Chemical Research in Toxicology

Article

pubs.acs.org/crt

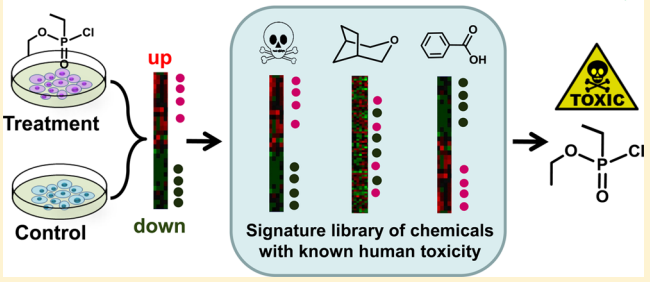
Using Chemical-Induced Gene Expression in Cultured Human Cells to Predict Chemical Toxicity

Ruifeng Liu,* Xueping Yu,† and Anders Wallqvist*

Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, Maryland 21702, United States

Supporting Information

ABSTRACT: Chemical toxicity is conventionally evaluated in animal models. However, animal models are resource intensive; moreover, they face ethical and scientific challenges because the outcomes obtained by animal testing may not correlate with human responses. To develop an alternative method for assessing chemical toxicity, we investigated the feasibility of using chemical-induced genome-wide expression changes in cultured human cells to predict the potential of a chemical to cause specific organ injuries in humans. We first created signatures of chemical-induced gene expression in a vertebral-cancer of the prostate cell line for ~15,000 chemicals tested in the US National Institutes of Health Library of Integrated Network-Based Cellular Signatures program. We then used the signatures to create naïve Bayesian prediction models for chemical-induced human liver cholestasis, interstitial nephritis, and long QT syndrome. Detailed cross-validation analyses indicated that the models were robust with respect to false positives and false negatives in the samples we used to train the models and could predict the likelihood that chemicals would cause specific organ injuries. In addition, we performed a literature search for drugs and dietary supplements, not formally categorized as causing organ injuries in humans but predicted by our models to be most likely to do so. We found a high percentage of these compounds associated with case reports of relevant organ injuries, lending support to the idea that *in vitro* cell-based experiments can be used to predict the toxic potential of chemicals. We believe that this approach, combined with a robust technique to model human exposure to chemicals, may serve as a promising alternative to animal-based chemical toxicity assessment.



1. INTRODUCTION

As new chemical synthesis and manufacturing technologies rapidly advance, industrial applications and household products employ an increasing number of new chemicals in higher quantities. To reduce the risk of exposure to toxic chemicals, the European Union enacted legislations on the Registration, Evaluation, Authorization, and Restriction of CHemical Substances (REACH) to ensure that chemicals do not adversely affect human health or the environment.[1] Other countries, including the U.S., have also implemented similar regulations.[2] However, toxicity information is currently available for only a small fraction of the chemicals currently on the market. A major reason for this is that conventional animal toxicity studies are resource intensive. They are also subject to both ethical and scientific challenges because the results may not be relevant to human health.[3] Thus, there is a need for alternative methods, including *in vitro* and *in silico* models, to fill the gap.[2,4]

In concert with the rapid advances in genomic technology, the cost of assessing genome-wide expression changes due to chemical exposure has markedly decreased. Many *in vitro* toxicogenomic assays have been developed and evaluated as alternatives to animal models for safety assessment.[5] In addition, as large toxicogenomic data sets have become publicly available, many researchers have investigated transcriptional signatures (e.g., coregulated genes) that correlate with specific toxicity end points[6−9] and developed prediction models based on chemical-induced gene expression changes and machine learning algorithms.[10−12] However, most of these studies have used small data sets, with the number of chemical compounds ranging from a few tens to less than 200. In addition, most studies have investigated the correlation between chemical-induced gene expression features and animal toxicity.

In this article, we report the feasibility of using chemical-induced genome-wide expression changes in cultured human cells to predict the potential of chemicals to cause specific organ injuries in humans. We hypothesized that if two chemicals cause similar genome-wide expression changes in a model cell line, they have the potential to induce similar human responses. This conjecture was the basis of a recent drug repurposing approach, which identified the previously unknown potential of existing drugs to treat some rare diseases of unmet medical needs.[13,14] Therefore, we reasoned that it might also be valid for chemical toxicity in humans because both desirable therapeutic effects and undesirable toxicities are chemical-

Received: August 24, 2016
Published: October 21, 2016

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893

induced effects in humans. To test the hypothesis, we created chemical-induced gene expression signatures for a large number of compounds and investigated the correlations between the signatures and specific types of human toxicity.

2. MATERIALS AND METHODS

2.1. Source of Chemical-Induced Gene Expression Data. A major challenge in this study was obtaining consistently generated high-quality data sets that encompass chemical-induced gene expression data for a large number of compounds. This remains prohibitively expensive despite the rapid advances in genomic technologies that have greatly reduced the cost of transcriptome studies. To overcome this issue, the Broad Institute of MIT developed a new technology that monitors expression levels of ~1,000 selected landmark genes from the human genome.[15] Expression levels of the remaining genes in the human genome can be inferred from these landmark genes. Such an approach makes it economically feasible to acquire data for tens of thousands of compounds with multiple replicates. As part of a U.S. National Institutes of Health Library of Integrated Network-based Cellular Signatures (LINCS) program (http://www.lincsproject.org/), the Broad Institute has used this technology to generate gene expression data for multiple human cell lines subjected to treatment by thousands of chemicals.

In the present study, we used data downloaded from the LINCS Cloud server (http://www.lincscloud.org/) in December 2014. The bulk of the LINCS transcriptome data was acquired from six cancer cell lines: A375 (human malignant melanoma), A549 (human lung adenocarcinoma), MCF7 (breast cancer), PC3 (prostate cancer), VCaP (prostate cancer cells derived from metastatic site), and HT29 (human colorectal adenocarcinoma). The number of data sets for these cell lines ranges from 10,913 for HT29 to 15,852 for VCaP, where each data set corresponds to a chemical treatment. We focused on the VCaP data sets because these contained the largest number of treatments and replicates.

To examine the feasibility of using chemical-induced gene expression in VCaP cells to predict chemical-induced organ injuries in humans, we first created a string consisting of 0s and 1s to represent the changes in gene expression caused by a chemical treatment. The response of each landmark gene is represented by the combination values at two adjacent positions in the string, one of which represents up-regulation of the gene and the other its down-regulation, as caused by a chemical treatment. If a chemical does not significantly alter the expression level of a landmark gene, a 0 is assigned to both positions. On the other hand, if a chemical causes significant up- or down-regulation of a gene, then a 1 is assigned to the corresponding position and a 0 to the other. Thus, we used a string of 0s and 1s to represent the gene expression signature associated with a chemical treatment. We refer to this string as the chemical genomic signature of the cell line.

2.2. Source of Data on Chemical-Induced Organ Injuries in Humans. To examine the feasibility of predicting specific organ injuries caused by exposure to a chemical, we need a large data set of chemical-induced organ injuries in humans. Although a wealth of detailed histopathological data from animal studies of chemical exposure exists,[16,17] comparable data on chemical-induced organ injuries in humans are scarce and usually associated with many confounding factors that make it challenging to unambiguously attribute an organ injury to a specific chemical.

To circumvent this limitation, we decided to use drugs that have been marketed for an extended period and positively identified as inducing organ injuries in humans. We searched databases of adverse drug reactions and identified 35 drugs that induce cholestasis (liver injury) from the Sider[18] and Offside databases;[19] 106 drugs reported to cause interstitial nephritis of the kidney (kidney injury);[20] and 65 drugs with a moderate to high risk of inducing long QT syndrome (heart injury) that may lead to cardiac arrest.[21−23] However, not all of these drugs were part of the chemical set tested in all cell lines of the LINCS program; hence, we selected only those drugs tested in VCaP cells as positive samples for causing liver, kidney, and heart injuries.

Training a prediction model for chemical toxicity requires both toxic compounds as positive samples and nontoxic compounds as negative samples. An even more daunting task is collecting a large set of nontoxic compounds. A drug can be classified as causing an organ injury based on case reports of treated patients developing a liver, kidney, or heart condition, but the absence of case reports does not allow us to positively conclude that it does not cause injury. Hence, in a strict sense, there are no true negative samples. To overcome this issue, we decided to compare and contrast drugs that cause a specific organ injury (positives) with all other compounds that possess VCaP-derived signatures. We denote the latter as *baseline* compounds to acknowledge that they may include untested or unrecognized positive samples. Thus, we trained our prediction models on a class of positive compounds and a class of baseline compounds. We hypothesized that the number of baseline compounds not recognized as being toxic to a specific human organ should be a small fraction of the total number (~15,000) of baseline compounds. If so, then to the extent that our prediction model is based on the genomic features of a group of positive compounds distinct from those of a group of baseline compounds, the impact of any misclassified compounds should be minimized. As described below, this is precisely what is achieved by our statistical method for correlating chemical signatures with specific organ injuries in humans.

