Multicenter Study of Acetaminophen Hepatotoxicity Reveals the Importance of Biological Endpoints in Genomic Analyses


*University of Washington, Seattle, Washington 98195; ‡Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; §Oregon Health & Science University, Portland, Oregon 97239; ||University of North Carolina, Chapel Hill, North Carolina 27599; *National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709; ||Cogenics, a Division of Clinical Data, Inc., Morrisville, North Carolina 27560; |||SAS Institute, Inc., Cary, North Carolina 27513; ||||Duke University Medical Center, Durham, North Carolina 27710; and *Fred Hutchinson Cancer Research Center, Seattle, Washington 98109

Received April 19, 2007; accepted June 5, 2007

Gene expression profiling is a widely used technique with data from the majority of published microarray studies being publicly available. These data are being used for meta-analyses and in silico discovery; however, the comparability of toxicogenomic data generated in multiple laboratories has not been critically evaluated. Using the power of prospective multilaboratory investigations, seven centers individually conducted a common toxicogenomic experiment designed to advance understanding of molecular pathways perturbed in liver by an acute toxic dose of N-acetyl-p-aminophenol (APAP) and to uncover reproducible genomic signatures of APAP-induced toxicity. The nonhepatotoxic APAP isomer N-acetyl-m-aminophenol was used to identify gene expression changes unique to APAP. Our data show that c-Myc is induced by APAP and that c-Myc–centered interactomes are the most significant networks of proteins associated with liver injury. Furthermore, sources of error and data variability among Centers and methods to accommodate this variability were identified by coupling gene expression with extensive toxicological evaluation of the toxic responses. We show that phenotypic anchoring of gene expression data is required for biologically meaningful analysis of toxicogenomic experiments.

Key Words: liver injury; toxicogenomics; phenotypic anchoring.

The biology of liver and other tissues in normal and disease states increasingly is being probed using global approaches such as microarray transcriptional profiling. Acceptance of this technology is based principally on a satisfactory level of reproducibility of data among laboratories and across platforms (Bammler et al., 2005; Shi et al., 2006). Achieving reasonable concordance in expression data among laboratories requires standardized protocols for RNA labeling, hybridization, microarray processing, data acquisition, and normalization (Bammler et al., 2005).

The majority of standardization efforts have focused on (1) comparing gene expression data from the analysis of reference samples distributed among several laboratories (Bammler et al., 2005; Pennie et al., 2004), or (2) developing guidelines for documenting experimental details and results when publishing microarray data (Ball et al., 2002). The issue of reproducibility and reliability of genomics data obtained from similar (standardized) biological experiments performed in different laboratories is crucial to the generation and utility of large databases of microarray results (Becker, 2001; Miles, 2001). While several recent studies uncovered important limitations of expression profiling of chemical injury to cells and tissues (Baker et al., 2004; Beekman et al., 2006; Ulrich et al., 2004), determining the effects of intralaboratory variables on the reproducibility, validity, and general applicability of the results that are generated by different laboratories and deposited into publicly available databases remains a gap.

The National Institute of Environmental Health Sciences (NIEHS) established the Toxicogenomics Research Consortium to apply the collective and specialized expertise from academic institutions to address issues in integrating gene expression profiling, bioinformatics, and general toxicology. Key elements include developing standardized practices for...
gene expression studies, and conducting systematic assessments of the reproducibility of traditional toxicity endpoints and microarray data within and among laboratories. To this end, the consortium selected the classical hepatotoxicant acetaminophen (APAP) (James et al., 2003; Kaplowitz, 2005), for its proof-of-concept experiments. Despite more than 30 years of research on APAP, we are far from a complete understanding of the mechanisms of liver injury, risk factors, and molecular markers that predict clinical outcome after poisoning (Watkins et al., 2006).

To evaluate both mechanisms and interlaboratory variation in injury response, the toxicogenomic analysis of APAP-induced hepatotoxicity was performed at seven geographically dispersed Centers. Parallel studies with N-acetyl-\(m\)-aminophenol (AMAP), the nonhepatotoxic isomer of APAP, provided a method to isolate transcripts associated with hepatotoxicity. This study identified potential sources of interlaboratory variability when microarray analyses are conducted on samples generated in different laboratories. The results also show that phenotypic anchoring of gene expression data is required for biologically meaningful meta-analysis of genomic experiments. This study reports an extensive temporal pathway map of APAP- and AMAP-specific, as well as aminophenol-nonspecific responses in mouse liver, and uncovers a central role of Myc-regulated signaling networks in APAP-induced acute liver injury. Furthermore, we demonstrate that transcriptional responses in livers of mice and rats are similar when equitoxic doses of APAP are considered.

