TOXICOLOGICAL SCIENCES **58,** 399–415 (2000)

Toxicogenomics-Based Discrimination of Toxic Mechanism in HepG2 Human Hepatoma Cells

Michael E. Burczynski,* Michael McMillian,* Joe Ciervo,* Li Li,*[,1] James B. Parker,* Robert T. Dunn II,† Sam Hicken,† Spencer Farr,† and Mark D. Johnson*[,2]

**Drug Safety Evaluation, Robert Wood Johnson Pharmaceutical Research Institute, P.O. Box 300, Route 202, Raritan, New Jersey 08869; and †Phase-1 Molecular Toxicology, Inc., Santa Fe, New Mexico 87505*

Received May 18, 2000; accepted August 14, 2000

The rapid discovery of sequence information from the Human Genome Project has exponentially increased the amount of data that can be retrieved from biomedical experiments. Gene expression profiling, through the use of microarray technology, is rapidly contributing to an improved understanding of global, coordinated cellular events in a variety of paradigms. In the field of toxicology, the potential application of toxicogenomics to indicate the toxicity of unknown compounds has been suggested but remains largely unsubstantiated to date. A major supposition of toxicogenomics is that global changes in the expression of individual mRNAs (i.e., the transcriptional responses of cells to toxicants) will be sufficiently distinct, robust, and reproducible to allow discrimination of toxicants from different classes. Definitive demonstration is still lacking for such specific "genetic fingerprints," as opposed to nonspecific general stress responses that may be indistinguishable between compounds and therefore not suitable as probes of toxic mechanisms. The present studies demonstrate a general application of toxicogenomics that distinguishes two mechanistically unrelated classes of toxicants (cytotoxic anti-inflammatory drugs and DNA-damaging agents) based solely upon a cluster-type analysis of genes differentially induced or repressed in cultured cells during exposure to these compounds. Initial comparisons of the expression patterns for 100 toxic compounds, using all ~ 250 genes on a DNA microarray (~ 2.5 million data points), failed to discriminate between toxicant classes. A major obstacle encountered in these studies was the lack of reproducible gene responses, presumably due to biological variability and technological limitations. Thus multiple replicate observations for the prototypical DNA damaging agent, cisplatin, and the non-steroidal anti-inflammatory drugs (NSAIDs) diflunisal and flufenamic acid were made, and a subset of genes yielding reproducible inductions/repressions was selected for comparison. Many of the "fingerprint genes" identified in these studies were consistent with previous observations reported in the literature (e.g., the well-characterized induction by cisplatin of p53-regulated transcripts such as $p21^{waf1/cip1}$ and PCNA [proliferating cell nuclear antigen]). These gene subsets not only discriminated among the three compounds in the learning set but also showed predictive value for the rest of the database (~ 100 compounds of various toxic mechanisms). Further refinement of the clustering strategy, using a computer-based optimization algorithm, yielded even better results and demonstrated that genes that ultimately best discriminated between DNA damage and NSAIDs were involved in such diverse processes as DNA repair, xenobiotic metabolism, transcriptional activation, structural maintenance, cell cycle control, signal transduction, and apoptosis. The determination of genes whose responses appropriately group and dissociate anti-inflammatory versus DNA-damaging agents provides an initial paradigm upon which to build for future, higher throughput-based identification of toxic compounds using gene expression patterns alone.

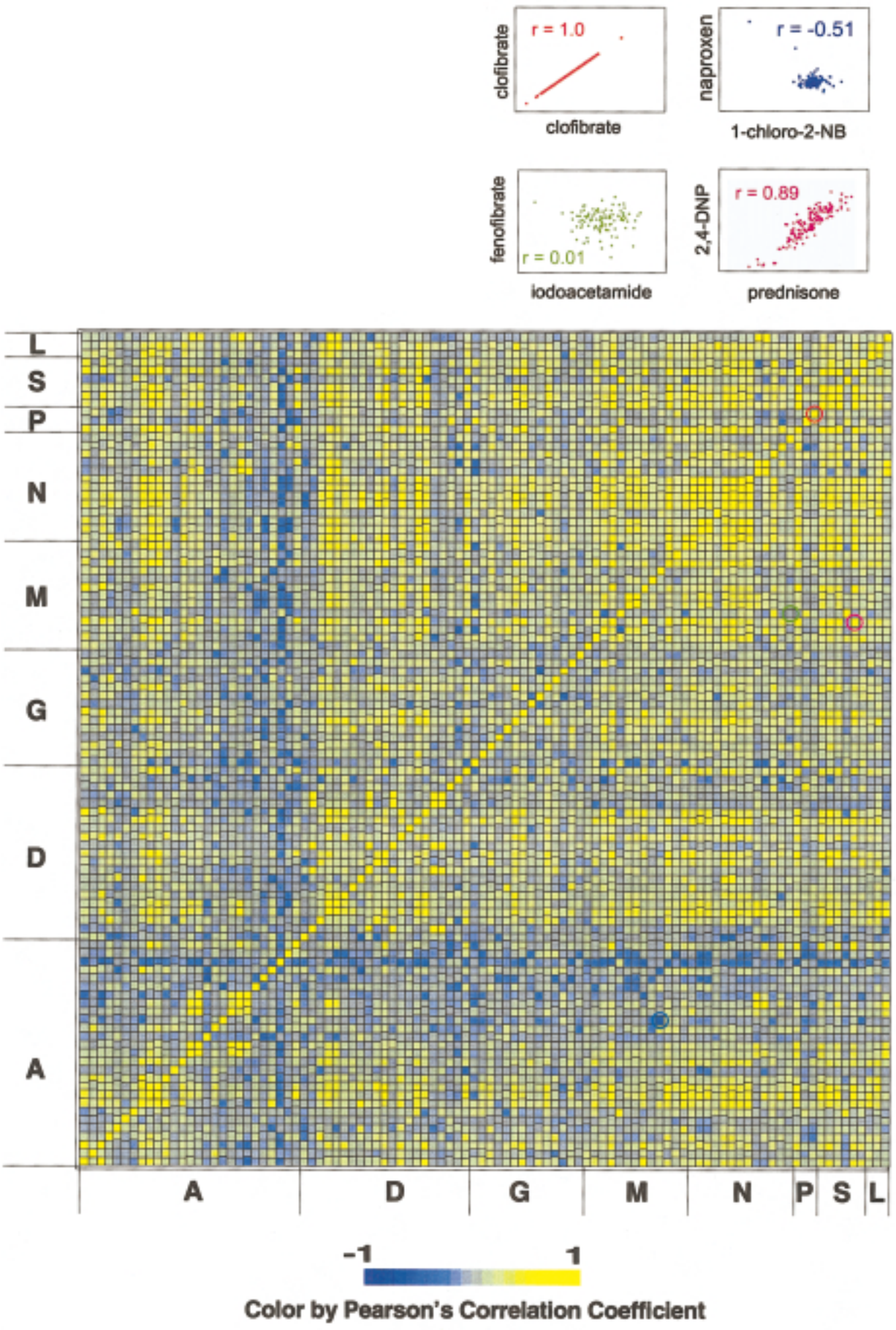
Key Words: **toxicogenomics; gene expression profiling; cDNA microarrays; non-steroidal anti-inflammatory agents; DNA damaging agents.**

The genome projects continue to produce gene sequence information at an accelerating rate (http://www.ncbi.nim.nih.gov/Web/Genbank). Sequencing of the human genome will soon be essentially complete (Marshall, 1998; Pennisi, 2000a; Ventner *et al.*, 1998) and the rat and mouse genomes will likely be among the next to receive priority (Pennisi, 1999, 2000b). Complementary technologies are under development that will take advantage of the new genomic information to explore biological processes. Gene microarray technology was developed as a way to simultaneously monitor the expression levels of large numbers of genes (Chee *et al.*, 1996; Schena *et al.*, 1995; Shalon *et al.*, 1996), eventually on a genome-wide scale. The ability to probe cell biology in this way is expected to produce functional insights that were not possible using previous approaches.

There is great interest in the application of this technology to toxicology, both as a powerful new tool for mechanistic studies and as a diagnostic parameter for toxicity screens that may far surpass traditional approaches in terms of sensitivity and speed. A main assumption in the use of toxicogenomics is that toxicity is accompanied by changes in gene expression that are either causally linked to the toxic outcome or are downstream sequelae of the toxic exposure. Monitoring gene expression

[1] Current address: General Toxicology, Schering-Plough Research Institute, Lafayette New Jersey 07848.

[2] To whom correspondence should be addressed. Fax: (908) 218-0668. E-mail: mjohnson@prius.jnj.com.


clofibrate
r = 1.0
clofibrate
naproxen
r = -0.51
1-chloro-2-NB
fenofibrate
r = 0.01
iodoacetamide
2,4-DNP
r = 0.89
prednisone
L
S
P
N
M
G
D
A
A
D
G
M
N
P
S
L
-1
1
Color by Pearson's Correlation Coefficient

profiles, induced directly or indirectly by different classes of toxicants, should eventually allow recognition of signature patterns that are representative of specific toxicities. Once recognized, these patterns could be used to evaluate new compounds (pharmaceutical candidates) possessing undefined toxicities. This is a compelling scenario that has received widespread attention, but to date there is little published data to support such a possibility (Afshari *et al.*, 1999; Braxton and Bedilion, 1998; Nuwaysir *et al.*, 1999).

The present studies were undertaken to evaluate the potential use of gene expression analysis to detect and distinguish toxicants with different mechanisms of action. HepG2 human hepatoma cells were selected as the model system in an attempt to minimize such complicating factors as cell type heterogeneity and interindividual differences. A single time point of 24 h was chosen to eliminate the potentially confusing contribution of nonspecific immediate early stress responses of cells exposed to toxic stimuli. To minimize the influence of potency differences, all compounds were first tested for cytotoxicity in HepG2 cells using a reductase activity assay at 72 h, and an ED_{30} dose was selected for monitoring gene expression at the earlier time point of 24 h.

