Toxicology and Applied Pharmacology 255 (2011) 297–306

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap

Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database

Takeki Uehara [a,b,*], Yohsuke Minowa [b], Yuji Morikawa [b], Chiaki Kondo [a], Toshiyuki Maruyama [a], Ikuo Kato [a], Noriyuki Nakatsu [b], Yoshinobu Igarashi [b], Atsushi Ono [b], Hitomi Hayashi [c,d], Kunitoshi Mitsumori [c], Hiroshi Yamada [b], Yasuo Ohno [e], Tetsuro Urushidani [b,f]

[a] Drug Developmental Research Laboratories, Shionogi & Co., Ltd., 3-1-1 Futaba-cho, Toyonaka, Osaka 561-0825, Japan
[b] Toxicogenomics Informatics Project, National Institute of Biomedical Innovation, 7-6-8 Asagi, Ibaraki, Osaka 567-0085, Japan
[c] Laboratory of Veterinary Pathology, Tokyo University of Agriculture and Technology, 3-5-8 Harumi-cho, Fuchu, Tokyo 183-8509, Japan
[d] Pathogenetic Veterinary Science, United Graduate School of Veterinary Sciences, Gifu University, 1-1 Yanagido, Gifu, 501-1193 Gifu, Japan
[e] National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan
[f] Department of Pathophysiology, Faculty of Pharmaceutical Sciences, Doshisha Women's College of Liberal Arts, Kodo, Kyotanabe, Kyoto 610-0395, Japan

ARTICLE INFO

Article history:
Received 7 April 2011
Revised 5 July 2011
Accepted 6 July 2011
Available online 19 July 2011

Keywords:
Toxicogenomics
Microarray
Hepatocarcinogenicity
TG-GATEs
Non-genotoxic hepatocarcinogens

ABSTRACT

The present study was performed to develop a robust gene-based prediction model for early assessment of potential hepatocarcinogenicity of chemicals in rats by using our toxicogenomics database, TG-GATEs (Genomics-Assisted Toxicity Evaluation System developed by the Toxicogenomics Project in Japan). The positive training set consisted of high- or middle-dose groups that received 6 different non-genotoxic hepatocarcinogens during a 28-day period. The negative training set consisted of high- or middle-dose groups of 54 non-carcinogens. Support vector machine combined with wrapper-type gene selection algorithms was used for modeling. Consequently, our best classifier yielded prediction accuracies for hepatocarcinogenicity of 99% sensitivity and 97% specificity in the training data set, and false positive prediction was almost completely eliminated. Pathway analysis of feature genes revealed that the mitogen-activated protein kinase p38- and phosphatidylinositol-3-kinase-centered interactome and the v-myc myelocytomatosis viral oncogene homolog-centered interactome were the 2 most significant networks. The usefulness and robustness of our predictor were further confirmed in an independent validation data set obtained from the public database. Interestingly, similar positive predictions were obtained in several genotoxic hepatocarcinogens as well as non-genotoxic hepatocarcinogens. These results indicate that the expression profiles of our newly selected candidate biomarker genes might be common characteristics in the early stage of carcinogenesis for both genotoxic and non-genotoxic carcinogens in the rat liver. Our toxicogenomic model might be useful for the prospective screening of hepatocarcinogenicity of compounds and prioritization of compounds for carcinogenicity testing.

Introduction

Carcinogenicity is one of the most serious side effects associated with new drug development. Thus, it is especially important for pharmaceutical companies to know as much as possible about the eventual carcinogenic properties of new drugs, even in the early stages of drug development. However, the current "gold standard" for carcinogenicity testing is a bioassay in which mice and rats are treated with a target compound for their entire 2-year lifespan. This carcinogenicity testing cannot be performed in the early stage of drug development because it is time-consuming and expensive and it requires the use of many animals and large amounts of chemicals. Additionally, while this assay provides evidence of carcinogenicity of the target chemicals in rodents, it provides only limited mechanistic information about carcinogenesis. Thus, the current strategy of a 2-year bioassay to evaluate in vivo carcinogenicity is not satisfactory. It should be replaced with better test systems that are cheaper and faster, use fewer animals, and provide the appropriate sensitivity and specificity desired for a screen of carcinogenic potential.

Toxicogenomics has been expected as a powerful approach for elucidating mechanisms underlying toxicological endpoints and a useful strategy for the early detection of potential chemical toxicity (Battershill, 2005; Heinloth et al., 2004; Irwin et al., 2004; Kiyosawa et al., 2009; Searfoss et al., 2005). Important scientific breakthroughs have been achieved by applying toxicogenomics to the early detection of chemical carcinogenicity in vivo. Studies have focused on hepatocarcinogenicity because the liver is the most common target organ for carcinogenesis. Kramer et al. (2004) studied microarray-derived comprehensive gene

* Corresponding author at: Developmental Research Laboratories, Shionogi & Co., Ltd., 3-1-1 Futaba-cho, Toyonaka, Osaka 561-0825, Japan. Fax: +81 6 6332 6385.
E-mail address: takeki.uehara@shionogi.co.jp (T. Uehara).

0041-008X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.taap.2011.07.001

expression data from the livers of rats treated with 10 non-genotoxic hepatocarcinogens and determined that the expression level of NAD(P)H P450 reductase was positively correlated with hepatocarcinogenicity and the expression level of transforming growth factor-β clone 22 was negatively correlated with hepatocarcinogenicity. Nie et al. (2006) demonstrated that 6 biomarker genes predict the carcinogenic potential of non-genotoxic hepatocarcinogens with a prediction accuracy of 88.5% by using expression data from the livers of rats treated with a single dose of 24 non-genotoxic chemicals and 28 non-hepatocarcinogens. Nakayama et al. (2006) successfully separated several isomers with or without hepatocarcinogenicity based on the expression profiles of selected genes, and they identified characteristic gene expression changes in hepatocarcinogenic isomers after up to 28 days of repeated dosing. Ellinger-Ziegelbauer et al. (2008) constructed support vector machine (SVM) prediction models by using gene expression data from rats treated for up to 14 days with 13 different chemicals for the training set and rats treated with 16 independent chemicals for the validation set; the resulting prediction models differentiated between genotoxic hepatocarcinogens, non-genotoxic hepatocarcinogens and non-hepatocarcinogens with up to 88% classification accuracy. Fielden et al. (2007) analyzed hepatic gene expression in rats treated with 25 non-genotoxic hepatocarcinogens and 75 non-hepatocarcinogens for 1 to 7 days. They constructed a SVM model consisting of 37 probes that yielded prediction accuracies for carcinogenicity with 86% specificity and 81% sensitivity. These researchers successfully constructed gene-based prediction models from data obtained from up to 1 month of repeated dosing. In contrast, Auerbach et al. (2010) recently demonstrated further evidence that the dosing period is an important factor in the construction of highly accurate prediction models based on toxicogenomics-derived gene expression profiles, especially in the case of weakly carcinogenic compounds. They concluded that a 90-day exposure period is needed to detect gene expression changes specifically related to carcinogenic compounds.

In Japan, the Toxicogenomics Project (TGP) has established a large-scale toxicogenomics database known as TG-GATEs (Genomics-Assisted Toxicity Evaluation System developed by the Toxicogenomics Project in Japan) (Uehara et al., 2010; Urushidani, 2010). In this project, rats were exposed to 3 different doses of 150 compounds for a period ranging from 1 to 28 days; the gene expression in the livers and kidneys of these animals at 8 different time points was comprehensively analyzed using microarrays. We have used this database to identify several different types of biomarker genes and construct prediction models for hepatotoxicity and nephrotoxicity (Gao et al., 2010; Hirode et al., 2008; Kondo et al., 2009; Uehara et al., 2010). Regarding hepatocarcinogenicity, we previously tried to build a gene-based predictor of non-genotoxic hepatocarcinogenicity in rats (Uehara et al., 2008). Consequently, we have successfully built a model, consisting of 112 probes, for the early detection of hepatocarcinogenesis based on gene expression changes that are commonly induced by compounds with hepatocarcinogenicity in rats. However, since this model was trained to achieve early and sensitive detection of potential carcinogenicity after a single exposure to a compound, false positive predictions occurred in some non-carcinogenic hepatotoxins. Moreover, a limited number of compounds were used for training of the model in the study since our database was under construction. In an effort to make high-quality predictive models based on gene expression data, a fairly extensive data set of several compounds with multiple time points and multiple dose levels is required. Now, our large-scale toxicogenomics database has been completed, and microarray data for all 150 compounds are available. In this research, we hypothesized that our large-scale toxicogenomics database might lead to the construction of a more robust and accurate prediction model of hepatocarcinogenicity in rats. By taking into account the findings of the latest work by Auerbach et al. (2010), we have trained a classifier by using data from our longest dosing period (28 days) to decrease the percentage of false-positive predictions. Consequently, our new SVM-based classifier yielded prediction accuracies for hepatocarcinogenicity with 99% sensitivity and 97% specificity in a training data set, and false-positive predictions were almost completely eliminated. The usefulness and robustness of our predictor were further confirmed in an independent validation data set obtained from a public database. Interestingly, similar positive predictions were obtained for several genotoxic hepatocarcinogens as well as non-genotoxic hepatocarcinogens. In the present report, we provide reliable candidate gene biomarkers in the early stages of the hepatocarcinogenesis that are predictive for both genotoxic and non-genotoxic hepatocarcinogens. Our present toxicogenomic model might be useful to reduce the dependence on 2-year rodent bioassays by instead using a short-term repeated dosing study.

Materials and methods

Animals and experimental design. Five-week-old male Sprague–Dawley rats were obtained from Charles River Japan, Inc. (Kanagawa, Japan). After a 7-day quarantine and acclimatization period, the 6-week-old animals were assigned to dosage groups (5 rats per group) by using a computerized stratified random grouping method based on individual body weight. The animals were individually housed in stainless-steel cages in an animal room that was lighted for 12 h (7:00–19:00) daily, ventilated with an air-exchange rate of 15 times per hour, and maintained at 21 °C–25 °C with a relative humidity of 40%–70%. Each animal was allowed free access to water and a pellet diet (CRF-1, sterilized by radiation, Oriental Yeast Co., Ltd., Tokyo, Japan).

