Is Toxicogenomics a More Reliable and Sensitive Biomarker than Conventional Indicators from Rats To Predict Drug-Induced Liver Injury in Humans?

ARTICLE *in* CHEMICAL RESEARCH IN TOXICOLOGY · NOVEMBER 2011

Impact Factor: 3.53 · DOI: 10.1021/tx200320e · Source: PubMed

3 AUTHORS, INCLUDING:

Minjun Chen
U.S. Department of Health and Human Services
40 PUBLICATIONS **1,333** CITATIONS
SEE PROFILE

Weida Tong
U.S. Food and Drug Administration
244 PUBLICATIONS **9,279** CITATIONS
SEE PROFILE

Chemical Research in Toxicology

Article

pubs.acs.org/crt

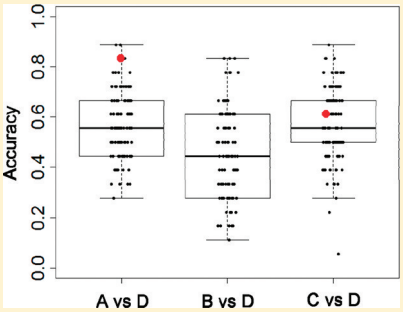
Is Toxicogenomics a More Reliable and Sensitive Biomarker than Conventional Indicators from Rats To Predict Drug-Induced Liver Injury in Humans?

Min Zhang, Minjun Chen, and Weida Tong*

Center of Excellence for Bioinformatics, Division of Systems Biology, National Center for Toxicological Research, U.S. Food and Drug Administration, 3900 NCTR Road, Jefferson, Arkansas 72079, United States

Supporting Information

ABSTRACT: Around 40% of drug-induced liver injury (DILI) cases are not detected in preclinical studies using the conventional indicators. It has been hypothesized that genomic biomarkers will be more sensitive than conventional markers in detecting human hepatotoxicity signals in preclinical studies. For example, it has been hypothesized and demonstrated in some cases that (1) genomic biomarkers from the rat liver can discriminate drug candidates that have a greater or lesser potential to cause DILI in susceptible patients despite no conventional indicators of liver toxicity being observed in preclinical studies, and (2) more sensitive biomarkers for early detection of DILI can be derived from a "subtoxic dose" at which the injury in the liver occurs at the molecular but not the phenotypic level. With a public TGx data set derived from short-term *in vivo* studies using rats, we divided drugs exhibiting human hepatotoxicity into three groups according to whether elevated alanine aminotransferase (ALT) or total bilirubin (TBL) were observed in the treated rats: (A) The elevation was observed in the treated rats, (B) no elevation was observed for all of the treated rats, and (C) no elevation could be observed at a lower dose and shorter duration but occur when a higher or longer treatment was applied. A control group (D) was comprised of drugs known not to cause human hepatotoxicity and for which no rats exhibited elevated ALT or TBL. We developed classifiers for groups A, B, and C against group D and found that the gene signature from scenario A could achieve 83% accuracy for human hepatotoxicity potential of drugs in a leave-one-compound-out cross-validation process, much higher than scenarios B (average 45%) and C (61%). Furthermore, the signature derived from scenario A exhibited relevance to hepatotoxicity in a pathway-based analysis and performed well on two independent public TGx data sets using different chemical treatments and profiled with different microarray platforms. Our study implied that the human hepatotoxicity potential of a drug can be reasonably assessed using TGx analysis of short-term *in vivo* studies only if it produces significant elevation of ALT or TBL in the treated rats. The study further revealed that the value of "sensitive" biomarkers derived from scenario C was not promising as expected for DILI assessment using the reported TGx design. The study will facilitate further research to understand the role of genomic biomarkers from rats for assessing human hepatotoxicity.



INTRODUCTION

Drug-induced liver injury (DILI) is the most common cause of acute liver failure,[1] and all forms of DILI have caused many drugs to fail during clinical trials or to be withdrawn from the market.[2] Although animal testing protects humans from exposure to potential risk associated with drugs, it is a not a fail-safe paradigm, particularly for DILI.[3,4] A consortium determined that about half of the drugs that cause human hepatotoxicity were not identified as having this potential in nonclinical animal testing.[3] Some reports suggest that liver toxicity accounts for ~40% of the drugs that fail during the clinical trial stage of development. DILI has been linked to approximately 1000 marketed drugs.[5] The failure to effectively identify the potential human hepatotoxicity of drug candidates during nonclinical testing has spawned a wide range of research.

In the assessment of liver injury, the elevation of alanine aminotransferase (ALT) or total bilirubin (TBL) is often used as a preclinical and clinical biomarker for DILI.[6,7] By evaluating the relationship of biochemical changes and hepatotoxicity in a preclinical study involving more than 3000 rats, Ennulat et al. concluded that elevated ALT or TBL had diagnostic utility for manifestation of hepatocellular necrosis and biliary injury.[8] Recently, it has been shown that there is significant association between increased ALT levels detected in premarketing clinical trials and postmarketing DILI case reports.[9] In 2009, "Hy's law" involving both ALT and TBL measurements was recommended by the FDA for DILI assessment of drugs in the premarketing stage.[10] However, biochemical biomarkers such as ALT lack specificity,[11] particularly with other factors that could also affect the enzyme level such as pre-existing disease, nondrug-related injury, or injury to other organs.[6,10,12] TBL is a measure of liver function and, unfortunately, is only elevated after pronounced