The ability to detect reproducible gene expression patterns that are consistent within a class of toxicants and different across classes is key to the emerging field of toxicogenomics. These studies attempt to determine whether such patterns can be observed and as such are an early step towards proving principle.

MATERIALS AND METHODS

Chemicals and reagents. Cell culture reagents were obtained from Gibco BRL (Frederick, MD). Chemicals were obtained from Sigma Chemical Company (St. Louis, MO) or Aldrich Chemical Company (Milwaukee, WI) and were of the highest purity commercially available.

Cell culture. HepG2 human hepatoma cells were obtained from ATCC (catalog number HB-8065). Cells were maintained in log growth phase in minimal essential media (MEM) supplemented with 10% fetal bovine serum. Antibiotic or antifungal agents were not used, to avoid the potential effects of these agents on gene expression and cytotoxicity assays. Cultures were re-established after 20 passages.

Cytotoxicity assay. HepG2 cells were grown in 96-well black, clear-bottom plates (Polyfiltronics, NUNC) at 37°C in a humidified cell culture incubator, in 5% CO_2 and minimal essential media with 10% fetal bovine serum. Cells at 30–50% confluence were treated with test compound (0.1 μM–100 μM, in 0.5 log concentration increments) or vehicle (media or <1% DMSO) for 72 h and cellular reductase activity was measured using an Alamar Blue assay. One hundred μl of Alamar Blue (diluted 1:50 in Hanks Buffer) was added and fluorescence readings of the 96-well plate were immediately recorded using a Wallac Victor II plate reader with excitation at 535 nm and emission at 580 nm. The initial zero time-point readings (which were essentially equal to Alamar Blue readings from an empty culture plate) were subtracted from readings at 1 h to determine cell viability. Control-cell wells showed a pronounced increase in fluorescence at this time, whereas dead-cell wells were essentially equal to background. This assay is extremely sensitive and detects responses unaccompanied by cell death as measured morphologically or by assays such as LDH release. Thus concentrations producing ~ 30% inhibition of fluorescence at 72 h were chosen to examine genomic effects of test compounds at the earlier, minimally cytotoxic time point of 24 h.

Data analysis overview. Multiple replicate RNA samples were obtained for the genotoxic compound, cisplatin, its relatively inactive stereoisomer, transplatin, and the hepatotoxic nonsteroidal anti-inflammatory compounds diflunisal and flufenamic acid. Expression analysis was performed using a cDNA-based DNA microarray containing approximately 250 inducible genes (see Appendix) that respond during expression of various toxic endpoints. Their identification was based on extensive reviews of the scientific literature and on unpublished pilot studies that were conducted during the past several years as part of the core business activities at Phase-1 Molecular Toxicology (Farr and Dunn, 1999). Expression patterns from replicate experiments were examined for reproducibility and the subset of significantly regulated genes was used for subsequent similarity metric-based correlational analyses described below.

A second selection at the end of the study was performed using a computational algorithm that identified the reproducibly regulated genes that best clustered the class of compounds in question and also distinguished the class from different compound classes (see Figure 12 for a detailed description). To determine whether the discovered relationships could be generalized, the responses of this gene set for cisplatin, flufenamic acid, and diflunisal were then examined across a broader database of expression profiles, to approximately 100 toxic compounds in the Phase-1 toxicology database (http://www.phase1tox.com). To facilitate the comparisons, compounds were grouped by mechanism of toxicity and/or mechanism of action based upon extensive reviews of the scientific literature. From these analyses a battery of genes was discovered whose expression pattern accurately discriminates DNA damaging agents versus anti-inflammatory drugs.

Cell treatments, RNA isolation, and cDNA synthesis. Log growth-phase HepG2 cells in T75 flasks (Corning) were treated with test compounds for 24 h. For the replicate determinations that were done for cisplatin, transplatin, diflunisal, and flufenamic acid, the RNA samples were blinded for labeling,

FIG. 1. Initial correlational analysis of all toxic compounds in the database using all genes on the microarrays. HepG2 cells were treated with the various cytotoxic compounds and relative changes in gene expression were ascertained using microarray analysis. Comparisons of the gene expression patterns between 2 treatments were made by Pearson's correlation coefficient as described in Materials and Methods. A perfect positive correlation for all genes on the microarray of "+1" is represented by yellow (see color key), a lack of correlation between all genes of "0" is gray, while a perfectly negative correlation for all genes of "–1" is represented by blue. Each small box in the plot represents the overall Pearson's correlation coefficient observed between 2 treatments. This results in 10,000 comparisons (each of the 100 compounds with every other). By comparing each compound's correlation with all others in this manner, a mirror-image correlation plot is obtained which is symmetrical about a diagonal line of identity (i.e., the first box in the lower left-hand corner represents a correlational comparison of the first compound with itself). The letters designate the following toxicant classes: A, anti-inflammatory; D, DNA-damagers; G, gene-synthesis inhibitors; L, low-dose DNA-damaging agents (non-toxic), M, metabolic poisons; N, non-genotoxic controls; O, other; P, peroxisomal proliferators; and S, steroids. Red, blue, green, and pink circles are drawn around 4 representative Pearson's correlations out of the 10,000 which illustrate the significance of the blue-yellow plots in further detail (see inset, upper right-hand corner). Inset: In the actual correlational analysis, each of the ~ 250 genes are plotted on both axes for each treatment, to determine whether the magnitude of induction/repression between 2 treatments are identical (linearity with a positive slope = 1.0), opposite (linearity with a negative slope =–1.0), or somewhere in between. Red, r = 1.0: Perfect positive correlation of clofibrate compared with itself; Green, r = 0.01: Lack of correlation between fenofibrate and iodoacetamide; Blue, r = –0.51: Relative negative correlation between naproxen and 1-chloro-2-nitrobenzene; and Pink, r = 0.89: Relative positive correlation between 2,4-dinitrophenol and prednisone.

hybridization, and scanning. At the end of each treatment, total RNA was isolated using RNA purification columns (Qiagen). Untreated sample RNA (20 μg) was mixed with anchored oligo dT, heated to 70°C for 10 min, cooled on ice, and reverse transcribed with Superscript II RnaseH$^-$ (Gibco BRL) reverse transcriptase in the presence of Cy5-dUTP. Treated sample RNA (20 μg) was reverse-transcribed with Cy3-dUTP as the fluorescent label. Reverse transcription reactions proceeded at 42^{0}C for 2 h and unincorporated fluorescent nucleotides were removed by centrifugation (20,800 $\times$ *g*). Fluorescence units remaining in the purified probes were determined for all samples and normalized so that equivalent amounts of each label were added to all slides in a hybridization experiment (below).

cDNA microarray hybridization and analysis. Purified, labeled cDNA was boiled for 5 min in 30 μl of hybridization buffer (50% formamide, 5$\times$ SSC, 0.1% SDS), then cooled and maintained at 70°C. The solution was applied to the microarray slide and hybridized in a humidified custom hybridization chamber overnight at 42°C. Slides were washed in 2 $\times$ SSC, 0.2% SDS for 5 min, then 0.05 $\times$ SSC for 1 min. Slides were dried and then scanned using a confocal laser scanner, and fluorescence intensities were recorded.

Data normalization. The data for each gene was normalized by dividing individual treated and untreated fluorescence values by the medians of the treated and untreated fluorescence values in each experiment, respectively. The expression ratio for each gene, determined by the ratio of treated to untreated values, was then log transformed.

Similarity matrix. All statistical analysis was carried out using algorithms written in Oracle PL/SQL and Java. To measure the degree of similarity between the gene expression profiles produced by different toxicant treatments, the Pearson correlation coefficient was chosen. The Pearson correlation coefficient is a common similarity metric for hierarchical and other types of cluster analysis applied to gene expression patterns (Alizadeh *et al*., 2000; Ben-Dor *et al*., 1999; Eisen *et al*., 1998; Ross *et al*., 2000; Scherf *et al*., 2000; Weinstein *et al*., 1997).

The Pearson correlation coefficient for experiment *i* and experiment *j*, r_{ij}, is given by:

$$r_{ij} = r_{ji} = \frac{\Sigma\ (x_{ig} - \bar{x_i})(x_{jg} - \bar{x_j})}{(n-1)s_{x_i}s_{x_j}}$$

where x_i and x_j are the log-transformed expression profiles for experiment *i* and experiment *j*, $(x_{ig} - \bar{x_i})$ is the deviation of gene *g* for experiment *i* from the mean expression value for experiment *i*, $(x_{jg} - \bar{x_j})$ is the deviation of the same gene for experiment *j* from the mean for experiment *j*, s_{x_i} and s_{x_j} are the sample standard deviations for experiment *i* and experiment *j*, *n* is the number of pairs of genes, and the summation Σ is across the $i = 1,2, \ldots, n$ pairs.

The correlation coefficient was calculated for all pairs of experiments. For *m* experiments, exactly m^2 coefficients were calculated. Since certain genes were not examined across all experiments, missing gene expression values in each experiment were ignored in the calculation.

Graphical representation of data. The similarity matrix was displayed using Spotfire Pro data visualization software (Spotfire Inc., http:// www.spotfire.com). The toxicant expression profiles were ordered on the *x* and *y* axes by grouping together toxicants by assigned mechanisms of action. Others have used similar types of visual representations; however in those cases the order of items on the *x* and *y* axes were determined by the application of a clustering algorithm (Ben-Dor *et al*., 1999; Weinstein *et al*., 1997).

Each of the 10,000 cells in a plot represents a comparison of the gene expression profile between 2 single toxicant treatments. The color in each cell of the plot reflects the similarity between the 2 experiments. The 2-color scale used to represent the correlation coefficient ranged from yellow for a perfect correlation coefficient of 1.0 to blue for an absolute negative correlation coefficient of –1.0. A grayish color, resulting from equal parts blue and yellow, signifies no correlation between the 2 samples ($r = 0$). Because of the nature of the plot, the similarity matrix is symmetric about the main diagonal and the correlation coefficients on the diagonal are unity, because each sample is 100% correlated with itself.