The compounds used in this study are summarized in Table 1 (for detailed experimental conditions, see Supplemental Table 1). A total of 150 compounds were used for training and testing of models. The training data set consisted of 6 positive compounds (necrogenic hepatocarcinogens with no evidence of genotoxicity, namely non-genotoxic hepatocarcinogens) and 54 negative compounds (non-hepatocarcinogens), and the test data set consisted of remaining 90 compounds (for more detailed information, see Supplemental Table 2). According to the standard protocol in our project (Uehara et al., 2010), 5 rats per group were treated with these compounds at 3 different dose levels (low: L, middle: M, and high: H). The maximum tolerated dose of each compound, which was estimated from a preliminary 7-day repeated dosing study, was chosen as the highest dose level. For single-dose studies, rats were euthanized at 3, 6, 9 and 24 h after dosing. For repeated dose studies, the animals were treated daily for 3, 7, 14 and 28 days and euthanized 24 h after the last dosing (4, 8, 15 and 29D). The animals were euthanized by exsanguination from the abdominal aorta under ether anesthesia, and liver samples were collected from the left lateral lobe of the liver immediately after the animals were euthanized. The experimental protocols were carefully reviewed and approved by the Ethics Review Committee for Animal Experimentation of the National Institute of Health Sciences.

RNA extraction and microarray analysis. An aliquot of the sample (about 30 mg) for microarray analysis was obtained from the left lateral lobe of the liver in each animal immediately after the animals were euthanized. The sample was kept in RNAlater® (Ambion, Austin, TX, USA) overnight at 4 °C and then frozen at −80 °C until use. Liver samples were homogenized with buffer RLT supplied in the RNeasy Mini Kit (Qiagen, Valencia, CA, USA), and total RNA was isolated according to the manufacturer's instructions.

Microarray analysis was conducted on 3 of 5 samples for each group by using Affymetrix Rat Genome 230 2.0 arrays (Affymetrix, Santa Clara, CA, USA). The procedure was basically conducted according to the manufacturer's instructions as previously reported (Uehara et al., 2010). The digital image files were preprocessed by Affymetrix Microarray Analysis Suite version 5.0 (MAS5.0). The expression signal values were scaled by the median of each chip sample. The normalized data sets were then converted into the log-ratio of base 2 to the means of the

Table 1
Compounds and carcinogenicity definitions used in this study.

Compound class	Compound name
Training set (positive)	
Non-genotoxic hepatocarcinogen (hepatotoxic oxidative stressor)	Carbon tetrachloride (CCL4), ethionine (ET), thioacetamide (TAA), methapyrilene (MP), coumarin (CMA), monocrotaline (MCT)
Training set (negative)	
Non-hepatocarcinogen	Acetaminophen (APAP), naphthyl isothiocyanate (ANIT), allyl alcohol (AA), theophylline (TEO), trimethadione (TMD), naproxen (NPX), methotrexate (MTX), aspirin (ASA), labetalol (LBT), ketoconazole (KC), tetracycline (TC), metformin (MFM), methyldopa (MDP), vitamin A (VA), chlorpropamide (CPP), nicotinic acid (NIC), famotidine (FAM), ranitidine (RAN), diltiazem (DIL), captopril (CAP), enalapril (ENA), mexiletine (MEX), meloxicam (MLX), lornoxicam (LNX), cyclosporine A (CSA), isoniazid (INAH), phenylbutazone (PhB), nitrofurantoin (NFT), propylthiouracil (PTU), amiodarone (AM), cimetidine (CIM), flutamide (FT), methimazole (MYZ), iproniazid (IPA), chloramphenicol (CMP), furosemide (FUR), chlorpheniramine (CHL), caffeine (CAF), sulpiride (SLP), simvastatin (SST), chlormadinone (CLM), carboplatin (CBP), bucetin (BCT), perhexiline (PH), pemoline (PML), ibuprofen (IBU), erythromycin ethylsuccinate (EME), nifedipine (NIF), sulindac (SUL), disopyramide (DIS), disulfiram (DSF), tolbutamide (TLB), acarbose (ACA), ajmaline (AJM)
Test set	
Genotoxic hepatocarcinogen	Lomustine (LS), acetamidofluorene (AAF), nitrosodiethylamine (DEN)
Non-genotoxic hepatocarcinogen (enzyme inducer)	Phenobarbital (PB), carbamazepine (CBZ), phenytoin (PHE), rifampicin (RIF), hexachlorobenzene (HCB), sulfasalazine (SS)
Non-genotoxic hepatocarcinogen (peroxisome proliferator)	Clofibrate (CFB), WY-14643 (WY), gemfibrozil (GFZ), fenofibrate (FFB)
Non-genotoxic hepatocarcinogen (hormonal modulator)	Ethinylestradiol (EE)
Non-hepatocarcinogen/unknown (non-genotoxicant)	Ethionamide (ETH), indomethacin (IM), bromobenzene (BBZ), ethambutol (EBU), colchicine (COL), clomipramine (CPM), puromycin aminonucleoside (PAN), methyltestosterone (MTS), valproic acid (VPA), chlorpromazine (CPZ), diclofenac (DFNa), benzbromarone (BBr), allopurinol (APL), fluphenazine (FP), thioridazine (TRZ), adapin (ADP), glibenclamide (GBC), chlormezanone (CMN), moxisylyte (MXS), imipramine (IMI), amitriptyline (AMT), hydroxyzine (HYZ), quinidine (QND), mefenamic acid (MEF), tiopronin (TIO), acetazolamide (ACZ), promethazine (PMZ), dantrolene (DTL), triazolam (TZM), terbinafine (TBF), danazol (DNZ), bendazac (BDZ), benziodarone (BZD), bromoethanamine (BEA), nimesulide (NIM), phenylanthranilic acid (NPAA), cephalothin (CLT), ticlopidine (TCP), gentamicin (GMC), vancomycin (VMC), omeprazole (OPZ), diazepam (DZP), haloperidol (HPL), griseofulvin (GF), tamoxifen (TMX), tannic acid (TAN), triamterene (TRI), ethanol (ETN), ciprofloxacin (CPX), tacrine (TAC), nitrofurazone (NFZ), papaverine (PAP), penicillamine (PEN), azathioprine (AZP), doxorubicin (DOX), cyclophosphamide (CPA), etoposide (ETP), cisplatin (CSP), phenacetin (PCT)
Unknown	K01, K02, K03, K04, K05, K06, K07, K08, K09, K10, K11, K12, K13, K14, K15, K16, K17

K01 to K17 were compounds synthesized in member companies.

corresponding control groups. Raw microarray data (CEL files) are available in Open TG-GATEs (http://toxico.nibio.go.jp/).

Gene selection and supervised classification. Among a total of 150 compounds in our database, we have selected 6 compounds consisting of carbon tetrachloride, ethionine, thioacetamide, methapyrilene, coumarin and monocrotaline, as positive training compounds for modeling, which are non-genotoxic hepatocarcinogens with hepatocellular necrosis/degeneration in histopathology following multiple dosing for up to 28 days in our experimental condition. Individual histopathological data of all compounds are available (http://toxico.nibio.go.jp/datalist.html). High-dose 29D groups treated with these compounds were used for the positive training data set, with the exception for monocrotaline; the middle-dose group of monocrotaline at 29D was included in the positive training set because all animals in the high-dose group were dead at 29D in the current experimental condition. High- or middle-dose groups at all time points (3 to 24H for single-dose studies, 4 to 29D for repeated-dose studies) of randomly selected 54 non-hepatocarcinogens were selected as the negative training data set. To exclude genes being transiently regulated by the treatment of non-carcinogenic compounds, data obtained from all time points were used as the negative training set. The remaining compounds were used as the independent test set as follows: (i) genotoxic hepatocarcinogens; (ii) non-genotoxic hepatocarcinogens whose carcinogenic mechanisms are thought to be related to hepatic enzyme induction, peroxisomal proliferation and hormonal modulation; and (iii) non-hepatocarcinogens (for more detailed information, see Supplemental Table 2).

SVM combined with wrapper-type gene selection algorithms was used to build a prediction model, as previously reported (Kondo et al., 2009). First, the probes that were judged as absent in all samples of the training data set using MAS5.0 P/A-call were excluded. Next, the following 3 statistical parameters were calculated for each probe: (i) fold change of gene expression between positive and negative training data sets; (ii) Mann–Whitney U value; (iii) confident margin of SVM classifier (normal margin corrected by classification accuracy). The probes were filtered by following criteria: (i) fold change >2 or <0.5; (ii) p-value <0.01; and (iii) confident margin >0.05. Then, the combined gene ranking based on the 3 parameters was calculated by using a layer ranking algorithm (Chen et al., 2007). To estimate the performance of our classifier, 5-fold cross-validation was executed as described in our previous report (Kondo et al., 2009). Finally, 9 of the 82 top-ranked probes (7 genes) were selected to maximize the classification accuracy and the area under the curve (AUC) of the receiver operating characteristic curve (ROC).

Ingenuity pathways analysis. The 82 top-ranked probes were analyzed by using Ingenuity Pathways Analysis software (v. 7.1; Ingenuity Systems, Redwood City, CA) to determine the biological networks that were enriched in selected feature genes.

Independent validation of our classifier by using an external data set from NEDO. An external microarray data set from the NEDO project, another toxicogenomics consortium in Japan, was used for independent validation of our classifier (Matsumoto et al., 2009). In this project, the NEDO-ToxArray III consists of 6709 genes, and hepatic gene expression data was comprehensively obtained from F344 rats treated with 88 compounds for up to 28 days. All of the microarray data are available in the public microarray database of Gene Expression Omnibus (GEO). SVM modeling and principal component analysis (PCA) were performed by using the expression data at 3 different time points (4, 15 and 29D).

Predictions using published biomarker genes. For comparison of prediction accuracy with previously published models, we built

SVM models with our training data set by using published gene lists and then compared the prediction performance of all models (Auerbach et al., 2010; Ellinger-Ziegelbauer et al., 2008; Fielden et al., 2007; Nakayama et al., 2006; Uehara et al., 2008). To estimate the performance of each model, 5-fold cross-validation was executed using the training data set.

Results

Gene selection and supervised classification. We trained a binary classifier by using an SVM algorithm combined with wrapper-type gene selection to construct a statistically reliable model without over-fitting to the profiles of training samples according to our previous report with minor modifications (Kondo et al., 2009). By applying statistical analysis for feature selection, 82 probes passed current statistical criteria, and these top-ranked probes are summarized in Supplemental Table 3. 1 to 82 of the top-ranked probes were used to construct the classifiers with further feature selection. A ROC curve and its AUC are plotted in Fig. 1. Although there were no big differences in the prediction accuracy of each classifier, a classifier consisting of 9 probes (7 genes; Table 2) achieved the best classification accuracy under the 5-fold cross-validation. The sensitivity and specificity of the prediction of the optimized classifier was 99% and 97%, respectively.