Received: August 2, 2011

Table 1. Drugs Known To Cause DILI in Humans Exhibit Three Typical Preclinical Scenarios According to Whether or Not the Elevated ALT or TBL Is Observed in a Short-Term Animal Study That Often Involves Multiple Doses and Treatment Durations

scenario	observations in the short-term animal study
A	The treated animals show elevated ALT or TBL. This scenario indicates that the DILI signal observed in animals might imply the drug's potential to cause human hepatotoxicity. Because both ALT and TBL are not specific, more reliable biomarkers such as genomic biomarkers for this category of drugs are still required.
B	None of treated animals exhibit increased level of ALT or TBL, even at a maximum tolerated dose. This scenario indicates that the human hepatotoxicity can not be detected in this assay. However, it has been argued that the DILI signal not apparent in the conventional measure could be detected using other means such as toxicogenomic methods.
C	Given a dose and treatment length, the treated animals have no elevation in ALT and TBL. However, the elevation was observed when a higher dose and/or longer treatment is applied. The question is whether a more sensitive biomarker (e.g., genomic biomarker) can be developed based on the former situation, which foresees the injury occurring in the latter situation. This scenario has an important clinical implication in early diagnosis.

injury. Thus, additional preclinical biomarkers are needed to predict drugs that cause DILI in the clinic.

Various methodologies have been investigated with particular focus on the identification of molecular biomarkers in animals for DILI using emerging technologies such as toxicogenomics (TGx).[13−17] TGx applies gene expression technologies to toxicology studies to elucidate molecular mechanisms involved in the toxicity.[15,18−21] In reality, however, it may be difficult to predict every drug causing DILI using preclinical animal models regardless of which types of technologies applied including TGx. For example, some suggest that existing differences in metabolic functions between rats and humans may render human-specific metabolites.[22] Thus, the real question is how to know which drug candidates can be reliably assessed in the preclinical stage for DILI and which can not and whether ALT and TBL can be useful to differentiate drugs with DILI potential that can be assessed in a preclinical TGx study from those that can not.

Most TGx studies use a short-term animal experimental design in conjunction with gene expression detection using microarrays.[18,19,23] These studies often involve using multiple doses (e.g., low, medium, and high doses) with multiple treatment periods (e.g., 3, 5, and 7 days). Drugs known to cause DILI in humans behave differently in these short-term animal assays, depending on whether or not the ALT or TBL level was significantly increased in the treated animals. Specifically, these drugs can be grouped into three distinct scenarios, as summarized in Table 1. Scenario A represents a situation where a drug's potential to cause DILI in humans is recognized in the animal model, but confirmatory studies need to follow since both ALT and TBL are not specific. Scenario B represents one of the most important motivations of using TGx in preclinical studies: Genomic biomarkers from the rat liver can discriminate drug candidates that have the potential to cause DILI in susceptible patients from drugs that do not have this potential despite no conventional indicators of liver toxicity being observed in preclinical studies. Scenario C is close to the "sensitive biomarker" concept, which announces the existence of gene expression signals before the conventional indicators of toxicity (e.g., clinical chemistry or histopathology) are manifested in animals. There is a lot of interest in identifying the sensitive biomarkers with TGx.[24−29] The sensitive biomarker phenomenon has been demonstrated in acetaminophen (APAP)-induced liver injury at the "subtoxic dose".[24,30]

In this study, we questioned which of the genomics signatures from the aforementioned scenarios have a satisfactory translational ability from animals to humans for DILI. This is important since understanding the application boundary of a preclinical TGx assay for humans is a necessary step to appropriately and effectively apply the TGx study in a "fit-for-purpose" manner. We identified a public TGx data set that contains all three scenarios along with a control group (group D) that was comprised of rats that showed no elevations of ALT or TBL and were treated with drugs known not to cause DILI. We conducted a comparative analysis for each scenario against group D to determine that the TGx biomarkers from which scenarios will yield a better predictivity for DILI potential in humans. The results suggested that the potential of human hepatotoxicity can be reasonably assessed for the drugs only causing elevated ALT or TBL in animals (scenario A) but not for those causing no biochemical changes in any of the treated animals (scenario B). Moreover, contrary to the common belief, we found that the sensitive biomarker identified in animals is not promising for assessing human hepatotoxicity (scenario C).

MATERIALS AND METHODS

Drugs with Human Hepatotoxic and Gene Expression Data. We identified 85 drugs with both human hepatotoxic and gene expression data. These human drugs have a single active molecule and are administered either orally or parentally. All of these drugs have FDA-approved labels and have been marketed for over 10 years. The DILI potential of these drugs is defined based on the adverse events described in three sections of their labels, namely, Boxed Warning, Warnings and Precautions, and Adverse Reactions. We applied a set of keywords developed in our previous work[31] to divide drugs into four lists. The first three lists contained drugs with at least one of these keywords presented in their respective labeling sections (i.e., Boxed Warning, Warnings and Precautions, and Adverse Reactions). Given the fact that the risk of an adverse effect follows the order of Boxed Warning > Warnings and Precautions > Adverse Reactions,[31] we only considered the drugs that were either withdrawn from the markets or mentioned in Boxed Warning (list 1) and Warning and Precautions (list 2) for DILI as positive (i.e., DILI positive drugs) in our analysis. The drugs with none of the keywords present in any of the labels (the fourth list) were considered as negative (DILI negative drugs). Nine drugs were DILI negative, while 76 were DILI positive.

The gene expression data of these 85 drugs are obtained from a public TGx data set, the Iconix data. The data set was downloaded from Gene Expression Omnibus (GEO) under the accession number GSE8858. The Iconix data set contains 5288 arrays of liver gene expression for rats treated with 344 different compounds.[32] The array data were generated using the CodeLink platform. We only included the 3 day and 5 day gene expression data for the 85 drugs since more than 90% of the samples with the ALT and TBL information were collected on these two time points in the original study. Both ALT/TBL and gene expression data were collected at the same time (3 or 5 day).

For 76 DILI positive drugs, nine showed ALT or TBL elevation in at least one of the treated rats, while 67 did not exhibit significant elevation in any of the treated rats. As expected, nine DILI negative drugs did not exhibit significant elevation in any of the treated rats. A rat with at least three times ALT or two times TBL elevation as compared to its matched control was considered as a significant or observable elevation.