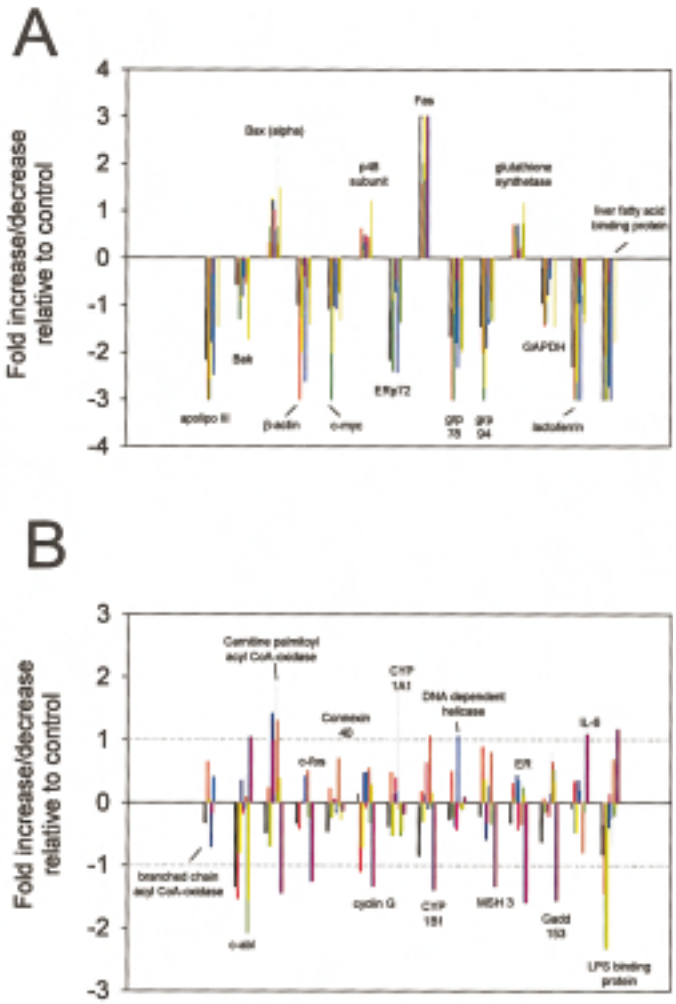


FIG. 2. Reproducibility of gene responses to cisplatin. HepG2 cells were treated on different occasions with identical concentrations of cisplatin (3 μg/ml; $n = 13$) and gene changes were monitored by microarray analysis. Each bar for a given gene group represents the relative induction/repression of that gene in a single cisplatin experiment. Results are expressed as a normalized value where a value of zero equals no change from control and a value of 1 equals 100% change, 2 equals 200% change, etc. (A) Induction/repression of transcripts that changed in an identical direction in all experiments; (B) variable induction/repression of transcripts that responded inconsistently to cisplatin.

The initial correlational analysis was performed across the entire database of 100 toxic compounds, using all genes on the microarray (~ 250). Subsequent correlations used smaller subsets of genes, which were found to be reproducibly and differentially expressed between a given set of treatments or based upon the computer optimization algorithm (see Fig. 12).

RESULTS

Gene Expression Patterns Do Not Discriminate Between Toxicant Classes, Using All DNA Elements on the Microarray

Following individual treatment of HepG2 cells with approximately 100 compounds having various mechanisms of action

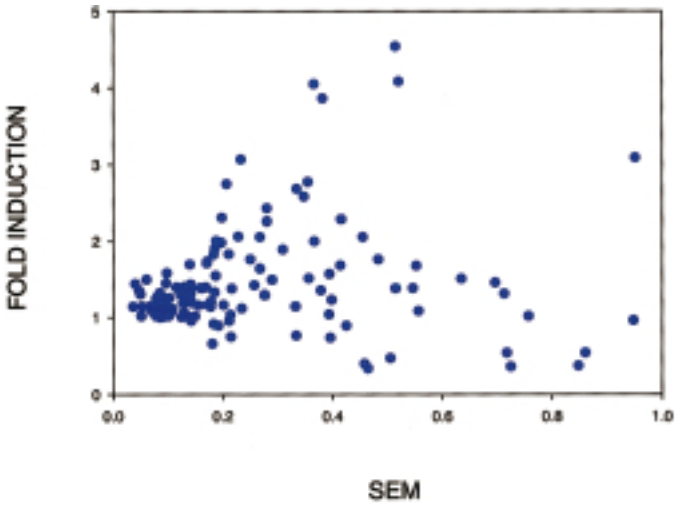


FIG. 3. Lack of relationship between the magnitude of a gene induction event and the standard error of the mean across all experiments. The fold induction/repression of each gene on the microarray for a single cisplatin experiment was calculated and then plotted against the SEM for each gene on the microarray over all experiments.

and toxicity (listed in Appendix), RNA from each sample was harvested and then analyzed using microarrays as described. The normalized expression levels of all genes following each treatment were entered into the Phase-1 database and then subjected to further analysis.

After RNA changes for each treatment were determined relative to control, the change in a gene's expression pattern following toxicant exposure (fold induction or repression) was compared to its change following exposure to every other compound in the database. Comparisons were performed, using the Pearson's correlation coefficient as described in Materials and Methods. This was repeated for ~ 250 genes from all 100 toxicant samples (10,000 comparisons, ~ 2.5 million data points), and a similarity score for the overall gene expression profile was then assigned pair-wise between compounds as described in Materials and Methods. By comparing each compound's correlation with all others, a mirror-image correlation plot was obtained that was symmetrical about a diagonal line of identity (Fig. 1). This type of analysis revealed that inclusion of all genes on the array in the expression comparisons failed to yield significant correlations between database compounds having similar toxic or pharmacologic actions.

Reproducibility of the Gene Expression Data between Experiments Using Cisplatin

Since the expression pattern for each of the ~ 100 compounds in the initial database was determined only once, an equivalent weighing of all gene inductions, whether representative or not, may have obscured the presence of significant correlations. To examine this possibility, multiple replicates were obtained for the genotoxic compound cisplatin by performing treatments at 13 separate times, with each experiment taken to completion before the next was begun.

This exercise showed quite clearly that certain gene induction events occurred consistently while others were highly variable. Although more than 200 genes were analyzed, only a small percentage appeared to respond similarly to cisplatin over all experiments (Fig. 2A), whereas most genes responded variably or with average fold inductions that were less than twice the standard error of the mean (Fig. 2B). Approximately 20% of the genes analyzed in each of the cisplatin experiments were induced or repressed more than 2-fold by cisplatin exposure after 24 h. No relationship was apparent between the magnitude of a gene induction event and its reproducibility in subsequent experiments (Fig. 3). Moreover, plotting intra-experimental coefficient of variation (COV) against inter-experimental SEM for each gene also failed to show a relationship (Fig. 4).

Comparison of Gene Induction Profiles for Cisplatin and Transplatin

Analysis of the genes that were consistently and significantly regulated by cisplatin (Fig. 5) demonstrated significant cisplatin-dependent induction of several subsets of genes with roles in oxidative stress (SOD [superoxide dismutase], glutathione-linked enzymes), DNA damage/repair ($p21^{waf1/cip1}$, PCNA, DNA polymerase beta) and apoptosis (Fas, BAK). Cisplatin also caused a notable repression of several co-regulated ER stress-response genes (grp78, grp 94 and disulfide isomerase-related protein 72 [Erp72]), which function as molecular chaperones and appear to inhibit Ca^{2+}-dependent cell death (Liu *et al.*, 1998; Yoshida *et al.*, 1998).

Further microarray analyses were also performed on transplatin as a non-genotoxic control, in order to identify gene