2.3. Statistical Method for Models to Predict Chemical-Induced Organ Injuries in Humans. We used the Laplacian-corrected Bayesian method of Xia et al.[24] to associate chemical signatures with organ damage. Figure 1 illustrates a collection of

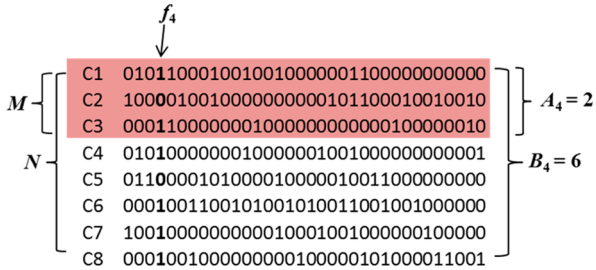


Figure 1. Schematic illustration of the chemical genomic signatures of N compounds, of which M cause an organ injury in humans. We used two positions to encode the chemical-induced up- or down-regulation of each landmark gene. If a chemical did not significantly change the expression level of a gene, a 0 was assigned to both positions. If a chemical significantly up- or down-regulated the expression level of a gene, a 1 was assigned to one of the corresponding positions of the chemical signature. In this illustration, $f_4 = 1$ encodes the down-regulation of gene 2; A_4 denotes the count of toxic compounds with $f_4 = 1$; and B_4 is the count of all compounds with $f_4 = 1$.

chemical signatures, where we designate M as the number of positive compounds that cause a specific organ injury out of a total of N chemicals. The number of compounds with 1 at the i-th position of their signatures ($f_i = 1$) in the positive class is denoted by A_i, and that with $f_i = 1$ among all compounds is denoted by B_i. To predict an organ injury from a chemical signature, we need to estimate the conditional probability that a chemical is positive given $f_i = 1$, i.e., $p(+|f_i = 1)$. Bayes theorem gives

$$p(+|f_i = 1) = \frac{P(+) \times p(f_i = 1|+)}{p(f_i = 1)} = \frac{\frac{M}{N} \times \frac{A_i}{M}}{\frac{B_i}{N}} = \frac{A_i}{B_i} \quad (1)$$

where $P(+)$ is the prior probability that a chemical is positive, $p(f_i = 1|+)$ is the conditional probability that $f_i = 1$ for the positive class, and $p(f_i = 1)$ is the probability that $f_i = 1$ for all compounds. In most cases, eq 1 provides a reasonable estimate of $p(+|f_i = 1)$. However, under conditions of severe under-sampling, such as when $B_i = A_i = 1$, then $p(+|f_i = 1) = 1.0$; that is, the method provides an overly optimistic probability estimate. To correct for the effect of under-sampling, we

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893

add K virtual compounds with $f_i = 1$ in their signatures. A reasonable estimate of the number of positive virtual drugs is then given by $K \times P(+)$. Thus, the corrected probability is as follows:

$$p_c(+|f_i = 1) = \frac{A_i + [K \times P(+)]}{B_i + K} \quad (2)$$

This correction ensures that as $B_i \to 0$ and $A_i \to 0$, $p(+|f_i = 1)$ approaches the prior probability $P(+)$. When K is set to $1/P(+)$, the adjustment corresponds to the Laplacian correction.[25]

Following Xia et al.,[24] we defined weight w_i to represent the contribution of the i-th feature in a chemical signature to an organ injury, as follows:

$$w_i = \log\left(\frac{p_c(+|f_i = 1)}{P(+)}\right) \quad (3)$$

Once the w_i values of all features in the chemical signatures have been determined from a training set consisting of positive and baseline compounds, the potential of a new chemical to cause an organ injury in humans can be estimated by a toxicity score defined as

$$\text{score} = \sum_{i=1}^{n} w_i \times f_i \quad (4)$$

where f_i equals 1 or 0, depending on whether a landmark gene is either up-/down-regulated or shows no significant change, respectively, upon exposure to the chemical. The higher the score of a compound, the higher is its potential to cause injury.

3. RESULTS AND DISCUSSION

3.1. Creating Chemical Signatures. *3.1.1. Overview of LINCS VCaP Data.* The VCaP data downloaded from the LINCS cloud server in December 2014 contained 120,358 data sets generated from 15,820 chemical treatments of VCaP cells at concentrations between 0 and 178 μM. There were three treatment durations: 6, 24, and 48 h. The number of data sets for treatment durations of 6 and 24 h were 55,986 and 64,154, respectively. Because the number of available data sets for 48-h treatment was small (218 data sets), we excluded them from this study.

Figure 2 shows the distribution of the number of data sets for the compounds tested. Most compounds were associated with five or six data sets, typically derived from testing three replicates for two treatment durations (6 and 24 h) for a total of six data sets, with the occasional quality control failure resulting in one fewer data set. For a small fraction of the compounds, the number of data sets was extremely high because the compounds were used as in-plate controls and repeatedly tested together with other samples.

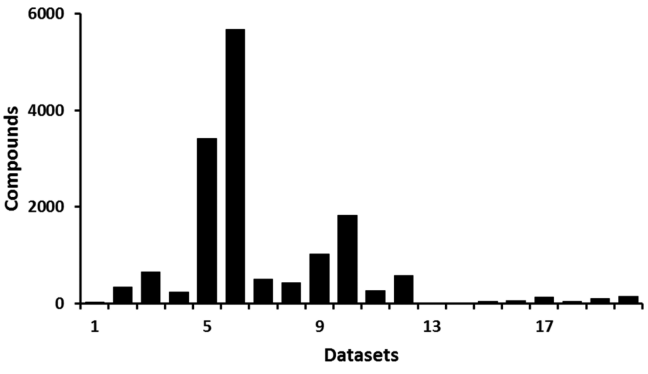


Figure 2. Histogram of the number of treatments applied to VCaP cells for each chemical (of 15,280 treatments extracted from LINCS VCaP data) showing that most compounds were tested five or six times.

Figure 3 shows a histogram of the chemical concentrations used. Most compounds were tested at around 5 and 10 μM, and only a small number were tested over a wider concentration range.

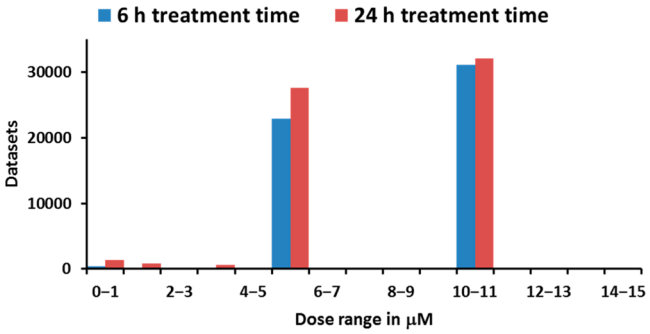


Figure 3. Histogram of data sets extracted from LINCS VCaP data showing that most compounds were tested at concentrations of 5−6 and 10−12 μM.

3.1.2. Importance of Replicates. The importance of replicates to ensure the statistical reliability of a biological discovery is well recognized.[26] However, because genomic technologies are still expensive, most studies using microarray or next-generation sequencing techniques only employ a limited number of replicates. For instance, in the LINCS program, most gene expression data were acquired with three replicates for each condition (chemical dose and treatment duration). In contrast, a large number of replicates are available for selected compounds used as in-plate controls. These data sets offer an opportunity to examine the impact of replicates on the identification of activated genes. For this purpose, we used compounds with more than 100 replicates for any treatment condition. We started with the z-values of the landmark genes provided in the LINCS data, which were calculated by comparing the expression level of a gene in response to a chemical treatment with expression levels of the same gene in response to all other chemical treatments. To create a chemical signature for a compound, we used the median z-value of a gene in all replicates as the z-value of the gene. We then used a z-value cutoff of ±2.0 to classify whether a gene was up- or down-regulated by a chemical treatment. That is, if the median z-value was 2.0 or higher, we considered the gene as up-regulated, and the corresponding position in the chemical signature of the compound was assigned a value of 1. Likewise, if the median z-value was −2.0 or lower, the gene was considered down-regulated, and the corresponding position of the chemical signature was assigned a value of 1. If a gene had a median z-value between −2.0 and 2.0, its expression level was considered unperturbed by the chemical treatment; therefore, the corresponding positions in the chemical signature were assigned a value of 0. We created chemical signatures for all compounds with 100 or more replicates, using all of the replicates. We considered these signatures to be statistically more robust than those generated from fewer replicates.

