setting Clinical Exposure Levels of Concern for Drug-Induced Liver Injury (DILI) Using Mechanistic in vitro Assays. *Toxicol. Sci. 147*, 500–514.

(10) Thompson, R. A., Isin, E. M., Li, Y., Weidolf, L., Page, K., Wilson, I., Swallow, S., Middleton, B., Stahl, S., Foster, A. J., Dolgos, H., Weaver, R., and Kenna, J. G. (2012) In vitro approach to assess the potential for risk of idiosyncratic adverse reactions caused by candidate drugs. *Chem. Res. Toxicol. 25*, 1616–1632.

(11) Sakatis, M. Z., Reese, M. J., Harrell, A. W., Taylor, M. A., Baines, I. A., Chen, L., Bloomer, J. C., Yang, E. Y., Ellens, H. M., Ambroso, J. L., Lovatt, C. A., Ayrton, A. D., and Clarke, S. E. (2012) Preclinical strategy to reduce clinical hepatotoxicity using in vitro bioactivation data for >200 compounds. *Chem. Res. Toxicol. 25*, 2067–2082.

(12) Kostadinova, R., Boess, F., Applegate, D., Suter, L., Weiser, T., Singer, T., Naughton, B., and Roth, A. (2013) A long-term three dimensional liver co-culture system for improved prediction of clinically relevant drug-induced hepatotoxicity. *Toxicol. Appl. Pharmacol. 268*, 1–16.

(13) O'Brien, P. J., Irwin, W., Diaz, D., Howard-Cofield, E., Krejsa, C. M., Slaughter, M. R., Gao, B., Kaludercic, N., Angeline, A., Bernardi, P., Brain, P., and Hougham, C. (2006) High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. *Arch. Toxicol. 80*, 580–604.

(14) Persson, M., Loye, A. F., Jacquet, M., Mow, N. S., Thougaard, A. V., Mow, T., and Hornberg, J. J. (2014) High-content analysis/screening for predictive toxicology: Application to hepatotoxicity and genotoxicity. *Basic Clin. Pharmacol. Toxicol. 115*, 18–23.

(15) Persson, M., Løye, A. F., Mow, T., and Hornberg, J. J. (2013) A high content screening assay to predict human drug-induced liver injury during drug discovery. *J. Pharmacol. Toxicol. Methods 68*, 302–313.

(16) Kleinstreuer, N. C., Yang, J., Berg, E. L., Knudsen, T. B., Richard, A. M., Martin, M. T., Reif, D. M., Judson, R. S., Polokoff, M., Dix, D. J., Kavlock, R. J., and Houck, K. A. (2014) Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat. Biotechnol. 32*, 583–91.

(17) Bisgin, H. (2014) Toward predictive models for drug-induced liver injury in humans : are we there yet ? *Biomarkers Med 8*, 201–213.

(18) Kumar, R., Sharma, A., and Varadwaj, P. K. (2011) A prediction model for oral bioavailability of drugs using physicochemical properties by support vector machine. *J. Nat. Sci. Biol. Med. 2*, 168–73.

(19) Khojasteh, S. C., Wong, H., and Hop, C. E. C. A. (2011) ADME Properties and Their Dependence on Physicochemical Properties, in *Drug Metabolism and Pharmacokinetics Quick Guide*, pp 165–181. Springer New York, New York, NY.

(20) Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W., and Kavlock, R. J. (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci. 95*, 5–12.

(21) Richard, A. M., Judson, R. S., Houck, K. A., Grulke, C. M., Volarath, P., Thillainadarajah, I., Yang, C., Rathman, J., Martin, M. T., Wambaugh, J. F., Knudsen, T. B., Kancherla, J., Mansouri, K., Patlewicz, G., Williams, A. J., Little, S. B., Crofton, K. M., and Thomas, R. S. (2016) ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. *Chem. Res. Toxicol. 29*, 1225–1251.

(22) Hu, B., Gifford, E., Wang, H., Bailey, W., and Johnson, T. (2015) Analysis of the ToxCast Chemical-Assay Space using the Comparative Toxicogenomics Database. *Chem. Res. Toxicol. 28*, 2210–2223.

