Chemical Research in Toxicology

ARTICLE

pubs.acs.org/crt

Predicting Drug-Induced Hepatotoxicity Using QSAR and Toxicogenomics Approaches

Yen Low,[†,‡,▽] Takeki Uehara,[‡,§,▽] Yohsuke Minowa,[§] Hiroshi Yamada,[§] Yasuo Ohno,[∥] Tetsuro Urushidani,[§,⊥] Alexander Sedykh,[†] Eugene Muratov,[†,#] Viktor Kuz'min,[#] Denis Fourches,[†] Hao Zhu,[†] Ivan Rusyn,[*,‡] and Alexander Tropsha[*,†]

[†]Laboratory for Molecular Modeling, University of North Carolina, Chapel Hill, North Carolina 27599, United States

[‡]Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, North Carolina 27599, United States

[§]Toxicogenomics Informatics Project, National Institute of Biomedical Innovation, Asagi, Osaka, Japan

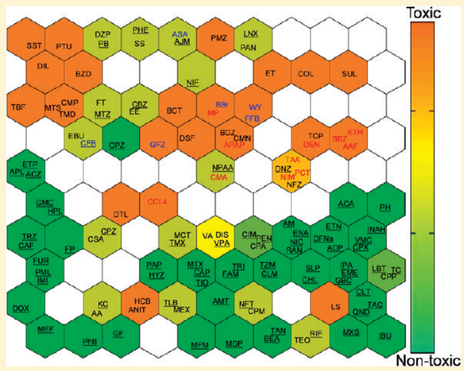
[∥]National Institute of Health Sciences, Kamiyoga, Tokyo, Japan

[⊥]Doshisha Women's College of Liberal Arts, Kodo, Kyoto, Japan

[#]A.V. Bogatsky Physical-Chemical Institute NAS of Ukraine, Odessa, Ukraine

S *Supporting Information*

ABSTRACT: Quantitative structure−activity relationship (QSAR) modeling and toxicogenomics are typically used independently as predictive tools in toxicology. In this study, we evaluated the power of several statistical models for predicting drug hepatotoxicity in rats using different descriptors of drug molecules, namely, their chemical descriptors and toxicogenomics profiles. The records were taken from the Toxicogenomics Project rat liver microarray database containing information on 127 drugs (http://toxico.nibio.go.jp/datalist.html). The model end point was hepatotoxicity in the rat following 28 days of continuous exposure, established by liver histopathology and serum chemistry. First, we developed multiple conventional QSAR classification models using a comprehensive set of chemical descriptors and several classification methods (*k* nearest neighbor, support vector machines, random forests, and distance weighted discrimination). With chemical descriptors alone, external predictivity (correct classification rate, CCR) from 5-fold external cross-validation was 61%. Next, the same classification methods were employed to build models using only toxicogenomics data (24 h after a single exposure) treated as biological descriptors. The optimized models used only 85 selected toxicogenomics descriptors and had CCR as high as 76%. Finally, hybrid models combining both chemical descriptors and transcripts were developed; their CCRs were between 68 and 77%. Although the accuracy of hybrid models did not exceed that of the models based on toxicogenomics data alone, the use of both chemical and biological descriptors enriched the interpretation of the models. In addition to finding 85 transcripts that were predictive and highly relevant to the mechanisms of drug-induced liver injury, chemical structural alerts for hepatotoxicity were identified. These results suggest that concurrent exploration of the chemical features and acute treatment-induced changes in transcript levels will both enrich the mechanistic understanding of subchronic liver injury and afford models capable of accurate prediction of hepatotoxicity from chemical structure and short-term assay results.



INTRODUCTION

Hepatotoxicity is a major factor contributing to the high attrition rate of drugs. At least a quarter of the drugs are prematurely terminated or withdrawn from the market due to liver-related liabilities.[1] As a result, modern drug development has evolved into a complex process relying on the iterative evaluation of multiple data sources to eliminate potentially harmful candidates as cheaply and as early as possible. In addition, high throughput, high content, and other data-rich experimental techniques, accompanied by the appropriate informatics tools, are rapidly incorporated into toxicity testing.

Quantitative structure−activity relationship (QSAR) modeling is widely used as a computational tool that allows one to relate the potential activity (e.g., toxicity) of an agent to its structural features represented by multiple chemical descriptors. As with any multivariate statistical modeling, rigorous validation procedures are necessary to guard against overfitting and overestimating model predictivity.[2] QSAR models have demonstrated good predictivity especially for specific end points such as solubility or binding affinity to a certain target. However, QSAR predictivity is generally poor in the case of a complex end point such as hepatotoxicity where the structure−activity relationship is less straightforward due to multiple mechanisms of action.[3]

Received: April 8, 2011
Published: June 23, 2011

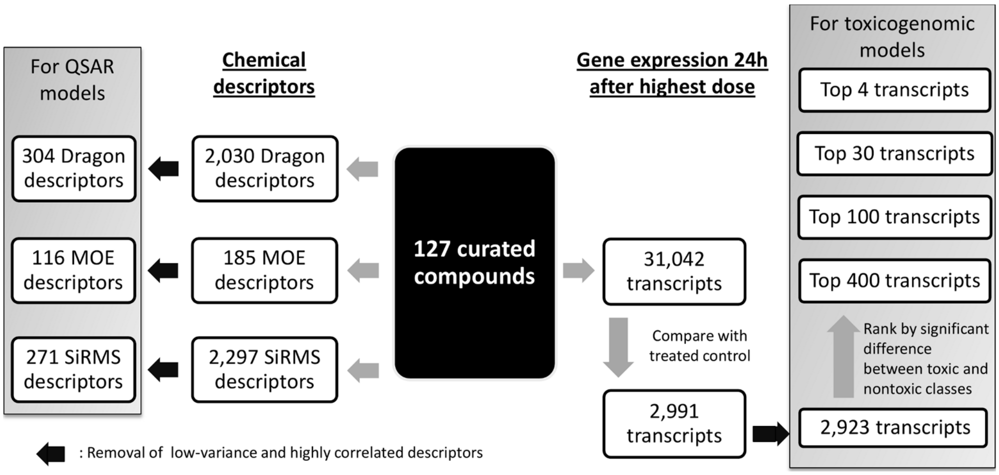


Figure 1. Workflow illustrating data curation and feature selection for modeling.

Toxicogenomics is now routinely used in drug and chemical safety evaluation, providing valuable mechanistic understanding of the molecular changes associated with the disease or treatment.[4] In addition, its utility for predicting toxicity has been explored. Blomme et al.[5] developed models using transcriptional changes after short-term (5 days) exposure to predict bile duct hyperplasia that otherwise required long-term *in vivo* experiments. Fielden et al.[6] developed a 37-gene classification model using microarray data following short-term (1−5 days) exposure to predict nongenotoxic hepatocarcinogenicity with over 80% accuracy. Zidek et al.[7] reported high accuracy with a 64-gene classifier for the prediction of acute hepatotoxicity. The Toxicogenomics Project in Japan, set up by the Ministry of Health, Labour and Welfare, National Institute of Health Sciences, and 15 pharmaceutical companies, has also identified several toxicogenomics signatures indicative of the various toxic modes of action such as phospholipidosis,[8] glutathione depletion,[9] bilirubin elevation,[10] nongenotoxic hepatocarcinogenesis,[11] and peroxisome proliferation.[12]

Most previous studies on statistical modeling of toxicity used either chemical descriptors (conventional QSAR) or toxicogenomics profiles independently for model development. However, in our recent studies, we have demonstrated the benefits of hybrid classification models of *in vivo* carcinogenicity[13] and toxicity,[14] and employing both chemical descriptors and biological assay data (treated as biological descriptors). In the first study of this type,[13] we used the results of high-throughput screening assays of environmental chemicals along with their chemical descriptors to arrive at improved models of rat carcinogenicity. This approach was extended to predicting acute toxicity half-maximal lethal dose in rats using dose−response *in vitro* data as quantitative biological descriptors.[14]

Following our hybrid (chemical and biological descriptors) data modeling paradigm, we sought to integrate QSAR and toxicogenomics data to develop classification models of hepatotoxicity using a data set of 127 drugs studied in the Japanese Toxicogenomics Project.[15] We built classifiers combining chemical descriptors and toxicogenomics data alongside the conventional QSAR, as well as toxicogenomics models. Our objective was to investigate if chemical descriptors and biological descriptors, such as gene expression, could be complementary. In addition, we sought to enhance the interpretation of the models in terms of elucidating the chemical structural features and biological mechanisms associated with hepatotoxicity. We show that statistically significant and externally predictive models can be developed by combining chemical and biological descriptors and can be used to predict hepatotoxicity and prioritize chemicals for toxicogenomics and other *in vivo* studies.