To examine the impact of the number of replicates in a study, we regenerated a series of chemical signatures by using randomly selected replicates whose numbers ranged from 3, 6, 9, etc., up to 60. For each chosen number of replicates, we regenerated the signature 50 times with randomly selected data

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893

sets. The results showed that for most chemicals, signatures generated from a small number of replicates have more activated genes and that the number of activated genes is highly variable. By increasing the number of replicates used, the average number of activated genes decreases and approaches the number of activated genes generated from all replicates.

Figure 4 shows the number of activated genes for wortmannin (BRD-A75409952) and geldanamycin (BRD-A19500257) as representative examples. Wortmannin and geldanamycin were associated with 164 and 160 replicates, respectively, under the treatment condition of 10 μM for 6 h. The numbers of activated genes in signatures generated using all replicates were 12 and 25 for wortmannin and geldanamycin, respectively. However, when using three replicates, the corresponding numbers on average increased to 45 and 103, with a high degree of variability. The value of repeated measurements is clear, given that the number of activated genes and its variability decreased as the number of replicates included in the analysis increased. Even when using nine replicates, the average number of activated genes was still at least twice that derived from all replicates. This indicates that even when using ~10 replicates, more than half of the discoveries could be false positives.

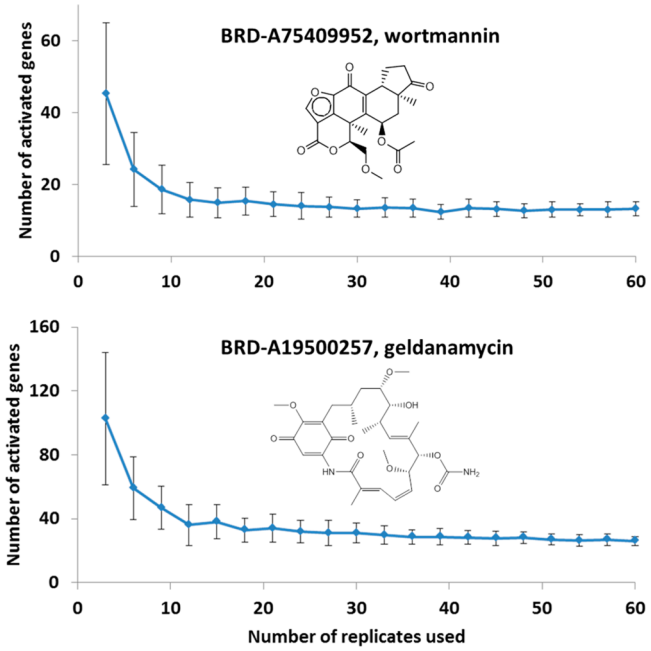


Figure 4. Number of activated genes as a function of replicates. Wortmannin and geldanamycin have 164 and 160 replicates, respectively, for VCaP cells treated at 10 μM for 6 h. The numbers of activated genes identified from all replicates were 12 and 25, respectively. The average number of activated genes identified from fewer replicates was higher and more variable, suggesting that data obtained with an insufficient number of replicates may contain a considerable number of false positives.

Figure 5 shows another measure that demonstrates the importance of the number of replicates. Here, we compared the similarity of signatures generated from a limited number of replicates to those generated from all replicates. Similarity was measured by the Tanimoto similarity, which is defined as

$$TS = \frac{a \cap b}{a \cup b} \quad (5)$$

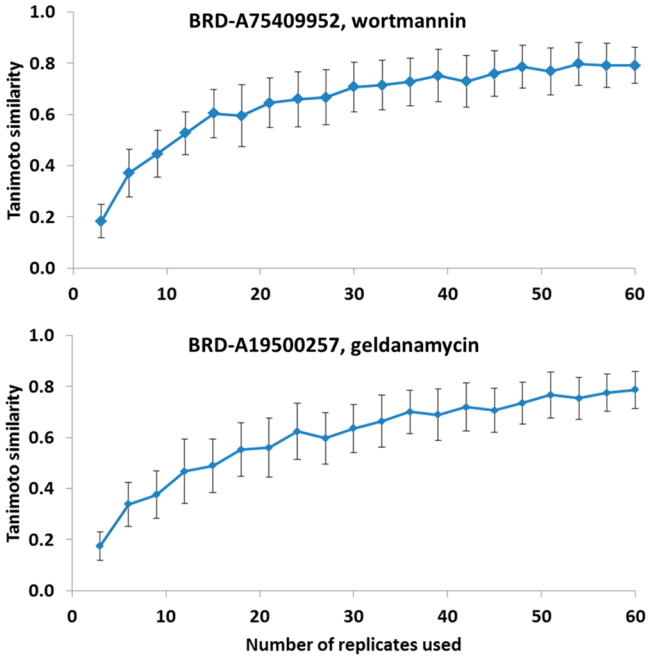


Figure 5. Gene signature similarity as a function of the number of replicates. Tanimoto similarity between chemical genomic signatures generated from a small number of replicates and those generated from all available replicates.

where $a \cap b$ denotes the number of commonly activated genes activated in signatures a and b, and $a \cup b$ denotes the total number of activated genes. The value of TS ranges from 0, which indicates no common genes activated in the two signatures, to 1, which indicates two identical signatures. Figure 5 shows that even with as many as 60 replicates, the generated signatures had a Tanimoto similarity of only ~0.8 to the signatures derived from all replicates. For both compounds, the Tanimoto similarity between signatures generated from nine replicates and those generated from all replicates was lower than 0.4. This is commensurate with a high degree of noise in high-throughput genomic data and associated risk of false discoveries when an insufficient number of biological replicates are used. However, using 10 or more replicates is prohibitively expensive for most studies that employ a large number compounds.

3.1.3. Creating Chemical Genomic Signatures. To partially alleviate the problem of insufficient replicates, we pooled the data from all treatments within the concentration range of 4–12 μM and treatment durations between 6 and 24 h. We used the median z-values of these data sets as the z-values of the genes to generate chemical signatures. When we set the z-value cutoff to ±2.0, the signatures of some chemicals were devoid of any activated genes (populated by 0s only), whereas those of other chemicals included over 500 (more than 50%) activated landmark genes. Because the chemicals were directly applied to the cells in the *in vitro* experiment, the null outcome of having no differentially expressed genes is biologically implausible. The approach of using a fixed z-value threshold may appear to introduce artifacts in determining the genes that respond significantly to chemical treatments. This issue can be addressed by means of ranking methods. For instance, a fixed number of genes with the highest z-values can be selected as the set of activated genes (e.g., considering the top 250 up- and down-regulated genes[27]).

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883–1893

Here, we adopted the top n-gene approach and explored the optimal n-value for signature generation. As a criterion to gauge whether parameter selection was optimal, we assessed the extent to which the selected top n signatures could be used to develop prediction models for compounds known to disrupt specific molecular pathways. We used the therapeutic target database (TTD)[28] to identify molecular targets with the highest number of ligands tested in the VCaP data. Table 1 shows the top 10 molecular targets, with the number of inhibitors for each target ranging from 10 to 32 drugs.

Table 1. Summary of Molecular Targets and Numbers of Their Antagonists[a] Used in This Study to Find Optimal n-Value for Creating Top n-Chemical Genomic Signatures by Using LINCS VCaP Data

molecular target	HUGO ID	number of antagonists
dopamine receptor	D2	24
histamine receptor	H1	29
type II topoisomerase	TOPO2	19
histone deacetylase 1	HDAC1	13
epidermal growth factor receptor	EGFR	17
vascular endothelial growth factor receptor	VEGFR	16
γ-aminobutyric acid receptor	GABAa	16
cytochrome *c* oxidase subunit 1 receptor	COX-1	20
cytochrome *c* oxidase subunit 2 receptor	COX-2	32
5-hydroxytryptamine receptor subtype 1a	5HT1a	10

[a]On the basis of data retrieved from the Therapeutic Target Database (TTD).