**MATERIALS AND METHODS**

**Animals and treatments.** To minimize the potential effects of environmental and genetic variables, all participating laboratories received animals, feed, bedding, and test article preparations from centralized sources. Each participating laboratory used 45 10- to 11-week-old male isogenic C57BL/6J mice (Jackson Labs, Bar Harbor, ME, or Sacramento, CA). The mice were housed individually in an environmentally controlled area with temperature (69–75°F), humidity control (35–65%), and a 12-h light/dark cycle under pathogen-free conditions. The bedding and animal chow used by all laboratories were prepared at Battelle and supplied to all laboratories to assure consistency among the treatment centers. Flash frozen, and stored at −80°C until analyses. Leftover solutions of the vehicle and test chemicals, serum samples, as well as liver samples designated for histopathology and glutathione measurements, were shipped on dry ice to Battelle for further processing. RNA was extracted from each liver sample at the treatment center, flash frozen, and stored at −80°C for shipment to Paradigm Array Labs (Now Cogenics, a Division of Clinical Data, Inc., Morrisville, NC) where microarray experiments were performed. All animal studies for this project were approved by each Institution’s respective Animal Care and Use Committee prior to the start of the study.

**Clinical chemistry.** All serum samples were analyzed with a centrifugal analyzer (Hitachi 911, Chula Vista, CA) at Battelle using procedures supplied by the manufacturer. Variables analyzed included alanine aminotransferase (ALT), aspartate aminotransferase (AST), bilirubin, and bile salts/acid (ALT, historical control values from male B6C3F1 mice: 51 ± 25 U/l, minimum/maximum: 21/112 U/l, n = 34), aspartate aminotransferase (no historical control available), sorbitol dehydrogenase (30 ± 9 U/l, min/max: 20/63 U/l, n = 36), and bile salts/acid (23 ± 5 μM/l, min/max: 15/35 μM/l, n = 35). The complete data set is provided in Supplemental Table 2. Serum enzyme measurements were performed on a small number of animals (15 in total) from which blood samples were not available.

**Histological methods.** Liver tissue samples were embedded in paraffin, sectioned, stained with hematoxylin and eosin (H&E), and examined for pathological changes at Battelle where the diagnoses were subject to Quality assurance and pathology peer review following standard National Toxicology Program (NTP) procedure (Boorman et al., 2002). In addition to standard qualitative pathology reads of the liver sections, a semiquantitative method of estimating a fraction of necrotic tissue in histopathological sections of the liver was applied (Mauton, 2002). For this purpose, a grid with 100 evenly spaced points was overlaid on printed images of H&E-stained liver sections taken at ×100 magnification. All of the tissues available on each slide were evaluated. The percent liver necrosis score was determined as the number of points lying within necrotic areas divided by a total number of points within the tissue section. Histopathological sections from all Centers were assessed for percent liver necrosis score in a blinded fashion by one trained individual. These results were independently verified by a board certified veterinary pathologist. These data are provided in Supplemental Table 2.

**Hepatic glutathione analysis.** The method was based on the determination of thiols by high-performance liquid chromatography with fluorescence detection (Toyokawa et al., 2001). Liver samples (median and right anterior lobes) from individual animals were analyzed for reduced and total glutathione at Battelle. Briefly, 1 mg of liver was placed in a test tube with 1 ml of 5% trichloroacetic acid containing 5 mM ethylenediaminetetraacetic acid (EDTA), homogenized, vortexed, and centrifuged. Next, 50 μl of homogenate was transferred to a test tube with 950 μl of 5 mM EDTA containing 0.1 M borax (pH 9.3). Then, 100-μl aliquots were placed in duplicate tubes for each liver sample. Fluorobenzofurazan-4-sulfonic acid ammonium salt was added for reduced glutathione and tri-butylinphosphate in acetonitrile was added to the other tube for total glutathione. High-performance liquid chromatography with fluorescence detection was used to determine total and reduced glutathione and compared against calibration standards that were prepared with each assay. These data are provided in Supplemental Table 2.
RNA isolation. Total RNA was isolated at each Center according to the standardized protocol. Frozen liver samples (~30 mg from the left lateral lobe) of four randomly selected mice from each study group were homogenized in the extraction buffer supplied in the RNeasy mini kit (Qiagen, Valencia, CA). Then, RNA was isolated using the spin column technology as detailed by the manufacturer. RNA quality and quantity were determined spectrophotometrically from the absorbance at 260 and 280 nm. Aliquots of 25 μg RNA were frozen at ~70°C and shipped to Cogenics for microarray analysis.