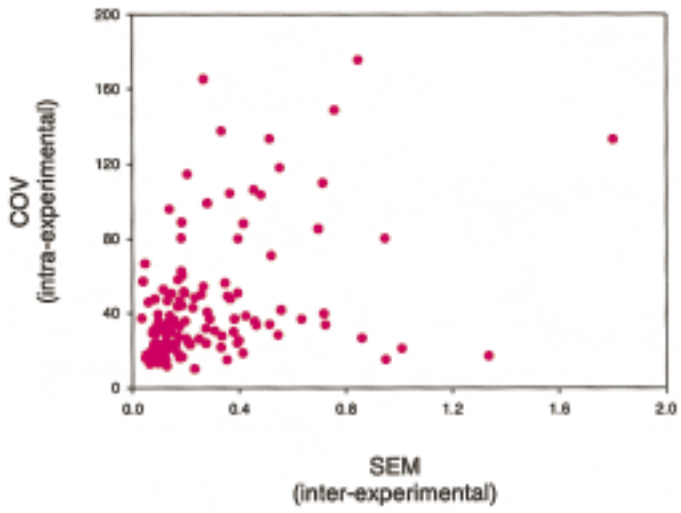
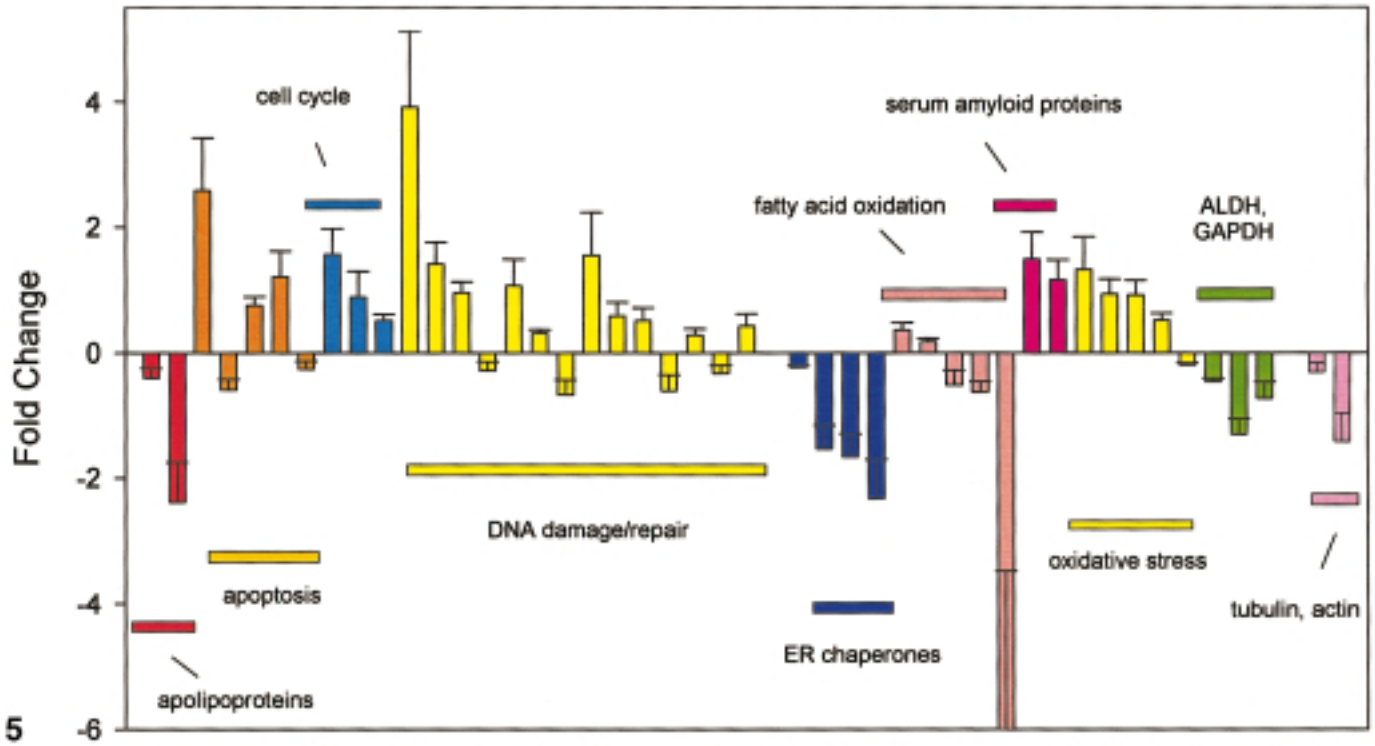
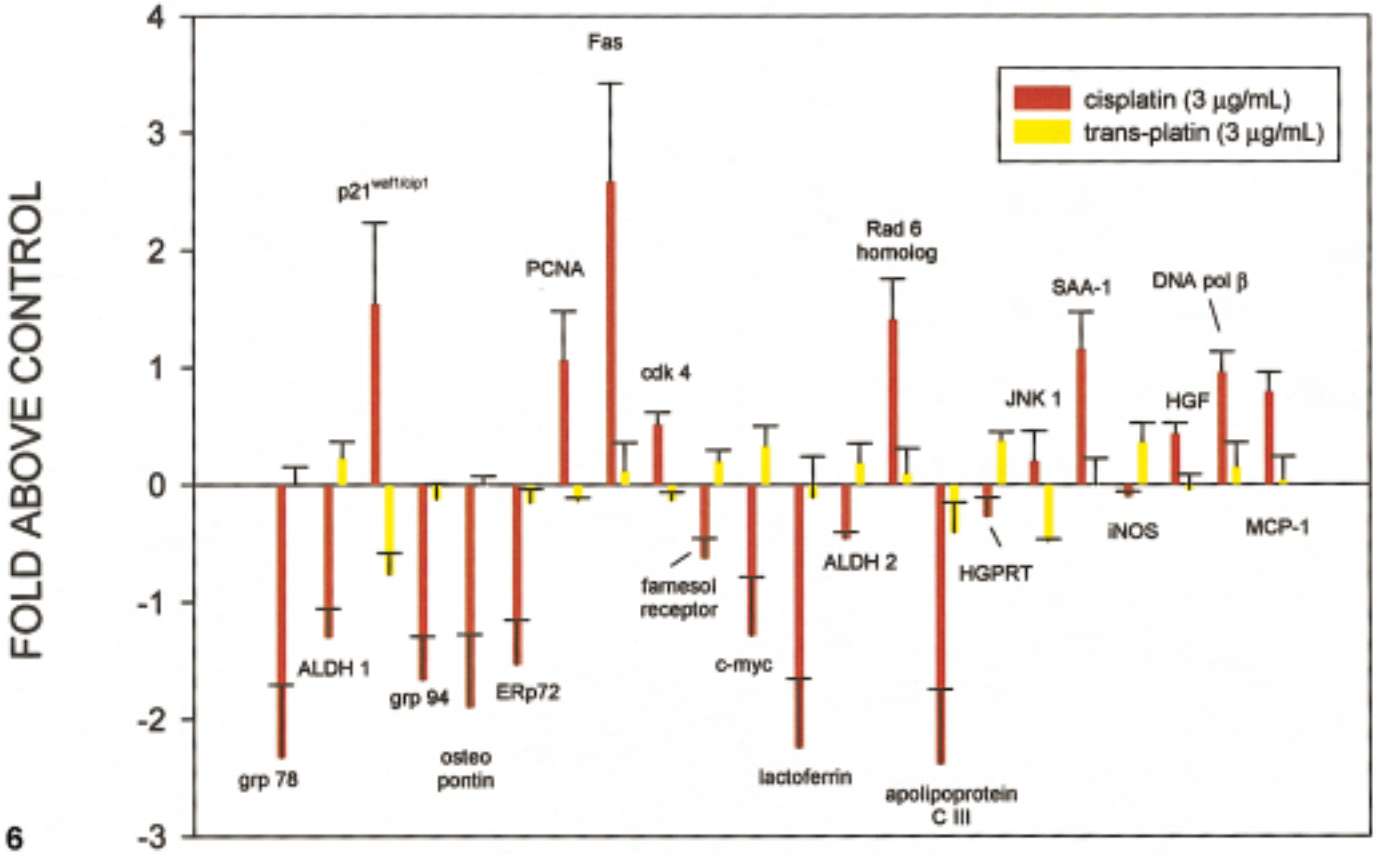


FIG. 4. Lack of relationship between intra-experimental coefficient of variation and inter-experimental standard error of the mean. The average COV value for each gene from a single experimental chip was plotted against the SEM for that gene across all experiments.


Fold Change
4
2
0
-2
-4
-6
cell cycle
serum amyloid proteins
fatty acid oxidation
ALDH, GAPDH
DNA damage/repair
oxidative stress
apoptosis
tubulin, actin
ER chaperones
apolipoproteins

5


FOLD ABOVE CONTROL
4
3
2
1
0
-1
-2
-3
cisplatin (3 µg/mL)
trans-platin (3 µg/mL)
Fas
p21 waf1/cip1
PCNA
Rad 6 homolog
SAA-1
DNA pol β
cdk 4
JNK 1
HGF
ALDH 1
grp 94
ERp72
farnesol receptor
ALDH 2
HGPRT
iNOS
MCP-1
c-myc
grp 78
osteo pontin
lactoferrin
apolipoprotein C III

6

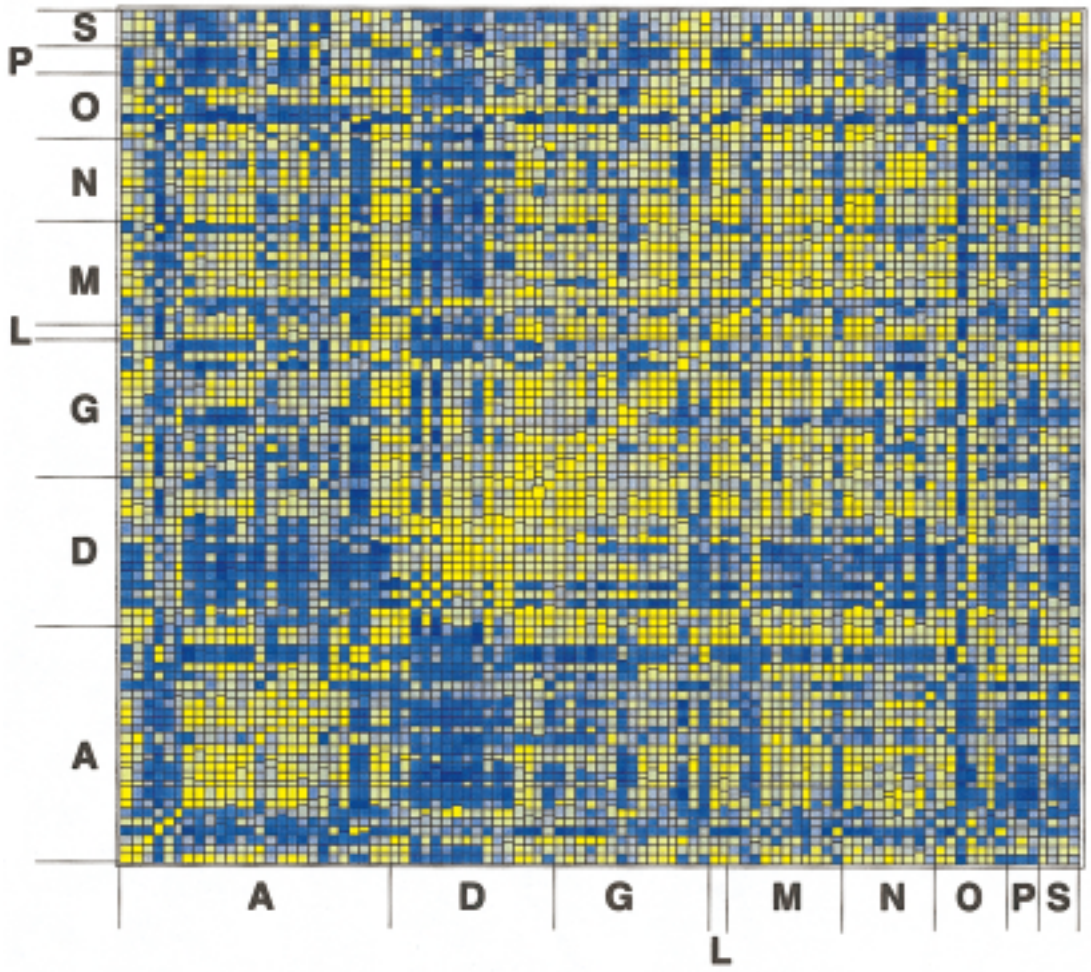


FIG. 7. Correlational analysis of all compounds in the database, using genes differentially expressed between cisplatin (3 μg/ml) and transplatin (3 μg/ml) on the microarrays. Comparisons of the gene expression patterns between all treatments were made by Pearson's correlation coefficient, as described in Figure 1, except that only genes differentially regulated between equimolar concentrations of cisplatin and transplatin were used for comparisons.

induction events specific to the DNA-damaging agent, cisplatin. In an initial experiment, equimolar concentrations of the 2 stereoisomers (3 μg/ml, based on the ED_{30} for cisplatin on reductase activity in HepG2 cells) were used and differential gene inductions were observed (Fig. 6).