■ MATERIALS AND METHODS

Data. The chemical name, dosage, administration route, and vehicle for the 127 compounds used in this study are summarized in Table 1 of the Supporting Information. The detailed protocol for the animal study was described previously.[15] Briefly, 6-week old male Sprague−Dawley rats (Charles River Japan, Inc., Kanagawa, Japan) with five animals per group were used in the study. Animals were sacrificed 24 h after a single dose or 24 h after repeat daily treatment for 28 days. Blood samples were collected from the abdominal aorta under ether anesthesia. Serum chemical indicators included alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), total bilirubin (TBIL), direct bilirubin (DBIL), and gamma-glutamyl transpeptidase (GGT). The livers were quickly removed following exsanguination and sections of the livers were placed in 10% phosphate-buffered formalin for histopathology. Formalin-fixed liver tissue was embedded in paraffin, and sections were stained with hematoxylin and eosin and examined histopathologically under light microscopy. Remaining liver tissues from left lateral lobes were soaked in RNALater (Ambion Inc., Austin, TX) and stored at −80 °C until used for microarray analysis. Detailed methods for microarray analysis were previously reported.[15] Raw microarray data files with individual animal histopathological data are available (http://toxico.nibio.go.jp/datalist.html). In this study, toxicogenomics data obtained from rats treated with a single dose of a drug or vehicle for 24 h was used. The experimental protocols were reviewed and approved by the Ethics Review Committee for Animal Experimentation of the National Institute of Health Sciences (Tokyo, Japan).

Liver histopathology and serum chemistry in animals treated for 28 days were assessed for the determination of the hepatotoxicity end point for prediction. Histopathology was graded by two trained pathologists in a blinded manner as follows: no change, very slight (minimal), slight, moderate, and severe. Spontaneously observed lesions (e.g., minimal focal necrosis and microgranuloma) were not used for grading. The results of a histopathology analysis were considered positive if the grade recorded was other than "no change." Table 1 of the Supporting Information lists serum chemistry and histopathology classification for each compound. A compound was denoted *hepatotoxic* if it exhibited histopathology characteristics of hepatotoxicity (e.g., hepatocellular necrosis/degeneration, inflammatory cell infiltration, bile duct proliferation, etc.) regardless of the findings from serum chemistry. Conversely, a compound was deemed *nonhepatotoxic* if it did not result in adverse histopathological features. When the histopathological observations were inconclusive (e.g., hepatocellular hypertrophy, vacuolization, etc.), serum chemistry data was considered. Under these circumstances, significant changes (Dunnett's test) in at least one enzyme marker would render the compound *hepatotoxic*. Otherwise, the compounds with inconclusive histopathology and normal serum chemistry were denoted *nonhepatotoxic*. In total, there were 53 (42%) hepatotoxic and 74 (58%) nonhepatotoxic compounds.

Curation of Chemical Data. The data set was curated according to the procedures described by Fourches et al.[16] Briefly, counterions and duplicates were removed, and specific chemotypes such as aromatic and nitro groups were normalized using several cheminformatics software such as ChemAxon Standardizer (v.5.3, ChemAxon, Budapest, Hungary), HiT QSAR,[17] and ISIDA.[18] Following the automated curation, the data set was inspected manually, and two metal-containing compounds for which most chemical descriptors cannot be calculated, cisplatin and carboplatin, were removed. Chemical descriptors were calculated with Dragon (v.5.5, Talete SRL, Milan, Italy) and Molecular Operating Environment (MOE, v.2009.10, Chemical Computing Group, Montreal, Canada) software. Simplex representation of molecular structure (SiRMS) descriptors were derived as detailed elsewhere.[19] After range scaling (from 0 to 1), low variance (SD < 10^{-6}) and highly correlated descriptors (if pairwise $r^2 > 0.9$, one of the pair was randomly removed) were removed. QSAR models were built separately using 304 Dragon, or 116 MOE, or 271 SiRMS descriptors (Figure 1).

Selection of Transcripts. Transcripts were selected for modeling using various feature selection methods. Of the 31,042 transcripts measured, we removed those consistently absent across all compounds. Then we extracted 2,991 transcripts with sufficient variation across all the compounds on the basis of the following criteria: the largest change of any transcript over its untreated equivalent must exceed 1.5-fold, and the smallest false discovery rate (Welch *t*-test) must be less than 0.05. Next, transcripts with low variance (all, or all but one value is constant) and high correlation (if pairwise $r^2 > 0.9$, one of the pair, chosen randomly, was removed) were excluded leaving 2,923 transcript variables (Figure 1) which were range scaled.

Then, supervised selection methods were used to filter genes differentially expressed between hepatotoxic and nonhepatotoxic compounds. Significance analysis of microarrays (SAM),[20] a permutation variant of the *t*-test commonly used for transcript selection, was used. Top ranked transcripts were retained for modeling. Different sets of transcripts were selected for each modeling set used in 5-fold external cross-validation to avoid selection bias introduced by a supervised selection process.

Modeling and Validation. *K* nearest neighbors (*k*NN),[21] support vector machines (SVM),[22] random forest (RF),[23] and distance weighted discrimination (DWD)[24] machine learning techniques, designed to effectively handle high dimension-low sample size data, were used for modeling. The modeling workflow[2,25] used both internal and external validation (Figure 1 of the Supporting Information). In a 5-fold external cross-validation, 127 compounds were randomly partitioned into 5 subsets of nearly equal size. Each subset was paired with the remaining 80% of the compounds to form a pair of external and modeling sets. The data within each modeling set were further divided into multiple pairs of training and test sets for internal validation.

Although models were built using the training set, model selection depended on their performance on both the training and test sets (i.e., internal validation) since training set accuracy alone is insufficient to establish robust and externally predictive models.[26] The prediction outcome for each model was categorized as "0" for nontoxic compounds or "1" for toxic ones. Selected models were then pooled into a consensus model by simple averaging and used to predict the hepatotoxicity of compounds in the external sets (i.e., external validation). The toxicity threshold was set at 0.5 unless otherwise mentioned, i.e., a compound is predicted to be nontoxic if a consensus mean is less than 0.5 and toxic otherwise.

Table 1. 5-Fold External Cross-Validation Prediction Performance of QSAR Models

descriptors	Dragon	Dragon	MOE	SiRMS
method	*k*NN	SVM	*k*NN	RF
specificity ± SD[a]	0.62 ± 0.17	0.62 ± 0.16	0.60 ± 0.18	0.77 ± 0.08
sensitivity ± SD	0.56 ± 0.14	0.48 ± 0.17	0.56 ± 0.16	0.45 ± 0.14
CCR ± SD	0.59 ± 0.11	0.55 ± 0.09	0.58 ± 0.12	0.61 ± 0.10
coverage (%)	98	98	98	100

[a] SD refers to the standard deviation of the external predictivity measures (e.g., specificity) across the 5 folds.

The Y-randomization test was employed to ensure that there was no chance correlation between selected descriptors and hepatotoxicity. After random permutation of the hepatotoxicity labels in the modeling sets, models were rebuilt following the same workflow, and their CCR values for both training and test sets were collected and compared. This test was repeated at least three times. Models generated from the randomized labels were expected to perform significantly worse than those derived from the original data set.

All reported model predictivity measures, specificity, sensitivity, and correct classification rate, were obtained from 5-fold external cross-validation. Specificity denotes the true negative rate, or the rate correctly predicted within the nonhepatotoxic class. Similarly, sensitivity, the true positive rate, measures the rate correctly predicted within the hepatotoxic class. CCR is the average of the rates correctly predicted within each class (CCR = [specificity + sensitivity]/2). Coverage is the percentage of compounds in the external set within the applicability domain (AD) of the model. The AD is a similarity threshold within which compounds can be reliably predicted.[27]

Chemical and toxicogenomics descriptors found to be predictive were subsequently analyzed. Ingenuity Pathway Analysis (Ingenuity Systems, Redwood City, CA) software was used for the functional analysis of the significant transcripts. The networks were constructed on the basis of predefined molecular interactions in the Ingenuity database, and the Ingenuity score was used to rank pathways for analysis. Chemicals were clustered by the selected toxicogenomics descriptors using an unsupervised self-organizing map (SOM) in R (Kohonen package). Chemical structural alerts for hepatotoxicity were identified using HiT QSAR[17] and verified with XCHEM.[28] Briefly, XCHEM searches for common structural motifs within each class and ranks them by their relative frequencies.

RESULTS

Model Development. First, we developed QSAR models of subchronic (28 days of treatment) hepatotoxicity using various types of chemical descriptors (Table 1). Prediction performance was generally poor (55−61% CCR) across all descriptor types and classification methods. Three compounds (tannic acid, vancomycin, and cyclosporine) with molecular weights exceeding 1,200 (median molecular weight of the data set was 285) were excluded from the data set, corresponding to a coverage of 98% for some of the models. Given the generally unpromising results of the QSAR models described in Table 1, further Combi-QSAR[29] efforts to systematically combine each descriptor type with each classification method were not attempted.