Using the LINCS VCaP data, we created chemical signatures for all compounds in Table 1 by considering the top 10, top 20, etc., up to the top 100 up- and down-regulated genes. We then trained Laplacian-corrected Bayesian models (see Materials and Methods) for each of the targets by using signatures derived from the top n-genes. We assessed the performance of the models in enriching ligands to the respective targets by using the areas under the receiver operating characteristic curves (AUCs) calculated from leave-one-out (LOO) cross-validation.[29] We evaluated the specificity of the models by their AUCs for enriching the ligands of other targets (off-target ligands). AUC values range from 0.5, which indicates a random model without any predictive power, to 1.0, which reflects a model that correctly separates all positives from other samples. Overall, a good model should have relatively high AUCs for ligands of its own target and low AUCs for off-target ligands.

Figure 6 shows the AUCs of the models for ligands of their targets (white cells on the diagonal) and AUCs for off-target predictions (colored off-diagonal cells). The cells in the tables are colored according to differences between the AUC values for predicting known targets and those for mis-predicting off-targets as true targets. The greater the number of green cells, the better is the prediction model and the higher the quality of the signatures used to build the models. Figure 6 shows that the number of green cells was lowest for models built on signatures consisting of the top 10 up- and down-regulated genes. By increasing the value of n, the models performed increasingly better, as indicated by the steady increase in the number of green cells. Performance was optimal around $n = 50$ and started to deteriorate around $n = 70$. Hence, we selected $n = 50$ as the threshold value to create signatures for all of the chemicals tested in VCaP cells.

3.2. Developing Models to Predict Organ Injury in Humans from Chemical Genomic Signatures. *3.2.1. Performance of Models as a Function of Positive Sample Size in Training Set.* As described above (see Materials and Methods), we identified 35 LiverTox, 106 KidneyTox, and 65 HeartTox drugs, although some of these were not tested in VCap cells. We created chemical genomic signatures based on unique LINCS compound IDs, resulting in 19 VCaP-based chemical genomic signatures of 15 unique LiverTox drugs, 83 signatures of 71 unique KidneyTox drugs, and 46 signatures of 38 unique HeartTox drugs. The drug names and their LINCS chemical IDs are given in Table S1 (see Supporting Information).

We first examined the similarity between genomic signatures of different samples of the same drug. For instance, both erythromycin and cyclosporin-a are associated with four distinct LINCS compound IDs, and samples identified by these IDs have been tested in the VCaP cell line. Table 2 shows the Tanimoto similarities, which range from 0.01 to 0.24 between the chemical genomic signatures of different samples of the same drugs. The low values were comparable to those expected between different compounds—an observation that highlights potential lot-effects or insufficient biological replicates. Because we could not disentangle these effects with the current data, we decided to treat signatures of the same drug as signatures of different compounds. We used the signatures of these drugs and the other chemicals to build and evaluate models of chemical-induced organ injuries in humans.

To examine the performance of these toxicity prediction models, we separated the baseline compounds randomly into two equal-sized baseline data sets, one in the training set and the other in the test set. We also randomly separated "positive" drugs known to cause organ injury into positive training and test sets. We varied the size of the positive test set from one positive compound to 10%, 20%, 30%, 40%, and 50% of the compounds being positive. This correspondingly determined the size of the positive training set as one less the number of positive compounds (i.e., LOO), and 90%, 80%, 70%, 60%, and 50% of the compounds being positive. We trained a Bayesian model on each training set and assessed model quality, as measured by the AUC, using the corresponding test set. To ensure that the results were statistically reliable, we repeatedly trained and evaluated the model 500 times for each training set composition. Each of the 500 models was built from a different, randomly selected subset of the training compounds matching the criteria of the composition, and the remaining compounds (the test set) were used to calculate the AUC. We then calculated the mean and standard deviation of the AUC.

Figure 7 shows the results for the LiverTox, HeartTox, and KidneyTox models. The best performance was achieved using the LOO training set (i.e., one positive compound left out of the model training and used as the positive sample for performance assessment). Reducing the size of the positive training set successively caused model performance to deteriorate. Because the size of the baseline set remained constant, the number of positive samples in the training set was a crucial factor in determining model quality. Hence, for the LiverTox model, which included the fewest compounds (19), the AUC was also the lowest among the toxicity models; for the KidneyTox model, which included the most compounds (83), the AUC was also the highest among the three toxicity models.

A major reason why increasing positive samples improves model performance is that the training data are highly unbalanced. That is, the ratio of samples in the two classes is

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893

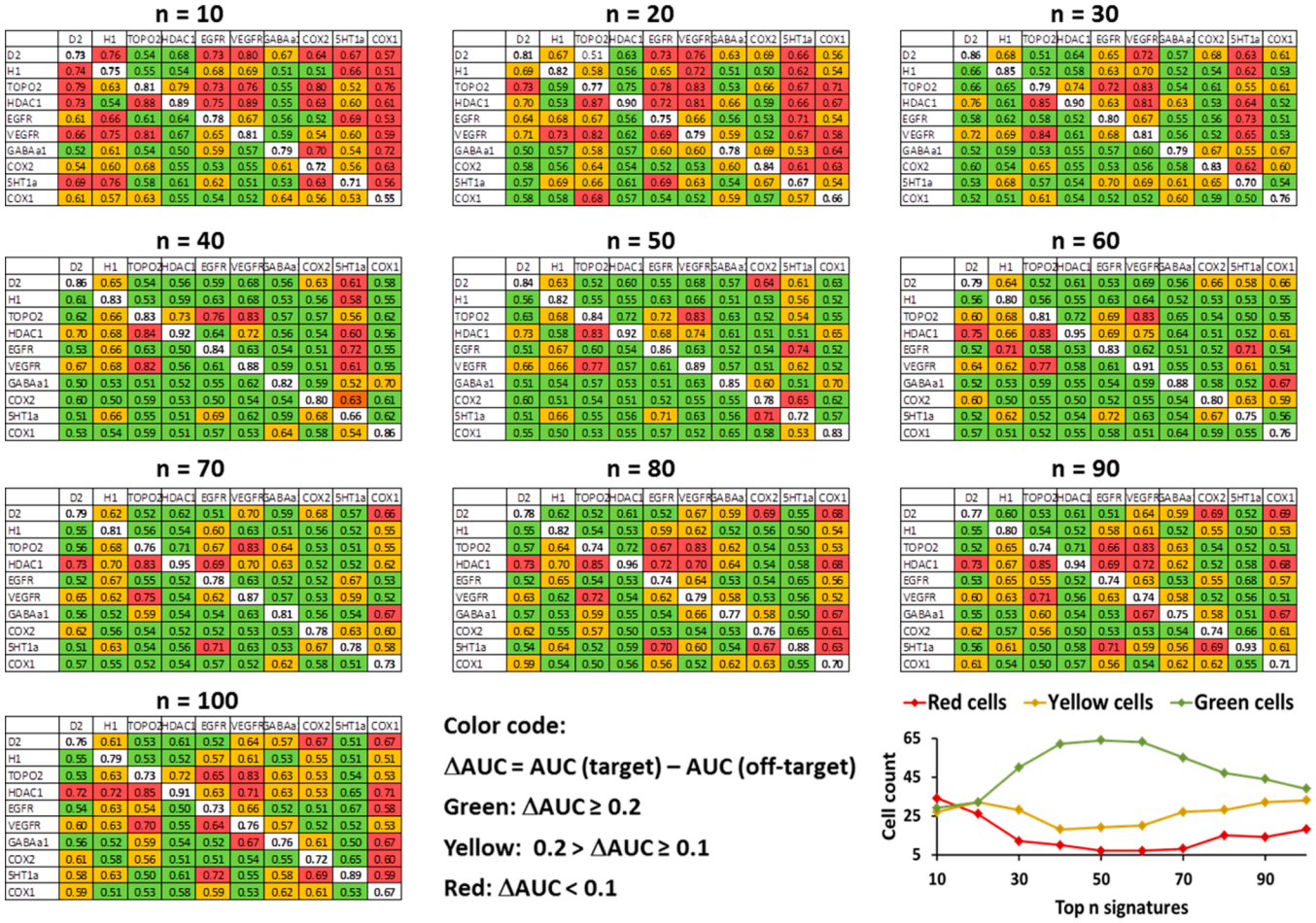


Figure 6. Influence of top n-signatures on model performance. We used target and off-target AUC values of 10 protein ligand models developed from chemical genomic signatures with different n-values to determine the optimal number of top up- and down-regulated genes to generate chemical genomic signatures.