(23) Liu, J., Mansouri, K., Judson, R. S., Martin, M. T., Hong, H., Chen, M., Xu, X., Thomas, R. S., and Shah, I. (2015) Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. *Chem. Res. Toxicol. 28*, 738−751.

(24) Plunkett, L. M., Kaplan, A. M., and Becker, R. A. (2015) Challenges in using the ToxRefDB as a resource for toxicity prediction modeling. *Regul. Toxicol. Pharmacol. 72*, 610–614.

(25) Hong-Geller, E., Micheva-Viteva, S., Benjamin, B., Kanta Barman, T., Chaira, T., and K. Paliwal, J. (2010) Integration of Physicochemical and Pharmacokinetic Parameters in Lead Optimization: A Physiological Pharmacokinetic Model Based Approach. *Curr. Drug Discov. Technol. 7*, 143–153.

(26) Fuernkranz, J., Gamberger, D., and Lavrač, N. (2012) Rule Learning in a Nutshell. *Found. Rule Learn.*

(27) Martin, M. T., Judson, R. S., Reif, D. M., Kavlock, R. J., and Dix, D. J. (2009) Profiling Chemicals Based on Chronic Toxicity Results from the U.S. EPA ToxRef Database. *Environ. Health Perspect. 117*, 392–399.

(28) EPA, U. S. (2016) Toxicity ForeCaster (ToxCastTM) Data. *https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data*.

(29) Fay, K. A., Villeneuve, D. L., Swintek, J., Edwards, S. W., Nelms, M. D., Blackwell, B. R., and Ankley, G. T. (2018) Differentiating Pathway-Specific From Nonspecific Effects in High-Throughput Toxicity Data: A Foundation for Prioritizing Adverse Outcome Pathway Development. *Toxicol. Sci. 163*, 500–515.

(30) Standardizer was used for structure canonicalization and transformation, JChem 15.12.14.0, 2015, ChemAxon (http://www.chemaxon.com).

(31) Landrum, G. RDKit: Open source chemoinformatics.

(32) Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., and Wiswedel, B. (2009) KNIME - the Konstanz information miner. *ACM SIGKDD Explor. Newsl. 11*, 26.

(33) Calculator Plugins: Calculator Plugins were used for structure property prediction and calculation, Marvin 15.12.14.0, 2015, ChemAxon (http://www.chemaxon.com).

(34) ChemAxon. Instant JChem was used for structure database management, search and prediction, Instant JChem 15.12.14.0, 2015, ChemAxon (http://www.chemaxon.com).

(35) Quinlan, J. R. (John R., and Ross, J. (1993) C4.5 : programs for machine learning. Morgan Kaufmann Publishers.

(36) Kuhn, M., Weston, S., Coulter, N., Culp, M. C., and Maintainer, R. Q. (2015) Package "C50" Title C5.0 Decision Trees and Rule-Based Models code for C5.0.

(37) Kuhn, M. (2008) Building Predictive Models in R Using the caret Package. *J. Stat. Softw. 28*, 1–26.

(38) SHANNON, C. E. (1948) The mathematical theory of communication. *Bell Syst. Tech. J. 27*, 379–423.

(39) Real, R., and Vargas, J. M. (1996) The probabilistic basis of Jaccard's index of similarity. *Syst. Biol. 45*, 380–385.

(40) Murtagh, F., and Legendre, P. (2014) Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif. 31*, 274–295.

(41) Computing, R Foundation for Statistical Vienna, A. (2008) R: A language and environment for statistical computing. Vienna, Austria.

(42) Wickham, H., and Chang, W. (2016) Package 'ggplot2.'

(43) Thomas, R. S., Black, M. B., Li, L., Healy, E., Chu, T.-M., Bao, W., Andersen, M. E., and Wolfinger, R. D. (2012) A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. *Toxicol. Sci. 128*, 398–417.

(44) Russmann, S., Kullak-Ublick, G., and Grattagliano, I. (2009) Current Concepts of Mechanisms in Drug-Induced Hepatotoxicity. *Curr. Med. Chem. 16*, 3041–3053.