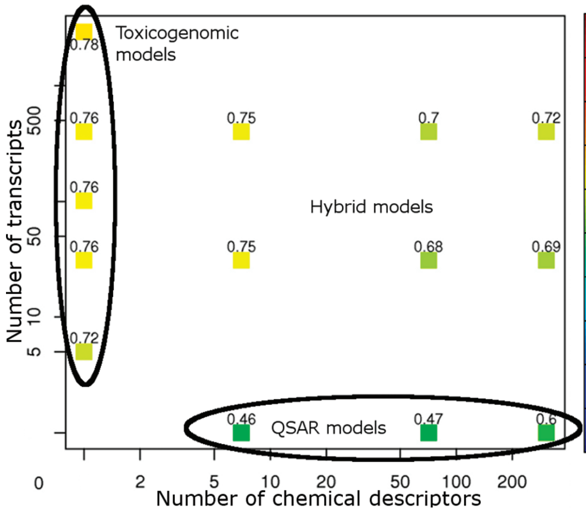


Figure 2. CCR accuracy of the models with respect to the number of chemical descriptors and transcripts used. All models were generated by SVM classification with 5-fold external cross-validation.

Second, we developed classification models of subchronic (28 days of treatment) hepatotoxicity using liver toxicogenomics data obtained after a single dose treatment as a predictor of future toxicity. To find the optimal number of variables (transcripts), several sets of top ranking transcripts were selected (based on SAM analysis) for modeling by SVM, and the outcomes were compared (Figure 2). CCR ranged from 72% with top 4 significant transcripts per modeling fold to 78% with all 2,923 significant transcripts. An optimal model with a CCR of 76% was achieved when 30 transcripts per fold were used. These 5 sets of 30 transcripts per fold comprised of 85 unique transcripts across all folds, which may serve as predictive biomarkers (Table 2 of the Supporting Information). We used these 85 transcripts to develop additional models employing other classification methods (Table 2). The RF model had the highest performance with a CCR of 76%. DWD was also applied to the full set of 2,923 transcripts and had a CCR of 73%. The difference in performance between the QSAR and the toxicogenomic models was significant ($p < 0.001$).

Third, we developed hybrid models of subchronic (28 days of treatment) hepatotoxicity using both chemical descriptors and single-dose treatment toxicogenomics data as biological descriptors. We studied how SVM model predictivity was affected when both the number of chemical descriptors and the number of transcripts were varied. To that effect, SAM was applied to independently rank chemical descriptors and transcripts, after which, different portions of top ranked variables were used for SVM modeling. Figure 2 shows that the CCR of the hybrid models did not exceed that of the models based on toxicogenomics data alone. However, hybrid models identified both important chemical descriptors and transcripts for the enhanced interpretation of the modeling outcomes. We could not have reliably detected the important chemical features from the relatively poorly fitted QSAR models. Adding transcripts boosted the predictivity of the hybrid models such that important chemical features were identified with greater confidence. Specifically, contributions of SiRMS descriptors used in RF hybrid models were interpreted using the approach of Polishchuk et al.[23] to uncover chemical substructures critical to hepatotoxicity. The substructures obtained through this analysis were compared to the alerts derived using XCHEM[28] and found to be concordant. The largest and most frequent substructures within each toxicity class are listed in Table 3 and provide evidence of the structure–activity relationship in the hybrid model. All QSAR, toxicogenomics, and hybrid models were significantly better than Y-randomized models ($p < 0.05$ by Z-test), indicating that our models were not the result of chance correlations.

Table 2. 5-Fold External Cross-Validation Prediction Performance of Toxicogenomics Models Based on the 85 Selected Transcripts[a]

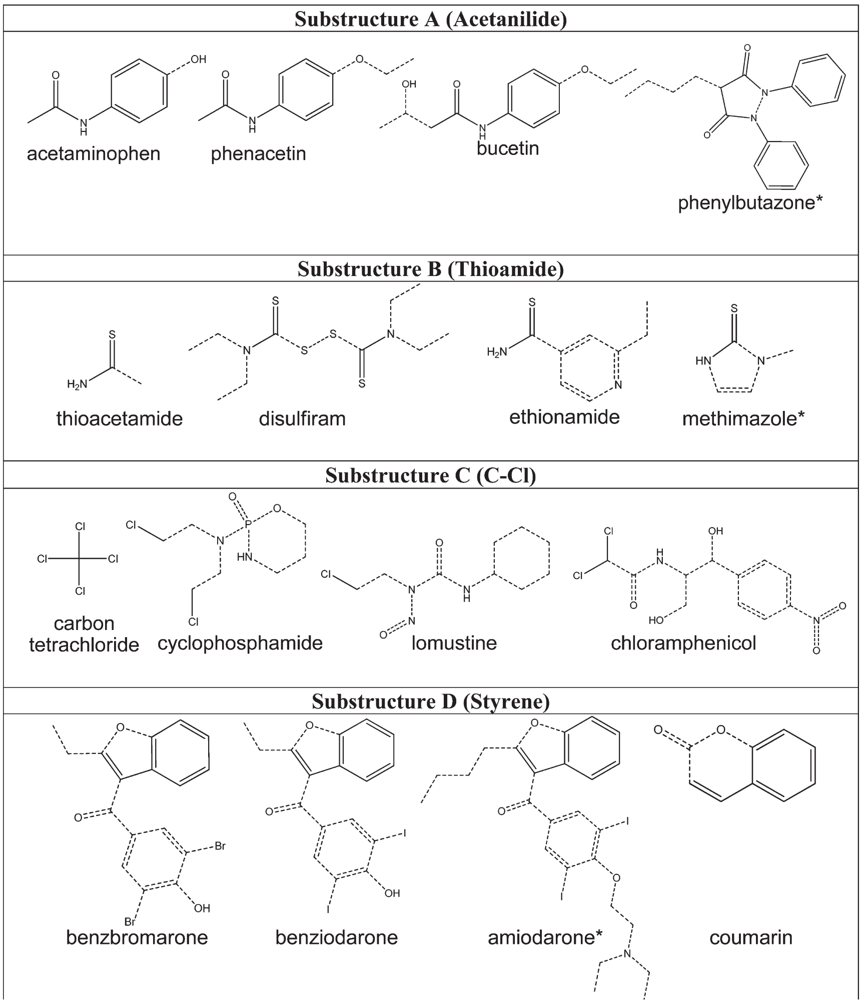
method	*k*NN	SVM	DWD	RF
specificity ± SD	0.82 ± 0.08	0.84 ± 0.10	0.77 ± 0.11	0.84 ± 0.05
sensitivity ± SD	0.57 ± 0.07	0.67 ± 0.12	0.62 ± 0.17	0.66 ± 0.20
CCR ± SD	0.70 ± 0.06	0.76 ± 0.09	0.69 ± 0.11	0.76 ± 0.10
coverage (%)	95	99	99	100

[a] See Table II of the Supporting Information for a complete list.

The toxicity threshold of the consensus models was set to 0.5, below which the compounds were classified as nontoxic and above which they were classified as toxic. Because the compounds on the margin are typically predicted with less confidence, we sought to determine the effect of adjusting the toxicity threshold on prediction performance. Figure 3A shows the distribution of QSAR-predicted values (using *k*NN method) for nontoxic and toxic compounds. Overall, the separation was poor due to a large proportion of nontoxic compounds that were predicted as toxic. While alternative thresholds yielding models with very high CCR may be selected (Figure 3C), severely reduced coverage of such models is a considerable drawback (Figure 3E). For example, setting two thresholds (dashed lines in Figure 3A), one at 0.36 (<0.36 are assigned nontoxic) and the second one at 0.56 (>0.56 are assigned as toxic) increased CCR to 68%, as compared to 59% with a single threshold of 0.5. However, the coverage of such a model was only 80% because the compounds whose predicted activities were between 0.36 and 0.56 could no longer be classified. Conversely, the toxicogenomics model developed with *k*NN showed good separation between toxic and nontoxic compounds (Figure 3B). A change in thresholds had a minor effect on the model's CCR and coverage (Figure 3D and F), showing that a single threshold was sufficient and that optimization of the activity thresholds would not be necessary. The optimal thresholds will be useful in the prediction of additional external compounds.