Table 2. Tanimoto Similarity[a] between Chemical Genomic Signatures of Different Samples of the Same Compounds

erythromycin	BRD-A44827100	BRD-K39746403	BRD-K51677086	BRD-K63550407
BRD-A44827100	1	0.02	0.05	0.03
BRD-K39746403	0.02	1	0.01	0.02
BRD-K51677086	0.05	0.01	1	0.06
BRD-K63550407	0.03	0.02	0.06	1
cyclosporin-a	**BRD-A38030642**	**BRD-A64290322**	**BRD-A69815203**	**BRD-K13533483**
BRD-A38030642	1	0.22	0.23	0.23
BRD-A64290322	0.22	1	0.24	0.18
BRD-A69815203	0.23	0.24	1	0.24
BRD-K13533483	0.23	0.18	0.24	1

[a]Tanimoto similarity is 0 when two signatures share no activated genes and 1 when they are identical.

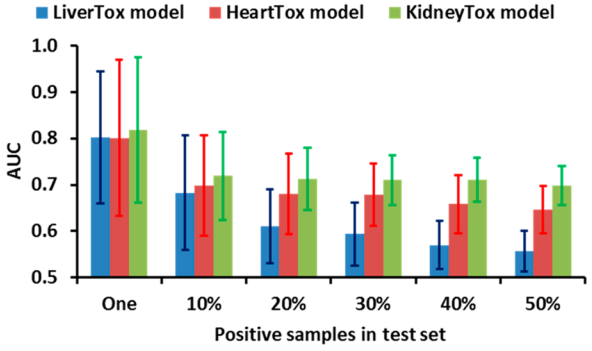


Figure 7. Average AUC values of 500 runs for human organ injury prediction models with randomly selected training and test set samples. Error bars represent ±1 standard deviation. Model performance deteriorates as the number of positive samples in the training set decreases.

far from 1:1. All machine-learning methods suffer from highly unbalanced data, and most fail on data sets with a few tens of samples in one class and thousands in the other. Some strategies, including stratified bagging, were developed to improve the performance on unbalanced data sets.[30] Interestingly, the results presented in Figure 7 show that even without using such strategies, the method we used here handles highly unbalanced data sets reasonably well. We will investigate in a future study whether such strategies improve the performance of our method and if so, to what extent.

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893

3.2.2. Impact of False Positives on Model Performance. The information on human organ toxicity came from adverse drug reaction reports in which the reporting threshold was relatively low. This suggests that the chance occurrence of false positives in the data set is non-negligible. If so, a practical model should to be able to handle false positives in the training set without significant reductions in performance.

To examine the impact of false positives on model performance, we retrained the models as described above but moved 10 compounds from the baseline training set to the positive training set and used them as putative false positives. This constitutes a high proportion of false positives relative to true positives in the data set because the total number of liver-toxic signatures is only 19. Figure 8 shows the performance of

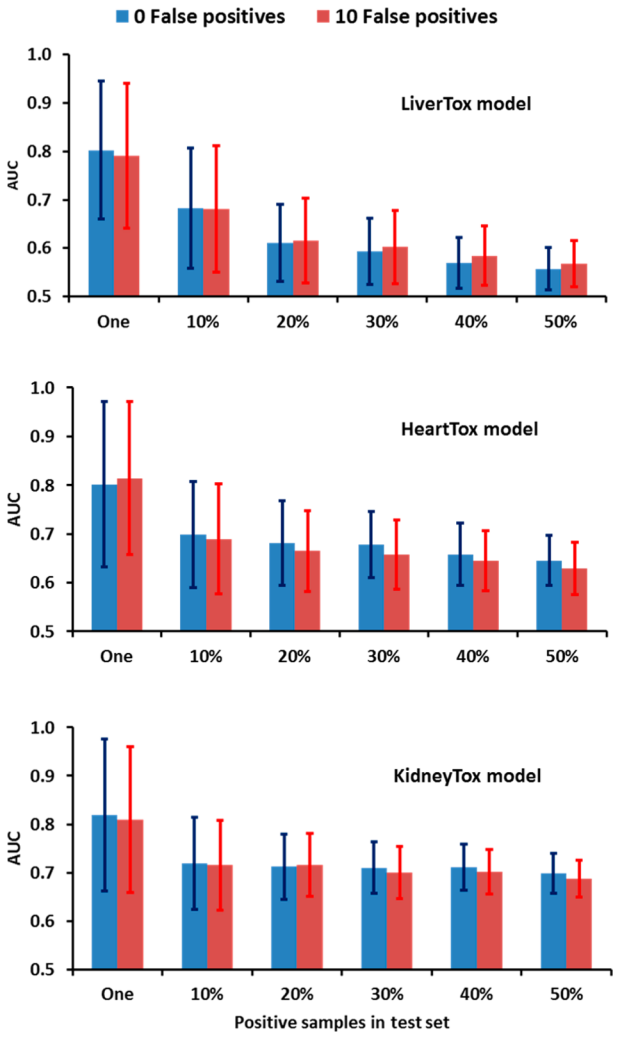


Figure 8. Performance of human organ injury prediction models trained with and without false positive samples. Bar heights indicate mean AUC values of 500 runs with randomly selected training and test set samples. Error bars represent ±1 standard deviation.

the organ-toxicity models trained with and without false positives, as assessed by the mean AUC values from 500 runs with randomly selected samples in the training and test sets. The impact of including false positives was negligible on model performance. The Bayesian model performed robustly because the chemical genomic signatures of the positive compounds shared a number of up- and down-regulated genes, whereas those of the false positives did not. Consequently, the false positives had little effect on the weights of the genes in eq 4.

3.2.3. Impact of False Negatives on Model Performance. The human organ toxicity of most compounds in the baseline data set is unknown. As such, some of these may cause human organ injuries. Models that predict toxicity, therefore, must be able to handle false negatives in the training set as well. Although many positives in the training set are needed to develop prediction models, the number of positive compounds with organ-specific toxicities is limited. This makes it challenging to develop a sound experimental design to examine the impact of false negatives on model performance. Such a design requires a reasonable number of positives used as false negatives, a reasonable number of positives used as test set positives, and a sufficient number of positives in the training set to draw statistically reliable conclusions.

To address these limitations, we first trained models that used a training set consisting of 70% positives. We then calculated the AUCs by using 50% of the left-out positive samples as test set positives. The mean AUCs of 500 runs with randomly selected positive and baseline compounds in each data set were considered as the AUCs of models without false negatives. We then trained another model with a training set consisting of 70% positives and moved 50% of the left-out positives to the baseline training set as false negatives. The AUCs were then calculated by using the remaining half of the left-out positives as the test set. As Figure 9 shows, false negatives did not impact model performance.

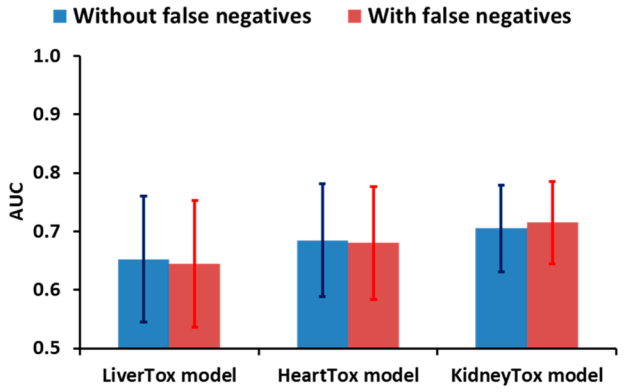


Figure 9. Performance of human organ injury prediction models trained with and without false negative samples. Bar heights indicate mean AUC values of 500 runs with randomly selected training and test set samples. Error bars represent ±1 standard deviation.

3.2.4. Chemical Genomic Signature-Based Models for Specific Organ Injury in Humans. To test whether the models predict general toxicity or specific organ toxicity, we constructed LiverTox, HeartTox, and KidneyTox models and calculated the AUC of each model by using LOO cross-validation. To examine the specificity of the LiverTox model, we calculated the AUC by using heart- and kidney-toxic compounds as positives. We repeated the same exercise, using the HeartTox model to calculate the AUC with liver- and kidney-toxic compounds and using the KidneyTox model to calculate the AUC with liver- and heart-toxic compounds. Figure 10 shows that the AUC of the LiverTox model was elevated for liver-toxic compounds and low for heart- and kidney-toxic compounds and that the AUCs of the HeartTox and KidneyTox models behaved similarly. The low off-target

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893

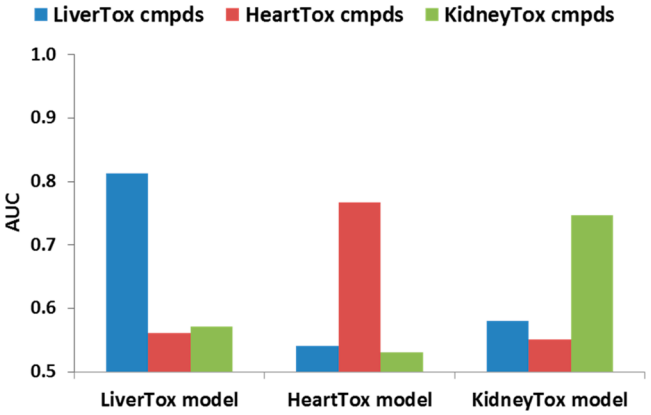


Figure 10. Model performance, as measured by AUC values, of human organ injury prediction models to identify compounds that induce specific human organ injuries.