Microarray hybridizations. All gene expression profiling was performed at Cogenics. One microgram of total RNA from individual mouse liver samples was amplified and labeled with a fluorescent dye (Cy3), and a common reference—pooled mouse liver messenger RNA (mRNA) isolated from the livers of 100 male C57BL/6J mice (see Bannmler et al., 2005 for details)—was amplified and labeled with Cy5 using Agilent Technologies (Palo Alto, CA). Low RNA Input Linear Amplification labeling kit following the manufacturer’s protocol. The quantity of the resulting fluorescently labeled cRNA was evaluated using a Nanodrop ND-100 spectrophotometer (Nanodrop Technologies, Wilmington, DE) and its integrity measured using an Agilent Bioanalyzer. Equal amounts of Cy3 and Cy5-labeled RNA (750 ng) from the individual animals and from the pooled control, respectively, were hybridized to an Agilent Mouse Oligo Microarray (~21,000 features, catalogue# G4121) for 17 h at 65°C. The hybridized microarrays were washed and scanned using an Agilent G2565BA scanner. Data were extracted from the scanned image using Agilent Feature Extraction software version 7.5. There were two noted study protocol deviations due to technical issues: Center #6 supplied RNA samples for only one time point (6 h), and two samples from AMAP-treated animals (one each for 12- and 24-h time points) from Center #3 were not included in follow-up analyses due to poor microarray data quality as determined by Feature Extraction software (excessive background signal intensity and large number of saturated features).

Data analysis. Raw microarray data were processed and analyzed with tools in the Bioconductor (Gentleman et al., 2004) software package. Following the recommendations of the Toxicogenomics Research Consortium’s standardization experiment (Bannmler et al., 2005; Qin and Kerr, 2004), median Cy3 and Cy5 fluorescence intensities for each spot on the array were used as the signal and no background adjustment was performed. Two microarrays (Center 3) were excluded from further processing and analysis due to poor quality. Intensities flagged as saturated signal by the Feature Extraction software (Agilent Technologies) were treated as missing values. This resulted in less than 1% missing data per array (median 0.1%, maximum 0.9%). Cy3/Cy5 ratios were generated for each spot and these values were normalized for intensity and spatially dependent systematic variation by regional lowess smoothing (Cui et al., 2003). This data set (raw and normalized data files) is publicly available from cebi.niehs.nih.gov (accession number 009-00001-0010-000-1). Using the normalized data, genes with significant evidence for differential expression were identified using the Significance Analysis of Microarrays (SAM) methodology (Tusher et al., 2001), as implemented in the Bioconductor package “samr.” The k-nearest-neighbor method was used to impute missing values (Troyanskaya et al., 2001). For each of the three time points (6, 12, and 24 h) we compared each pair of the three treatment groups for a total of nine pairwise group comparisons. For each pairwise comparison, we treated “Center” as a blocking factor and used 500 permutations (permuting the treatment labels within Center) to estimate the reference null distributions for the significance tests.

Correlations between liver toxicity and gene expression. The Spearman correlation coefficient between gene expression and percent liver necrosis was computed separately for each Center. A weighted median (Birkes and Dodge, 1993) of these individual correlation coefficients was computed with weights proportional to the sample size from each Center (giving greater weight to correlation coefficients from Centers with more data). Significance for this (weighted) median correlation was determined by an approximate permutation test (199 random permutations plus the original configuration). The p value summarizes the proportion of random permutations that produced a (weighted) median correlation coefficient at least as extreme as the one observed. The number of permutations considered led to a resolution for the p value of 1/200 = 0.005; a 95% confidence interval (Wilson’s interval; Agresti and Coull, 1998) for p was computed to provide limits on the actual p value. A (weighted) median correlation was deemed significant if p < 0.01 together with an upper confidence limit for p strictly less than 0.05, a condition that provides some assurance that a higher number of permutations would still produce small p values. The dashed-gray lines on Figure 4a show a robust linear fit (Frees, 1991) applied to each Center’s data.

Variability analysis. An analysis of the variability of the microarray data was conducted by fitting a random effects analysis of variance (ANOVA) model on a feature-by-feature basis to the normalized log ratios (Jin et al., 2001; Wolfinger et al., 2001). Eight variance components in the model correspond to treatment, time, Center, their two- and three-way interactions, and the residual variance. Since each animal was observed only once, the latter is a confounded effect of animal-to-animal and technical variability. All effects are considered random in this analysis in order to obtain a balanced partitioning of estimated total variability. All effects are assumed to be normally distributed, and variance components were estimated using restricted maximum likelihood via JMP Microarray software (SAS Institute Inc., Cary, NC).

Analysis of biological networks. Network analysis was performed on lists of genes generated from the earlier analyses described above as being differentially expressed. Ingenuity Pathways Knowledge Base software (Ingenuity Systems, Inc., Redwood City, CA) was used for the analysis of significant interactions. Ingenuity Pathways Knowledge Base is a database that contains data of individually curated relationships between gene objects (e.g., genes, mRNAs, and proteins) to generate significant biological networks and pathways. A significance score (p value) for each network and/or canonical pathway is calculated according to the fit of the test gene set versus genes in the particular pathway. The score of each identified network is calculated as the negative log p value which indicates the likelihood that the genes in a test gene list can be found together in a network due to random chance. In this study, a score corresponding to p < 10^-20 or lower was used to select highly significant biological networks.