Model Interpretation. Toxicogenomics data-based models were the most predictive of hepatotoxicity. To explore the biological significance and the mechanistic relevance of the selected 85 transcripts (64 up-regulated and 21 down-regulated), functional pathway analysis was performed. Hepatic nuclear factor 4α (*Hnf4a*)- and v-myc myelocytomatosis viral oncogene homologue (*Myc*)-centered interactomes were the two highest ranked networks involving large numbers of the 64 selected up-regulated genes (Figure 4A–B and Table IIIa of the Supporting Information). Canonical pathway analysis revealed that the eukaryotic initiation factor (*Eif*) 2 signaling pathway responsible for protein translation was up-regulated (Table IIIb of the Supporting Information). Among the down-regulated genes, the network involving cellular function and maintenance including transporters and inflammatory responses was the highest ranked network (Figure 4C and Table IIIc of the Supporting Information). Canonical pathway analysis also revealed that

Table 3. Structural Alerts Mapped onto Example Compounds[a]



[a] All compounds are toxic unless denoted with an asterisk.

many down-regulated genes were involved in the complement pathway (Table IIId of the Supporting Information).

In addition, we used an unsupervised self-organizing map to cluster chemicals on the basis of their gene expression profiles (Figures 5 and 2 of the Supporting Information). The objective was to uncover commonalities within clusters with similar gene expression profiles. As expected, the nonhepatotoxic agents were tightly clustered (green background). Among the hepatotoxic drugs (orange background), there were several clusters of compounds which may act through similar mechanisms of action. For example, oxidative stress-inducing agents (red text) such as acetaminophen, methapyriline, and nimesulide, and peroxisome proliferator-activated alpha (PPARα) agonists (blue text) such as fenofibrate, WY-14643, benzbromarone, clofibrate, and gemfibrozil formed two subclusters among the hepatotoxicants. The model-selected 85 transcripts were sufficient to cluster the drugs into toxicologically meaningful groups with similar modes of hepatotoxicity.

Understanding this difference in performance between the QSAR and the toxicogenomics models warrants an in-depth examination of the spatial distribution of compounds in their chemical and toxicogenomics descriptor space. Principal component analysis of the chemical features (Dragon descriptors, Figure 6A) and toxicogenomics data (85 selected transcripts, Figure 6B) demonstrated that the separation between nontoxic and toxic classes was poor in the chemical space. Table IVa of the Supporting Information lists 40 most chemically similar pairs of compounds. Half of them had opposite toxicities. Conversely, among pairs of compounds with the most similar gene expression profiles, only 23% exhibited opposite toxicities (Table IVb of the Supporting Information). In other words, pairs of compounds with similar gene expression profiles were more likely to have the same hepatotoxicity than pairs of chemically similar compounds.

The best hybrid model had similar performance to the best toxicogenomics model (76−77% CCR), differing only in the predictions of three compounds (ajmaline, griseofulvin, propylthiouracil). Examining QSAR and toxicogenomics models in comparison with each other revealed instances for which the models were complementary. When both QSAR and toxicogenomics models were in agreement, it implied greater reliability of the prediction (Table 4). When predictions made with these two types of models were in disagreement, deferring to the toxicogenomics model (statistically superior to the QSAR model) would more likely return correct predictions. However, of note

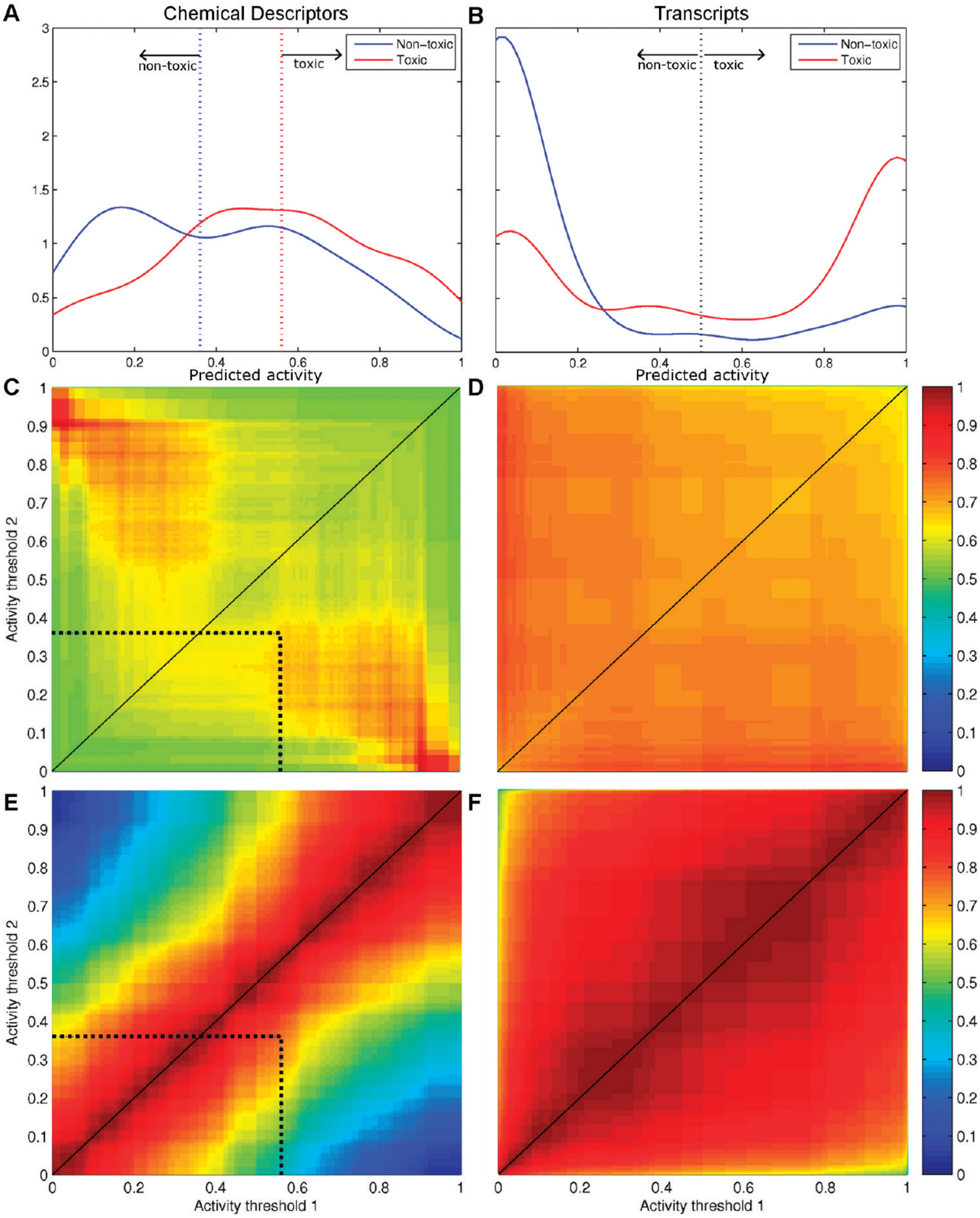


Figure 3. External prediction results of the QSAR (A, C, and E) and toxicogenomics (B, D, and F) models by *k*NN using different classification criteria. Distribution of the predicted values (A and B) and heat maps illustrating classification accuracy (C and D, CCR) and coverage (E and F, percent chemicals within the applicability domain) results are shown. Dashed (A and B) and diagonal (B−F) lines denote a default single-threshold classification (threshold = 0.5). An example of a double-threshold classification (nontoxic if activity <0.36; toxic if activity >0.56) is shown by the dashed lines (C and E).

were 19 compounds (italicized in Table 4) mis-predicted by the toxicogenomics model but correctly predicted by the QSAR model. The PCA plot shows that many of these compounds (denoted by crosses in Figures 6A and B) had neighbors in the multidimensional toxicogenomics descriptor space of opposite toxicities (Figure 6B), but their neighbors in the chemistry space had similar toxicities (Figure 6A). For example, nontoxic danazol has toxic neighbors in the toxicogenomics descriptor space (Figure 6B) but nontoxic neighbors in the chemistry space (Figure 6A). Some of these mis-predicted compounds, e.g., gemfibrozil (PPARα activator) and lomustine (genotoxic hepatocarcinogen), exhibit late-onset toxicity which could explain the failure of 24 h