AUCs indicate that the models predict organ-specific toxicities rather than a general toxic response.

3.2.5. Reports of High-Scoring Baseline Drugs Associated with Organ Injuries in Humans. To further assess the performance of the toxicity models, we examined baseline compounds with the highest scores (eq 4, Materials and Methods) to gauge whether the models could provide prospective predictions. We developed our final models by using all drugs known to cause organ injury as the positive training samples and all other compounds as the baseline training samples. We used the models to calculate scores for all compounds and plotted their distribution. Figure 11 shows that

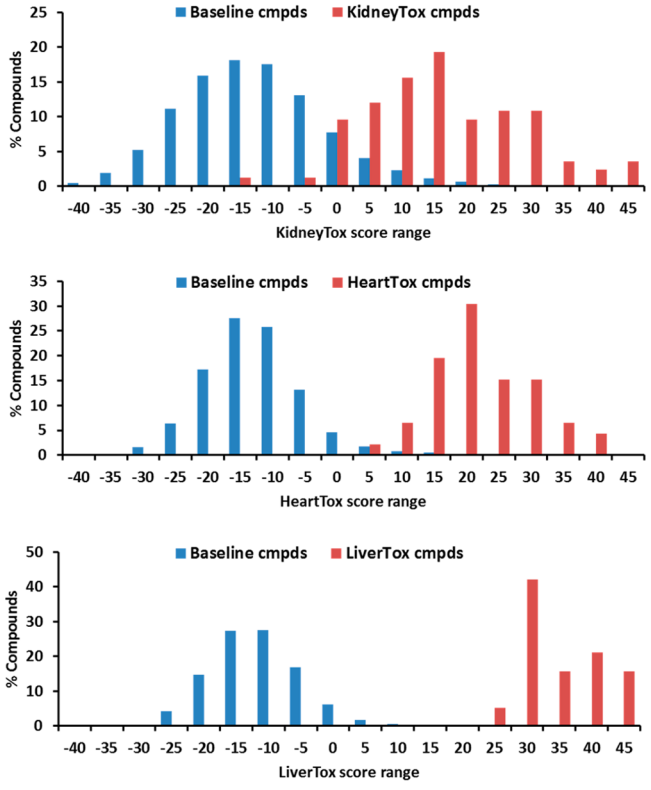


Figure 11. Histograms of toxicity scores for known toxic and baseline compounds in the three organ injury prediction models developed in this study.

the score distribution of organ-toxic compounds was different from that of the baseline compounds, with high scores indicative of potential toxicity. The normalized distributions overlapped to some degree for all three models; the largest overlap occurred for the KidneyTox model. To some extent, this normalization conceals the presence of baseline compounds that score within the range of most toxic compounds.

To evaluate baseline compounds with high scores, we sorted them by their scores. We excluded compounds without brand or generic names because information on organ toxicity in humans is unavailable for most such compounds. The remaining compounds were drugs or dietary supplements. We hypothesized that if a drug or supplement has the potential to cause organ injury in humans and has been on the market for a sufficient time, then case reports and investigations that associate the compound with organ toxicity could exist. Therefore, we searched the literature by using the names of compounds and organ injuries as keywords.

Tables 3–5 show the highest-scoring drugs or supplements of the baseline set along with relevant literature citations. Table 3 shows that the baseline drugs and dietary supplements with the highest interstitial nephritis (KidneyTox) scores were huperzine-A, sibutramine, retinol, xaliproden, tienilic acid, and diltiazem. The score for huperzine-A (38.1) was higher than the scores of 80% of the positives in the training set of the interstitial nephritis model (Figure 11, top panel). Huperzine-A is a naturally occurring sesquiterpene alkaloid found in the extracts of the firmoss *Huperzia serrate*.[31] It is available in the U.S. as a nutraceutical and is currently being investigated as a possible treatment for diseases characterized by neurodegeneration, including Alzheimer's disease. We found no link between huperzine-A and interstitial nephritis. Reports of its adverse effects may be limited because this nutraceutical, which is not a drug approved by the U.S. Food and Drug Administration (ADA), is infrequently prescribed by physicians.

Table 3. Highest Scoring Baseline Drugs or Dietary Supplements in a Model of Chemical-Induced Interstitial Nephritis (KidneyTox) with Relevant Literature Reports of Links between the Compounds and Interstitial Nephritis

name	LINCS ID	score	relevant literature
huperzine-a	BRD-K45888792	38.1	N/A
sibutramine	BRD-A23359898	36.7	32
retinol	BRD-K22429181	32.5	N/A
xaliproden	BRD-K88358234	32.1	N/A
tienilic acid	BRD-K34098590	31.2	39
diltiazem	BRD-A69636825	30.7	40

The compound with the next highest interstitial nephritis score was sibutramine, a drug used to treat obesity. We identified a case report of chronic interstitial nephritis associated with the Chinese herbal supplement Zi Xiu Tang Bee Pollen.[32] The report describes at least one other case of bee pollen associated with acute interstitial nephritis, and a 2010 FDA study found hidden drug ingredients, in particular sibutramines, in this herbal supplement. The case reports, coupled with a particularly high interstitial nephritis score, support an association between interstitial nephritis and sibutramine.

Our literature search revealed no association between interstitial nephritis and retinol or xaliproden. Retinol, also known as vitamin A, is an important vitamin whose deficiency

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883–1893

Table 4. Highest Scoring Baseline Drugs in a Model of Chemical-Induced Long QT Syndrome (HeartTox) with Relevant Literature Reports of Links between the Compounds and Long QT Syndrome

name	LINCS ID	score	relevant literature
sertraline	BRD-K82036761	38.1	N/A
promazine	BRD-K06980535	31.2	http://factmed.com/report-PROMAZINE%20HYDROCHLORIDE-causing-LONG%20QT%20SYNDROME.php
trifluoperzaine	BRD-K89732114	28.8	42
perphenazine	BRD-K10995081	26.1	43
mibefradil	BRD-K09549677	25.4	44

Table 5. Highest Scoring Baseline Drugs in a Model of Chemical-Induced Human Liver Cholestasis (LiverTox) with Relevant Literature Reports of Links between the Compounds and Human Liver Cholestasis

name	LINCS ID	score	relevant literature
norethisterone	BRD-K92073408	33.2	46
flubendazole	BRD-K86003836	33.1	N/A
danazol	BRD-K48970916	31.4	50
levonorgestrel	BRD-K35189033	29.6	45
testosterone	BRD-A55393291	29.5	N/A

is a public health problem in preschool-age children of developing countries and a leading cause of preventable childhood blindness and many other health issues.[33,34] Vitamin A is also known for its acute toxicity; it is potentially teratogenic,[35] and diets high in vitamin A are associated with an increased risk of hip fracture in postmenopausal women.[36] For this reason, dietary supplements either contain low doses of vitamin A, or more commonly contain provitamin A (β-carotene), which is converted into vitamin A in human intestinal cells.[37] β-Carotene is a safe source because its intestinal conversion to vitamin A decreases as its oral dose increases.[38] This may limit the chance of renal injury due to exposure to high levels of vitamin A.