Gene ontology analysis. Lists of genes generated by various statistical analyses of the microarray data were submitted to the GoMiner web-based (http://discover.nci.nih.gov/gominer/) tool (Zeeberg et al., 2003) for identification of the biological processes that are overrepresented. Statistical significance of the biological process enrichment was determined using Fisher’s exact p value based on the relative overrepresentation of genes within a given gene ontology (GO) category as compared to a random sample of genes probed on the array that was used in these experiments (see above). GO pathways with a p < 0.05 and a minimum of five gene transcripts represented from that category were considered for further analysis.

Species comparison of gene expression and toxicity data. Publicly available (ntp.niehs.nih.gov/go/15716) gene expression and liver toxicity data generated from F344/N rats that were exposed to either subtoxic (50 or 150 mg/kg) or toxic (1500 or 2000 mg/kg) doses of APAP for 6, 18, 24, or 48 h were used. In order to identify APAP-responsive genes in both species, each of the probes present on the Rat Oligonucleotide Array (Agilent G4130A) used in the above study, and the Mouse Oligonucleotide Array used in this work, were mapped to Entrez Gene IDs (where possible). Homologous genes present on both microarray platforms were identified by using National Center for Biotechnology Information HomoloGene tool by requiring that both Entrez Gene IDs provided reciprocal BLAST hits. This resulted in the identification in 3076 candidate homologs that were measured on both microarrays. These results were further restricted to 962 candidates by selecting genes with a twofold or higher APAP-induced expression difference (as compared to time-matched vehicle control animals separately for each species/time/dose group) and a log ratio p value less than 0.001 in at least six treated animals (either rat or mouse, see Supplemental Table 7 for a complete list of transcripts).
RESULTS

Reducing Variability and Assuring Confidence in Gene Expression Studies

This study was designed to (1) advance understanding of the molecular pathways perturbed in the liver by a toxic dose of APAP by contrasting toxic compound-specific responses with those of a nontoxic structural analog (AMAP), and (2) apply the power of a multiple laboratory investigation to uncover a stereotypic genomic signature of APAP-induced toxicity. The experiment achieved the overall goal of inducing liver injury with APAP (300 mg/kg, ig), while both vehicle and AMAP showed no adverse effects (Figs. 1a and 1b, Supplemental Table 2). Greater than 20% necrosis was observed in ~55% of the APAP-treated animals at each time point and correlated well ($r_{\text{Pearson}} = 0.846$) with serum ALT (Fig. 1c). However, toxic injury induced by APAP showed variability across Centers and animals. Animal-to-animal variability within individual Centers was less than variation among centers, and one Center (#2) did not induce liver injury in 12 out of 15 mice treated with APAP. Furthermore, serum ALT levels in vehicle-treated animals from Center 3 were elevated, about threefold higher than in other Centers. Other clinical chemistry data largely followed the trends of changes in ALT (Supplemental Table 2). A decrease in reduced glutathione content at 6 h following APAP treatment was observed (Fig. 1d, samples from Center #2 were not provided). However, the magnitude of glutathione depletion varied considerably from < 5% (Center #5)

![FIG. 1. Toxicological assessment of APAP-induced injury in mouse liver. Severity and time-course (6, 12, and 24 h) of liver necrosis (a) and elevations in serum ALT activity (b) associated with APAP treatment across Centers that performed the in vivo portion of the study. Mean values for up to five animals per treatment group are shown for APAP-treated animals. An average of up to 15 measurements (all time points) is shown for the vehicle category. (c) Correlation plot between liver necrosis and serum ALT measurements at the level of the individual animal. Data from vehicle- and APAP-treated animals (all centers combined) are plotted according to the time point (● = 6 h, + = 12 h, × = 24 h). (d) Fraction (mean values for up to five animals) of reduced glutathione (GSH, reported as % of glutathione content in vehicle-treated mice at each corresponding time point) in livers of APAP-treated animals. See Supplemental Table 2 for a complete data table on the clinical chemistry and histopathology measurements collected in this study.]
A weak negative correlation was observed between the fraction of the reduced glutathione and necrosis ($r_{\text{Pearson}}$: $-0.537$ [6 h], $-0.413$ [12 h], and $-0.377$ [24 h]); however, it should be noted that glutathione depletion in mouse liver after a hepatotoxic dose of APAP is known to occur within 60 min of dosing (Mitchell et al., 1973) and the time points in this study may be less informative.

The statistical significance of changes in gene expression was determined in pairwise comparisons between treatments for each time point and a Venn diagram of the overlap was constructed (Fig. 2a, Table 1, Supplemental Table 3). Genes in the m5 sector, those differentially expressed between APAP- and vehicle-treated mice and also between the APAP- and AMAP-treated mice, were considered to represent an APAP-selective transcriptional response in mouse liver. Thus, a non-redundant list of genes (7806 transcripts) whose expression changed significantly and specifically due to APAP treatment at any of the time points was compiled and the transcriptional response was visualized (Fig. 2b, top). When measurements of liver necrosis and serum ALT in the same animals are also displayed in a yellow–blue heat map (Fig. 2b), a pronounced concordance between APAP hepatotoxicity and transcriptional response is evident.