dx.doi.org/10.1021/tx200148a | *Chem. Res. Toxicol.* 2011, 24, 1251–1262

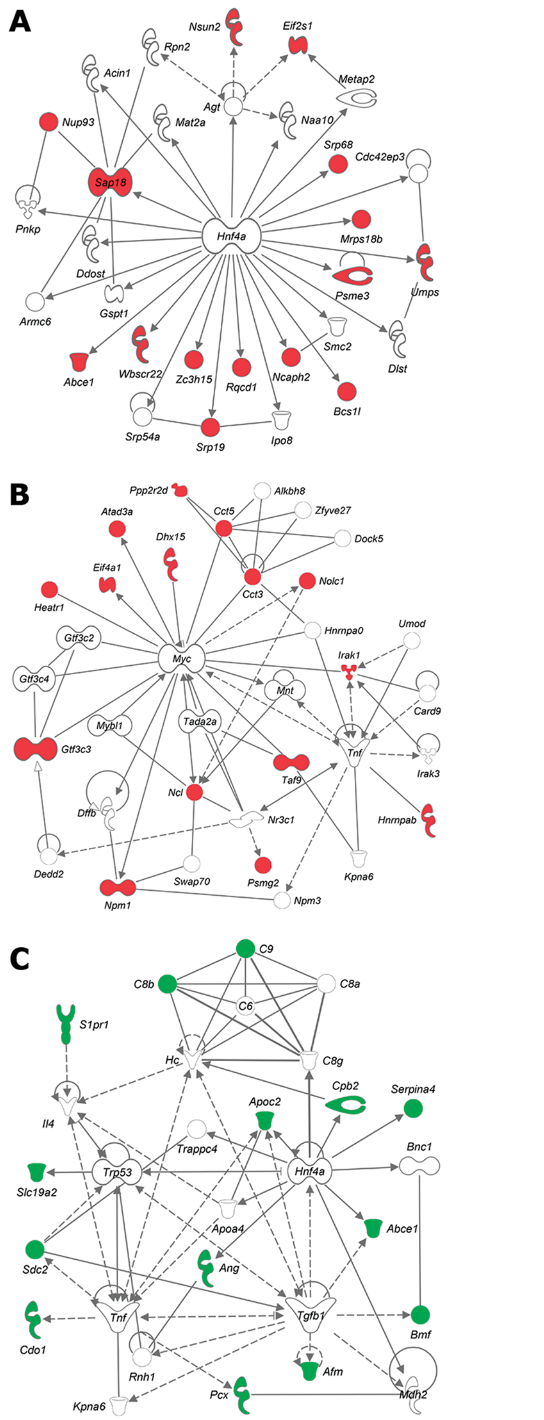


Figure 4. Molecular networks representing the toxicogenomics predictors of hepatotoxicity. *Hnf4a*-centered (A), *Myc*-centered (B), and cellular function, and maintenance-related (C) interactomes were selected as the highest ranked networks among the 64 up- or 21 down-regulated genes used in modeling. Red and green represent molecules up-regulated or down-regulated, respectively, by the hepatotoxic compounds. Ellipses, squares, triangles, trapezoids, lozenges, and circles represent transcription regulator, cytokine, kinase, transporter, enzyme, and other molecules, respectively. Arrows indicate molecular interactions, while lines indicate binding. Dashed arrows or lines indicate indirect interactions or binding. See Tables IIIa-d in the Supporting Information for a complete list of networks.

expression profiles to capture relevant changes and consequently to predict their 28-day hepatotoxicity.

DISCUSSION

Our study showed that chemical features and toxicogenomics data were useful and relevant for the development of classification models for understanding and predicting hepatotoxicity. The high classification accuracy of toxicogenomics models supports the use of early transcriptional response as an indicator for long-term toxicity and for understanding a potential mode of action. Even though QSAR models were less predictive, they will continue to be used for initial virtual screening in cases where no experimental data (e.g., toxicogenomics) are available. By developing hybrid models using both chemical descriptors and toxicogenomics data, we identified both chemical features and transcripts, which provided additional insights into understanding drug-induced liver injury.

Biological Pathways Involved in Liver Injury. Toxicogenomics data from single exposure were not only useful for the classification of 28-day liver injury phenotype but also provided important mechanistic insights into pathways that may lead to long-term toxicity. Pathway analysis showed that the 85 most predictive transcripts were in Hnf4α-, Myc-, and Eif2-centered networks, all of which have been implicated in hepatotoxicity. Hnf4α, a transcriptional factor of the nuclear hormone receptor family, is known to play an important role in liver function, morphological and functional differentiation of hepatocytes, cell proliferation, and detoxification.[30] Although the *Hnf4α* gene itself was not among the selected transcripts, Hnf4α-regulated genes were up-regulated in the early stage of hepatocellular injury.

In addition, Hnf4α is essential for controlling the acute phase response of the liver induced by endoplasmic reticulum (ER) stress.[31] ER stress is a common response to many toxicants, and under conditions of severe or prolonged ER stress, apoptosis is triggered by accumulation of incompletely assembled or misfolded proteins.[32] Activation of Eif2 signaling pathway is widely recognized as a key contributor to ER stress. In the present study, we found the characteristic up-regulation of several genes involved in Eif2 signaling pathway after treatment with several hepatotoxicants, such as Eif2 subunit 1 alpha (*Eif2s1*), Eif3 subunits G (*Eif3G*) and J (*Eif3J*), and *Eif4a1*. Thus, our analysis provided additional supporting evidence that the Eif2 signaling pathway may be a common mechanism involved in early liver damage through ER stress.

Myc is a transcription factor which regulates cell proliferation, differentiation, and apoptosis.[33] In the present study, we found up-regulations of several genes in the Myc-centered network including transcription factors nucleophosmin 1 (*Npm1*), TAF9 RNA polymerase II, TATA box binding protein (TBP)-associated factor (*Taf9*), *Eif4a1*, and general transcription factor IIIC polypeptide 3 (*Gtf3c3*). While further studies are needed to link the effects of individual chemicals to transcriptional changes in the Myc-centered network, our analysis shows that these transcripts may be important early predictive biomarkers for subchronic hepatocellular injury.

Biological pathway analysis revealed the down-regulation of genes involved in cellular function and maintenance, consisting of transporters and inflammatory response, such as the complement system pathway. Abnormal homeostasis and cellular function are often associated with hepatotoxicity. In particular, coagulopathy is often involved because many factors in the coagulation system are synthesized in the liver. Recently, toxicogenomics biomarkers for diagnosis and prognosis of hepatotoxicity-related

dx.doi.org/10.1021/tx200148a |Chem. Res. Toxicol. 2011, 24, 1251–1262

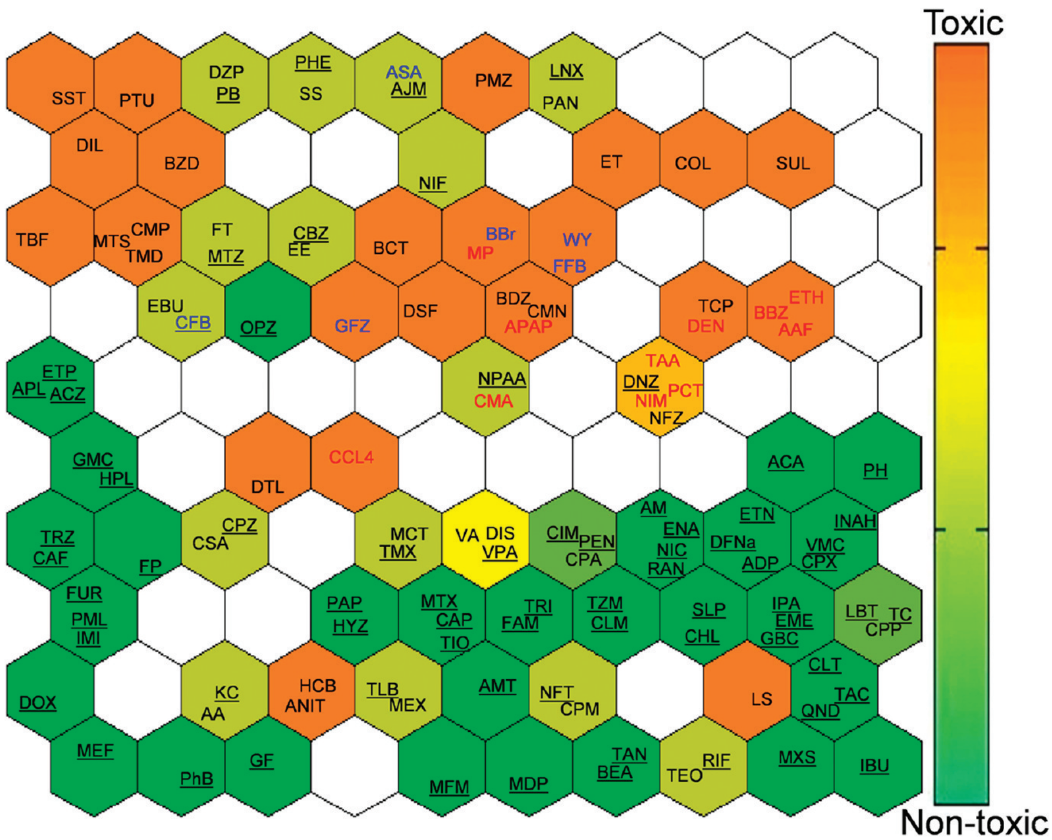


Figure 5. Self-organizing map of the compounds clustered by the expression of the 85 selected transcripts. Nontoxic (underlined) compounds are tightly clustered in the bottom right. PPARα activating and oxidative stress-inducing chemicals are colored in blue and red, respectively.