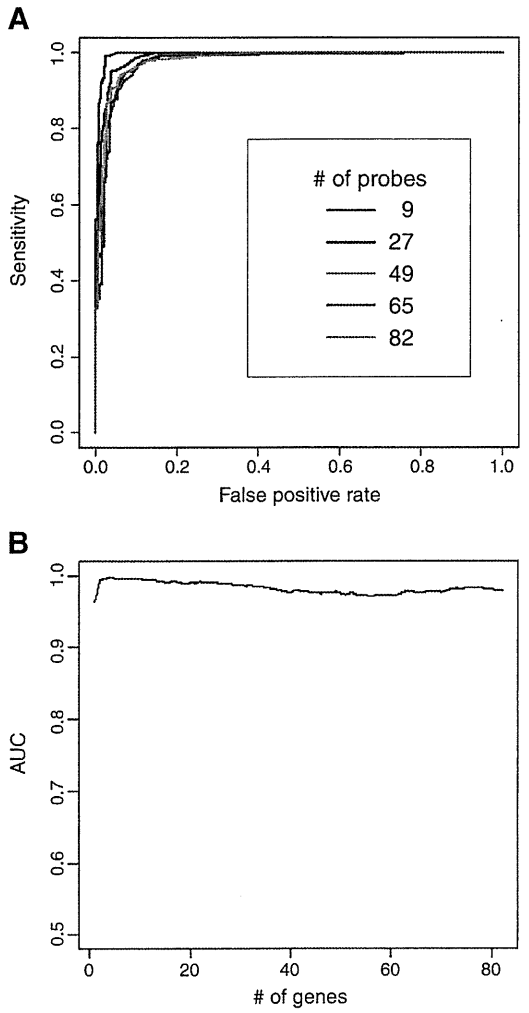


Fig. 1. Receiver operating characteristic analysis of prediction models. ROC curve (A) and area under the ROC curve (B) of prediction models are plotted. A model consisting of 9 probes was selected as the best model in our training data set.

Table 2
The 9 top-ranked probes in the optimized model.

Affymetrix probe ID	Gene title	Gene abbreviation	GO biological process (selected)
1370583_s_at	ATP-binding cassette, sub-family B (MDR/TAP), member 1A/1B	*Abcb1a/1b*	Nucleotide binding/transporter activity/protein binding/ATP binding/ATPase activity/drug transporter activity/hydrolase activity/nucleoside-triphosphatase activity
1395737_at /1374198_at	Cd276 molecule	*Cd276*	Receptor binding/protein binding
1367787_at	Islet cell autoantigen 1	*Ica1*	Protein binding/protein domain specific binding
1379419_at	Transmembrane protein 184C	*Tmem184c*	Unknown
1379262_at	Acyl-CoA thioesterase 9	*Acot9*	Unknown
1383401_at	Testis derived transcript	*Tes*	Zinc ion binding/metal ion binding
1375719_s_at/1373102_at	Cadherin 13	*Cdh13*	Calcium ion binding/protein binding/low-density lipoprotein binding/protein homodimerization activity/cadherin binding/adiponectin binding

Gene expression profiles of selected feature genes. Fig. 2 shows a heat map of the selected 9 probes and all 82 top-ranked probes. Among the 82 probes, 44 probes were upregulated, and 38 probes were downregulated in positive training compounds. All 9 probes involved in the optimized model were upregulated in positive training compounds. The extent of upregulation of genes in positive compounds was clearly higher than that in negative compounds and changed in a time- and dose-dependent manner (Fig. 3).

Ingenuity pathways analysis. To further characterize the biological significance of alterations in gene expression, functional pathway analysis was performed by using the 82 selected probes. The mitogen-activated protein kinase p38 (*p38 Mapk*)- and phosphatidylinositol-3-kinase (*Pl3k*)-centered interactome (Fig. 4A) and the v-myc myelocytomatosis viral oncogene homolog (*Myc*)-centered interactome (Fig. 4B) were the 2 most significant networks. Among 9 probes (7 genes) of the best classifier, the following 5 genes were involved in these networks: ATP-binding cassette, sub-family B (MDR/TAP), member 1A/B (*Abcb1a/b*), Cd276 molecule (*Cd276*), islet cell autoantigen 1 (*Ica1*), testis-derived transcript (*Tes*), and cadherin 13 (*Cdh13*).

Prediction results of all compounds. The SVM classification scores of all 150 compounds are summarized in Supplemental Table 2. All 3 dose groups of 90 test compounds that had not been used as the training set and the remaining groups of the 60 training compounds were used as a test data set for external validation of the classifier. The classifier predicted the following samples as positive: thioacetamide (H: 8D and 15D; M: 29D), methapyrilene (H: 8D and 15D), carbon tetrachloride (M: 29D) and monocrotaline (H: 15D). As expected, positive predictions for several hepatocarcinogens were observed only after long-term repeated dosing. There were no positive predictions in the low-dose groups of these positive-training compounds. Among the

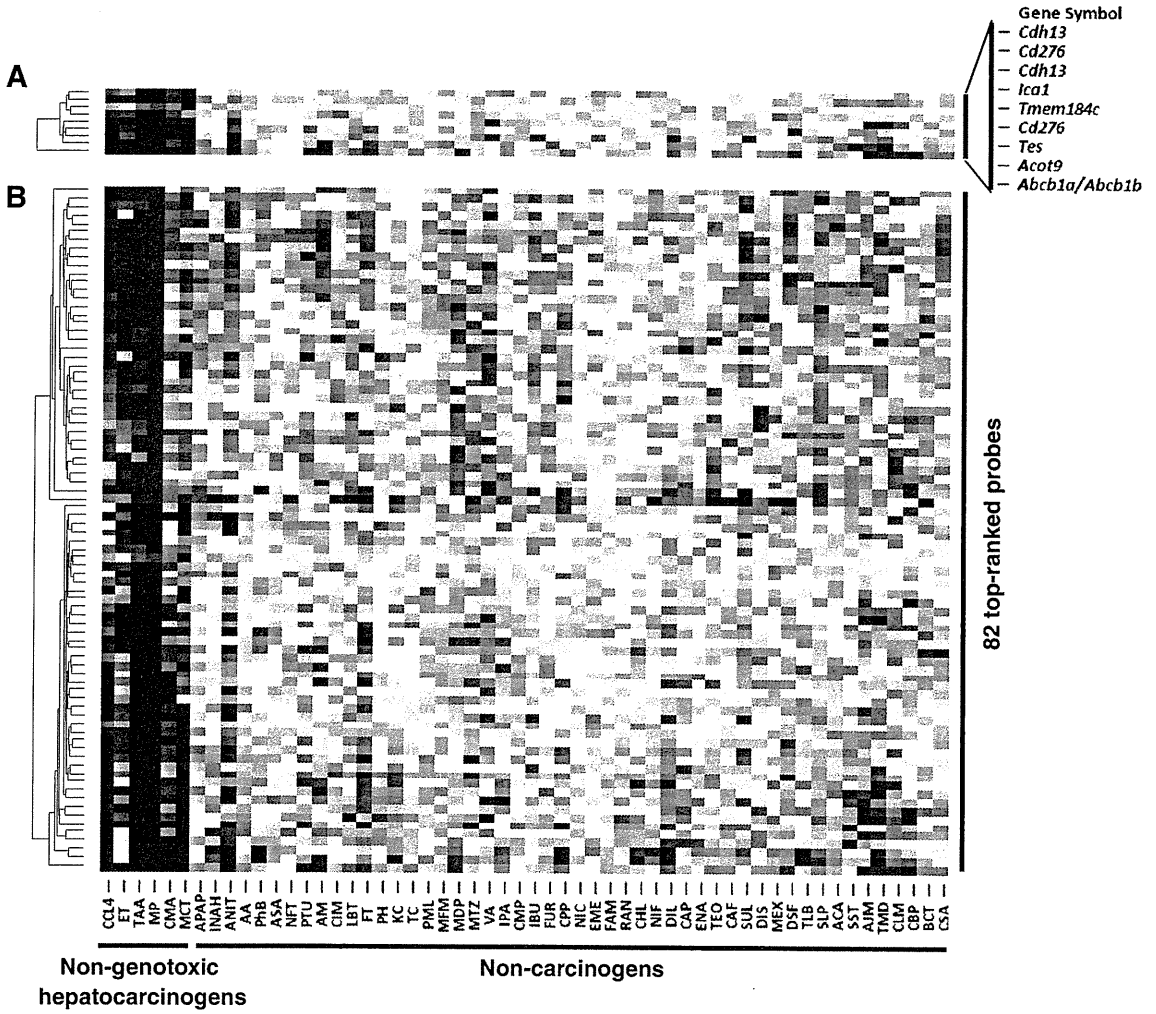


Fig. 2. Hierarchical clustering analysis of expression changes of selected feature genes. Gene expression changes of the 9 (A) and 82 top-ranked probes (B) were analyzed by hierarchical clustering. Clustering method: UPGMA (unweighted average); similarity measure: Euclidean distance; ordering function: average value. Heat map shows log-ratio of base 2 to the means of the corresponding control groups (red: 5-fold higher, blue: -5-fold lower expression of log 2 ratio). Symbols of selected genes are shown on the right side of the heat map.

genotoxins, lomustine (H: 29D), acetamidofluorene (L, M and H: 15 and 29D) and nitrosodiethylamine (M: 29D and H: 8 and 15D), for which the target of carcinogenicity is the liver, were correctly predicted as positive. Several genotoxic compounds that have potential to elicit cancer in other target organs, such as cyclophosphamide, etoposide, cisplatin and carboplatin, colchicine, phenacetin and doxorubicin, were predicted as negative at all time points of the 3 dose groups. Non-genotoxic hepatocarcinogens, including enzyme inducers (phenobarbital, carbamazepine, phenytoin, rifampicin, hexachlorobenzene and sulfasalazine), peroxisome proliferators (clofibrate, Wy-14643, gemfibrozil and fenofibrate) and hormonal modulators (ethinylestradiol), were classified as negative in all groups. The priority focus on this study was to build a prediction model with improved false-positive classification of non-hepatocarcinogens reflecting temporal gene expression changes at early time points after single or short-term repeated dosing. Among the 46 compounds in the negative/unknown test set, almost all compounds were classified as negative except for 2 compounds that were falsely classified as positive, ethionamide (H: 24H, 4D, and 15D; M: 24H) and etambutol (H: 8D).