To assess the relative contributions of the different sources of biological, technical, and experimental variability, we fitted a random effects ANOVA model separately to data from each feature in the normalized microarray data. Through this statistical test, overall observed variability for each feature on the microarray was partitioned into components corresponding to the effects of treatment, time, Center, two- and three-way interactions, and the residual variability (a combination of biological and technical variability). Figure 3 shows the distribution of $R^2$ (the proportion of total variability explained for each feature), as well as box plots for each of the eight sources of variability detailed above. Results are shown for (1) all 21,318 features on the Agilent array (Figs. 3a and 3b), and (2) the 7806 transcripts (see Fig. 2) representing the APAP-specific expression profile in liver (Figs. 3a inset and 3c). When all transcripts on the array are considered, $R^2$ distribution is bimodal and centered on 20% and 70% of explained variability. On average, less than half of the overall variability observed in this standardization experiment can be attributed to the effects of time, treatments, and Center, or their combinations. However, genes in the selectively APAP-hepatotoxic signature fall clearly in the upper $R^2$ mode where considerably more variation is due to the effects of time,
treatments, and Center, or their combinations. Furthermore, the residual variability in these APAP hepatotoxicity–associated genes is reduced to 40% and treatment- and Center-specific effects are more pronounced than those of time and the two- and three-way interactions.

### Probing Mechanisms by Linking Gene Expression with Phenotypic Endpoints

A Spearman correlation analysis of all features on the array identified 3505 (6 h), 6085 (12 h), and 5988 (24 h) transcripts with a significant association between gene expression and percent liver necrosis (Table 1, Fig. 4a, and Supplemental Table 4). The majority of these transcripts are found in sector m5 (75% [6 h], 57% [12 h], and 72% [24 h]). Figure 4a graphically illustrates the Spearman correlation-based analysis by displaying two representative examples of the positive and negative association between gene expression and the ‘phenotypic anchor’—percent liver necrosis—for two genes known to be involved in the mechanism of liver toxicity by APAP. The dashed-gray lines show a robust linear fit applied to each Center’s data and provide a visual guide of the consistency of the association.

Data mining of this signature using GO annotations was used to assign functional classes and to determine the biological processes defining the APAP-selective transcriptional response associated with necrosis (Table 1, p < 0.01 columns). These lists were segregated by time point and into positively and negatively correlating groups. Temporal changes in biological processes identified by this analysis are shown in Figures 4b and 4c and Supplemental Table 5. To further characterize the APAP-selective hepatotoxic response, we used the Ingenuity protein interactome analysis (Supplemental Table 6). In all time points, Myc-regulated interactomes were identified as the most significant (p < 10^{-37}, Fig. 4d).

Furthermore, a majority of the members of the Myc-centric interactomes were shared between time points with most pronounced overlap at 12 and 24 h after APAP treatment.

### Determining Toxicant-Specific Gene Expression through Structural Analognes

Since it has been suggested that gene expression profiling may mis-classify transcriptional changes occurring in response to xenobiotic exposure as signs of toxicity (Hayes and Bradfield, 2005), this experiment also included a treatment with the nontoxic stereo-isomer of APAP, AMAP. AMAP caused no observable adverse health outcomes in any of the seven Centers; however, it did evoke an AMAP-selective transcriptional response (sector m3, Fig. 2b, middle and Table 1). While the total number of transcripts modulated by AMAP is much smaller than those changed by APAP, GO-based analysis of the biological processes identified several significant categories at 12 and 24 h (Fig. 4e). Further analysis identified genes that respond to both APAP and AMAP similarly. These 397 transcripts (sector m6, Fig. 2b, bottom and Table 1) comprise a nonspecific response to aminophenol-like xenobiotics and thus should not be included in the analysis of the drug-specific response. These genes lack a consistent pattern in expression changes, nonetheless catabolism and protein localization appear to be small but significant biological processes (Fig. 4f).

### Validating Biological Responses across Rodent Models

To compare the transcriptional responses to hepatotoxic doses of APAP between mouse and rat we combined the gene expression data obtained in this study to previously published data (http://ntp.niehs.nih.gov/go/15716) on expression profiling of rat liver samples from time-course (2–48 h) and dose–response (50, 150, 1500, and 2000 mg/kg) analysis of APAP.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>6 h</th>
<th>12 h</th>
<th>24 h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Significance of correlation</td>
<td>Total # of transcripts</td>
<td>Significance of correlation</td>
</tr>
<tr>
<td>m0</td>
<td>p* &lt; 0.01</td>
<td>314</td>
<td>14,658</td>
</tr>
<tr>
<td>m1</td>
<td>p &gt; 0.01</td>
<td>1088</td>
<td>1400</td>
</tr>
<tr>
<td>m2</td>
<td>1</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>m3</td>
<td>11</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>m4</td>
<td>209</td>
<td>867</td>
<td>1076</td>
</tr>
<tr>
<td>m5</td>
<td>2619</td>
<td>1116</td>
<td>3735</td>
</tr>
<tr>
<td>m6</td>
<td>3</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td>m7</td>
<td>36</td>
<td>11</td>
<td>47</td>
</tr>
<tr>
<td>Total</td>
<td>3505</td>
<td>17,813</td>
<td>21,318</td>
</tr>
</tbody>
</table>