(45) Huang, R., Xia, M., Sakamuru, S., Zhao, J., Shahane, S. A., Attene-Ramos, M., Zhao, T., Austin, C. P., and Simeonov, A. (2016) Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. *Nat. Commun. 7*, 10425.

(46) Wienkers, L. C., and Heath, T. G. (2005) Predicting in vivo drug interactions from in vitro drug discovery data. *Nat. Rev. Drug Discov. 4*, 825–833.

(47) Feng, S., and He, X. (2013) Mechanism-based inhibition of CYP450: an indicator of drug-induced hepatotoxicity. *Curr. Drug Metab. 14*, 921–45.

(48) Murugaiyan, G., Martin, S., and Saha, B. (2007) CD40-induced countercurrent conduits for tumor escape or elimination? *Trends Immunol. 28*, 467–473.

(49) Zhai, Y., Shen, X. D., Gao, F., Zhao, A., Freitas, M. C., Lassman, C., Luster, A. D., Busuttil, R. W., and Kupiec-Weglinski, J. W. (2008) CXCL10 regulates liver innate immune response against ischemia and reperfusion injury. *Hepatology 47*, 207–214.

(50) Schmitz, V., Dombrowski, F., Prieto, J., Qian, C., Diehl, L., Knolle, P., Sauerbruch, T., Caselmann, W. H., Spengler, U., and Leifeld, L. (2006) Induction of murine liver damage by overexpression of CD40 ligand provides an experimental model to study fulminant hepatic failure. *Hepatology 44*, 430–439.

(51) Shu, U., Kiniwa, M., Wu, C. Y., Maliszewski, C., Vezzio, N., Hakimi, J., Gately, M., and Delespesse, G. (1995) Activated T cells induce interleukin-12 production by monocytes via

CD40-CD40 ligand interaction. *Eur. J. Immunol. 25*, 1125–1128.

(52) Karthivashan, G., Arulselvan, P., and Fakurazi, S. (2015) Pathways involved in acetaminophen hepatotoxicity with specific targets for inhibition/downregulation. *RSC Adv. 5*, 62040–62051.

(53) Bone-Larson, C. L., Hogaboam, C. M., Evanhoff, H., Strieter, R. M., and Kunkel, S. L. (2001) IFN-gamma-inducible protein-10 (CXCL10) is hepatoprotective during acute liver injury through the induction of CXCR2 on hepatocytes. *J. Immunol. 167*, 7077–83.

(54) Jaeschke, H., Williams, C. D., Ramachandran, A., and Bajt, M. L. (2012) Acetaminophen hepatotoxicity and repair: the role of sterile inflammation and innate immunity. *Liver Int. 32*, 8–20.

(55) Saiman, Y., and Friedman, S. L. (2012) The role of chemokines in acute liver injury. *Front. Physiol. 3*, 213.

(56) McMahan, R. H., Porsche, C. E., Edwards, M. G., and Rosen, H. R. (2016) Free fatty acids differentially downregulate chemokines in liver sinusoidal endothelial cells: Insights into non-alcoholic fatty liver disease. *PLoS One 11*, e0159217.

(57) Yamamoto, Y. (2006) Estrogen Receptor Mediates 17 -Ethynylestradiol Causing Hepatotoxicity. *J. Biol. Chem. 281*, 16625–16631.

(58) Mueller, K. M., Kornfeld, J.-W., Friedbichler, K., Blaas, L., Egger, G., Esterbauer, H., Hasselblatt, P., Schlederer, M., Haindl, S., Wagner, K.-U., Engblom, D., Haemmerle, G., Kratky, D., Sexl, V., Kenner, L., Kozlov, A. V, Terracciano, L., Zechner, R., Schuetz, G., Casanova, E., Pospisilik, J. A., Heim, M. H., and Moriggl, R. (2011) Impairment of hepatic growth hormone and glucocorticoid receptor signaling causes steatosis and hepatocellular carcinoma in mice. *Hepatology 54*, 1398–409.