Predictive Model Development. The 85 drugs were divided into four groups, summarized in Table 2 (more details in Table S1 in

dx.doi.org/10.1021/tx200320e | Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

Table 2. Overview of the Study Design

group	hepatotoxicity observed in: human (based on the drug labeling)	hepatotoxicity observed in: rat (based on elevated ALT or TBL)	no. of drugs	no. of arrays	description of the gene expression data
A	yes	yes	9	25	the gene expression data at the dose and time when the elevation in either ALT or TBL was observed
B	yes	no	67	194	the gene expression data at the highest available dose and longest duration
C	yes	no	9	26	the same set of drugs presented in group A but only contains arrays associated with rats with no elevation in both ALT and TBL
D	no	no	9	26	the expression data at the highest available dose and longest duration

the Supporting Information). The 76 drugs known to cause human hepatotoxicity (DILI positives) were divided into three groups (groups A, B, and C) corresponding to the definition in Table 1. Group D consists of nine DILI negative drugs (rats treated with group D drugs did not demonstrate any significant elevations in ALT and TBL).

Three sets of classifiers were developed, namely, "A vs D", "B vs D", and "C vs D" to determine which scenario yields the best predictive model. The *K*-nearest neighbor (*K*NN, $K = 3$ or 5) classification method was used to develop gene signatures for each scenario with a leave-one-compound-out cross-validation (LOCOCV) procedure (Figure 1). In this method, each drug (denoted as "test set" in Figure 1)

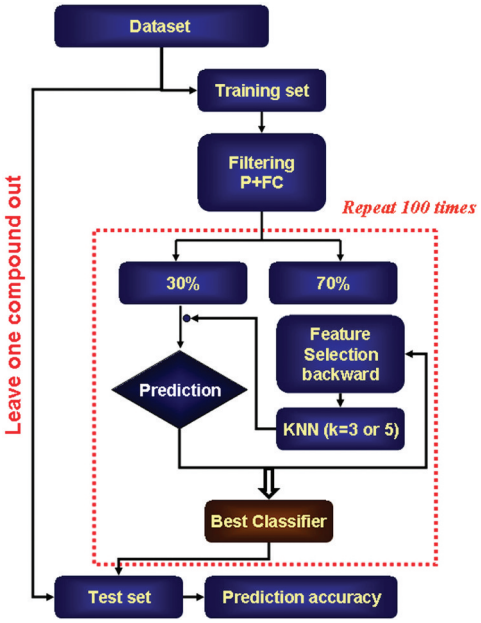


Figure 1. Overview of the LOCOCV process.

was systematically excluded once in turn and only once from the data set, and its DILI potential was predicted by a classifier derived from the remaining drugs (denoted as "training set" in Figure 1) in the data set. For the training set, genes were first filtered by their fold-change (FC > 1.5) and *t* test *P* values ($P < 10 \times 10^{-5}$) by comparing the DILI positive drugs versus DILI negative drugs.[34,35] Then, the training set was randomly split into two parts, and 70% of the drugs were used to determine the parameters (i.e., the number of features and *K* value) that lead to a best classifier in prediction of the rest of 30% drugs in a process that was repeated 100 times to avoid the chance correlation. The backward feature selection was used in this process.[36] The identified parameters were then used to construct the *K*NN model using all of the drugs in the training set, and its prediction accuracy to the test set was recorded along with the signature genes.

A permutation test was also carried out to determine whether predictive performance observed in the LOCOCV exceeded what would be expected by chance alone.[37] In this process, 100 permuted data sets were generated from the original data set, in which the drugs' class labels were randomly scrambled. For each permuted data set, the same classification procedure described above was applied.

Independent Validation Sets. Two publicly available TGx data sets, both generated by National Institute of Environmental Health (NIEHS), were used to assess the models derived from the Iconix data set. NIEHS data set 1 (GSE24363) was downloaded from GEO, while NIEHS data set 2 was downloaded from Chemical Effects in Biological Systems (002-00001-0011-000-5).

NIEHS data set 1 is comprised of gene expression data on the liver sample of rats treated with eight chemical compounds, including 1,2-dichlorobenzene, 1,4-dichlorobenzene, bromobenzene, diquat dibromide, galactosamine, monocrotaline, *N*-nitrosomorpholine, and thioacetamide. They were selected based on published literature regarding the differences that exist in the cell types and liver regions that are injured in response to exposure. For each compound, doses that elicited a subtoxic ("low"), a moderately toxic ("medium"), or a overtly toxic ("high") response 24 h after treatment were selected. Samples were collected for gene expression profiling, clinical chemistry (e.g., ALT and TBL), hematology, and histopathology at 6, 24, and 48 h postexposure. The array data were generated using the Affymetrix platform. We used 318 arrays associated with the treated rats by these eight hepatotoxicants and 114 arrays from the control animals.

NIEHS data set 2 contained gene expression data on the liver samples of rats treated with APAP.[33] Five doses of treatment were selected (i.e., 0, 50, 150, 1500, and 2000 mg/kg). Samples were collected for gene expression profiling, clinical chemistry (e.g., ALT and TBL), hematology, and histopathology at 6, 18, 24, and 48 h postexposure. The data were generated using the Agilent Rat Oligo Microarray (22K) platform. A total of 64 arrays from the treated rats and 12 arrays from the control animals were used in this study.

RESULTS

We identified 85 drugs that had the human hepatotoxic data (based on the drug labeling), the gene expression data from the rat liver (profiled using the CodeLink platform from the Iconix data set), and the clinical chemistry data (i.e., the ALT/TBL measure from the Iconix data set). These drugs were divided into four groups (Table 2), matching three scenarios listed in Table 1 (DILI positive drugs; denoted as A, B, and C hereafter) along with a control group (DILI negative drugs that were denoted as D). The detailed information of the four groups is summarized in Table S1 in the Supporting Information. Three separate classifiers were developed by comparing groups A, B, and C against D, respectively. The signature derived from the predictive model of A vs D was applied to two independent data sets (i.e., NIEHS data sets 1 and 2).