Further independent validation of the optimized predictor. Since interlaboratory difference is an important issue in the field of toxicogenomics research (Fielden et al., 2008), a further independent validation was performed by using microarray data obtained by independent laboratories. In the present study, we used an external data set obtained in the NEDO project for further validation (Matsumoto et al., 2009). The NEDO study used custom microarrays consisting of 6709 probes, and gene expression data was comprehensively obtained from the livers of rats treated with several compounds for up to 28 days. Among their data set, we used following 14 compounds commonly involved in our data set: carbon tetrachloride, ethionine, thioacetamide, methapyrilene and nitrosodiethylamine (positive test set), and phenobarbital, phenytoin, hexachlorobenzene, clofibrate, ethinylestradiol, indomethacin, acetaminophen, aspirin and tannic acid (negative test set). For the purpose of a comprehensive comparison, we used 82 top-ranked probes for this analysis. Due to differences in the microarray platform, only 53 out of 82 probes were shared by both microarray platforms. By using the gene expression data for 53 probes, a SVM classifier was built without further feature selection using our training data set and then analyzed prediction accuracy of these test chemicals. As a result, all hepatocarcinogens included in the test data set were correctly predicted as positive regardless of using microarray data obtained by the different platform in independent laboratories. The overall sensitivity and specificity of the prediction by this classifier consisting of 53 probes was 100% and 89%, respectively (Supplemental Table 4).

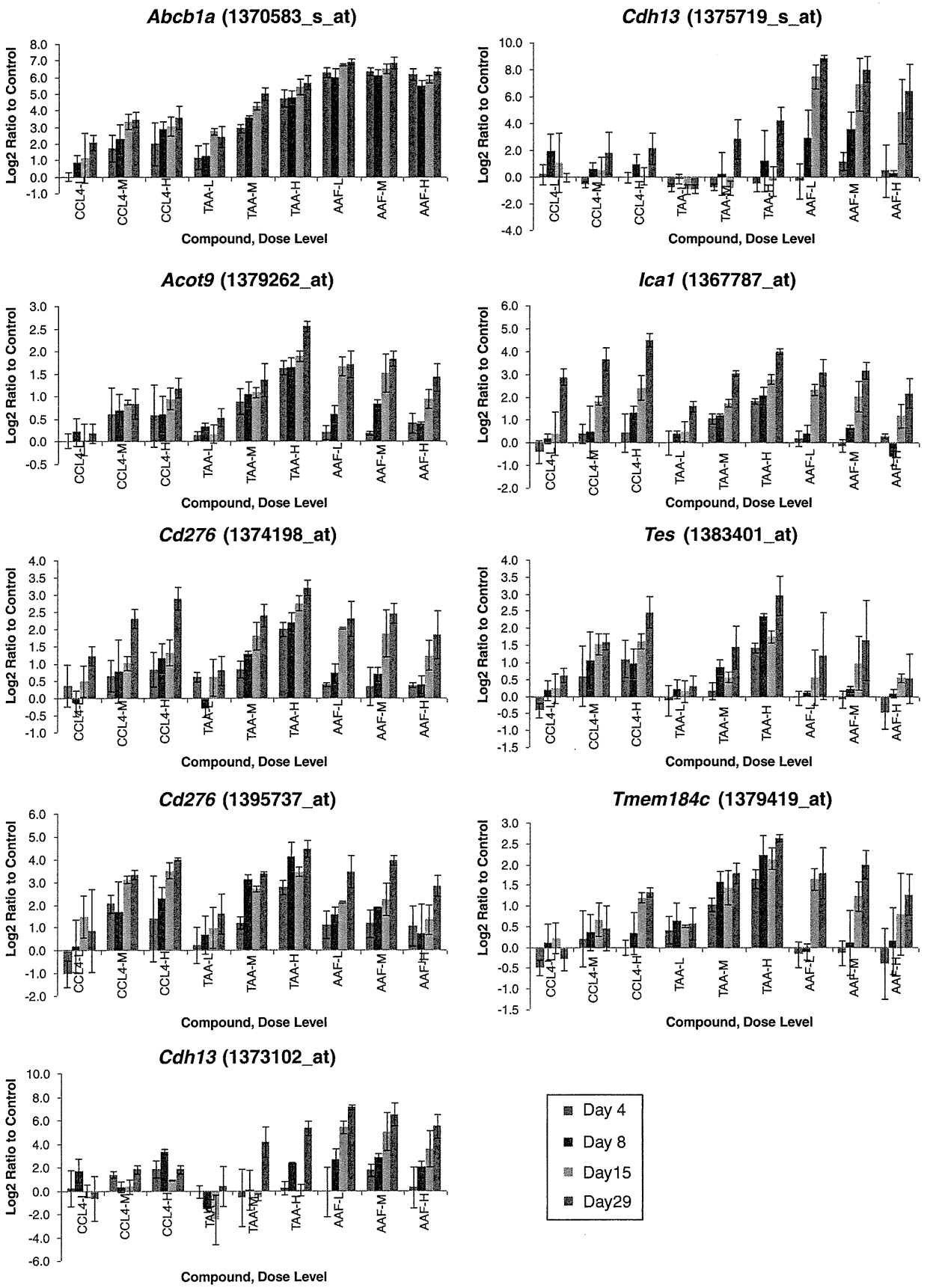


Fig. 3. Effect of dose and exposure duration on the expression of candidate biomarker genes for hepatocarcinogenesis. Log-ratio of base 2 to the means of the corresponding control groups with standard deviation are shown in 3 representative hepatocarcinogens, carbon tetrachloride, thioacetamide and acetamidofluorene.

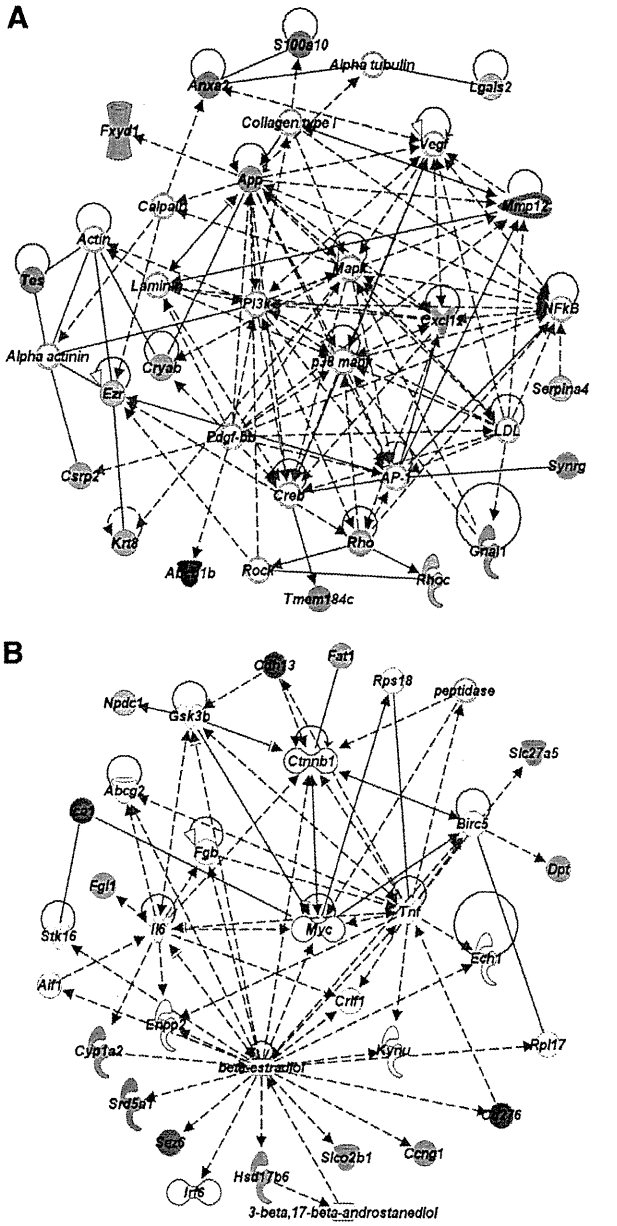


Fig. 4. Molecular networks representing selected feature genes. *p38 Mapk-* (A) and *Myc*-centered interactomes (B) were selected as significant network in the selected 82 genes. Red and green represent molecules upregulated or downregulated in positive compounds compared to negative compounds, respectively. Molecules incorporated into the network are shown in white. Ellipse, square, triangle, trapezoid, lozenge and circle represent transcription regulator, cytokine, kinase, transporter, enzyme and other molecules, respectively. Arrows connecting molecules indicate one molecule acts on another, and lines indicate one molecule binds to another. Dashed arrows or lines indicate indirect interactions of 2 molecules.

Furthermore, unsupervised analysis, PCA, was also performed based on the expression profiles of 53 probes in all of the NEDO data set. The scores of the first and second principal components (PC1 and PC2) at 4, 15 and 29D were separately plotted in Fig. 5. Several genotoxic hepatocarcinogens, such as nitrosodiethylamine, N-nitrosomorpholine, 2-nitropropane, furan and N-nitrosopiperidine, which show positive findings in in vitro genotoxicity assays and in vivo carcinogenicity assays in rats, were separately plotted from non-hepatocarcinogens toward the PC1 direction. In addition, non-genotoxic hepatocarcinogens, methapyrilene, thioacetamide and carbon tetrachloride, which are also included in our data set, showed clear separations in the PC1 direction. In almost all of these compounds, a distinct separation was observed in the data set after long-term repeated dosing.

Predictions using published biomarker genes. ROC curves of all models are plotted in Fig. 6. Although there was no large overlap of genes among the models, all models showed high prediction performance in our data set. Among all models, our new model achieved the best classification accuracy under the 5-fold cross-validation (Table 3).

Discussion

The goal of the present study was to develop a robust toxicogenomic model for the early assessment of potential hepatocarcinogenicity of chemicals based on gene expression changes stored in our toxicogenomics database, TG-GATEs. Carcinogenesis occurs via multiple mechanisms. Based on the properties of DNA damage, carcinogens are classified as genotoxins or non-genotoxins. Genotoxic carcinogens induce direct DNA damage, that is detectable by in vitro genotoxicity assays (e.g., Ames test and chromosomal aberration assays) and in vivo micronucleus test. Non-genotoxic carcinogens act through various modes of action that do not involve direct DNA damage (e.g., hepatocellular necrosis and regenerative proliferation, xenobiotic receptor agonists, peroxisome proliferation, or hormonal-mediated processes). Although it would be useful to detect the potential hepatocarcinogenicity of these different classes of non-genotoxic hepatocarcinogens by a single model, it is generally believed that a mechanism-based approach would be a promising strategy for robust toxicogenomic modeling since different classes of compounds generally show different gene expression profiles. In the present study, we have tried to establish reliable candidate gene biomarkers to assess the potential risks of hepatocarcinogenicity in the early stage of drug development with a particular focus on liver necrogenic compounds. Consequently, we have successfully identified robust biomarker genes specifically upregulated in necrogenic compounds with hepatocarcinogenicity. In contrast, no positive predictions were observed in the other class of non-genotoxic hepatocarcinogens involved in hepatic enzyme inducers, peroxisome proliferators and hormonal modulators. Therefore, these characteristics of prediction profile might help to elucidate the mechanisms of action involved in non-genotoxic hepatocarcinogenesis.