(59) Thole, Z., Manso, G., Salgueiro, E., Revuelta, P., and Hidalgo, A. (2004) Hepatotoxicity induced by antiandrogens: A review of the literature. *Urol. Int. 73*, 289–295.

(60) Zhang, J., Huang, W., Qatanani, M., Evans, R. M., and Moore, D. D. (2004) The constitutive androstane receptor and pregnane X receptor function coordinately to prevent bile acid-induced hepatotoxicity. *J. Biol. Chem. 279*, 49517–49522.

(61) Lee, F. Y., de Aguiar Vallim, T. Q., Chong, H. K., Zhang, Y., Liu, Y., Jones, S. A., Osborne, T. F., and Edwards, P. A. (2010) Activation of the Farnesoid X Receptor Provides Protection against Acetaminophen-Induced Hepatic Toxicity. *Mol. Endocrinol. 24*, 1626–1636.

(62) Collin, M., Abdelrahman, M., and Thiemermann, C. (2004) Endogenous ligands of PPAR-γ reduce the liver injury in haemorrhagic shock. *Eur. J. Pharmacol. 486*, 233–235.

(63) Shan, W., Nicol, C. J., Ito, S., Bility, M. T., Kennett, M. J., Ward, J. M., Gonzalez, F. J., and Peters, J. M. (2007) Peroxisome proliferator-activated receptor-β/δ protects against chemically induced liver toxicity in mice. *Hepatology 47*, 225–235.

(64) Osawa, Y., Oboki, K., Imamura, J., Kojika, E., Hayashi, Y., Hishima, T., Saibara, T.,

Shibasaki, F., Kohara, M., and Kimura, K. (2015) Inhibition of Cyclic Adenosine Monophosphate (cAMP)-response Element-binding Protein (CREB)-binding Protein (CBP)/β-Catenin Reduces Liver Fibrosis in Mice. *EBIOM 2*, 1751–1758.

(65) AlSharari, S. D., Al-Rejaie, S. S., Abuohashish, H. M., Ahmed, M. M., and Hafez, M. M. (2016) Rutin Attenuates Hepatotoxicity in High-Cholesterol-Diet-Fed Rats. *Oxid. Med. Cell. Longev. 2016*, 1–11.

(66) Dooley, S., and Ten Dijke, P. (2012) TGF-β in progression of liver disease. *Cell Tissue Res. 347*, 245–256.

(67) Wagner, M., Zollner, G., and Trauner, M. (2011) Nuclear receptors in liver disease. *Hepatology 53*, 1023–1034.

(68) Zuniga, S., Firrincieli, D., Housset, C., and Chignard, N. (2011) Vitamin D and the vitamin D receptor in liver pathophysiology. *Clin. Res. Hepatol. Gastroenterol. 35*, 295–302.

(69) Saini, S. P. S., Zhang, B., Niu, Y., Jiang, M., Gao, J., Zhai, Y., Hoon Lee, J., Uppal, H., Tian, H., Tortorici, M. A., Poloyac, S. M., Qin, W., Venkataramanan, R., and Xie, W. (2011) Activation of liver X receptor increases acetaminophen clearance and prevents its toxicity in mice. *Hepatology 54*, 2208–17.

(70) Lemasters, J. J., Qian, T., Bradham, C. A., Brenner, D. A., Cascio, W. E., Trost, L. C., Nishimura, Y., Nieminen, A. L., and Herman, B. (1999) Mitochondrial dysfunction in the pathogenesis of necrotic and apoptotic cell death. *J. Bioenerg. Biomembr. 31*, 305–19.

(71) Jaeschke, H., Gores, G. J., Cederbaum, A. I., Hinson, J. A., Pessayre, D., and Lemasters, J. J. (2002) Mechanisms of hepatotoxicity. *Toxicol. Sci. 65*, 166–176.