Effect of Elevated ALT or TBL in Rats To Predict DILI in Humans. We first examined whether the observable change of either ALT or TBL is essential to develop a gene signature in rat for predicting human hepatotoxicity. As shown in Figure 2, the predictive model of A vs D (see Materials and Methods) yielded 83% accuracy (red dot, the left side of the figure) in the LOCOCV process. We further confirmed that the obtained accuracy was not due to chance (*P* value < 0.05) by comparing

dx.doi.org/10.1021/tx200320e | Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

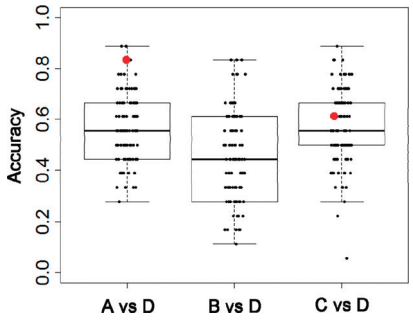


Figure 2. Effect of ALT or TBL elevation on the prediction accuracy of DILI. The large red dots represent the prediction accuracy in the LOCOCV for the classification of A vs D and C vs D. The left and right box plots show the distribution of the accuracy rates from 100 permutated data for A vs D and C vs D. The middle box plot shows the distribution of the accuracy rates from 100 B vs D models.

it to the permutation test results (box plot in the left side of the figure). Similarly, we also developed predictive models for B vs D. To achieve a fair comparison with the results from A vs D, we selected only nine drugs (i.e., the same number as group A) from group B (that contains 67 drugs, Table 2), and the selection process was random with repeating 100 times (a bootstrapping method). The process yielded 100 B vs D models, and the results were depicted in a box plot in the middle section of Figure 2. The average accuracy of the 100 models is 45% with only a few having accuracy above 60%, which was no better than chance. The results demonstrated that TGx biomarkers for predicting DILI in use in humans could only be reliably developed for drugs causing either ALT or TBL elevation in the rat.

Next, we used the same method to evaluate whether sensitive biomarkers derived from the "subtoxic dose" study could be of any value for DILI identification (shown in the right section of Figure 2). The C vs D model only yielded 61% accuracy (red dot, the right side of Figure 2), which was no better than chance (P value = 0.27) when compared to the permutation result (box plot in the right side of Figure 2). The results demonstrated that sensitive biomarkers performed poorly as compared to those derived from the "toxic dose" (represented in group A) in discriminating DILI potential of drugs in human.

For each classification model (i.e., A vs D, B vs D, and C vs D), a signature gene list was obtained that contained all of the genes used in the LOCOCV process. The signature genes identified from all three scenarios were quite different. Although scenarios A and C involved the same drugs, only seven genes were identified in common between the 190 genes in scenario A and the 60 genes in scenario C. For scenario B, we generated 100 bootstrapping-derived models, and little between-model overlap in signature gene was found with the average pairwise overlaps of around 10%. Out of the total 991 signature genes from all of the 100 models in scenario B, only small portions of genes were found to overlap with those from scenarios A and C; the average overlap percentage was 6 and 8%, respectively. Such a small number of common signatures among these three scenarios indicates distinct biological phenomena manifested at different stages of DILI development.

Evaluation of the Identified Signature Genes on Independent Data Sets. To assess how effective the signature genes identified in the Iconix rat data set were at predicting human hepatotoxicity, two other data sets were evaluated. Specifically, the signature genes from scenario A vs D were tested on two NIEHS data sets, which were profiled with two different microarray platforms (i.e., Affymetrix and Agilent), time frames, and benchmarked with the presence or absence of observed liver histopathology changes and ALT/TBL elevations.

Eight chemicals in the NIEHS data set 1 are well-known hepatotoxicants in rats. It reasonably assumes that these chemicals could cause DILI in humans as well. Consequently, we reclassified these data according to the scenarios defined in Table 2: (1) Scenario A contains the arrays from the treated rats with elevated ALT or TBL, (2) scenario C contains the arrays from the treated rats with no ALT or TBL elevation, and (3) scenario D contains the arrays from the control animals. Because each of eight chemicals provoke ALT or TBL elevation in at least one of the treated rats, we do not have scenario B for this data set. We identified 62 out of the 190 signature genes (generated from the CodeLink platform) that were matched to genes from the Affymetrix platform based on the GenBank accession number (Table 3). Principal component analysis (PCA) was conducted using these 62 genes, and the results are depicted in Figure 3. Each chemical exhibited the same pattern in the PCA plot; the control samples (black circle) were tightly grouped together, and the samples with no elevation in ALT or TBL (red circle) tended to cluster with the control samples, while the samples with elevated ALT or TBL (blue triangle) were scattered away from the control samples. Because most hepatotoxicants in NIEHS data set 1 have not been evaluated in humans, we selected APAP (NIEHS data set 2) as a positive control to repeat the analysis. The results were consistent with the findings from NIEHS data set 1 (Figure S1 in the Supporting Information). The results indicate that the samples with no elevated ALT or TBL behave more like the control samples and thus suggest that it might be difficult to develop sensitive biomarkers from the treated animals at the doses that do not elicit the changes of traditional indicators.

We also examined the functions of the 62 signature genes using Ingenuity Pathway Analysis software.[38] Table 4 shows the over-representation of canonical and toxicity pathways in the 62 signature genes. The identified pathways are all related to liver functions and liver injury, such as "xenobiotic metabolism signaling", "LPS/IL-1 mediated inhibition of RXR function", "glutathione depletion", and "metabolism of xenobiotics by cytochrome P450".