We also found reports of tienilic acid and diltiazem associated with interstitial nephritis. In these cases, one patient taking tienilic acid developed acute renal failure due to acute allergic interstitial nephritis,[39] and another developed acute interstitial nephritis soon after the administration of diltiazem.[40]

Table 4 shows the baseline drugs with the highest HeartTox scores. Sertraline, the highest-scoring compound, does not cause long QT syndrome at approved doses and is an alternative to other antidepressants with long QT concerns. There are case reports of all other drugs in Table 4 linked to long QT syndrome. For instance, an analysis of data in the FDA adverse drug event reporting system, between January 2004 and October 2012, revealed that three individuals taking promazine hydrochloride reported long QT syndrome (http://factmed.com/report-PROMAZINE%20HYDROCHLORIDE-causing-LONG%20QT%20SYNDROME.php). A widened QRS interval and left ventricular systolic depression have also been reported after propafenone and promazine exposure.[41] Many reports suggest that trifluoperazine is associated with a risk of developing long QT syndrome.[42] Long QT syndrome has also been attributed to perphenazine[43] and mibefradil usage.[44]

For compounds with the potential to induce liver cholestasis (LiverTox), Table 5 shows that the highest scoring baseline drugs were norethisterone, flubendazole, danazol, levonorgestrel, and testosterone, with scores of 33.2, 33.1, 31.4, 29.6, and 29.5, respectively. Among these, the progestational hormones norethisterone and levonorgestrel are two of the main components in oral contraceptives. The incidence of liver cholestasis due to oral contraceptives is low in Western Europe (approximately 1 in 10,000) but as high as 1 in 4,000 in Chile and Scandinavia.[45] In addition, case reports have specifically associated norethisterone with liver cholestasis.[46,47] A study of 1,267 people who experienced side effects while taking levonorgestrel, based on FDA and social media reports, found 10 cases with intrahepatic cholestasis (http://www.ehealthme.com/ds/levonorgestrel/intrahepatic%20cholestasis/). The high liver cholestasis scores of these compounds suggest that progestational hormones may be responsible for women developing this condition while taking oral contraceptives.

Danazol and testosterone are anabolic steroids; danazol has been associated with liver cholestasis in multiple case reports.[48−50] Many reports link structurally modified testosterone with multiple human liver diseases (http://livertox.nlm.nih.gov/AndrogenicSteroids.htm) including liver cholestasis.[51] However, structurally unmodified testosterone is not orally bioavailable to humans. This may explain why we did not find case reports associating cholestasis with testosterone itself. The only top-scoring drug for which our search found no association to liver cholestasis was flubendazole, an anthelmintic used by veterinarians for protection against parasites and worms in dogs and cats.

Although a case report of a drug associated with an organ injury should not be considered as proof of a causal relationship, the high percentage of high-scoring drugs linked to reports of relevant organ injuries is consistent with our models to predict chemical-induced organ injuries.

SUMMARY

In this study, we evaluated the feasibility of predicting chemical-induced human organ injuries from chemical-induced genome-wide expression changes in cultured human cells. We used a large amount of publicly available transcriptional data derived from chemical treatment of a prostate cancer cell line to create gene expression signatures for nearly 15,000 compounds. We then used these signatures to develop models that predict human organ injury. The results of detailed cross-validation calculations indicated that the models were robust with respect to false positives and false negatives in the training data and could predict organ-specific injuries. We also performed a literature search of drugs that were not used as toxic compounds in training the models but predicted by the models to have injury-inducing potential. The search identified multiple case reports that linked these drugs with the relevant predicted human injuries. Thus, our study demonstrated the feasibility of using *in vitro* cell-based experiments to predict the potential of a chemical to cause human organ injuries.

This study relied on validating our predictions with independently generated *in vivo* toxicity data collected from many studies. The availability of human toxicity data is a crucial requirement for developing robust toxicity prediction models

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893

that can bridge the divide between *in vitro* or animal data and human data. Thus, we strongly recommend that such data be made readily available for general use by the community.

We also showed that our approach must overcome some practical challenges. The first is the requirement of a large number of biological replicates to generate reliable chemical genomic signatures. Given the cost of genomic sequencing techniques, acquiring transcriptome data for a large number of samples remains prohibitively expensive. Consequently, most published studies use only a few biological replicates for each sample. Our analysis indicates that with insufficient replicates, many genomic studies could be contaminated with high false discovery rates.

The second is the issue of how to estimate the degree of chemical exposure, which we did not address here. Bioavailability is not an issue in the LINCS data, which were generated by directly applying chemicals to human cancer cells. However, the human response to chemical exposure depends on both the pharmacokinetic profile of the chemical and its ability to activate molecular toxicity pathways at specific sites (organs or tissues) in the body. If a chemical is not easily absorbed by the human body, is readily metabolized or excreted once absorbed, or both, its likelihood of causing human organ injury is low even if it potently activates molecular toxicity pathways *in vitro*. The approach developed here predicts only the intrinsic potential of a chemical to cause human organ injuries. On the other hand, because a multitude of factors determine the bioavailability of industrial chemicals to humans, we will need to develop models of chemical exposure and combine the modeling results with predictions based on chemical genomic signatures to reliably assess human responses to chemical exposure.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemrestox.6b00287.

List of drugs causing human organ injuries (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*(R.L.) E-mail: rliu@bhsai.org.

*(A.W.) E-mail: sven.a.wallqvist.civ@mail.mil.

Funding

The research was supported by the US Army Medical Research and Materiel Command (Ft. Detrick, MD) as part of the US Army's Network Science Initiative, and by the Defense Threat Reduction Agency (Grant No. CBCall14-CBS-05-2-0007).

Notes

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the US Army or of the US Department of Defense.

The authors declare no competing financial interest.

[†]X.Y.: deceased.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Dr. Tatsuya Oyama in drafting and editing the manuscript.

■ ABBREVIATIONS

LINCS, Library of Integrated Network-based Cellular Signatures; REACH, Registration, Evaluation, Authorization, and Restriction of CHemical Substances; Long QT syndrome, a congenital disorder characterized by a prolongation of the QT interval on electrocardiograms and a propensity to ventricular tachyarrhythmias, which may lead to syncope, cardiac arrest, or sudden death; AUC, area under the receiver operating characteristic curve

■ REFERENCES

(1) Grindon, C., and Combes, R. (2008) Introduction to the EU REACH legislation. *Altern. Lab. Anim 36* (Suppl 1), 1−6.

(2) Sun, H., Xia, M., Austin, C. P., and Huang, R. (2012) Paradigm shift in toxicity testing and modeling. *AAPS J. 14*, 473−480.

(3) Ferdowsian, H. R., and Beck, N. (2011) Ethical and scientific considerations regarding animal testing and research. *PLoS One 6*, e24059.

(4) Worth, A. P., Bassan, A., DeBruijn, J., Gallegos-Saliner, A., Netzeva, G., Patlewicz, G., Pavan, M., Tsakovska, I., and Eisenreich, S. (2007) The role of the European chemicals bureau in promoting the regulatory use of (Q)SAR methods. *SAR QSAR Environ. Res. 18*, 111−125.

(5) Kleinjans, J. (2014) *Toxicogenomics-Based Cellular Models: Alternatives to Animal Testing for Safety Assessment*, Elsevier/AP, Amsterdam, The Netherlands.

(6) Cheng, F., Theodorescu, D., Schulman, I. G., and Lee, J. K. (2011) In vitro transcriptomic prediction of hepatotoxicity for early drug discovery. *J. Theor. Biol. 290*, 27−36.

(7) Tawa, G. J., AbdulHameed, M. D., Yu, X., Kumar, K., Ippolito, D. L., Lewis, J. A., Stallings, J. D., and Wallqvist, A. (2014) Characterization of chemically induced liver injuries using gene co-expression modules. *PLoS One 9*, e107230.

(8) Ippolito, D. L., AbdulHameed, M. D., Tawa, G. J., Baer, C. E., Permenter, M. G., McDyre, B. C., Dennis, W. E., Boyle, M. H., Hobbs, C. A., Streicker, M. A., Snowden, B. S., Lewis, J. A., Wallqvist, A., and Stallings, J. D. (2016) Gene Expression Patterns Associated With Histopathology in Toxic Liver Fibrosis. *Toxicol. Sci. 149*, 67−88.

(9) Te, J. A., AbdulHameed, M. D., and Wallqvist, A. (2016) Systems toxicology of chemically induced liver and kidney injuries: histopathology-associated gene co-expression modules. *J. Appl. Toxicol. 36*, 1137−1149.

(10) Burczynski, M. E., McMillian, M., Ciervo, J., Li, L., Parker, J. B., Dunn, R. T., 2nd, Hicken, S., Farr, S., and Johnson, M. D. (2000) Toxicogenomics-based discrimination of toxic mechanism in HepG2 human hepatoma cells. *Toxicol. Sci. 58*, 399−415.

(11) Uehara, T., Minowa, Y., Morikawa, Y., Kondo, C., Maruyama, T., Kato, I., Nakatsu, N., Igarashi, Y., Ono, A., Hayashi, H., Mitsumori, K., Yamada, H., Ohno, Y., and Urushidani, T. (2011) Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. *Toxicol. Appl. Pharmacol. 255*, 297−306.