**Note.** The number of genes that are significantly associated (p < 0.01 and *upper confidence interval ≤ 0.05, see Fig. 3 for examples) between gene expression and percent necrosis in the liver at each time point (6, 12, or 24 h) and within a particular m sector (see Fig. 2a) is shown.
Even though the data sets were generated using two different microarrays, albeit of the same platform and at the same reference laboratory, it was possible to compare and contrast gene expression results at the level of individual genes (Fig. 5). A yellow–blue heat map in Figure 5 displays measurements of necrosis and serum ALT activity from the same animals in both data sets (Supplemental Table 8). Similar to the data that were obtained with mice (Fig. 2b), a high degree of concordance between APAP-induced toxicity and transcriptional response is evident in individual animals, both mice and rats. Seventy-one transcripts ($r = 0.83$) constitute a species-independent universal transcriptional response to APAP (Fig. 5, red box). The transcripts for both Myc and Jun oncogenes are located within this signature. Several GO processes were downregulated following treatment with APAP as exemplified by a set of 52 transcripts ($r = 0.76$, Fig. 5, green box).

**DISCUSSION**

The use of genomics has created both opportunities and challenges. The concept of gene expression profiles as signatures of toxicant classes, disease subtypes, or other biological endpoints (Hamadeh et al., 2002a,b), together with public availability of expression data from multiple studies, has spurred the development of multiple databases and meta-analysis for comparative study of toxicity based on reference compounds (Waters et al., 2003). Disparate data streams may include gene expression, body and organ-weight measurements, clinical chemistry, and histopathology findings for several tissues and require careful collection, management, and integration (Mattes et al., 2004). While it has been recognized that validated biological endpoints, phenotypic anchors, are essential for interpreting toxicological outcomes (Paules, 2003; Powell et al., 2006), no study has systematically approached the issue of the variability in expression and toxicity data through multilaboratory comparisons.

The effects of early and/or low-dose effects of APAP in the rodent liver were considered recently and it was shown that gene expression is a sensitive means of identifying indicators of adverse effects that precede overt toxicity (Heinloth et al., 2004). The design of the current study was weighted toward uncovering robust genomic signatures of APAP-induced toxicity that were reproducible in studies conducted across multiple laboratories by focusing on doses and time points that were selected to result in a clear phenotype (i.e., liver injury) that would facilitate anchoring pathobiological component estimates, standardized ($\sum = 1$) within each transcript (21,318 transcripts). (c) Box plots of standardized variance component estimates for the 7806 m5 transcripts. Animal/Tech—the residual variance, which is a combination of animal-to-animal and technical variability. Other variance components correspond to the effects of Treatment, Time, Center, and their two- and three-way interactions.
endpoints with microarray data. Furthermore, this study firmly establishes that careful consideration of the traditional toxicity measurements, so-called “phenotypic anchors,” adds confidence to the interpretation of data from novel technologies, such as gene expression profiling. However, while “phenotypic anchoring” provides insights regarding mechanism of

FIG. 4. Phenotypic anchoring of gene expression and liver pathology reveals common biological pathways across Centers. (a) Plots show examples of the positive (top, glutamate-cysteine ligase, catalytic subunit, Gclc) and negative (bottom, cytochrome P450 2e1, Cyp2e1) association between gene expression and necrosis in the liver at 6 h after treatment. Points show individual animal’s response to APAP (red), AMAP (blue), or vehicle (green) treatment; gray dotted lines show a robust linear fit to data from each of seven Centers. The (weighted) median Spearman correlation value ($\tilde{r}_s$) is indicated together with a $p$ value determined from a permutation test (95% confidence interval for $p$ is provided in parentheses). (b, c) Gene ontology analysis of transcripts in m5 sector that significantly correlate with liver necrosis. The number of genes associated with each process is given in parentheses. Biological pathways affected in all time points are highlighted by a box. (d) Ingenuity pathway analysis of transcripts in m5 sector that significantly correlate (positively, red, or negatively, green) with liver necrosis identifies Myc-centered interactome as the top scoring network. Significant network interactions were identified for gene sets (same as in c and d) representing 6-h post-APAP treatment (inner blue circle), 12-h post-APAP treatment (middle circle), and 24-h post-APAP treatment (outer circle). Temporality of the response is displayed by line connections. Proteins connected to Myc by lines were found to be common between two and three time points. Solid lines represent physical interactions (i.e., binding), whereas arrows represent any other type of indirect cellular interactions. Proteins that are not connected to Myc were found in the interactomes specific for each particular time point. Different shapes of the nodes represent functional classification of the proteins that are shown: square, cytokines; vertical rhombus, enzymes; horizontal rhombus, peptides; triangle, phosphatases; inverted triangle, kinases; vertical oval, transmembrane receptors; horizontal oval, transcription regulators; trapezoid, transporters; circle, other. (e, f) Gene ontology analysis of transcripts significantly correlating with AMAP (m3) or aminophenol-like compound response (m6), respectively.
Despite our best efforts to standardize the study a considerable degree of animal-to-animal variability in the extent of liver injury was observed. While the overwhelming majority of studies in the literature report data on toxic effects in the form of mean ± standard error, we chose to consider and display clinical chemistry and histopathology data for individual animals, which clearly demonstrates that a treatment-specific gene expression signature is only present in subjects that exhibited the pathologic effect of treatment. Furthermore, our study used a correlation-based analytical approach to associate the extent of liver toxicity caused by APAP (as % liver necrosis) with gene expression. We argue that such correlation-based analytical approach to associate the extent of liver toxicity with gene expression is especially powerful for the selection of genes that change in expression at a certain statistical significance and that respond to treatment and/or pathophysiological endpoints. Thus, despite the lack of evident toxicity in some APAP-treated animals, a signature could be determined and used with confidence to map gene expression profiles to pathways.