■ DISCUSSION

We investigated the application boundary of a preclinical TGx study for prediction of DILI in humans. Three preclinical TGx scenarios according to ALT or TBL elevation in rats treated with drugs that cause DILI in humans were evaluated as follows: (A) The elevation was observed, (B) no elevation was observed, and (C) no elevation was observed until a higher or longer treatment was applied. Our study is the first published attempt to assess which scenario could yield a reliable translational biomarker from rats to humans with TGx.

It has been hypothesized and demonstrated in some cases that (1) genomic biomarkers from the rat liver can discriminate between drugs that can potentially cause DILI in susceptible patients and drugs that cannot, even in the absence of conventional indicators of liver toxicity in preclinical studies; and (2) more sensitive biomarkers for early detection of DILI can be derived from a "subtoxic dose" at which the injury in the liver occurs at the molecular level but not at the phenotypic level. However, using a publically available data set, we found that relevant genomic biomarkers can be identified in a

dx.doi.org/10.1021/tx200320e | Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

Table 3. Sixty-Two Signature Genes Identified for the Classification Model of A vs D

CodeLink	GeneBank accession no.	Affymetrix	gene symbol
AF097723_PROBE1	NM_031640	1368399_a_at	Pgcp
L22339_PROBE1	NM_031732	1369296_at	Sult1c3
NM_012811_PROBE1	NM_012811	1386860_at	Mfge8
NM_017246_PROBE1	NM_017246	1386959_a_at	Map2k5
NM_019153_PROBE1	NM_019153	1367866_at	Fbln5
NM_019341_PROBE1	NM_019341	1369957_at	Rgs5
U75394_PROBE1	NM_053752	1367642_at	Suclg1
X91234_PROBE1	NM_024391	1387156_at	Hsd17b2
M23601_PROBE1	NM_013198	1368514_at	Maob
AF065147_PROBE1	NM_012924	1368921_a_at	Cd44
AF087433_PROBE1	NM_053667	1367967_at	Lepre1
NM_019283_PROBE1	NM_019283	1398771_at	Slc3a2
L20900_PROBE1	NM_030844	1367787_at	Ica1
BE115626_PROBE1	BE115626	1379260_at	Heca
BF281848_PROBE1	BF281848	1390141_at	Mthfd1l
J02657_PROBE1	NM_019184	1387328_at	Cyp2c11
NM_013052_PROBE1	NM_013052	1367693_at	Ywhah
NM_012657_PROBE1	NM_012657	1368048_at	Serpina3k
Y08172_PROBE1	NM_012816	1367775_at	Amacr
NM_017125_PROBE1	NM_017125	1367709_at	Cd63
M35266_PROBE1	NM_052809	1367755_at	Cdo1
NM_013186_PROBE1	NM_013186	1368242_at	Kcnb1
AW251324_PROBE1	AW251324	1372808_at	Mthfd2
U05014_PROBE1	NM_053857	1386888_at	Eif4ebp1
BE108905_PROBE1	BE108905	1371872_at	Nap1l1
BF416285_PROBE1	BF416285	1376481_at	Adamts9
NM_013145_PROBE1	NM_013145	1387505_at	Gnai1
AF285078_PROBE1	NM_053431	1368023_at	Qsox1
AW524559_PROBE1	AW524559	1389150_at	
NM_021653_PROBE1	NM_021653	1369259_at	Dio1
AB012759_PROBE1	NM_031324	1368234_at	Prep
AF169409_PROBE1	NM_080786	1368295_at	Slco2b1
BE111762_PROBE1	BE111762	1375879_at	
BF555947_PROBE1	BF555947	1373023_at	
NM_017270_PROBE1	NM_017270	1369863_at	Adh4
X67654_PROBE1	NM_053293	1368354_at	Gstt1
AI170382_PROBE1	AI170382	1389648_at	Ripk4
AI412889_PROBE1	BI296275	1389221_at	Mmd2
AW915454_PROBE1	AW915454	1376394_at	Clec9a
L27081_PROBE1	NM_013182	1369498_at	Mc5r
NM_012770_PROBE1	NM_012770	1368779_a_at	Gucy1b2
NM_019290_PROBE1	NM_019290	1368072_at	Btg3
AF015949_PROBE1	NM_031714	1368060_at	Hrsp12
AF285631_PROBE1	NM_031725	1367688_at	Scamp4
AI171656_PROBE1	AI171656	1373696_at	RGD1564859
AI408557_PROBE1	AI408557	1373478_at	
AI548730_PROBE1	AI548730	1375022_at	Afg3l2
BE106523_PROBE1	BM389496	1389678_at	
BF282573_PROBE1	BF282573	1374489_at	Gtpbp2
BF392344_PROBE1	BF392344	1374883_at	Mtmr7
L46865_PROBE1	NM_017255	1368940_at	P2ry2
M77479_PROBE1	NM_017047	1368609_at	Slc10a1
NM_012552_PROBE1	NM_012552	1387819_at	Cela1
NM_012561_PROBE1	NM_012561	1387843_at	Fst
NM_012733_PROBE1	NM_012733	1367939_at	Rbp1
NM_013161_PROBE1	NM_013161	1368554_at	Pnlip
NM_017214_PROBE1	NM_017214	1368505_at	Rgs4
NM_021593_PROBE1	NM_021593	1368915_at	Kmo
NM_021661_PROBE1	NM_021661	1370133_at	Rgs19
NM_021997_PROBE1	NM_021997	1368571_at	Clip2
U40819_PROBE1	NM_019142	1369104_at	Prkaa1
U62667_PROBE1	NM_031123	1387623_at	Stc1