(72) Judson, R., Houck, K., Martin, M., Richard, A. M., Knudsen, T. B., Shah, I., Little, S., Wambaugh, J., Woodrow Setzer, R., Kothya, P., Phuong, J., Filer, D., Smith, D., Reif, D., Rotroff, D., Kleinstreuer, N., Sipes, N., Xia, M., Huang, R., Crofton, K., and Thomas, R. S. (2016) Analysis of the Effects of Cell Stress and Cytotoxicity on In Vitro Assay Activity Across a Diverse Chemical and Assay Space. *Toxicol. Sci. 152*, 323–339.

(73) Vairetti, M., Ferrigno, A., Rizzo, V., Ambrosi, G., Bianchi, A., Richelmi, P., Blandini, F., and Armentero, M. T. (2012) Impaired hepatic function and central dopaminergic denervation in a rodent model of Parkinson's disease: A self-perpetuating crosstalk? *Biochim. Biophys. Acta - Mol. Basis Dis. 1822*, 176–184.

(74) Veenman, L., and Gavish, M. (2012) The Role of 18 kDa Mitochondrial Translocator Protein (TSPO) in Programmed Cell Death, and Effects of Steroids on TSPO Expression. *Curr. Mol. Med. 12*, 398–412.

(75) Hatori, A., Yui, J., Xie, L., Yamasaki, T., Kumata, K., Fujinaga, M., Wakizaka, H., Ogawa, M., Nengaki, N., Kawamura, K., and Zhang, M. R. (2014) Visualization of acute liver damage induced by cycloheximide in rats using PET with [18F]FEDAC, a radiotracer for translocator protein (18 kDa). *PLoS One 9*, e86625.

(76) Rincón-Sánchez, A. R., Covarrubias, A., Rivas-Estilla, A. M., Pedraza-Chaverrí, J., Cruz,

C., Islas-Carbajal, M. C., Panduro, A., Estanes, A., and Armendáriz-Borunda, J. (2005) PGE2 alleviates kidney and liver damage, decreases plasma renin activity and acute phase response in cirrhotic rats with acute liver damage. *Exp. Toxicol. Pathol. 56*, 291–303.

(77) Takano, M., Nishimura, H., Kimura, Y., Washizu, J., Mokuno, Y., Nimura, Y., and Yoshikai, Y. (1998) Prostaglandin E2 protects against liver injury after Escherichia coli infection but hampers the resolution of the infection in mice. *J Immunol 161*, 3019–3025.

(78) Rudnick, D. A., Perlmutter, D. H., and Muglia, L. J. (2001) Prostaglandins are required for CREB activation and cellular proliferation during liver regeneration. *Proc. Natl. Acad. Sci. U. S. A. 98*, 8885–8890.

(79) Zhang, H., Liu, Y., Wang, L., Li, Z., Zhang, H., Wu, J., Rahman, N., Guo, Y., Li, D., Li, N., Huhtaniemi, I., Tsang, S. Y., Gao, G. F., and Li, X. (2013) Differential effects of estrogen/androgen on the prevention of nonalcoholic fatty liver disease in the male rat. *J. Lipid Res. 54*, 345–57.

(80) Hanioka, N., Gonzalez, F. J., Lindberg, N. A., Liu, G., Gelboin, H. V., and Korzekwa, K. R. (1992) Site-directed mutagenesis of cytochrome P450s CYP2A1 and CYP2A2: influence of the distal helix on the kinetics of testosterone hydroxylation. *Biochemistry 31*, 3364–3370.

(81) Williams, T. M., and Borghoff, S. J. (2000) Induction of Testosterone Biotransformation Enzymes following Oral Administration of Methyl tert-Butyl Ether to Male Sprague-Dawley Rats. *Toxicol. Sci. 57*, 147–155.

(82) Olokpa, E., Bolden, A., and Stewart, L. V. (2016) The Androgen Receptor Regulates PPARγ Expression and Activity in Human Prostate Cancer Cells. *J. Cell. Physiol. 231*, 2664–72.

(83) Olokpa, E., Moss, P. E., and Stewart, L. V. (2017) Crosstalk between the Androgen Receptor and PPAR Gamma Signaling Pathways in the Prostate. *PPAR Res. 2017*, 1–13.