In contrast to our previous classifier (Uehara et al., 2008), our current model successfully achieved robust detection of non-genotoxic hepatocarcinogenicity of compounds by using hepatic gene expression data after 28 days of repeated dosing. There were some differences in the prediction properties of these 2 models. Namely, the current model enables us to detect robust gene expression changes possibly related to hepatocarcinogenic process following chronic doses in contrast to the previous model, which is useful to detect early temporal signals after a single dose as well as repeated doses; therefore, the combined use of both models for comprehensive judgment is thought to be the best strategy for sensitive and robust detection of hepatocarcinogenicity in short-term repeated dosing studies.

Interestingly, our current classifier as well as the previous classifier (data not shown) provided positive prediction results for not only non-genotoxic hepatocarcinogens but genotoxic hepatocarcinogens as well. This observation was further confirmed by the external test data set from the NEDO project (Matsumoto et al., 2009). It is generally believed that genotoxic compounds directly induce DNA damage by themselves or their reactive metabolites and that sufficient initiation is achieved after a single dose or short-term repeated doses due to the strong potency of their genotoxic stimulus. In contrast, non-genotoxic carcinogens generally require repeated dosing for sufficient initiation. Although the mechanisms involved in hepatocarcinogenic process are not the same between genotoxic and non-genotoxic hepatocarcinogens, our selected genes showed similar

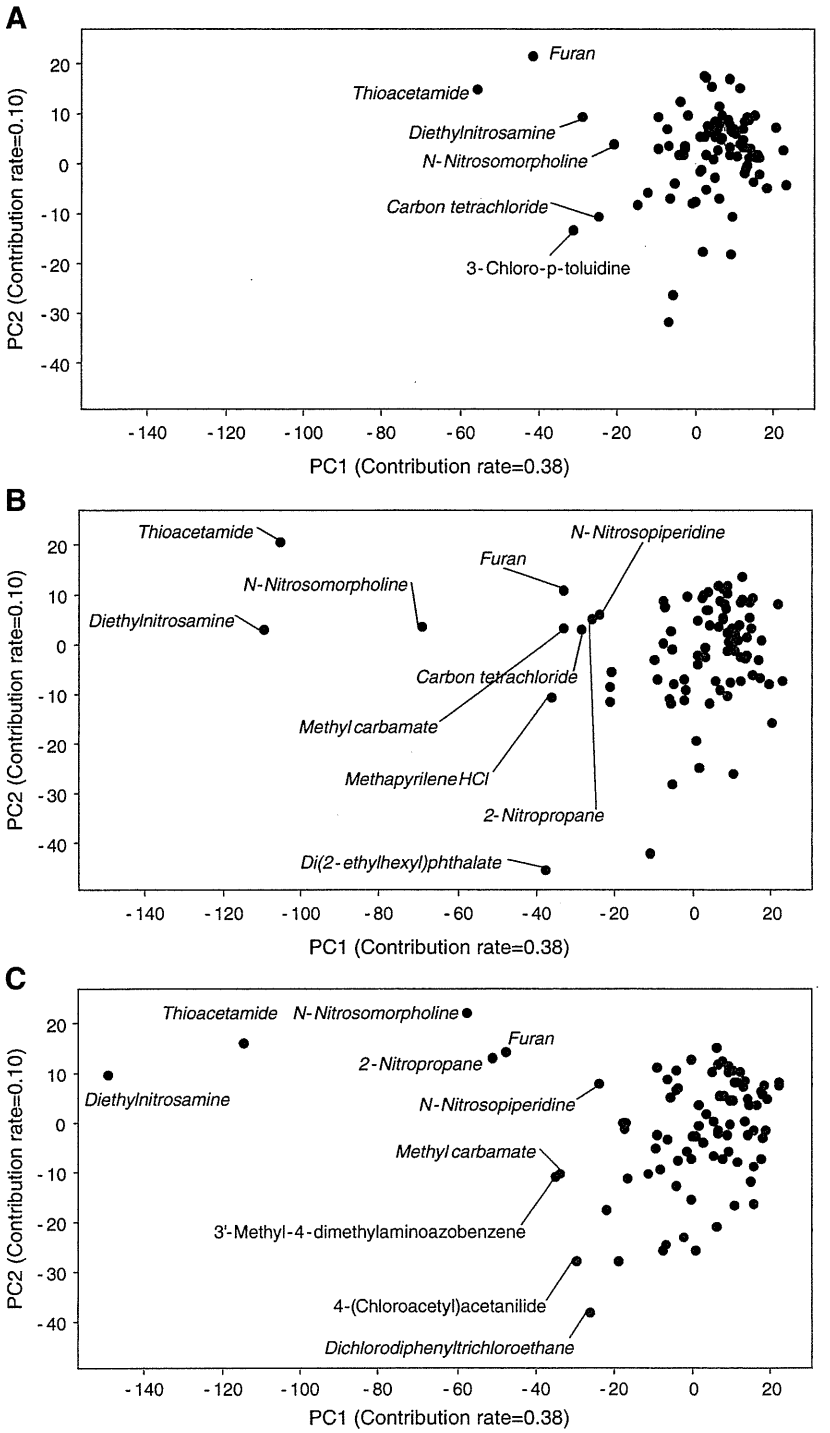


Fig. 5. Principal component analysis in independent validation. An external microarray data set obtained from the NEDO project was used for independent validation of our model. PCA was performed on all compounds by using the expression data at 3 different time points of 4 (A), 15 (B) and 29D (C). Italic font indicates hepatocarcinogens. Several hepatocarcinogens were separated from other chemicals including non-carcinogens toward the direction of PC1.

expression profiles after repeated dosing. Our results indicate that the expression profiles of our newly selected candidate biomarker genes might be common characteristics in the early stage of hepatocarcinogenic process, regardless of the type of carcinogens.

Among the test compounds, ethionamide and etambutol showed positive predictions in the current model, although there is no direct evidence supporting the hepatocarcinogenicity of these compounds in rodents. In our experiments, repeated doses of ethionamide caused anisokaryosis of centrilobular hepatocytes. In the etambutol-treated liver, karyomegaly was observed in hepatocytes with distinct nucleoli. Although it is hard to conclude that these morphological changes in the nucleus of hepatocytes are directly connected to hepatocarcinogenicity, these morphological changes might be early indicative changes of hepatocarcinogenesis and reflect nuclear DNA damage caused by these compounds. Thus, gene expression changes after further long-term repeated dosing as well as a 2-year bioassay study would be of particular

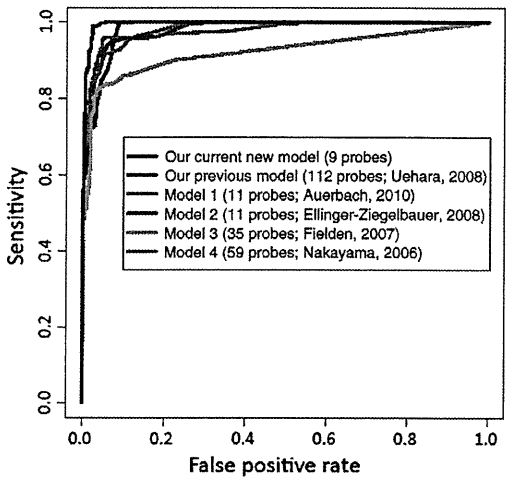


Fig. 6. Receiver operating characteristic analysis of previously published prediction models. ROC curves of prediction models are plotted. Our current new model: prediction results obtained by the best model consisting of 9 probes are shown. Model 1: a model consisting of 11 probes in the 90-day optimal-model, which showed the best prediction performance in our data set among their models, was used for this analysis. Model 2: among their models, commonly selected 101 probes were used for modeling. Models 3 and 4: 35 and 56 probes were used for modeling. Lists of genes used for each modeling are shown in Supplemental Table 5.

interest to determine the carcinogenic potentials of these compounds as well as to confirm the biological significance of these gene expression changes.

This study also provided the novel biological information that modulation of *p38 Mapk*- and *Myc*-centered networks is a common characteristic of gene expression in the early stage of hepatocarcinogenesis. There is increasing evidence that *p38 Mapk*, which is a stress-activated kinase, also participates in cell-cycle regulation, functioning as a suppressor of cell proliferation and tumorigenesis (Campbell et al., 2011). *p38* plays essential roles in modulating chronic inflammation-related cytokine expression, such as tumor necrosis factor α, interleukin 6 and 12, which might act as promoters of cancer growth and progression (Karin et al., 2006; Kumar et al., 2003; Naugler et al., 2007). *p38alpha*-deficient mice are susceptible to carcinogens, and *p38alpha* negatively regulates cell proliferation by antagonizing the *JNK-c-Jun* pathway in multiple cell types and in liver cancer development (Hui et al., 2007a, 2007b). Therefore, upregulation of genes involved in the *p38 Mapk*-network might be a tumor-suppressive action against DNA damage stimulus in the liver following repeated doses of carcinogens. Furthermore, these genes might be useful indicators in the early stage of hepatocarcinogenesis. The *Myc* oncogene is a transcription factor that plays an important role in the pathogenesis of hepatocellular carcinoma (Calvisi and Thorgeirsson, 2005; Thorgeirsson and Grisham, 2002). *Myc* overexpression is believed to exert its neoplastic function via several mechanisms, such as inducing autonomous cell proliferation and growth, blocking differentiation, and causing genomic destabilization (Dang, 1999; Felsher and Bishop, 1999; Grandori et al., 2000; Oster et al., 2002; Pelengaris et al., 2002). While the detailed biological significance should be determined in further studies, several genes were up or downregulated in this network. Among the genes involved in these networks, there is increasing evidence supporting their involvement in carcinogenesis. *Cdh13* (T-cadherin), an atypical member of the cadherin family, is thought to affect cellular biological processes largely via its signaling properties. It is often downregulated in cancerous cells, which is associated with a poor prognosis in various human carcinomas (Andreeva and Kutuzov, 2010). It is also reported that *Cdh13* re-expression in cancer cell lines inhibits cell proliferation and invasiveness, increases susceptibility to apoptosis, and reduces tumor growth (Andreeva and Kutuzov, 2010). Approximately 50% of human hepatocellular carcinomas overexpress *Cdh13* (Riou et al., 2006). While *Cdh13* expression is restricted to endothelial cells from large blood vessels in normal livers, it is upregulated in sinusoidal endothelial cells in invasive hepatocellular carcinomas (Riou et al., 2006). B7-H3 (*Cd276* protein), a surface immunomodulatory glycoprotein, inhibits the functions of natural killer cells and T cells. Whereas the B7-H3 transcript is ubiquitously expressed in various types of human solid tumors as well as normal tissues, the B7-H3 protein is preferentially expressed only in tumor tissues (Xu et al., 2009). While additional biological studies would be required for further confirmation of the biological significance of the modulation of genes involved in these networks, these gene expression changes might play pivotal roles in the early stage of hepatocarcinogenesis.