Previous studies indicated that comparability of gene expression experiments among laboratories can be affected by numerous technical variables (Bammler et al., 2005). An additional layer of complexity arises when entire biological experiments are performed in parallel, introducing additional sources of variability stemming from technical, genetic, epigenetic, or stochastic factors. While this study did not consider effects of genetics (since all mice were from an inbred line) or microarray platform (since all experiments were performed on the same commercial platform at a centralized location), we found that the largest source of variability in a real-life biological experiment was technical/animal whereby animal-to-animal technical or biological deviations affected the outcome. Possible contributors to this variability include differences in animal husbandry, duration, and dose of anesthesia, methods, and timing of blood draws, timing of tissue excision, storage, and preparation, etc. Overall, this result underscores the convolution of biological responses in complex systems and the importance of taking each step of the experiment into consideration. Inevitably, the high sensitivity of new assays creates additional noise in the measurements and makes it more challenging to extract meaningful knowledge with confidence. However, when only those genes that constitute a phenotypically anchored signature are considered, as shown in this work, the relative proportion of uncontrolled variability diminishes. Thus, our results argue strongly that the biological response–based filters increase confidence and reduce variability in the microarray data even if animal-to-animal variability is considerable.

This work, with the added confidence of a multilaboratory-based investigation, reveals a number of biological pathways that are responsive to APAP in mouse liver; however, current study design does not allow one to conclude that other centrilobular hepatotoxics might produce a similar or overlapping pattern of gene expression as seen for APAP (i.e., the specificity of the response). Furthermore, our analysis identifies genes that are not only consistently up-/down-regulated, but also are significantly correlated with the toxicity...
phenotype. While this work considered transcriptional responses to APAP-induced liver injury and did not include proteomic analysis, we argue that gene expression changes are key secondary indicators of the cell’s response to stress and thus can be used successfully to identify important biological pathways involved in causation of, or response to, tissue damage. It has been postulated that loss of mitochondrial function (Donnelly et al., 1994) and concomitant generation of oxidative stress (Jaeschke, 1990), as well as activation of inflammatory cytokine and chemokine cascades, are all involved in APAP-induced hepatotoxicity (James et al., 2003; Kaplowitz, 2005; Watkins et al., 2006). We show that APAP-related induction of programmed cell death accompanied by inflammatory response and the parallel upregulation of the cell cycle, activation of kinases, and metabolism are prominent at all time points considered in this study. A number of inflammatory cells and mediator molecules are thought to be involved in parenchymal cell damage. Neutrophils are an essential component of the innate immune response at both the initiation and progression levels. Recently, direct evidence that neutrophil accumulation in the liver contributes to the progression and severity of APAP-induced liver injury was presented (Liu et al., 2006). Our data confirm that neutrophil chemotaxis (i.e., Fcgr3, Il1b2, and Fcer1g) is one of the significant biological processes that is induced by APAP treatment and correlates significantly with liver necrosis.

The c-Jun N-terminal kinase (Jnk) signaling pathway is primarily activated by proinflammatory cytokines and stress stimuli. In vitro and in vivo studies have shown that APAP induces a sustained activation of Jnk that contributes to hepatocellular necrosis (Gunawan et al., 2006). The propensity for Jnk to mediate cell death has been postulated to be mediated by the antagonistic cross-talk between Jnk and nuclear factor kappa B (NF-kB). APAP has been shown to inhibit NF-kB transactivation in the liver and in turn amplify Jnk tipping the scale from cell survival to cell death (Matsumaru et al., 2003). Our data show that MAP kinases, including members of the Jnk signaling pathway (Traf2, Mapk8ip3, and Dusp10), are positively associated with APAP-induced liver necrosis. As expected, this pathway is most active at 6 h when oxidative stress and the production of proinflammatory cytokines from the influx of inflammatory cells are at their peak.