(84) Lin, H.-Y., Yu, I.-C., Wang, R.-S., Chen, Y.-T., Liu, N.-C., Altuwaijri, S., Hsu, C.-L., Ma, W.-L., Jokinen, J., Sparks, J. D., Yeh, S., and Chang, C. (2008) Increased hepatic steatosis and insulin resistance in mice lacking hepatic androgen receptor. *Hepatology 47*, 1924–1935.

(85) Daskalopoulos, E. P., Lang, M. A., Marselos, M., Malliou, F., and Konstandi, M. (2012) D2-Dopaminergic Receptor-Linked Pathways: Critical Regulators of CYP3A, CYP2C, and CYP2D. *Mol. Pharmacol. 82*, 668–678.

(86) Harkitis, P., Daskalopoulos, E. P., Malliou, F., Lang, M. A., Marselos, M., Fotopoulos, A., Albucharali, G., and Konstandi, M. (2015) Dopamine D2-receptor antagonists down-regulate CYP1A1/2 and CYP1B1 in the rat liver. *PLoS One 10*, e0128708.

(87) Schwabe, R. F., Schnabl, B., Kweon, Y. O., and Brenner, D. A. (2001) CD40 Activates NF-B and c-Jun N-Terminal Kinase and Enhances Chemokine Secretion on Activated Human Hepatic Stellate Cells. *J. Immunol. 166*, 6812–6819.

(88) Yamaguchi, J., Tanaka, T., Eto, N., and Nangaku, M. (2015) Inflammation and hypoxia linked to renal injury by CCAAT/enhancer-binding protein δ. *Kidney Int. 88*, 262–275.

(89) Anderson, E. R., Taylor, M., Xue, X., Martin, A., Moons, D. S., Omary, M. B., and Shah, Y. M. (2012) The hypoxia-inducible factor-C/EBPα axis controls ethanol-mediated hepcidin repression. *Mol. Cell. Biol. 32*, 4068–77.

(90) Jakobsen, J. S., Waage, J., Rapin, N., Bisgaard, H. C., Larsen, F. S., and Porse, B. T. (2013) Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries. *Genome Res. 23*, 592–603.

(91) Barchetta, I., Carotti, S., Labbadia, G., Gentilucci, U. V., Muda, A. O., Angelico, F., Silecchia, G., Leonetti, F., Fraioli, A., Picardi, A., Morini, S., and Cavallo, M. G. (2012) Liver vitamin D receptor, CYP2R1, and CYP27A1 expression: relationship with liver histology and vitamin D3 levels in patients with nonalcoholic steatohepatitis or hepatitis C virus. *Hepatology 56*, 2180–2187.

(92) Petta, S., Grimaudo, S., Tripodo, C., Cabibi, D., Calvaruso, M., Di Cristina, A., Guarnotta, C., Macaluso, F. S. alvatore, Minissale, M. G. iovanna, Marchesini, G., and Craxì, A. (2015) The hepatic expression of vitamin D receptor is inversely associated with the severity of liver damage in genotype 1 chronic hepatitis C patients. *J. Clin. Endocrinol. Metab. 100*, 193–200.

(93) Chen, Y., and Goldstein, J. A. (2009) The transcriptional regulation of the human CYP2C genes. *Curr. Drug Metab. 10*, 567–78.

(94) Cook, L. C., Hillhouse, A. E., Myles, M. H., Lubahn, D. B., Bryda, E. C., Davis, J. W., and Franklin, C. L. (2014) The role of estrogen signaling in a mouse model of inflammatory bowel disease: a Helicobacter hepaticus model. *PLoS One 9*, e94209.

(95) Zeremski, M., Dimova, R., Astemborski, J., Thomas, D. L., and Talal, A. H. (2011) CXCL9 and CXCL10 chemokines as predictors of liver fibrosis in a cohort of primarily African-American injection drug users with chronic hepatitis C. *J. Infect. Dis. 204*, 832–6.

(96) Jaeschke, H. (2007) Troglitazone hepatotoxicity: Are we getting closer to understanding idiosyncratic liver injury? *Toxicol. Sci. 97*, 1–3.