It is well-known that carcinogenicity is dependent on total exposure levels of compounds; therefore, we have compared total exposure levels of compounds between our experimental conditions and the 2-year bioassay for a better understanding of the toxicological significance of positive results by our prediction model. Due to the abundant literature on carcinogenesis studies, we selected 2 representative hepatocarcinogens, nitrosodiethylamine and acetamidofluorene, for detailed discussions. In our experimental condition, positive predictions for the nitrosodiethylamine and acetamidofluorene groups were observed at cumulative total doses of 240 to 450 mg/kg for nitrosodiethylamine and 420 to 8400 mg/kg for acetamidofluorene. Williams et al. (2004) performed a series of dose-response investigations with these 2 compounds to characterize differences in hepatocarcinogenic effects with respect to exposures. In their two-stage hepatocarcinogenesis study of rats, nitrosodiethylamine was weekly dosed at 40.9 mg/kg for 10 weeks, and a cumulative nitrosodiethylamine dose of 409 mg/kg yielded an 80% hepatocellular tumor incidence followed by 4 weeks of recovery and 24 weeks of promotion with phenobarbital. In contrast, another study demonstrated that a cumulative dose of 409 mg/kg induced a 45% hepatic tumor incidence after 130 weeks of exposure in drinking water (Peto et al., 1991a, 1991b). Regarding acetamidofluorene, a 100% incidence of hepatic tumors was reported at cumulative doses of 282 mg/kg (3.4 mg/kg/day for 12 weeks) followed by 24 weeks of promotion with phenobarbital, while the dose of 1772 mg/kg (3.3 mg/kg/day for 76 weeks) was needed without the promotion (Williams et al., 1991). Although there were some differences in the susceptibility to tumor occurrence in these 2 reports, the cumulative doses of nitrosodiethylamine and acetamidofluorene in our present study were sufficiently above the cumulative dose needed for hepatocarcinogenesis in rats. Taken together, it might be reasonable to conclude that sufficient initiation effects have been achieved under our current dosing conditions of hepatocarcinogens with positive prediction. For building

Table 3
Prediction performance of previously published models.

Models	# of probes	AUC	Sensitivity	False positive rate	Specificity	Correct classification rate
Our current new model	9	0.996	0.990	0.033	0.967	0.978
Our previous model	112	0.983	0.997	0.087	0.913	0.955
Model 1 (Auerbach)	11	0.977	0.960	0.053	0.947	0.953
Model 2 (Ellinger)	101	0.980	0.937	0.057	0.943	0.940
Model 3 (Fielden)	35	0.923	0.830	0.040	0.960	0.895
Model 4(Nakayama)	56	0.982	0.917	0.053	0.947	0.932

models, we have anchored on available carcinogenicity results from 2-year rat bioassays of each compound. Our toxicogenomics project does not only focus on the carcinogenicity of compounds; therefore, SD rats were used as experimental animals in our project, while carcinogenicity tests generally use F344 rats. In addition, rats received orally or intravenously administered compound at the 1-month maximum tolerated doses. However, other dosing methods, generally in the diet or water, are also used for repeated exposure during the 2-year lifespan of rats. Therefore, differences in experimental conditions should be taken into account for a precise comparison. Since we hypothesized that expression changes in our feature genes might reflect the initiated condition of the liver following large doses of carcinogens for up to 1 month, we are conducting further biological studies by using a 2-step carcinogenicity study model in rats. Further data will be published in the near future.

In conclusion, we constructed a new toxicogenomic model for early prediction for both genotoxic and non-genotoxic hepatocarcinogens by using comprehensive gene expression data stored in our large-scale toxicogenomics database. The usefulness and robustness of our predictor were further confirmed in an independent validation data set obtained from a public database. Our toxicogenomic model might be useful for the prospective screening of hepatocarcinogenicity of compounds and prioritization of compounds for carcinogenicity testing. Furthermore, these characteristics of gene expression changes would help to elucidate the mechanisms of action involved in hepatocarcinogenesis.

Conflict of interest statement

The authors have declared no conflict of interest.

Acknowledgment

These studies were supported by a grant from the Ministry of Health, Labour and Welfare of Japan (H14-Toxico-001 and H19-Toxico-001).

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.taap.2011.07.001.

References

Andreeva, A.V., Kutuzov, M.A., 2010. Cadherin 13 in cancer. Genes Chromosomes Cancer 49, 775–790.

Auerbach, S.S., Shah, R.R., Mav, D., Smith, C.S., Walker, N.J., Vallant, M.K., Boorman, G.A., Irwin, R.D., 2010. Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning. Toxicol. Appl. Pharmacol. 243, 300–314.

Battershill, J.M., 2005. Toxicogenomics: regulatory perspective on current position. Hum. Exp. Toxicol. 24, 35–40.

Calvisi, D.F., Thorgeirsson, S.S., 2005. Molecular mechanisms of hepatocarcinogenesis in transgenic mouse models of liver cancer. Toxicol. Pathol. 33, 181–184.

Campbell, J.S., Argast, G.M., Yuen, S.Y., Hayes, B., Fausto, N., 2011. Inactivation of p38 MAPK during liver regeneration. Int. J. Biochem. Cell Biol. 43, 180–188.

Chen, J.J., Tsai, C., Tzeng, S., Chen, C., 2007. Gene selection with multiple ordering criteria. BMC Bioinformatics 8, 74.

Dang, C.V., 1999. c-Myc target genes involved in cell growth, apoptosis, and metabolism. Mol. Cell. Biol. 19, 1–11.

Ellinger-Ziegelbauer, H., Gmuender, H., Bandenburg, A., Ahr, H.J., 2008. Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. Mutat. Res. 637, 23–39.

Felsher, D.W., Bishop, J.M., 1999. Transient excess of MYC activity can elicit genomic instability and tumorigenesis. Proc. Natl. Acad. Sci. U.S.A. 96, 3940–3944.

Fielden, M.R., Brennan, R., 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.

Fielden, M.R., Nie, A., McMillian, M., Elangbam, C.S., Trela, B.A., Yang, Y., Dunn II, R.T., Dragan, Y., Fransson-Stehen, R., Bogdanffy, M., Adams, S.P., Foster, W.R., Chen, S.J., Rossi, P., Kasper, P., Jacobson-Kram, D., Tatsuoka, K.S., Wier, P.J., Gollub, J., Halbert, D.N., Roter, A., Young, J.K., Sina, J.F., Marlowe, J., Martus, H.J., Aubrecht, J., Olaharski, A.J., Roome, N., Nioi, P., Pardo, I., Snyder, R., Perry, R., Lord, P., Mattes, W., Car, B.D., Predictive Safety Testing Consortium, Carcinogenicity Working Group, 2008. Interlaboratory evaluation of genomic signatures for predicting carcinogenicity in the rat. Toxicol. Sci. 103, 28–34.

Gao, W., Mizukawa, Y., Nakatsu, N., Minowa, Y., Yamada, H., Ohno, Y., Urushidani, T., 2010. Mechanism-based biomarker gene sets for glutathione depletion-related hepatotoxicity in rats. Toxicol. Appl. Pharmacol. 247, 211–221.

Grandori, C., Cowley, S.M., James, L.P., Eisenman, R.N., 2000. The Myc/Max/Mad network and the transcriptional control of cell behavior. Annu. Rev. Cell. Dev. Biol. 16, 653–699.

Heinloth, A.N., Irwin, R.D., Boorman, G.A., Nettesheim, P., Fannin, R.D., Sieber, S.O., Snell, M.L., Tucker, C.J., Li, L., Travlos, G.S., Vansant, G., Blackshear, P.E., Tennant, R.W., Cunningham, M.L., Paules, R.S., 2004. Gene expression profiling of rat livers reveals indicators of potential adverse effects. Toxicol. Sci. 80, 193–202.

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

Hui, L., Bakiri, L., Stepniak, E., Wagner, E.F., 2007a. p38alpha: a suppressor of cell proliferation and tumorigenesis. Cell Cycle 6, 2429–2433.

Hui, L., Bakiri, L., Mairhorfer, A., Schweifer, N., Haslinger, C., Kenner, L., Komnenovic, V., Scheuch, H., Beug, H., Wagner, E.F., 2007b. p38alpha suppresses normal and cancer cell proliferation by antagonizing the JNK-c-Jun pathway. Nat. Genet. 39, 741–749.

Irwin, R.D., Boorman, G.A., Cunningham, M.L., Heinloth, A.N., Malarkey, D.E., Paules, R.S., 2004. Application of toxicogenomics to toxicology: basic concepts in the analysis of microarray data. Toxicol. Pathol. 32, 72–83.

Karin, M., Lawrence, T., Nizet, V., 2006. Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer. Cell 124, 823–835.

Kiyosawa, N., Ando, Y., Manabe, S., Yamoto, T., 2009. Toxicogenomic biomarkers for liver toxicity. J. Toxicol. Pathol. 22, 35–52.

Kondo, C., Minowa, Y., Uehara, T., Okuno, Y., Nakatsu, N., Ono, A., Maruyama, T., Kato, I., Yamate, J., Yamada, H., Ohno, Y., Urushidani, T., 2009. Identification of genomic biomarkers for concurrent diagnosis of drug-induced renal tubular injury using a large-scale toxicogenomics database. Toxicology 265, 15–26.

Kramer, J.A., Curtiss, S.W., Kolaja, K.L., Alden, C.L., Blomme, E.A., Curtiss, W.C., Davila, J.C., Jackson, C.J., Bunch, R.T., 2004. Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. Chem. Res. Toxicol. 17, 463–470.

Kumar, S., Boehm, J., Lee, J.C., 2003. p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. Nat. Rev. Drug Discov. 2, 717–726.

Matsumoto, H., Yakabe, Y., Saito, K., Sumida, K., Sekijima, M., Nakayama, K., Miyaura, H., Saito, F., Otsuka, M., Shirai, T., 2009. Discrimination of carcinogens by hepatic transcript profiling in rats following 28-day administration. Cancer Inform. 7, 253–269.

Nakayama, K., Kawano, Y., Kawakami, Y., Moriwaki, N., Sekijima, M., Otsuka, M., Yakabe, Y., Miyaura, H., Saito, K., Sumida, K., Shirai, T., 2006. Differences in gene expression profiles in the liver between carcinogenic and non-carcinogenic isomers of compounds given to rats in a 28-day repeat-dose toxicity study. Toxicol. Appl. Pharmacol. 217, 299–307.