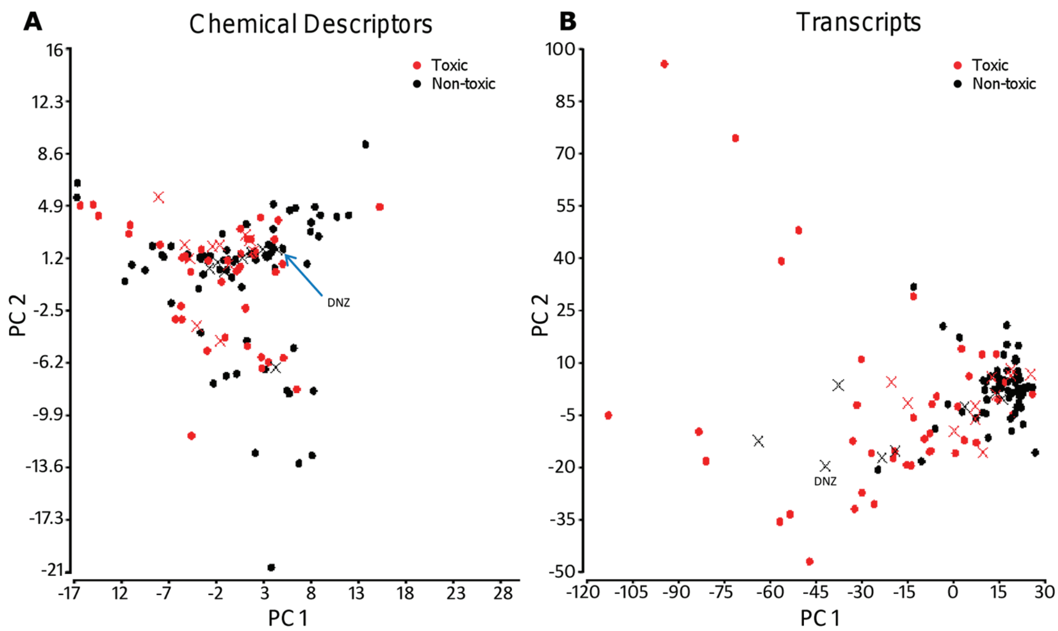


Figure 6. Principal component analysis of the chemical (A) and toxicogenomics (B) descriptors. Toxic and nontoxic compounds are colored red and black, respectively. Compounds mis-predicted by the toxicogenomics model but correctly predicted by the QSAR model are marked as crosses (×). An example of a nontoxic compound (danazol, DNZ) which has distant toxic toxicogenomic neighbors but close nontoxic chemical neighbors is shown.

coagulation abnormalities have been reported.[34] Our results further support that malfunction of the coagulation system is a common feature in liver injury and that the down-regulation of complement 8, β-polypeptide (*C8b*), complement 9 (*C9*), and

Table 4. Confusion Matrix Showing Predictions by the QSAR Model and Toxicogenomics Model[a]

Toxicogenomic model (85 transcripts, kNN)	QSAR model (Dragon descriptors, kNN): Predicted as non-toxic		QSAR model (Dragon descriptors, kNN): Predicted as toxic	
	Actually non-toxic	Actually toxic	Actually non-toxic	Actually toxic
Predicted as toxic	1. *carbamazepine* 2. *danazol* 3. *nitrofurazone* 4. *omeprazole* 5. *papaverine* 6. *phenylanthranilic acid* 7. *phenytoin* 8. *tamoxifen*	1. bendazac 2. chloramphenicol 3. colchicine 4. dantrolene 5. diltiazem 6. ethambutol 7. ethionine 8. fenofibrate 9. monocrotaline 10. propylthiouracil 11. terbinafine 12. trimethadione 13. WY-14643	1. bromoethanamine 2. clofibrate 3. griseofulvin 4. methimazole 5. nifedipine	1. acetaminophen 2. benzbromarone 3. bucetin 4. carbon tetrachloride 5. chlormezanone 6. coumarin 7. disulfiram 8. flutamide 9. methapyrilene 10. methyltestosterone 11. nimesulide 12. phenacetin 13. simvastatin 14. thioacetamide
Predicted as non-toxic	1. acarbose 2. adapin 3. amiodarone 4. amitriptyline 5. chlorpheniramine 6. cimetidine 7. ciprofloxacin 8. doxorubicin 9. enalapril 10. erythromycin ethylsuccinate 11. famotidine 12. fluphenazine 13. furosemide 14. gentamicin 15. glibenclamide 16. hydroxyzine 17. imipramine 18. iproniazid 19. ketoconazole 20. labetalol 21. mefenamic acid 22. metformin 23. methotrexate 24. moxisylyte 25. nicotinic acid 26. nitrofurantoin 27. pemoline 28. penicillamine 29. phenobarbital 30. quinidine 31. ranitidine 32. rifampicin 33. sulpiride 34. tacrine 35. tetracycline 36. thioridazine 37. triamterene	1. allyl alcohol 2. chlorpropamide 3. clomipramine 4. cyclosporine A 5. disopyramide 6. mexiletine 7. puromycin aminonucleoside 8. sulfasalazine 9. theophylline	1. acetazolamide 2. ajmaline 3. allopurinol 4. caffeine 5. captopril 6. cephalothin 7. chlormadinone 8. chlorpromazine 9. diclofenac 10. ethanol 11. etoposide 12. haloperidol 13. ibuprofen 14. isoniazid 15. lornoxicam 16. methyldopa 17. perhexiline 18. phenylbutazone 19. tannic acid 20. tiopronin 21. tolbutamide 22. triazolam 23. valproic acid 24. vancomycin	1. *aspirin* 2. *benziodarone* 3. *cyclophosphamide* 4. *diazepam* 5. *ethinylestradiol* 6. *gemfibrozil* 7. *hexachlorobenzene* 8. *lomustine* 9. *naphthyl isothiocyanate* 10. *promethazine* 11. *vitamin A*

[a] Compounds mis-predicted by the toxicogenomics model but correctly predicted by the QSAR model are identified in italicized font. Compounds mis-predicted by both the QSAR model and by the toxicogenomics model are underlined.

complement factor B (*Cfb*) may be an early indicator of impaired liver function by different types of drugs.

Many of the 85 selected transcripts have also been previously implicated with liver diseases by the same chemicals in the Comparative Toxicogenomics Database (http://ctd.mdibl.org/). For instance, ubiquitin specific peptidase 10 (*Usp10*) has been associated with the Myc-centered network in acetaminophen-induced liver toxicity.[35] It is also closely related to ubiquitin specific peptidase 2 (*Usp2*) which is among the 37 genes used to derive a toxicogenomics model for hepatotumorigenesis by Fielden et al.[6] The agreement with previous findings lends credence to our selected list of transcripts as biomarkers for hepatotoxicity.

Hybrid Models Afford More Reliable Exploration of Chemical Structural Alerts. Development of QSAR models of hepatotoxicity for structurally diverse chemicals is a challenge,[36] and the results of this study show that a correct classification rate of such models ranged between 55 and 61%. Thus, interpretation of such models with regards to the potential chemical "structural alerts" for hepatotoxicity may be futile. However, when chemical descriptors and toxicogenomics data were used together to develop hybrid models, significantly higher predictive accuracy (as high as 77%) of the models provided additional confidence for considering the chemical fragments selected by the models as potentially predictive of an increased risk of liver toxicity. By examining the chemical substructures suggested by the hybrid models (see Table 3), we observe that features selected through the modeling procedure are several well-known toxicophores. This finding provides a strong indication of the value of hybrid modeling for identification of the toxicophores as compared to the traditional QSAR, which is plagued by a weaker predictive power.

Substructure A (Acetanilide): Toxic Species Formed, N-Hydroxylamines and Nitroso Compounds. The acetanilide substructure was present in several hepatotoxic drugs, as well as the nontoxic phenylbutazone. The acetanilide substructure is especially susceptible to *N*-oxidation.[37] The *N*-hydroxylamine and nitroso products are highly reactive. However, some compounds may be toxic due to activation at sites outside of the acetanilide substructure. For example, acetaminophen owes much of its toxicity to the quinone imine metabolite despite its chemical similarity with phenacetin. Its only difference from phenacetin is its 4-hydroxyl group, which is preferentially oxidized by CYP2E1 to the reactive quinone imine. In phenacetin and bucetin, the 4-hydroxyl group is replaced by an alkoxyl substituent which renders them less susceptible to quinone formation and more likely to be activated by *N*-hydroxylation.[38] Phenylbutazone also undergoes another transformation (aromatic hydroxylation) instead of *N*-hydroxylation.[39] This probably explains its lack of rat hepatotoxicity in this study despite containing the acetanilide substructure.