(97) Masubuchi, Y. (2006) Metabolic and Non-Metabolic Factors Determining Troglitazone Hepatotoxicity: A Review. *Drug Metab. Pharmacokinet. 21*, 347–356.

(98) Vansant, G., Pezzoli, P., Saiz, R., Birch, A., Duffy, C., Ferre, F., and Monforte, J. (2006) Gene Expression Analysis of Troglitazone Reveals Its Impact on Multiple Pathways in Cell Culture: A Case for In Vitro Platforms Combined with Gene Expression Analysis for Early (Idiosyncratic) Toxicity Screening. *Int. J. Toxicol. 25*, 85–94.

(99) Ong, M. M. K., Latchoumycandane, C., and Boelsterli, U. A. (2007) Troglitazone-Induced Hepatic Necrosis in an Animal Model of Silent Genetic Mitochondrial Abnormalities. *Toxicol. Sci. 97*, 205–213.

(100) Giuliano, K. A., Gough, A. H., Taylor, D. L., Vernetti, L. A., and Johnston, P. A. (2010) Early safety assessment using cellular systems biology yields insights into mechanisms of action. *J. Biomol. Screen. 15*, 783–797.

(101) Messner, S., Agarkova, I., Moritz, W., and Kelm, J. M. (2013) Multi-cell type human liver microtissues for hepatotoxicity testing. *Arch. Toxicol. 87*, 209–213.

(102) Khetani, S., and Bhatia, S. (2007) Microscale culture of human liver cells for drug development. *Nat. Biotechnol. 26*, 120–6.

(103) Hop, C. E. C. A. (2012) Role of ADME Studies in Selecting Drug Candidates : Dependence of ADME Parameters on Physicochemical Properties. *Encycl. Drug Metab. Interact.* 1– 43.

(104) Veber, D. F., Johnson, S. R., Cheng, H., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002) Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem. 45*, 2615–2623.

(105) Ritchie, T. J., and Macdonald, S. J. F. (2009) The impact of aromatic ring count on compound developability - are too many aromatic rings a liability in drug design? *Drug Discov. Today 14*, 1011–1020.

(106) Trainor, G. L. (2007) The importance of plasma protein binding in drug discovery. *Expert Opin. Drug Discov. 2*, 51–64.

(107) Pellegatti, M., Pagliarusco, S., Solazzo, L., and Colato, D. (2011) Plasma protein binding and blood-free concentrations: which studies are needed to develop a drug? *Expert Opin. Drug Metab. Toxicol. 7*, 1009–1020.

(108) Chen, M., Borlak, J., and Tong, W. (2013) High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury. *Hepatology 58*, 388–396.

(109) Degoey, D. A., Chen, H. J., Cox, P. B., and Wendt, M. D. (2018) Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection. *J. Med. Chem. 61*, 2636–2651.

(110) Chen, M., Suzuki, A., Thakkar, S., Yu, K., Hu, C., and Tong, W. (2016) DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discov. Today 21*, 648–653.

(111) Wittwehr, C., Aladjov, H., Ankley, G., Byrne, H. J., de Knecht, J., Heinzle, E., Klambauer, G., Landesmann, B., Luijten, M., MacKay, C., Maxwell, G., Meek, M. E. B., Paini, A., Perkins, E., Sobanski, T., Villeneuve, D., Waters, K. M., and Whelan, M. (2017) How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. *Toxicol. Sci. 155*, 326–336.

(112) Benigni, R., Bossa, C., Giuliani, A., and Tcheremenskaia, O. (2010) Exploring in vitro/in vivo correlation: lessons learned from analyzing phase I results of the US EPA's ToxCast Project. *J. Environ. Sci. Health. C. Environ. Carcinog. Ecotoxicol. Rev. 28*, 272–86.

(113) Benigni, R., Bossa, C., Tcheremenskaya, O., and Giuliani, A. (2010) In Vitro / In Vivo Relationship in the Light of Toxcast Phase 1. *ALTEX Altern. zu Tierexperimenten 27*, 269–274.