When this subset of genes was used to repeat correlational comparisons across the database of toxic compounds, the similarity matrix revealed evidence of weak "clustering" for several DNA-damaging agents (Fig. 7). However, we noted that 3 μg/ml transplatin was non-cytotoxic in HepG2 cells. To rule

FIG. 5. Induction/repression of genes involved in diverse processes by cisplatin in HepG2 cells. Genes undergoing statistically significant (mean – 2 SEM > 0) induction/repression by cisplatin were classified according to function and are presented. In order (from left to right), the following genes were significantly changed following exposure to 3 μg/ml cisplatin for 24 h. Apolipoproteins: apolipoprotein A1, apolipoprotein C III. Apoptosis: Fas antigen, BAK, bax alpha, BAG-1, TRADD. Cell cycle: c-jun, MDM-2, cdk-4 (cyclin-dependent kinase 4). DNA damage/repair: GADD 45, rad 6 homolog, DNA polymerase beta, ATP-dependent helicase II, PCNA, damage-specfic DNA binding protein p48 subunit, DNA dependent protein kinase, $p21^{waf1/cip1}$, ERCC1, RAD, Rad 51 homolog, ERCC3, DNA mismatch repair protein (PMS2). ER molecular chaperones: calnexin, ERp72, grp 94, grp 78 (glucose regulated proteins 78 and 94). Fatty acid oxidation: peroxisomal fatty acyl coA oxidase, very long chain acyl coA dehydrogenase, lysyl oxidase, farnesol receptor, liver fatty acid binding protein. Serum amyloid proteins: SAA-2, SAA-1 alpha. Oxidative stress response: Mn^{2+}-SOD, epoxide hydrolase, glutathione peroxidase, glutathione synthetase, and iNOS (inducible nitric oxide synthase). Dehydrogenases: ALDH 1, ALDH 2 (aldehyde dehydrogenases 1 and 2), and GADPH (glyceraldehyde-3-phosphate dehydrogenase). Structural proteins: alpha-tubulin, beta-actin.

FIG. 6. Differential expression of genes following 3 μg/ml cisplatin versus 3 μg/ml transplatin, after 24 h. HepG2 cells were treated as described and the isolated RNA was subjected to microarray analysis. Genes undergoing statistically significant induction/repression were identified and compared, to identify genes differentially regulated between the 2 treatments.

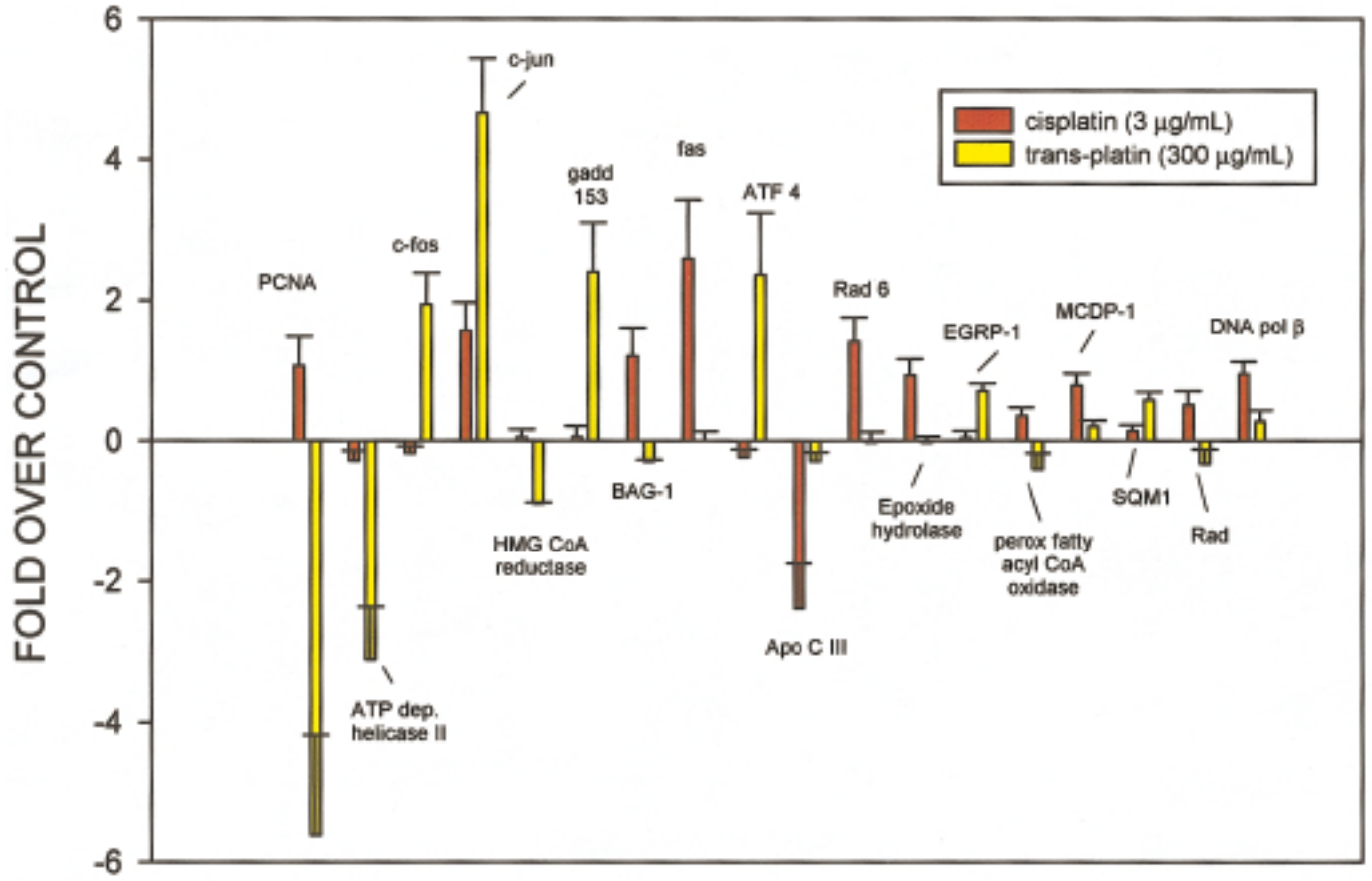


FIG. 8. Differential expression of genes following 3 μg/ml cisplatin versus an equally cytotoxic dose of transplatin (300 μg/ml) after 24 h. HepG2 cells were treated as described and the isolated RNA was subjected to microarray analysis. Genes undergoing statistically significant induction/repression were identified and compared, to identify genes differentially regulated between equitoxic concentrations of the DNA-damaging agent and its inactive stereoisomer.

out general (nonspecific) cellular responses to toxic insult, a 100-fold higher concentration of transplatin that gave a similar cytotoxic response in HepG2 cells was chosen for further microarray comparison. This exercise detected genes that were specifically induced or repressed by the DNA-damaging agent cisplatin relative to a similarly cytotoxic dose of its non-genotoxic stereoisomer transplatin (Fig. 8). While several of the genes remained differentially elevated by cisplatin (PCNA, Fas, RAD 6 homolog, DNA polymerase beta), several others were no longer differentially expressed between the two treatments. Indeed, one of the most surprising absences in Figure 8 is the p53-dependent signaling molecule, $p21^{waf1/cip1}$, which was induced to a similar extent by both treatments and was therefore removed from the differentially expressed gene subset. Whether the elevation of $p21^{waf1/cip1}$ is due to DNA damage at higher concentrations of transplatin is unclear at present, but is a subject of current interest.

Repeating the correlational comparisons across the database with the set of differentially regulated genes (from Fig. 8) maintained the similarity among DNA-damaging agents but failed to further dissociate DNA-damaging agents from anti-inflammatory compounds (Fig. 9). This exercise did, however, begin to discriminate among certain DNA damaging agents from select non-toxic controls and metabolic poisons (Classes N and M). While these correlations appear quite robust, it should be noted that the compounds from these latter classes in the database were analyzed by microarray only once. It is unknown whether these correlations would remain as striking following replicate observations.

Detection of Statistically Significant Gene Sets for Distinguishing Anti-inflammatory Compounds Versus the DNA Damaging Agent, Cisplatin

To identify genes that responded consistently and significantly to hepatotoxic anti-inflammatory drugs, diflunisal and flufenamic acid were analyzed in replicate experiments, and genes that responded differentially, following cisplatin and anti-inflammatory treatment, were identified (Fig. 10). This subset of genes that were differentially and consistently regulated between cisplatin and anti-inflammatories was used to repeat the correlational comparisons across the database (Fig.