Substructure B (Thioamide): Toxic Species Formed, Sulfur Species of Various Oxidation States. Our models showed that the presence of thioamide (Table 3, substructure B) is associated with hepatotoxicity. Thiocarbonyls are often oxidized or desulfurated to produce toxic sulfur-containing species. Thioacetamide S-oxide is highly polar and forms adducts with proteins.[40] Disulfiram, despite being a dithiocarbamate instead of a thioamide, also forms a sulfoxide that binds to proteins and inhibits their activity. Such protein binding is also responsible for disulfiram's therapeutic inhibition of aldehyde dehydrogenase.[41] The only nontoxic drug that has this substructure was methimazole. Although methimazole was defined as nonhepatotoxic in this study, it has been reported to yield atomic sulfur species that bind and inhibit P450 activity, possibly leading to liver necrosis.[42]

Substructure C (Alkyl Chloride): Toxic Species Formed, Alkyl Radicals. Hepatotoxicity of alkyl chloride compounds has been attributed to the homolytic cleavage of the C−Cl bond which produces damaging free radicals. This is a well-studied phenomenon best exemplified by carbon tetrachloride and its alkyl halide analogues such as chloroform and bromotrichloromethane.[43] However, other chlorinated alkanes studied here, cyclophosphamide, lomustine and chloramphenicol, do not share the same toxic mechanism as carbon tetrachloride and cannot be attributed to the C−Cl bond. For instance, the ultimate toxicant responsible for cyclophosphamide hepatotoxicity is acrolein, which is formed independently of the alkyl chloride group.

Substructure D (Styrene): Toxic Species Formed: Epoxides. The nonaryl double bond in substructure D when it is part of a benzofuran or benzopyran is especially prone to epoxide formation.[44] Such epoxides often form DNA and protein adducts.[45] Coumarin's toxicity requires the formation of an epoxide, which is followed by subsequent rearrangement of the epoxide to *o*-hydroxyphenylacetaldehyde, which is considered to be the hepatotoxic intermediate.[46] Hence, it is comparatively more toxic in rats than in humans because of the rat's metabolism via the 3,4-epoxide,[47] while in humans, coumarin primarily undergoes aromatic hydroxylation instead of forming the above-mentioned epoxide.[46,48] The three benzofurans in our study, benziodarone, benzbromarone, and amiodarone, are known hepatotoxic agents whose toxicity has been attributed to the 2-substituted benzofuran.[44] Although amiodarone was not found to be hepatotoxic on the basis of its 28-day histopathology and serum chemistry results, hepatocellular vacuolization indicative of phospholipidosis was noted (Table 1 of the Supporting Information).

Limitations. The performance of QSAR models generally suffers when predicting complex toxicity end points such as hepatotoxicity, a phenotype with several complex mechanisms. There are numerous examples of chemically similar compounds with widely divergent liver effects. While ibuprofen is safe in humans, ibufenac, lacking a methyl group, is toxic.[36] In our data set, nontoxic caffeine and toxic theophylline differ by a methyl group. This phenomenon is known as an "activity cliff" where very similar molecules possess disparate activities, such that the profile of activity plotted against compound's similarity is akin to a rugged landscape with many cliffs.[49] QSAR can be realistically applied if there are enough compounds to adequately represent the complex activity landscape. Unfortunately, this was not the case for our data set. The high proportion (50%) of opposite activities among chemically similar pairs compounded by the lack of congeners in our chemically diverse set posed further challenges to QSAR modeling. Hence, it was not surprising that the CCR of the QSAR models could barely exceed 60% in predicting the biologically complex hepatotoxicity end point.

In conclusion, this study shows that while QSAR and toxicogenomics are both important predictive tools on their own, concomitant exploration in chemical and toxicogenomics descriptor spaces, through hybrid models, will elicit deeper insight. Consistent with results from other toxicogenomics studies, we showed that toxicogenomics is predictive and provides valuable mechanistic information. The pathways suggested several mechanisms such as ER stress and coagulopathy that could be related to hepatotoxicity. As QSAR is entirely computational and obviates the need for experiments, it will remain an important virtual screening tool. Importantly, structural alerts can be identified with greater confidence from the better fitted hybrid models. In addition, hybrid models improve and refine the interpretation of the data in terms of chemical alerts for hepatotoxicity. Additional studies using methodologies and descriptors that can handle activity cliffs in both chemical and toxicogenomics descriptor spaces may improve the predictive power of models developed in this study and exploit further the complementarities between QSAR and toxicogenomics models of hepatotoxicity.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. Entire data set of compounds (identified by their CAS numbers) with the dosage information, histopathology, and serum chemistry results; list of predictive gene biomarkers; list of pathways involving the predictive gene biomarkers; and a list of compounds paired by chemical and transcriptional similarities. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ivan_rusyn@unc.edu (I.R.); alex_tropsha@unc.edu (A.T.).

Author Contributions

▽Equally contributing first coauthors.

Funding Sources

This study was supported by grants from the Ministry of Health, Labor and Welfare of Japan (H14-Toxico-001, H19-Toxico-001), NIH (GM66940, ES015241), and EPA (RD 83382501, RD83272001).

■ DISCLOSURE

The research described in this article has not been subjected to each agency's policy review and therefore does not necessarily reflect their views, and no official endorsement should be inferred.

■ REFERENCES

(1) Schuster, D., Laggner, C., and Langer, T. (2005) Why drugs fail--a study on side effects in new chemical entities. *Curr. Pharm. Des. 11*, 3545–3559.

(2) Tropsha, A. (2010) Best practices for QSAR model development, validation, and exploitation. *Mol. Inf. 29*, 1868–1751.

(3) Hou, T., and Wang, J. (2008) Structure-ADME relationship: still a long way to go? *Expert Opin. Drug Metab. Toxicol. 4*, 759–770.

(4) Cui, Y., and Paules, R. S. (2010) Use of transcriptomics in understanding mechanisms of drug-induced toxicity. *Pharmacogenomics 11*, 573–585.

(5) Blomme, E. A., Yang, Y., and Waring, J. F. (2009) Use of toxicogenomics to understand mechanisms of drug-induced hepatotoxicity during drug discovery and development. *Toxicol. Lett. 186*, 22–31.

(6) Fielden, M. R., Brennan, R., and Gollub, J. (2007) A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. *Toxicol. Sci. 99*, 90–100.

(7) Zidek, N., Hellmann, J., Kramer, P. J., and Hewitt, P. G. (2007) Acute hepatotoxicity: a predictive model based on focused illumina microarrays. *Toxicol. Sci. 99*, 289–302.

(8) Hirode, M., Ono, A., Miyagishima, T., Nagao, T., Ohno, Y., and Urushidani, T. (2008) Gene expression profiling in rat liver treated with compounds inducing phospholipidosis. *Toxicol. Appl. Pharmacol. 229*, 290–299.

(9) Kiyosawa, N., Uehara, T., Gao, W., Omura, K., Hirode, M., Shimizu, T., Mizukawa, Y., Ono, A., Miyagishima, T., Nagao, T., and Urushidani, T. (2007) Identification of glutathione depletion-responsive genes using phorone-treated rat liver. *J. Toxicol. Sci. 32*, 469–486.

(10) Hirode, M., Horinouchi, A., Uehara, T., Ono, A., Miyagishima, T., Yamada, H., Nagao, T., Ohno, Y., and Urushidani, T. (2009) Gene expression profiling in rat liver treated with compounds inducing elevation of bilirubin. *Hum. Exp. Toxicol. 28*, 231–244.

(11) Uehara, T., Hirode, M., Ono, A., Kiyosawa, N., Omura, K., Shimizu, T., Mizukawa, Y., Miyagishima, T., Nagao, T., and Urushidani, T. (2008) A toxicogenomics approach for early assessment of potential non-genotoxic hepatocarcinogenicity of chemicals in rats. *Toxicology 250*, 15–26.

(12) Tamura, K., Ono, A., Miyagishima, T., Nagao, T., and Urushidani, T. (2006) Profiling of gene expression in rat liver and rat primary cultured hepatocytes treated with peroxisome proliferators. *J. Toxicol. Sci. 31*, 471–490.

(13) Zhu, H., Rusyn, I., Richard, A., and Tropsha, A. (2008) Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. *Environ. Health Perspect. 116*, 506–513.

(14) Sedykh, A., Zhu, H., Tang, H., Zhang, L., Richard, A., Rusyn, I., and Tropsha, A. (2011) Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. *Environ. Health Perspect. 119*, 364–370.

(15) Uehara, T., Ono, A., Maruyama, T., Kato, I., Yamada, H., Ohno, Y., and Urushidani, T. (2010) The Japanese toxicogenomics project: application of toxicogenomics. *Mol. Nutr. Food Res. 54*, 218–227.

(16) Fourches, D., Muratov, E., and Tropsha, A. (2010) Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model. 50*, 1189–1204.