(12) Yamane, J., Aburatani, S., Imanishi, S., Akanuma, H., Nagano, R., Kato, T., Sone, H., Ohsako, S., and Fujibuchi, W. (2016) Prediction of developmental chemical toxicity based on gene networks of human embryonic stem cells. *Nucleic Acids Res. 44*, 5515−5528.

(13) Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., Sage, J., and Butte, A. J. (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med. 3*, 96ra77.

(14) Dudley, J. T., Sirota, M., Shenoy, M., Pai, R. K., Roedder, S., Chiang, A. P., Morgan, A. A., Sarwal, M. M., Pasricha, P. J., and Butte, A. J. (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med. 3*, 96ra76.

(15) Duan, Q., Flynn, C., Niepel, M., Hafner, M., Muhlich, J. L., Fernandez, N. F., Rouillard, A. D., Tan, C. M., Chen, E. Y., Golub, T. R., Sorger, P. K., Subramanian, A., and Ma'ayan, A. (2014) LINCS

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893

Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res. 42*, W449−460.
(16) Ganter, B., Tugendreich, S., Pearson, C. I., Ayanoglu, E., Baumhueter, S., Bostian, K. A., Brady, L., Browne, L. J., Calvin, J. T., Day, G. J., Breckenridge, N., Dunlea, S., Eynon, B. P., Furness, L. M., Ferng, J., Fielden, M. R., Fujimoto, S. Y., Gong, L., Hu, C., Idury, R., Judo, M. S., Kolaja, K. L., Lee, M. D., McSorley, C., Minor, J. M., Nair, R. V., Natsoulis, G., Nguyen, P., Nicholson, S. M., Pham, H., Roter, A. H., Sun, D., Tan, S., Thode, S., Tolley, A. M., Vladimirova, A., Yang, J., Zhou, Z., and Jarnagin, K. (2005) Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol. 119*, 219−244.
(17) Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., and Yamada, H. (2015) Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res. 43*, D921−927.
(18) Kuhn, M., Letunic, I., Jensen, L. J., and Bork, P. (2016) The SIDER database of drugs and side effects. *Nucleic Acids Res. 44*, D1075−1079.
(19) Tatonetti, N. P., Ye, P. P., Daneshjou, R., and Altman, R. B. (2012) Data-driven prediction of drug effects and interactions. *Sci. Transl. Med. 4*, 125ra31.
(20) Naughton, C. A. (2008) Drug-induced nephrotoxicity. *Am.Fam. Physician 78*, 743−750.
(21) Barnes, B. J., and Hollands, J. M. (2010) Drug-induced arrhythmias. *Crit. Care Med. 38*, S188−197.
(22) Behr, E. R., and Roden, D. (2013) Drug-induced arrhythmia: pharmacogenomic prescribing? *Eur. Heart J. 34*, 89−95.
(23) Viskin, S., Justo, D., Halkin, A., and Zeltser, D. (2003) Long QT syndrome caused by noncardiac drugs. *Prog. Cardiovasc. Dis. 45*, 415−427.
(24) Xia, X., Maliski, E. G., Gallant, P., and Rogers, D. (2004) Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem. 47*, 4463−4470.
(25) (2010) Laplace Estimate, in *Encyclopedia of Machine Learning* (Sammut, C., and Webb, G. I., Eds.) p 571, Springer, New York.
(26) Lee, M. L., Kuo, F. C., Whitmore, G. A., and Sklar, J. (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl. Acad. Sci. U. S. A. 97*, 9834−9839.
(27) Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., Murino, L., Tagliaferri, R., Brunetti-Pierri, N., Isacchi, A., and di Bernardo, D. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U. S. A. 107*, 14621−14626.
(28) Zhu, F., Han, B., Kumar, P., Liu, X., Ma, X., Wei, X., Huang, L., Guo, Y., Han, L., Zheng, C., and Chen, Y. (2010) Update of TTD: Therapeutic Target Database. *Nucleic Acids Res. 38*, D787−791.
(29) Geisser, S. (1993) *Predictive Inference: An Introduction*, Chapman & Hall, New York.
(30) Kotsiantis, S., Kanellopoulos, D., and Pintelas, P. (2006) Handling imbalanced datasets: A review. *GESTS Int. Trans. Comput. Sci. Eng. 30*, 25−36.
(31) Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res. 34*, D668−672.
(32) Joshi, H. J., and Obi, R. (2014) A case report of chronic interstitial nephritis associated with Chinese herbal supplement Zi Xiu Tang Bee Pollen. *CRIM 1*, 65−70.
(33) Maida, J. M., Mathers, K., and Alley, C. L. (2008) Pediatric ophthalmology in the developing world. *Current opinion in ophthalmology 19*, 403−408.
(34) Beaton, G. H. (1993) *Effectiveness of Vitamin A Supplementation in the Control of Young Child Morbidity and Mortality in Developing Countries*, International Nutrition Program, Dept. of Nutritional Sciences, Faculty of Medicine, Technical Secretary, ACC/SCN, Toronto, Canada.
(35) Teratology Society. (1987) Teratology Society Position Paper: Recommendations for Vitamin A Use during Pregnancy, *Teratology, 35*, 269−275.
(36) Feskanich, D., Singh, V., Willett, W. C., and Colditz, G. A. (2002) Vitamin A intake and hip fractures among postmenopausal women. *JAMA 287*, 47−54.
(37) Yeum, K. J., and Russell, R. M. (2002) Carotenoid bioavailability and bioconversion. *Annu. Rev. Nutr. 22*, 483−504.
(38) Novotny, J. A., Harrison, D. J., Pawlosky, R., Flanagan, V. P., Harrison, E. H., and Kurilich, A. C. (2010) Beta-carotene conversion to vitamin A decreases as the dietary dose increases in humans. *J. Nutr. 140*, 915−918.
(39) Walker, R. G., Whitworth, J. A., and Kincaid-Smith, P. (1980) Acute interstitial nephritis in a patient taking tienilic acid. *British medical journal 280*, 1212.
(40) Abadin, J. A., Duran, J. A., and Perez de Leon, J. A. (1998) Probable diltiazem-induced acute interstitial nephritis. *Ann. Pharmacother. 32*, 656−658.
(41) La Rocca, R., Ferrari-Toninelli, G., and Patane, S. (2014) Widened QRS interval and left ventricular systolic depression after propafenone and promazine exposure. *Int. J. Cardiol. 177*, 57−60.
(42) Yap, Y. G., and Camm, A. J. (2003) Drug-induced QT prolongation and torsades de pointes. *Br. Heart J. 89*, 1363−1372.
(43) Stollberger, C., Huber, J. O., and Finsterer, J. (2005) Antipsychotic drugs and QT prolongation. *International clinical psychopharmacology 20*, 243−251.
(44) Glaser, S., Steinbach, M., Opitz, C., Wruck, U., and Kleber, F. X. (2001) Torsades de pointes caused by Mibefradil. *Eur. J. Heart Failure 3*, 627−630.
(45) Kreek, M. J. (1987) Female sex steroids and cholestasis. *Semin. Liver Dis. 7*, 8−23.
(46) Anand, V., and Gorard, D. A. (2005) Norethisterone-induced cholestasis. *QJM 98*, 232−234.
(47) Yorulmaz, E., Özçelik, S., Atay, A. E., Kızılgül, M., Peker, Ö., and Tuncer, I. (2014) Norethisteron-induced cholestasis: a case report. *Acıbadem Üniversitesi Sağlık Bilimleri Dergisi 5*, 171−173.
(48) Qaseem, T., Jafri, W., Khurshid, M., and Khan, H. (1992) Cholestatic jaundice associated with danazol therapy. *Postgrad. Med. J. 68*, 984−985.
(49) Boue, F., Coffin, B., and Delfraissy, J. F. (1986) Danazol and cholestatic hepatitis. *Ann. Intern. Med. 105*, 139−140.
(50) Rafailidis, P., Dourakis, S. P., Petraki, K., and Hadziyannis, S. J. (2003) Danazol-induced acute icteric cholestatic hepatitis. *Ann. Gastroenterol 16*, 80−83.
(51) Singh, C., Bishop, P., and Willson, R. (1996) Extreme hyperbilirubinemia associated with the use of anabolic steroids, health/nutritional supplements, and ethanol: response to ursodeoxycholic acid treatment. *Am. J. Gastroenterol. 91*, 783−785.

DOI: 10.1021/acs.chemrestox.6b00287
Chem. Res. Toxicol. 2016, 29, 1883−1893