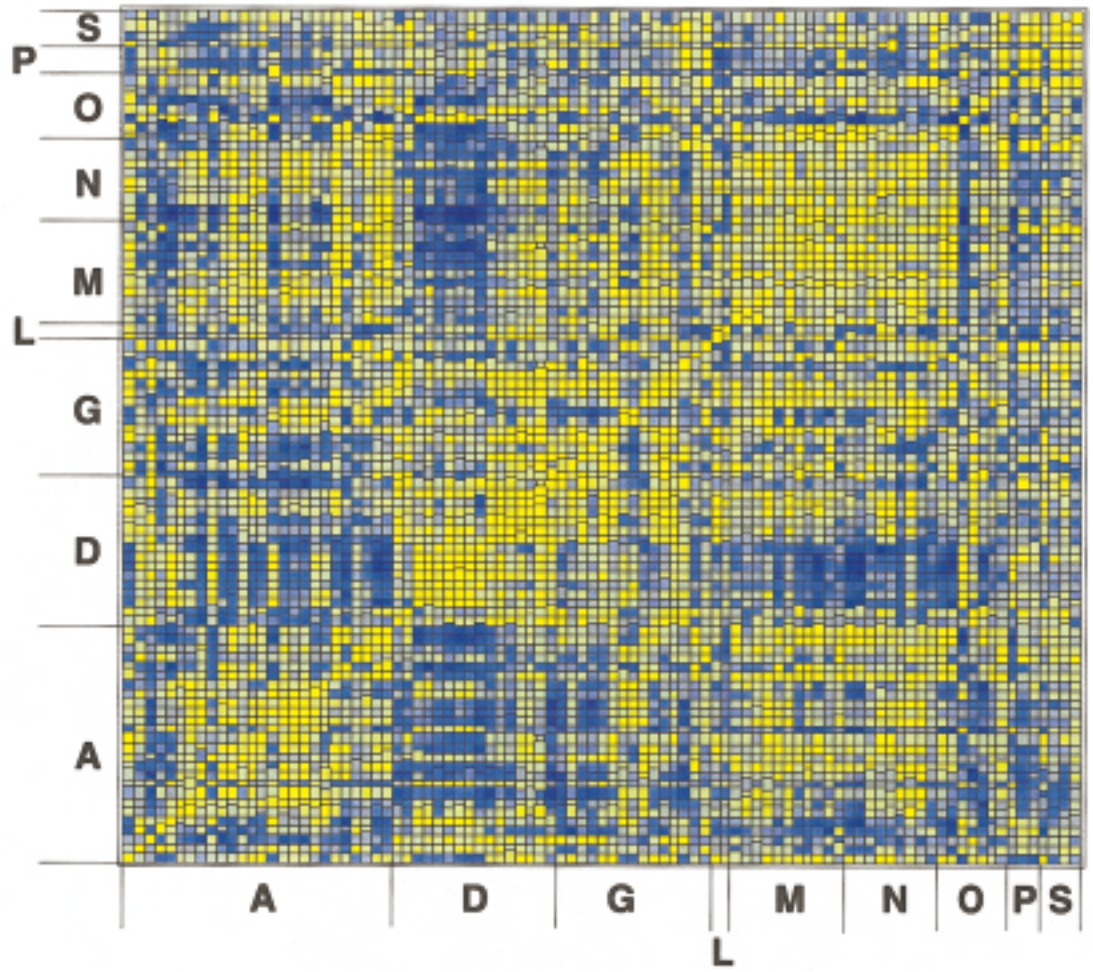


FIG. 9. Correlational analysis of all compounds in the database using genes differentially expressed between cisplatin (3 μg/ml) and transplatin (300 μg/ml) on the microarrays. Comparisons of the gene expression patterns between all treatments were made by Pearson's correlation coefficient, as described in Figure 1, except that only genes differentially regulated between equitoxic concentrations of cisplatin (3 μg/ml) and transplatin (300 μg/ml) were used for comparisons.

11). The improved predictability of this gene set for the majority of DNA-damaging agents and anti-inflammatory compounds substantiates this approach.

Algorithm-based Optimization of Gene Sets for Clustering Compounds of Known Toxicity

To confirm the above results, a second type of analysis was performed. This analysis used a computational algorithm (Fig. 12) to select the set of genes that were (1) most similar in their induction responses to DNA damaging agents, (2) similar in response to non steroidal anti-inflammatory compounds, and (3) also different in response between these 2 classes. A final correlational analysis using this algorithm-optimized gene set (Fig. 13) showed the strongest evidence of clustering between DNA-damaging agents and cytotoxic anti-inflammatory agents, including clusters for mechanisms (metabolic poisons, steroids) outside of the learning set. The set of genes selected with this approach (Fig. 13, legend) had some overlap with the previous set, but also showed differences. Overall, 7 of the 20 genes were common between the 2 approaches.

DISCUSSION

The rapid and highly publicized development of gene-expression microarray technology during the past 5 years has led to a proliferation of potential applications. Considerable efforts are underway to utilize microarray technology to detect biologically relevant patterns of gene expression, and initial results are beginning to appear (DeRisi *et al.*, 1997; Iyer *et al.*, 1999; Marton *et al.*, 1998). Within the pharmaceutical industry and elsewhere there is widespread interest in applying this technology to facilitate toxicity testing, in either a research mode for mechanistic investigations or in a diagnostic mode to screen preemptively for expression patterns associated with established toxicities (Nuwaysir *et al.*, 1999, Rodi *et al.*, 1999). Output of prospective pharmaceuticals from drug discovery

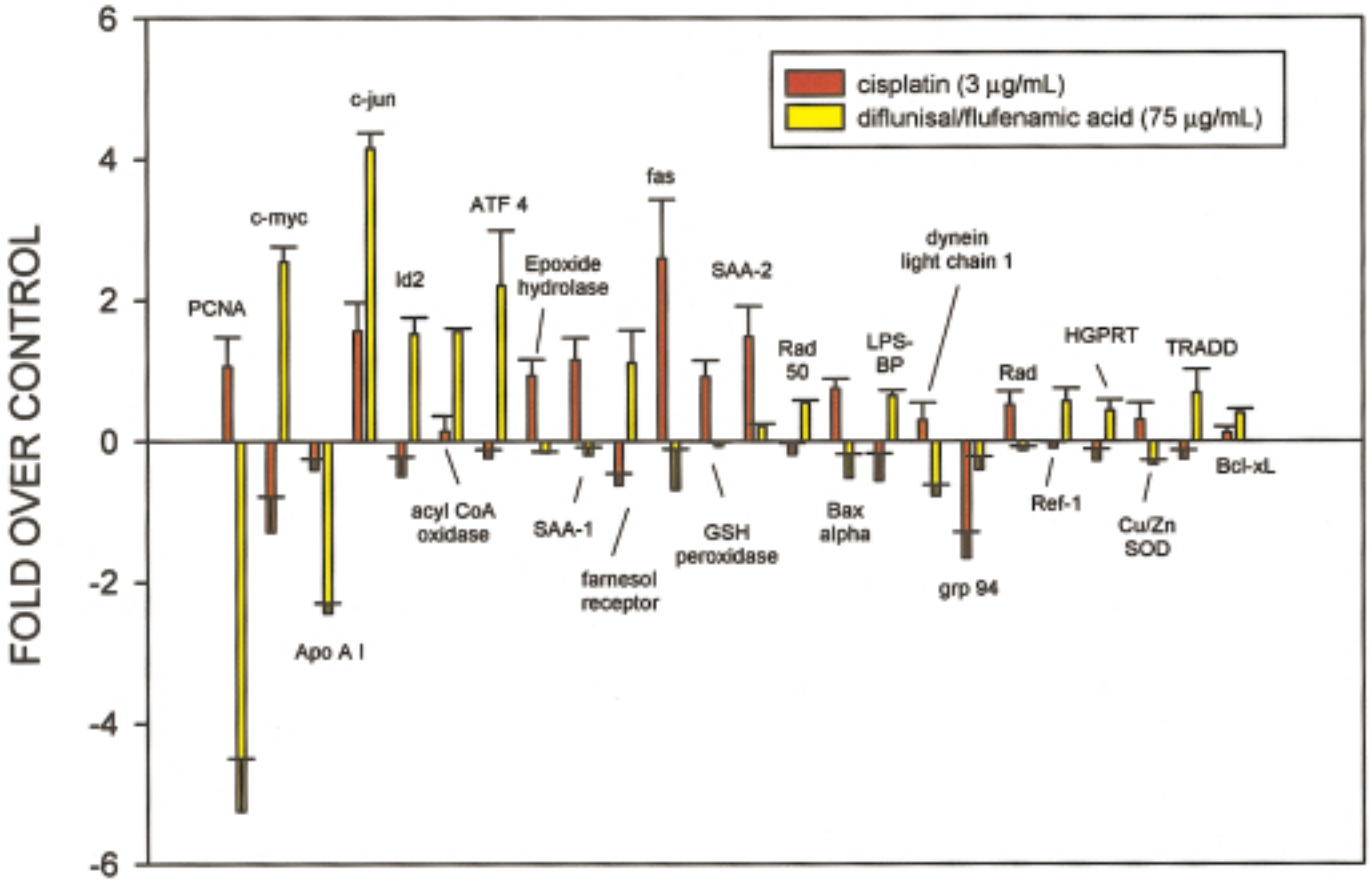


FIG. 10. Differential expression of genes following 3 μg/ml cisplatin versus equally cytotoxic doses of diflunisal and flufenamic acid (75 μg/ml) after 24 h. HepG2 cells were treated as described and the isolated RNA was subjected to microarray analysis. Genes undergoing statistically significant induction/repression were identified and compared to identify genes differentially regulated between equitoxic concentrations of a representative DNA damaging agent and 2 representative NSAIDs.

organizations is increasing throughout the industry, due in large part to application of enabling technologies such as genomics, high throughput screening, and combinatorial chemistry (Drews, 2000). Downstream resources are strained because drug development cannot keep pace using current methodologies. Thus, powerful drivers exist for improving the rate and efficiency of the drug development process. The utilization of gene-expression approaches appears to be a feasible solution, albeit one whose promise is as yet unrealized.