dx.doi.org/10.1021/tx200320e | *Chem. Res. Toxicol.* XXXX, XXX, XXX–XXX

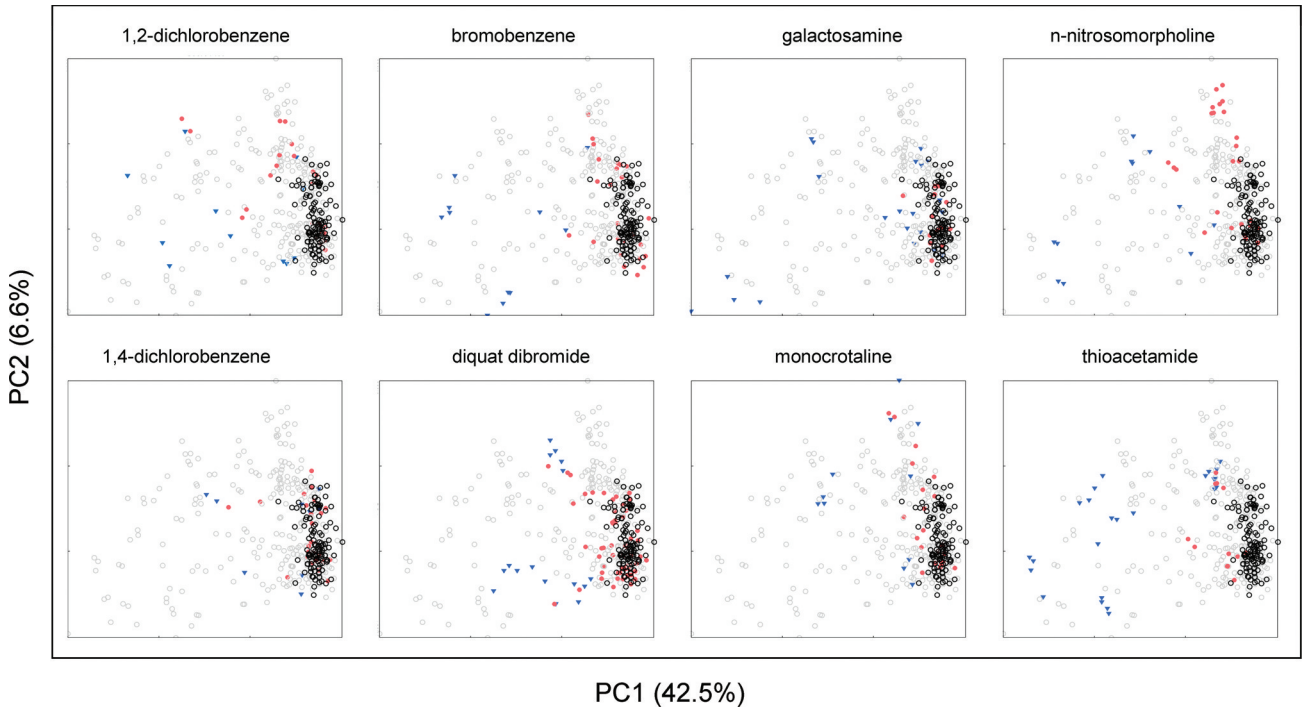


Figure 3. PCA based on the 62 signature genes (from CodeLink platform) for animals treated by eight hepatotoxicants and profiled with the Affymetrix platform. The figure is divided into eight panels corresponding to eight hepatotoxicants. All of the panels contain the same sets of black and gray circles that represent the control samples (118 arrays) and treated samples (318 arrays), respectively. In a panel associated with a specific chemical, the results for that chemical are highlighted as the blue triangles representing the treated samples with elevated ALT or TBL, while the red circles denote the treated samples with no elevation in ALT or TBL.

Table 4. Significantly Altered Functional Pathways (P Value < 0.05) in Gene Signature from Scenario A Using Ingenuity Pathway Analysis

category	pathway	P value	molecules
ingenuity toxicity lists	xenobiotic metabolism signaling	0.002	Gstt1, Maob, Cyp2c11, Sult1c1, Map2k5, Adh4
	LPS/IL-1 mediated inhibition of RXR function	0.003	Gstt1, Maob, Slc10a1, Cyp2c11, Sult1c1
	glutathione depletion−CYP induction and reactive metabolites	0.048	Gstt1
ingenuity canonical pathways	tryptophan metabolism	0.002	Maob, Kmo, Cyp2c11, Adh4
	LPS/IL-1 mediated inhibition of RXR function	0.010	Gstt1, Maob, Slc10a1, Cyp2c11
	metabolism of xenobiotics by cytochrome P450	0.011	Gstt1, Cyp2c11, Adh4
	xenobiotic metabolism signaling	0.021	Gstt1, Maob, Cyp2c11, Map2k5
	tyrosine metabolism	0.025	Maob, Adh4
	ERK5 signaling	0.028	Ywhah, Map2k5
	glycerolipid metabolism	0.031	Pnlip, Adh4
	acute myeloid leukemia signaling	0.035	Map2k5, Eif4ebp1
	taurine and hypotaurine metabolism	0.036	Cdo1
	lysine biosynthesis	0.048	Adh4

TGx study only for the drugs that cause significant ALT or TBL elevation in at least one treated animal, and the value of more sensitive biomarkers for DILI assessment using the TGx design was limited.

This study required a list of drugs with known DILI classification in humans to develop a model for classifying drugs based on a gene expression signature. However, classifying a drug for DILI potential is still a challenge since there is no commonly accepted practice in the research community. We used the curated drug-labeling data to determine a drug's hepatotoxicity potential in human,[31] which only identified 85 drugs that had both gene expression and clinical chemistry data. Although the identified drugs cover a broad range of pharmacological indications with diverse chemical structures, it might not be sufficient in size to draw a reliable conclusion. To ensure the validity of the analysis, a permutation test was performed to assess the reliability of the signature genes. Furthermore, a subset of the gene signature from scenario A was assessed in two independent data sets with gene expression data generated from different platforms along with the pathway analysis. The results indicated that the signature genes were reasonably robust despite the fact that only a small number of samples were used in this study.