Liver regeneration after damage depends on the ability of cells to progress through the cell cycle and proliferate (Fausto et al., 1995). In vitro, it has been shown that APAP interferes with the ability of cells to respond to growth factor stimuli since APAP prevented degradation of inhibitor of kB, thus blocking activation of NF-kB, phosphorylation of Raf1, and upregulation of Myc (Boulaires et al., 1999). While in vitro data should be considered with caution, data from this in vivo study also show that c-Myc is induced by APAP and that c-Myc–centered interactomes are the most significant networks of proteins associated with liver injury. Jun and Fos are also induced in liver by APAP treatment, and networks that encompass these two key regulators of AP-1–specific transcriptional response are second-most significant at all three time points. Indeed, the relationship between chemical stress and the synthesis of glutathione, including the role of these genes, transcription factors, and glutathione-synthesizing enzymes (e.g., glutamate-cysteine ligase) has been documented (Kitteringham et al., 2000).

Most of the significantly downregulated biological processes were found as early as 6 h after treatment with APAP. Lipid metabolism, electron transport, and cofactor metabolism are repressed at all time points. This observation likely reflects a temporary suspension of physiological functioning of the liver that occurs under conditions of massive necrosis. Indeed, it was previously suggested that transcript levels of enzymes involved in major energy-consuming biochemical pathways are downregulated after APAP treatment (Heinloth et al., 2004), and the mitochondrial dysfunction in liver after APAP treatment is well known (Hinson et al., 2004; Knight and Jaeschke, 2002; Ruepp et al., 2002).

Use of toxic and nontoxic structural analogues is a helpful method to isolate gene expression changes that are most likely to be associated with the phenotype of interest—in this case, liver damage. AMAP and APAP are structurally similar compounds and both are metabolized to reactive intermediates (Rashed and Nelson, 1989). AMAP is believed to be nontoxic (Nelson, 1980), even though it is known to produce protein adducts, glutathione depletion, and other effects (Rashed et al., 1990) postulated as contributors to APAP-mediated liver injury. The difference in the response to these isomers is thought to reflect lower and shorter depletion of glutathione by AMAP (Rashed et al., 1990). AMAP was also shown to have an effect on the liver proteome qualitatively similar to APAP (Fountoulakis et al., 2000). Even though the response to AMAP was less pronounced, it was most extensive at 12 and 24 h after treatment and many of the biological processes are shared between the two chemicals.

Rat and mouse models have been used extensively to investigate organ-specific effects and mechanisms of toxicity. The availability of detailed genetic information in these model systems and humans provides greater utility of the toxicity data obtained in animals (Cohen, 2004). Since regulatory elements of evolutionarily related species are conserved, gene expression signatures reflecting similar phenotypes in the species are also likely to be conserved. Indeed, the cross-species comparison of gene expression patterns revealed great similarity in the effects of toxic doses of APAP in both species. When the extent of liver injury was compared to the transcriptional response, it was clear that similar expression changes occur in rat and mouse liver under conditions of acute organ damage by APAP. Apoptosis, cell cycle, stress, and transcriptional response, coupled with downregulation of metabolism and transport, are common biological processes between mouse and rat in response to APAP and are likely to be in humans as well. For example, Myc and Jun signify a representative cross-species APAP-induced response. Although it is yet to be determined whether these
molecular events drive liver injury in humans as well, the relative similarity of the sequence of liver injury events, its dependency on glutathione content and inflammatory response, and the overall phenotype shared among rodents and man indicates that the signaling pathways driven by the mobilization of energy resources and activation of tissue repair may have a role in prognosis for human acute drug-induced liver injury. The clear gain realized from this comparative genomics approach is connecting molecular and histopathological features of human liver injury to mouse and rat models with greater confidence. Establishing this molecular relationship between species should provide new opportunities to explore research avenues into molecular pathogenesis, treatment, and prevention of adverse drug-induced liver toxicity in humans.

SUPPLEMENTARY DATA

Supplementary data are available online at http://toxsci.oxfordjournals.org/.

FUNDING


ACKNOWLEDGMENTS

The authors wish to thank David Schwartz (NIEHS) for contributing to the study design. The members of the Toxicogenomics Research Consortium also acknowledge the efforts of David Balshaw and Michael Humble of the NIEHS, and support of Molly Vantall of the NTP and the technical assistance of Laurene Fomby, Tammy Wheat, Daphne Vassconcelos, Michael Ryan, Brian Burback, and Milton Hejtcik (all from Battelle), and the Investigation Genomics Group of Cogenics. This work was supported (in part) by the Intramural Research Program of the National Institutes of Health and NIEHS.

REFERENCES