A variety of "toxicology arrays" are now available from several commercial vendors, consisting of genes with demonstrated or presumed relevance to toxic responses. These arrays can and will be used to produce large amounts of gene-expression data. However, utilization of the data for toxicity evaluation requires that the relevance of the gene expression patterns can be ascertained. Currently the application of toxicogenomics depends less on the availability of suitable gene arrays than on the existence of a reliable gene-expression database consisting of responses to prototypical toxic compounds. Indeed, the rate of application of toxicogenomics within the pharmaceutical industry will likely be directly related to the degree of complexity that is encountered in developing such a database. If expression patterns are found to be robust and reproducible, then meaningful and representative response patterns may be found rapidly. On the other hand, if expression patterns are found to be of lesser magnitude and reproducibility, or if they show high variability for a given biological system, depending on influences such as time or exposure conditions, then obtaining representative response patterns may become a long-term goal.

The studies described here were undertaken to determine whether toxicologically meaningful gene-expression responses could be detected in a model system. The key element was the availability of a database containing response patterns to various toxic compounds in a single-cell line under standardized conditions. It is encouraging that clustering can be observed so readily at this early stage of database development. Clusters of positive correlations were observed for compounds classified as DNA-damaging agents and for compounds classified as cytotoxic NSAIDs. For most compounds in the database $n = 1$, which may be suboptimal, as evidenced by the variability observed for the replicate treatments in this study. The subsets

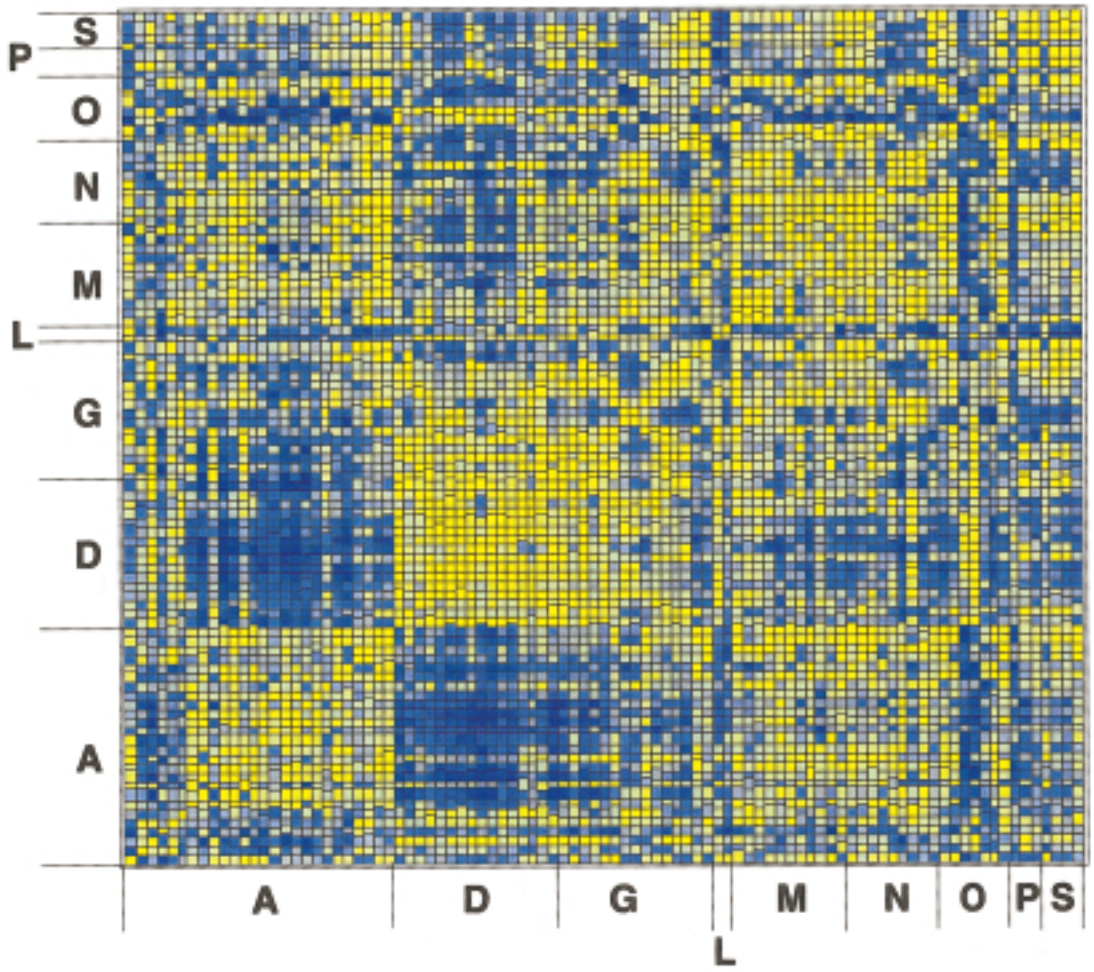


FIG. 11. Correlational analysis of all compounds in the database, using genes differentially expressed between cisplatin (3 μg/ml) and diflunisal/flufenamic acid (75 μg/ml) on the microarrays, after 24 h. Comparisons of the gene expression patterns between all treatments were made by Pearson's correlation coefficient, as described in Figure 1, except that only genes differentially regulated between equitoxic concentrations of cisplatin (3 μg/ml) and diflunisal and flufenamic acid (75 μg/ml) were used for comparisons.

of genes used for the correlational analyses were selected using compounds for which replicate data were available (cisplatin, flufenamic acid, and diflunisal), but the strong clustering for other DNA-damaging agents and NSAIDs are based on single microarray determinations. It is interesting to speculate that the gene subsets identified in replicate analyses may represent those genes that demonstrate consistent responses for the toxicant class in question, which may explain why clustering could be observed for other compounds analyzed only once. However, until existing methodologies improve, replicate analyses may be necessary to identify significantly regulated genes with sufficient certainty for assignment of toxic mechanism. Even with improved methodology, the biological variability in the chosen model system may still require replicate analyses to determine genes that are significantly and reproducibly changed. The ability to observe correlations for mechanistic classes outside the learning set suggests that relatively small sets of genes may be sufficient to distinguish a variety of different toxic mechanisms. In the future, small "information-rich" gene chips may provide greater utility in assigning/discriminating toxic mechanisms. To identify the smaller subsets of genes useful for toxicant classification in any given system, it may be necessary to first perform transcriptional profiling using gene chips that span entire genomes. Once these genetic relationships are established, the use of gene microarrays may ultimately diminish and be replaced by more manageable and cost-effective platforms to identify unknown toxicants, based upon their effects on a limited number of predictively "valuable" genes.

The experiments in this study did detect numerous cisplatin-inducible genes that are in agreement with previous observations in the literature. These include induction of several p53-responsive genes (Aubrecht *et al.*, 1999) including $p21^{waf1/cip1}$ (Zamble *et al.*, 1998), GADD45 (Sun *et al.*, 1995) and PCNA (Shivakuvar *et al.*, 1995). We also detected potent induction of Fas, which has been shown to mediate apoptosis in HepG2