Naugler, W.E., Sakurai, T., Kim, S., Maeda, S., Kim, K., Elsharkawy, A.M., Karin, M., 2007. Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production. Science 317, 121–124.

Nie, A.Y., McMillian, M., Parker, J.B., Leone, A., Bryant, S., Yieh, L., Bittner, A., Nelson, J., Carmen, A., Wan, J., Lord, P.G., 2006. Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. Mol. Carcinog. 45, 914–933.

Oster, S.K., Ho, C.S., Soucie, E.L., Penn, L.Z., 2002. The myc oncogene: MarvelouslY Complex. Adv. Cancer Res. 84, 81–154.

Pelengaris, S., Khan, M., Evan, G., 2002. c-MYC: more than just a matter of life and death. Nat. Rev. Cancer 2, 764–776.

Peto, R., Gray, R., Brantom, P., Grasso, P., 1991a. Effects on 4080 rats of chronic ingestion of N-nitrosodiethylamine or N-nitrosodimethylamine: a detailed dose-response study. Cancer Res. 51, 6415–6451.

Peto, R., Gray, R., Brantom, P., Grasso, P., 1991b. Dose and time relationships for tumor induction in the liver and esophagus of 4080 inbred rats by chronic ingestion of N-nitrosodiethylamine or N-nitrosodimethylamine. Cancer Res. 51, 6452–6469.

Riou, P., Saffroy, R., Chenailler, C., Franc, B., Gentile, C., Rubinstein, E., Resink, T., Debuire, B., Piatier-Tonneau, D., Lemoine, A., 2006. Expression of T-cadherin in tumor cells influences invasive potential of human hepatocellular carcinoma. FASEB J. 20, 2291–2301.

Searfoss, G.H., Ryan, T.P., Jolly, R.A., 2005. The role of transcriptome analysis in preclinical toxicology. Curr. Mol. Med. 5, 53–64.

Thorgeirsson, S.S., Grisham, J.W., 2002. Molecular pathogenesis of human hepatocellular carcinoma. Nat. Genet. 31, 339–346.

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

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

Urushidani, T., 2010. Toxicogenomics project and drug safety evaluation. Nippon Yakurigaku Zasshi 136, 46–49.

Williams, G.M., Tanaka, T., Maruyama, H., Maeura, Y., Weisburger, J.H., Zang, E., 1991. Modulation by butylated hydroxytoluene of liver and bladder carcinogenesis induced by chronic low level exposure to 2-acetylaminofluorene. Cancer Res. 51, 6224–6230.

Williams, G.M., Iatropoulous, M.J., Jeffrey, A.M., 2004. Thresholds for the effects of 2-acetylaminofluorene in rat liver. Toxicol. Pathol. 32, 85–91.

Xu, H., Cheung, I.Y., Guo, H.F., Cheung, N.K., 2009. MicroRNA miR-29 modulates expression of immunoinhibitory molecule B7-H3: potential implications for immune based therapy of human solid tumors. Cancer Res. 69, 6275–6281.

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

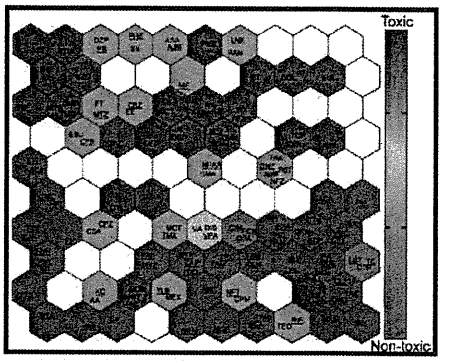
[∥]National Institute of Health Sciences, Kamiyoga, Tokyo, Japan

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

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

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

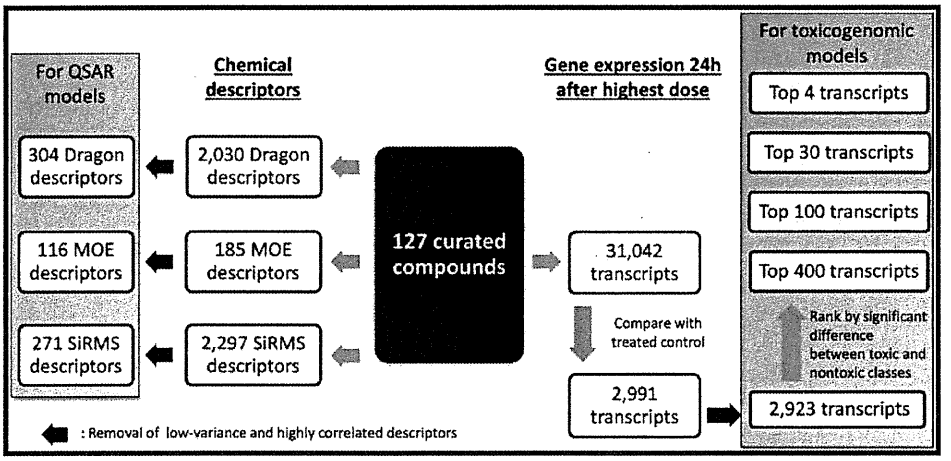


Figure 1. Workflow illustrating data curation and feature selection for modeling.

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.

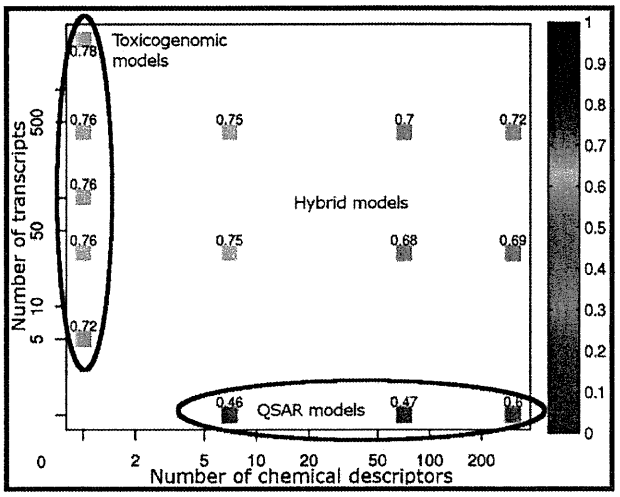


Figure 2. CCR accuracy of the models with respect to the number of chemical descriptors and transcripts used. All models were generated by SVM classification with 5-fold external cross-validation.

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]

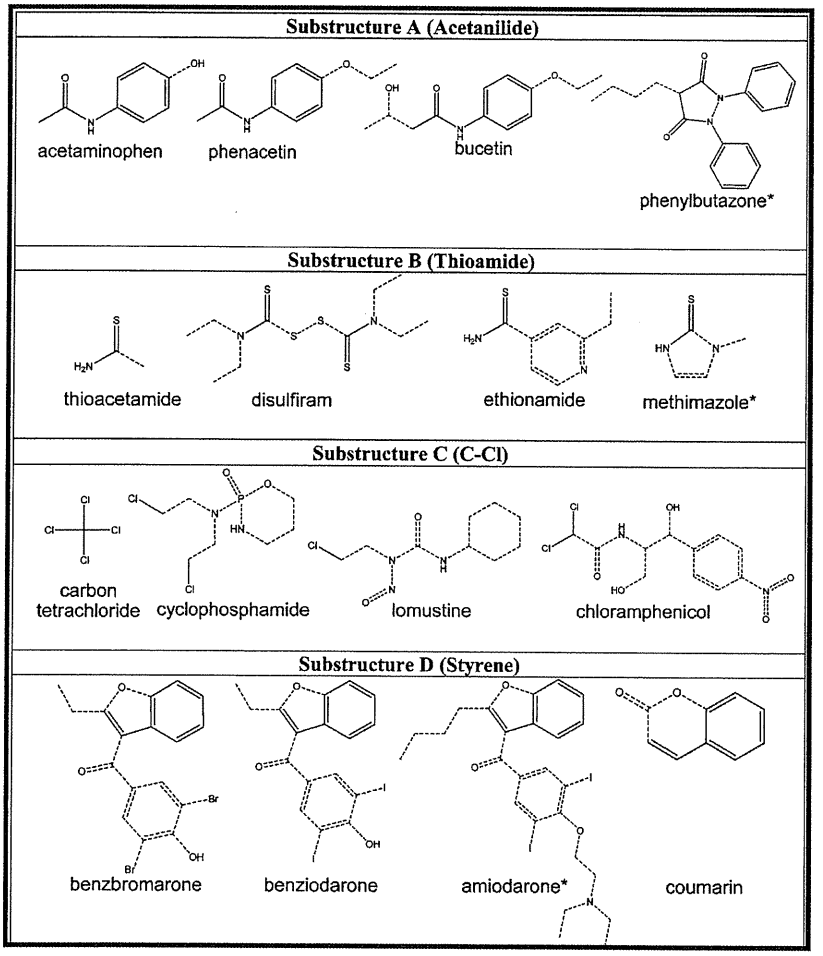
method	*k*NN	SVM	DWD	RF
specificity ± SD	0.82 ± 0.08	0.84 ± 0.10	0.77 ± 0.11	0.84 ± 0.05
sensitivity ± SD	0.57 ± 0.07	0.67 ± 0.12	0.62 ± 0.17	0.66 ± 0.20
CCR ± SD	0.70 ± 0.06	0.76 ± 0.09	0.69 ± 0.11	0.76 ± 0.10
coverage (%)	95	99	99	100