In both traditional toxicology and TGx studies, the dose selection is a very important factor in the experimental design. In several recent TGx studies, the corresponding alterations in gene expression patterns at a "subtoxic dose" revealed signals of injury and provided early gene expression signatures for specific compounds even without any injury observable by traditional approaches.[24,30,39,40] Several studies discovered DILI biomarkers from gene expression studies.[16,17,41−43] However, two questions remain as follows: whether the gene expression is more sensitive and specific than conventional biomarkers such as ALT and/or TBL and whether the gene expression could predict DILI without any signals detected by conventional methods.

dx.doi.org/10.1021/tx200320e | Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

It has been demonstrated that more sensitive biomarkers could predict liver injury for APAP at the "subtoxic dose".[24] APAP induces liver injury in a classic dose-dependent manner, which might be an attribute to the discriminating ability of the gene expression alteration at a "subtoxic dose" to predict the liver toxicity observed at a higher dose. It is important to note that the exact choice of a "subtoxic dose" for a specific drug is difficult to be defined. For example, even for a compound like APAP with a well-studied dose range, different doses have been defined as a "subtoxic dose" (150[24,30] and 500 mg/kg[44]). It is likely that "subtoxic dose" covers a range of doses at which a drug does not induce apparent toxicity by classical means, but alterations in gene expression carry sufficient signals for DILI identification. In this study, we defined a "subtoxic dose" at which no significant changes in ALT or TBL were observed at this dose but were seen at the next higher dose or longer treatment period. Even with such a "stringent" definition of "subtoxic dose", we still found that the derived biomarkers performed poorly and no better than random chance in discriminating DILI potential of drugs in humans.

The main conclusions of this study were based on a single TGx design with a specific array platform (i.e., CodeLink) and using our drug labeling-based classification schema to define DILI potential of drugs. To verify the findings, we applied the signature genes from the A vs D classifiers to two different TGx data sets profiled by different array platforms (i.e., Affymetrix and Agilent). As demonstrated in Figure 3 and Figure S1 in the Supporting Information, the samples with no elevated ALT or TBL (red circle) tended to closely group with the control samples, illustrating the challenge of developing biomarkers for the drugs that do not induce elevation of ALT or TBL in the treated rats.

It is necessary to point out that this study is driven by the availability of the data. For example, the choice of drugs was driven by availability of all three sets of data (i.e., human hepatotoxicity, gene expression from rat liver, and clinical chemistry measure in rats) rather by selection of a training set of compounds fit for purpose to test the hypotheses. In addition, the studied TGx data do not necessary represent the typical TGx design for study of liver injury. We suspect that certain forms of DILI may well be associated with gene expression that is more informative than ALT or TBL. Some of limitations in this study could have masked such associations. Therefore, a systematic evaluation of a large number of drugs with a well-defined DILI classification in different TGx settings is warranted to further verify the findings and advance our understanding of the role of preclinical TGx in DILI assessment for humans.

■ ASSOCIATED CONTENT

Supporting Information

Drugs and associated data used in this study (Table S1) and figure mentioned in the manuscript. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: 870-543-7142. Fax: 870-543-7854. E-mail: weida.tong@fda.hhs.gov.

Author Contributions

M.Z. developed the methods, performed most analysis, and wrote the first draft of manuscript. M.C. performed analysis related to validation study. W.T. had the original idea and guided the data analysis and presentation of results and finalized the manuscript. All of the authors read and approved the final manuscript.

Funding

M.Z. is grateful to the National Center for Toxicological Research (NCTR) of the U.S. Food and Drug Administration (FDA) for postdoctoral support through the Oak Ridge Institute for Science and Education (ORISE). This study was supported by the FDA's Critical Path Initiative, Office of Women's Health, and the Chief Scientist Challenge Grant.

Disclosure

The views presented in this article do not necessarily reflect those of the U.S. Food and Drug Administration.

■ REFERENCES

(1) Lee, W. M. (2003) Drug-induced hepatotoxicity. *N. Engl. J. Med. 349* (5), 474−485.

(2) Kaplowitz, N. (2005) Idiosyncratic drug hepatotoxicity. *Nat. Rev. Drug Discovery 4* (6), 489−499.

(3) Olson, H., et al. (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol. 32* (1), 56−67.

(4) Peters, T. S. (2005) Do preclinical testing strategies help predict human hepatotoxic potentials? *Toxicol. Pathol. 33* (1), 146−154.

(5) Abboud, G., and Kaplowitz, N. (2007) Drug-induced liver injury. *Drug Saf. 30* (4), 277−294.

(6) Ozer, J. S., et al. (2009) Enhancing the utility of alanine aminotransferase as a reference standard biomarker for drug-induced liver injury. *Regul. Toxicol. Pharmacol. 56* (3), 237−246.

(7) Senior, J. R. Drug hepatotoxicity from a regulatory perspective. *Clin. Liver Dis.* 2007, *11* (3), 507−524, vi.

(8) Ennulat, D., et al. (2010) Diagnostic performance of traditional hepatobiliary biomarkers of drug-induced liver injury in the rat. *Toxicol. Sci. 116* (2), 397−412.

(9) Llanos, L., et al. (2010) The existence of a relationship between increased serum alanine aminotransferase levels detected in premarketing clinical trials and postmarketing published hepatotoxicity case reports. *Aliment. Pharmacol. Ther. 31* (12), 1337−1345.

(10) FDA. *Guidance for Industry: Drug-Induced Liver Injury: Premarketing Clinical Evaluation*. Available from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174090.pdf, 2009.

(11) Committee for Proprietary Medicinal Products (CPMP). *CPMP/ICH/375/95. Note for Guidance on Population Exposure: The Extent of Population Exposure to Assess Clinical Safety*; European Medicines Agency: London, 1995.