(17) Kuz'min, V. E., Artemenko, A. G., and Muratov, E. N. (2008) Hierarchical QSAR technology based on the Simplex representation of molecular structure. *J. Comput.-Aided Mol. Des. 22*, 403–421.

(18) Varnek, A., Fourches, D., Horvath, D., Klimchuk, O., Gaudin, C., Vayer, P., Solov'yev, V., Hoonakker, F., Tetko, I. V., and Marcou, G. (2008) ISIDA: platform for virtual screening based on fragment and pharmacophoric descriptors. *Curr. Comput. Aided Drug Des. 4*, 191–198.

(19) Muratov, E. N., Artemenko, A. G., Varlamova, E. V., Polischuk, P. G., Lozitsky, V. P., Fedchuk, A. S., Lozitska, R. L., Gridina, T. L., Koroleva, L. S., Sil'nikov, V. N., Galabov, A. S., Makarov, V. A., Riabova, O. B., Wutzler, P., Schmidtke, M., and Kuz'min, V. E. (2010) Per aspera ad astra: application of Simplex QSAR approach in antiviral research. *Future. Med. Chem. 2*, 1205–1226.

(20) Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A. 98*, 5116–5121.

(21) Zheng, W., and Tropsha, A. (2000) Novel variable selection quantitative structure−property relationship approach based on the k-nearest-neighbor principle. *J. Chem. Inf. Comput. Sci. 40*, 185–194.

(22) Fan, R. E., Chen, P. H., and Lin, C. J. (2005) Working set selection using the second order information for training SVM. *J. Mach. Leaning Res. 6*, 1889–1918.

(23) Polishchuk, P. G., Muratov, E. N., Artemenko, A. G., Kolumbin, O. G., Muratov, N. N., and Kuz'min, V. E. (2009) Application of random forest approach to QSAR prediction of aquatic toxicity. *J. Chem. Inf. Model. 49*, 2481–2488.

(24) Marron, J. S., Todd, M. J., and Ahn, J. (2007) Distance weighted discrimination. *J. Am. Stat. Assoc. 102*, 1267–1271.

(25) Tropsha, A., and Golbraikh, A. (2007) Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr. Pharm. Des. 13*, 3494–3504.

(26) Golbraikh, A., and Tropsha, A. (2002) Beware of q2!. *J. Mol. Graphics Modell. 20*, 269–276.

(27) Tropsha, A., Gramatica, P., and Gombar, V. K. (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *Quant. Struct. Act. Relat. Comb. Sci. 22*, 69–77.

(28) Sedykh, A. Y., and Klopman, G. (2006) A structural analogue approach to the prediction of the octanol-water partition coefficient. *J. Chem Inf. Model. 46*, 1598–1603.

(29) Kovatcheva, A., Golbraikh, A., Oloff, S., Xiao, Y. D., Zheng, W., Wolschann, P., Buchbauer, G., and Tropsha, A. (2004) Combinatorial QSAR of ambergris fragrance compounds. *J. Chem. Inf. Comput. Sci. 44*, 582–595.

(30) Parviz, F., Matullo, C., Garrison, W. D., Savatski, L., Adamson, J. W., Ning, G., Kaestner, K. H., Rossi, J. M., Zaret, K. S., and Duncan, S. A. (2003) Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nat. Genet. 34*, 292–296.

(31) Luebke-Wheeler, J., Zhang, K., Battle, M., Si-Tayeb, K., Garrison, W., Chhinder, S., Li, J., Kaufman, R. J., and Duncan, S. A. (2008) Hepatocyte nuclear factor 4alpha is implicated in endoplasmic reticulum stress-induced acute phase response by regulating expression of cyclic adenosine monophosphate responsive element binding protein H. *Hepatology 48*, 1242–1250.

(32) Ji, C., and Kaplowitz, N. (2006) ER stress: can the liver cope? *J. Hepatol. 45*, 321–333.

(33) Lin, C. J., Malina, A., and Pelletier, J. (2009) c-Myc and eIF4F constitute a feedforward loop that regulates cell growth: implications for anticancer therapy. *Cancer Res. 69*, 7491–7494.

(34) Hirode, M., Omura, K., Kiyosawa, N., Uehara, T., Shimuzu, T., Ono, A., Miyagishima, T., Nagao, T., Ohno, Y., and Urushidani, T. (2009) Gene expression profiling in rat liver treated with various hepatotoxic-compounds inducing coagulopathy. *J. Toxicol. Sci. 34*, 281–293.

(35) Beyer, R. P., Fry, R. C., Lasarev, M. R., McConnachie, L. A., Meira, L. B., Palmer, V. S., Powell, C. L., Ross, P. K., Bammler, T. K., Bradford, B. U., Cranson, A. B., Cunningham, M. L., Fannin, R. D., Higgins, G. M., Hurban, P., Kayton, R. J., Kerr, K. F., Kosyk, O., Lobenhofer, E. K., Sieber, S. O., Vliet, P. A., Weis, B. K., Wolfinger, R., Woods, C. G., Freedman, J. H., Linney, E., Kaufmann, W. K., Kavanagh, T. J., Paules, R. S., Rusyn, I., Samson, L. D., Spencer, P. S., Suk, W., Tennant, R. J., and Zarbl, H. (2007) Multicenter study of acetaminophen hepatotoxicity reveals the importance of biological endpoints in genomic analyses. *Toxicol. Sci. 99*, 326–337.

(36) Rodgers, A. D., Zhu, H., Fourches, D., Rusyn, I., and Tropsha, A. (2010) Modeling liver-related adverse effects of drugs using knearest neighbor quantitative structure-activity relationship method. *Chem. Res. Toxicol. 23*, 724–732.

(37) Loew, G. H., and Goldblum, A. (1985) Metabolic activation and toxicity of acetaminophen and related analogs. A theoretical study. *Mol. Pharmacol. 27*, 375–386.

(38) Peters, J. M., Morishima, H., Ward, J. M., Coakley, C. J., Kimura, S., and Gonzalez, F. J. (1999) Role of CYP1A2 in the toxicity of long-term phenacetin feeding in mice. *Toxicol. Sci. 50*, 82–89.

(39) Aarbakke, J., Bakke, O. M., Milde, E. J., and Davies, D. S. (1977) Disposition and oxidative metabolism of phenylbutazone in man. *Eur. J. Clin. Pharmacol. 11*, 359–366.

(40) Porter, W. R., and Neal, R. A. (1978) Metabolism of thioacetamide and thioacetamide S-oxide by rat liver microsomes. *Drug Metab. Dispos. 6*, 379–388.

(41) Shen, M. L., Johnson, K. L., Mays, D. C., Lipsky, J. J., and Naylor, S. (2001) Determination of in vivo adducts of disulfiram with mitochondrial aldehyde dehydrogenase. *Biochem. Pharmacol. 61*, 537–545.

(42) Lee, P. W., and Neal, R. A. (1978) Metabolism of methimazole by rat liver cytochrome P-450-containing monoxygenases. *Drug Metab. Dispos. 6*, 591–600.

(43) Rechnagel, R. O., and Glende, E. A., Jr. (1973) Carbon tetrachloride hepatotoxicity: an example of lethal cleavage. *CRC Crit. Rev. Toxicol. 2*, 263–297.

(44) Kaufmann, P., Torok, M., Hanni, A., Roberts, P., Gasser, R., and Krahenbuhl, S. (2005) Mechanisms of benzarone and benzbromarone-induced hepatic toxicity. *Hepatology 41*, 925–935.

(45) Adam, W., Ahrweiler, M., Saha-Moller, C. R., Sauter, M., Schonberger, A., Epe, B., Muller, E., Schiffmann, D., Stopper, H., and Wild, D. (1993) Genotoxicity studies of benzofuran dioxetanes and epoxides with isolated DNA, bacteria and mammalian cells. *Toxicol. Lett. 67*, 41–55.

(46) Vassallo, J. D., Hicks, S. M., Daston, G. P., and Lehman-McKeeman, L. D. (2004) Metabolic detoxification determines species differences in coumarin-induced hepatotoxicity. *Toxicol. Sci. 80*, 249–257.

(47) Lake, B. G., Gray, T. J., Evans, J. G., Lewis, D. F., Beamand, J. A., and Hue, K. L. (1989) Studies on the mechanism of coumarin-induced toxicity in rat hepatocytes: comparison with dihydrocoumarin and other coumarin metabolites. *Toxicol. Appl. Pharmacol. 97*, 311–323.

(48) Felter, S. P., Vassallo, J. D., Carlton, B. D., and Daston, G. P. (2006) A safety assessment of coumarin taking into account species-specificity of toxicokinetics. *Food Chem. Toxicol. 44*, 462–475.

(49) Maggiora, G. M. (2006) On outliers and activity cliffs--why QSAR often disappoints. *J. Chem. Inf. Model. 46*, 1535.