[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

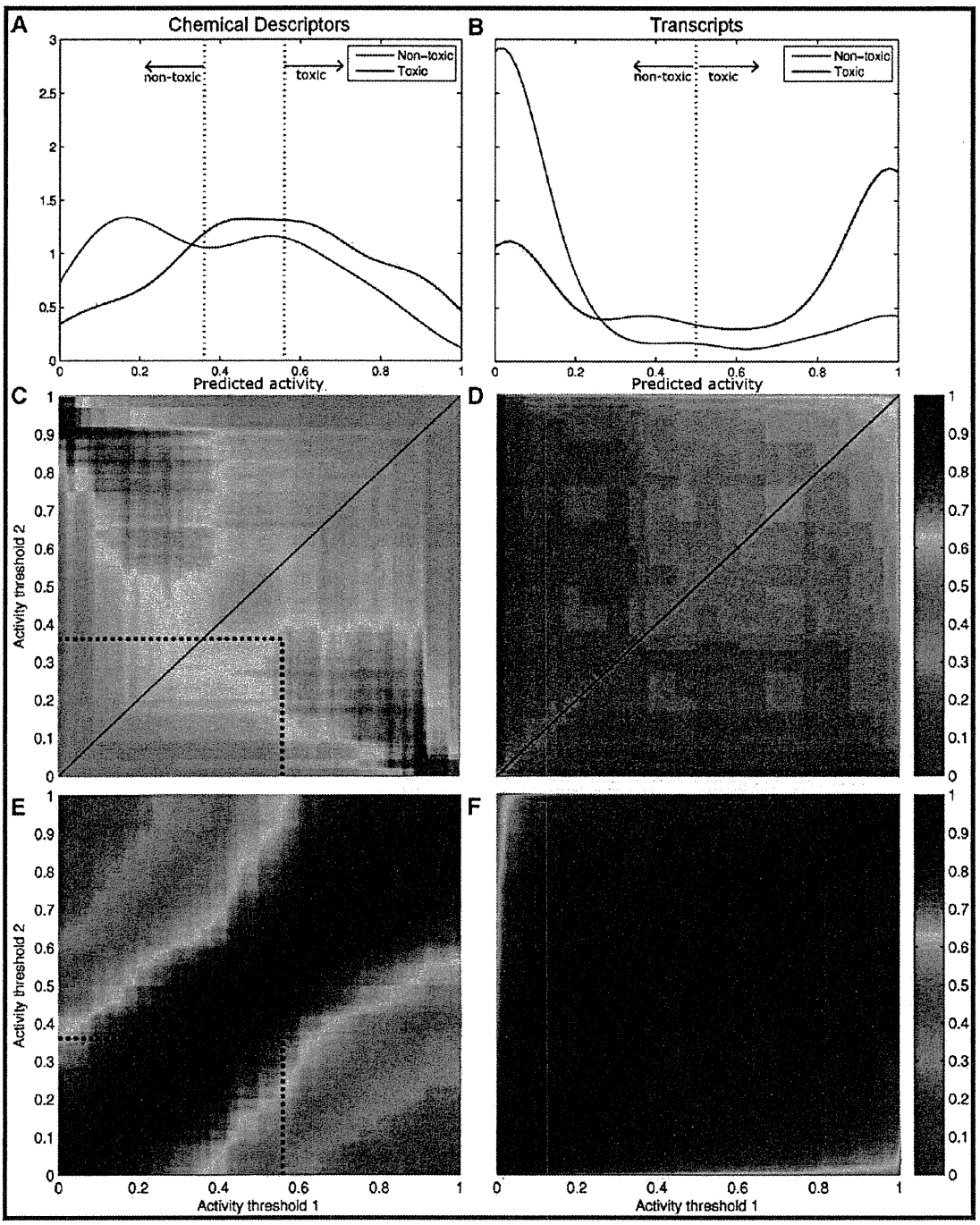


Figure 3. External prediction results of the QSAR (A, C, and E) and toxicogenomics (B, D, and F) models by *k*NN using different classification criteria. Distribution of the predicted values (A and B) and heat maps illustrating classification accuracy (C and D, CCR) and coverage (E and F, percent chemicals 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

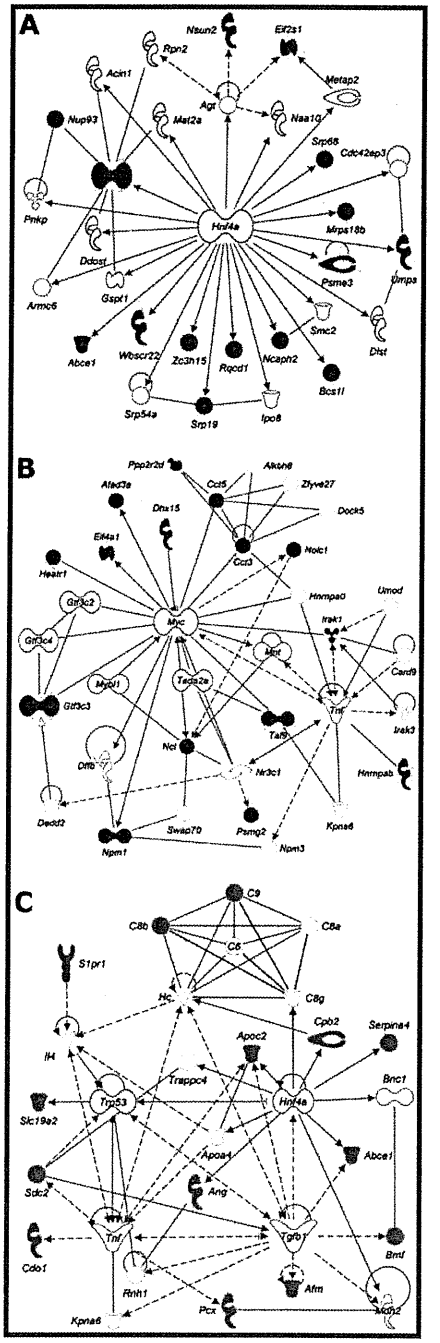


Figure 4. Molecular networks representing the toxicogenomics predictors of hepatotoxicity. *Hnf4a*-centered (A), *Myc*-centered (B), and cellular function, and maintenance-related (C) interactomes were selected as the highest ranked networks among the 64 up- or 21 down-regulated genes used in modeling. 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

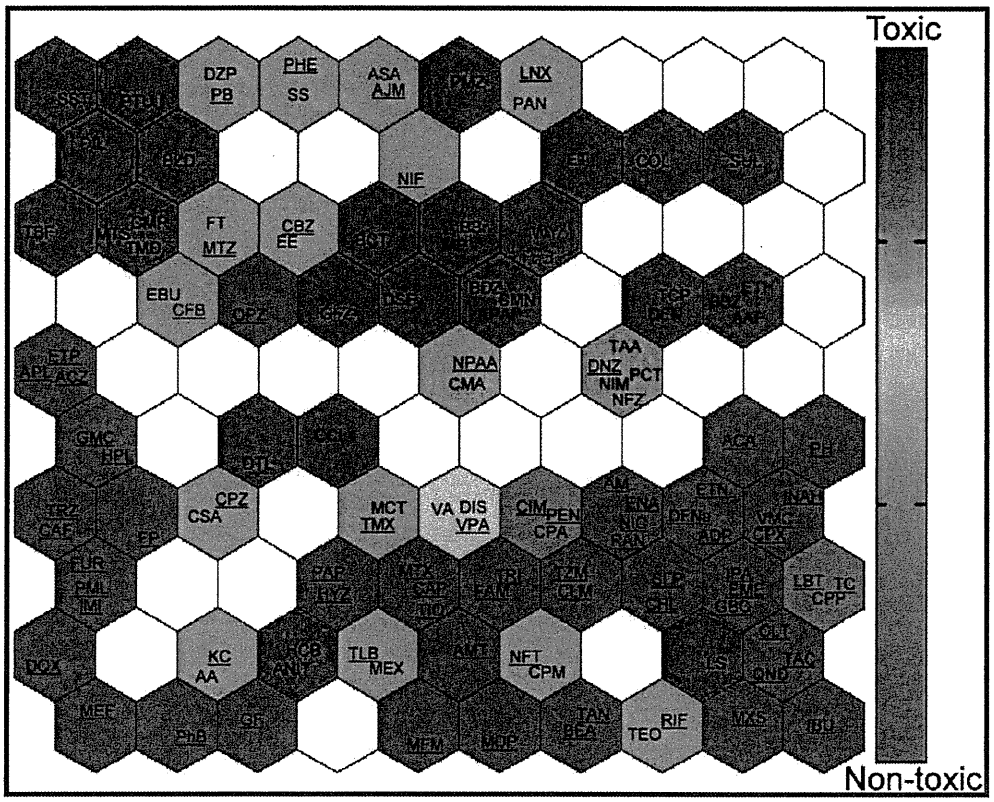


Figure 5. Self-organizing map of the compounds clustered by the expression of the 85 selected transcripts. Nontoxic (underlined) compounds are tightly clustered in the bottom right. PPARα activating and oxidative stress-inducing chemicals are colored in blue and red, respectively.

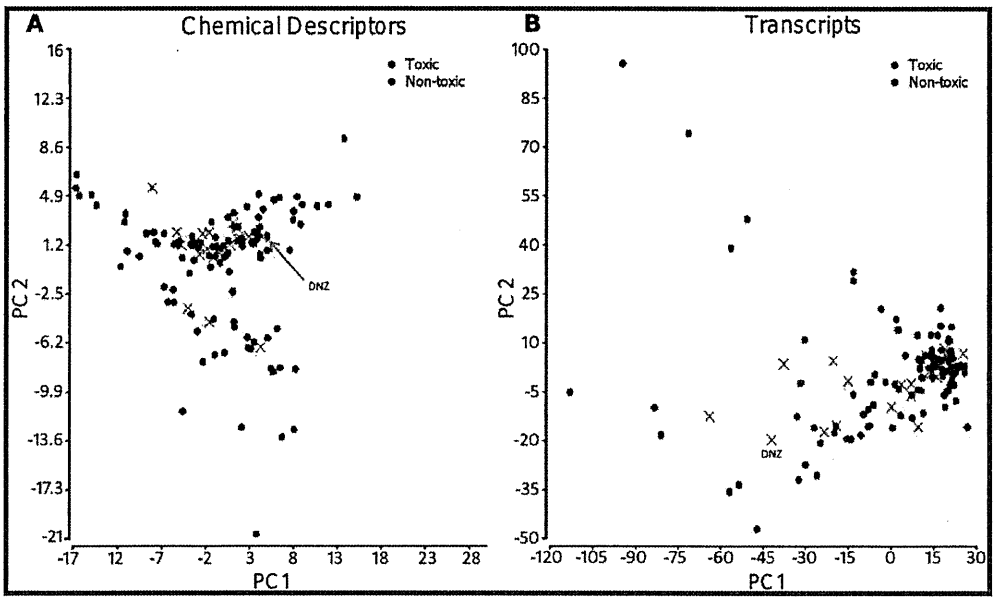


Figure 6. Principal component analysis of the chemical (A) and toxicogenomics (B) descriptors. Toxic and nontoxic compounds are colored red and black, respectively. Compounds mis-predicted by the toxicogenomics model but correctly predicted by the QSAR model are marked as crosses (×). An example of a 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	
Predicted as toxic	**Actually non-toxic** 1. *carbamazepine* 2. *danazol* 3. *nitrofurazone* 4. *omeprazole* 5. *papaverine* 6. *phenylanthranilic acid* 7. *phenytoin* 8. *tamoxifen*	**Actually toxic** 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	**Actually non-toxic** 1. bromoethanamine 2. clofibrate 3. griseofulvin 4. methimazole 5. nifedipine	**Actually toxic** 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	**Actually 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	**Actually toxic** 1. allyl alcohol 2. chlorpropamide 3. clomipramine 4. cyclosporine A 5. disopyramide 6. mexiletine 7. puromycin aminonucleoside 8. sulfasalazine 9. theophylline	**Actually non-toxic** 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	**Actually toxic** 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).