(12) Navarro, V. J., and Senior, J. R. (2006) Drug-related hepatotoxicity. *N. Engl. J. Med. 354* (7), 731−739.

(13) Wang, K., et al. (2009) Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc. Natl. Acad. Sci. U.S.A. 106* (11), 4402−4407.

(14) O'Connell, T. M., and Watkins, P. B. (2010) The application of metabonomics to predict drug-induced liver injury. *Clin. Pharmacol. Ther. 88* (3), 394−399.

(15) Huang, J., et al. (2010) Genomic indicators in the blood predict drug-induced liver injury. *Pharmacogenomics J. 10* (4), 267−277.

(16) Hirode, M., et al. (2008) Gene expression profiling in rat liver treated with compounds inducing phospholipidosis. *Toxicol. Appl. Pharmacol. 229* (3), 290−299.

(17) Kiyosawa, N., et al. (2008) Identification of glutathione depletion-responsive genes using phorone-treated rat liver. *J. Toxicol. Sci. 32* (5), 469−486.

(18) Blomme, E. A., Yang, Y., and Waring, J. F. (2008) Use of toxicogenomics to understand mechanisms of drug-induced hepatotoxicity during drug discovery and development. *Toxicol. Lett. 186* (1), 22−31.

(19) Cui, Y., and Paules, R. S. (2010) Use of transcriptomics in understanding mechanisms of drug induced toxicity. *Pharmacogenomics 11* (4), 573−585.
(20) Huang, T., et al. (2009) Prediction of pharmacological and xenobiotic responses to drugs based on time course gene expression profiles. *PLoS One 4* (12), e8126.
(21) Ulrich, R., and Friend, S. H. (2002) Toxicogenomics and drug discovery: Will new technologies help us produce better drugs? *Nat. Rev. Drug Discovery 1* (1), 84−88.
(22) Martignoni, M., Groothuis, G. M., and de Kanter, R. (2006) Species differences between mouse, rat, dog, monkey and human CYP-mediated drug metabolism, inhibition and induction. *Expert Opin. Drug Metab. Toxicol. 2* (6), 875−894.
(23) Afshari, C. A., Hamadeh, H. K., and Bushel, P. R. (2011) The evolution of bioinformatics in toxicology: Advancing toxicogenomics. *Toxicol. Sci. 120* (Suppl. 1), S225−S237.
(24) Heinloth, A. N., et al. (2004) Gene expression profiling of rat livers reveals indicators of potential adverse effects. *Toxicol. Sci. 80* (1), 193−202.
(25) Mori, Y., et al. (2010) Identification of potential genomic biomarkers for early detection of chemically induced cardiotoxicity in rats. *Toxicology 271* (1−2), 36−44.
(26) Wallace, M. J., et al. (2009) Early biomarkers and potential mediators of ventilation-induced lung injury in very preterm lambs. *Respir. Res. 10*, 19.
(27) Searfoss, G. H., Ryan, T. P., and Jolly, R. A. (2005) The role of transcriptome analysis in pre-clinical toxicology. *Curr. Mol. Med. 5* (1), 53−64.
(28) Ge, F., and He, Q. Y. (2009) Genomic and proteomic approaches for predicting toxicity and adverse drug reactions. *Expert Opin. Drug Metab. Toxicol. 5* (1), 29−37.
(29) Wang, E. J., et al. (2008) Validation of putative genomic biomarkers of nephrotoxicity in rats. *Toxicology 246* (2−3), 91−100.
(30) Bushel, P. R., et al. (2007) Blood gene expression signatures predict exposure levels. *Proc. Natl. Acad. Sci. U.S.A. 104* (46), 18211−18216.
(31) Chen, M., et al. (2011) FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discovery Today 16* (15−16), 697−703.
(32) Natsoulis, G., et al. (2008) The liver pharmacological and xenobiotic gene response repertoire. *Mol. Syst. Biol. 4*, 175.
(33) Boorman, G. A., et al. (2005) Hepatic gene expression changes throughout the day in the Fischer rat: implications for toxicogenomic experiments. *Toxicol. Sci. 86* (1), 185−193.
(34) MAQC Consortium (2010) The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol. 28* (8), 827−838.
(35) MAQC Consortium (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol. 24* (9), 1151−1161.
(36) Romanski, P. *R Package: FSelector*, 2009.
(37) Fan, X., et al. (2010) DNA microarrays are predictive of cancer prognosis: A re-evaluation. *Clin. Cancer Res. 16* (2), 629−636.
(38) http://www.ingenuity.com/.
(39) Chargui, A., et al. (2011) Cadmium-induced autophagy in rat kidney: An early biomarker of subtoxic exposure. *Toxicol. Sci. 121* (1), 31−42.
(40) Powell, C. L., et al. (2006) Phenotypic Anchoring of Acetaminophen-Induced Oxidative Stress with Gene Expression Profiles in Rat Liver. *Toxicol. Sci. 93* (1), 213−222.
(41) McMillian, M., et al. (2005) Drug-induced oxidative stress in rat liver from a toxicogenomics perspective. *Toxicol. Appl. Pharmacol. 207* (2 Suppl.), 171−178.
(42) Yuan, L., and Kaplowitz, N. (2008) Glutathione in liver diseases and hepatotoxicity. *Mol. Aspects Med. 30* (1−2), 29−41.
(43) Thomas, R. S., et al. (2009) Use of short-term transcriptional profiles to assess the long-term cancer-related safety of environmental and industrial chemicals. *Toxicol. Sci. 112* (2), 311−321.
(44) Kim, S. N., et al. (2007) Induction of hepatic CYP2E1 by a subtoxic dose of acetaminophen in rats: Increase in dichloromethane metabolism and carboxyhemoglobin elevation. *Drug Metab. Dispos. 35* (10), 1754−1758.