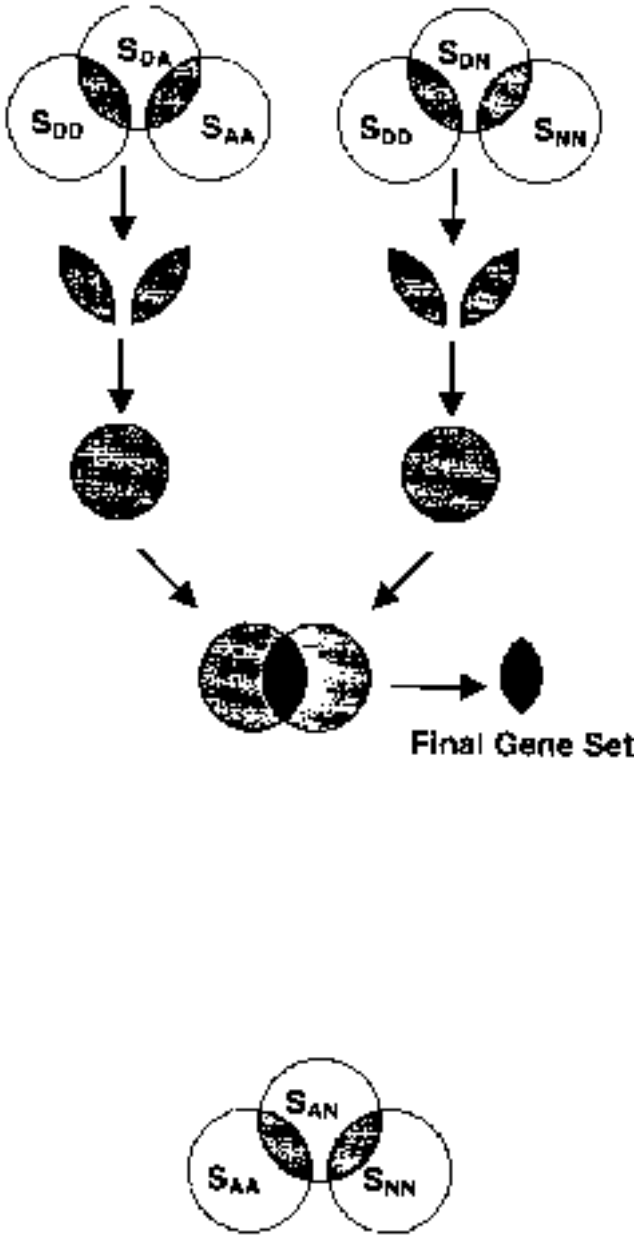


FIG. 12. Correlation optimization algorithm. The optimization algorithm was a "brute force" type, supervised-learning algorithm that relies on the computer's processing power to perform calculations on all pairs of experiments in all pairs of 2 groups. The groups of treatments from which data were used for algorithm optimization were DNA damaging-agents, anti-inflammatories, and non-genotoxic controls. By focusing on one gene at a time and determining each gene's effect on average correlation coefficient, the algorithm selected genes that maximized the average intra-group correlation coefficient and minimized the average inter-group correlation coefficient for all pairs of groups in our analysis. Application of the algorithm resulted in 6 group/group comparisons and their associated "optimized" gene sets. Gene $Set_{DNA\ damagers,\ DNA\ damagers,}$ $S_{DD;}$ gene $Set_{DNA\ damagers,\ antiinflammatories,}$ $S_{DA;}$ gene $Set_{DNA\ damagers,\ notoxic\ controls,}$ $S_{DN;}$ gene $Set_{antiinflammatories,\ antiinflammatories,}$ $S_{AA;}$ gene $Set_{antiinflammatories,\ notoxic\ controls,}$ $S_{AN;}$ gene $Set_{notoxic\ controls,\ notoxic\ controls,}$ $S_{NN.}$ The genes in the resulting 6 gene sets were reduced further by taking the intersection of the union of the intersection of the initial gene sets as displayed graphically in the Venn diagrams or by the following notation: $[(S_{DD} \cap S_{DA}) \cup (S_{AA} \cap S_{DA})] \cap [(S_{DD} \cap S_{DN}) \cup (S_{NN} \cap S_{DN})]$, where $\cup$ and $\cap$ represent union and intersection, respectively. The intersections involving the anti-inflammatory and nontoxic control groups (lower diagram) were ignored in the determination of the final gene set, because they repeatedly produced a null set for the final gene set. No genes resulting from this union of intersects were common to the genes sets derived from the analyses involving DNA damagers. Therefore, genes that minimized the similarity between anti-inflammatories and DNA-damaging agents and between nontoxic controls and DNA damagers maximized the similarity between anti-inflammatories and nontoxic controls.

cells exposed to cisplatin (Muller *et al.*, 1997). JNK activation and prolonged c-Jun induction have also been proposed to mediate apoptosis following cisplatin exposure (Nehme *et al.*, 1997; Sanchez-Perez and Perona, 1999), and c-jun mRNA was robustly induced by cisplatin in the present studies, even after 24 h. Cisplatin also resulted in the repression of several genes, most notably a family of molecular chaperones which appear to be co-regulated by ATF-6 and have been implicated in preventing Ca^{2+}-dependent cell death (Liu *et al.*, 1998; Yoshida *et al.*, 1998).

Taken together, these rather preliminary observations suggest that the ability of cisplatin to cause apoptosis in HepG2 cells may be determined by the relative levels of apoptotic-regulated versus anti-apoptotic–regulated genes and activities. For instance, Fas causes apoptosis in HepG2 cells (Muller *et al.*, 1997) but Fas-dependent apoptosis is blocked by the formation of an inactive complex between procaspase 3 and $p21^{waf1/cip1}$ (Suzuki *et al.*, 1998), which was also induced by cisplatin in these studies. In addition, both pro-apoptotic (Bax, Bak) and anti-apoptotic (BAG-1) proteins were also induced by cisplatin in these studies (Adams and Cory, 1998). In the future, more extensive microarray experiments could actually determine the ratios of different critical gene products that may decide cellular fate following exposure to DNA damaging agents. Although additional morphological and biochemical experiments to determine the nature of cell death were outside the scope of the present study, research correlating gene expression with cytotoxic mechanisms (for example, apoptosis vs. necrosis) could lead to improved understanding of these complex phenomena.

One illustration of the benefit of this type of microarray experiment is evidenced by the detection of cisplatin-inducible transcripts in these studies that had not been previously described, but which are, nevertheless, consistent with recent findings in the literature. For instance, these studies detected rather robust and consistent induction of a rad6 homolog in HepG2 cells exposed to cisplatin, suggesting that this mammalian homolog might play a role in the response to cisplatin-induced DNA damage. During the preparation of this manuscript, Simon *et al.* demonstrated that *S. cerevisiae* strains lacking the *rad6* allele are extremely hypersensitive to the toxic effects of cisplatin compared with wild-type or other mutant *rad* strains (Simon *et al.*, 2000). Thus, results from a purely

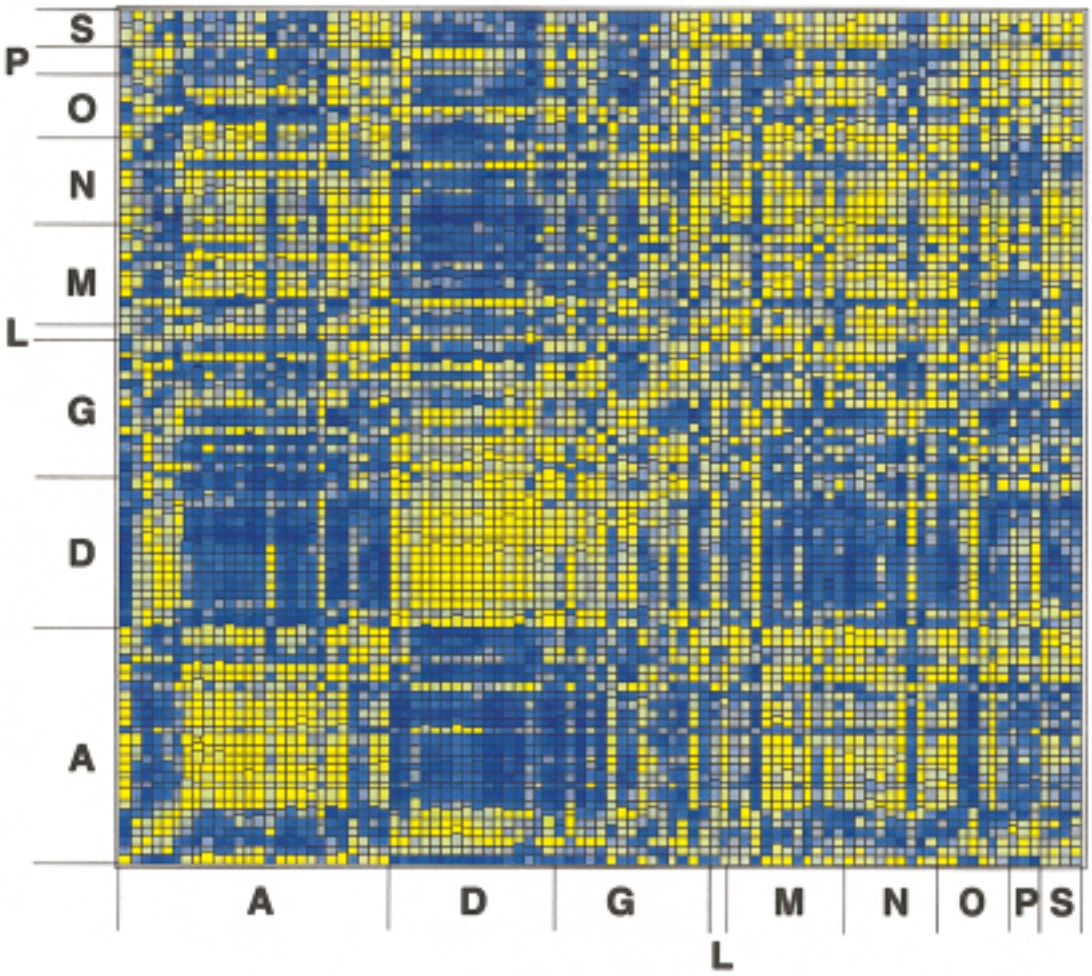


FIG. 13. Correlational analysis of all compounds in the database, using genes identified by the algorithm-based optimization gene set for distinguishing cisplatin and diflunisal/flufenamic acid. The following genes were identified by the algorithm (those in common with statistically based approaches are italicized): *ATP-dependent helicase II (Ku80)*; c-abl; *epoxide hydrolase*; GOS24 (NUP475); *alpha-tubulin*; *cyclin-dependent kinase 4*; cytochrome p450 1A1; ERK 1*; glutathione peroxidase*; *PCNA*; IB-a; replication protein A (70 kDa subunit); thymidylate synthase; hsp90; caspase 3; NADPH quinone oxidoreductase-1 (DT-diaphorase); catalase; *damage-specific DNA binding protein p48 subunit*; endothelin-converting enzyme; MET proto-oncogene.

functional genetic screen in yeast are consistent with our observation of mammalian rad6 homolog induction in HepG2 cells following exposure to cisplatin.

The supervised methods employed here (statistical brute force versus a computational algorithm for selecting gene sets that maximize/minimize Pearson's correlations) produced overlapping but distinct subsets of genes for discriminating between DNA-damaging agents and anti-inflammatory compounds. It is unclear at present whether either of these methods reflects the better approach until they are tested for assignment of mechanism to additional unknowns. Larger scale experiments on gene chips containing greater numbers of genes should allow the use of unsupervised clustering methods (hierarchical, K-means, neural networks) that will likely provide more powerful approaches to unbiased toxicant classification. Utilization of these types of learning methods would also allow the analysis of multiple time points to group toxicants by sets of genes that are similarly expressed in temporal fashion between treatments, likely improving pattern recognition in the future.

In the end, it will be important to compare these and other responses in HepG2 cells to other cellular and *in vivo* systems to determine whether response patterns are similar across different systems. The occurrence of characteristic responses across a variety of systems could allow the more rapid development of analytical capabilities. Because of an apparent deficiency in C/EBPα, HepG2 cells lack several key CYPs and other enzymes responsible for metabolism (Jover *et al.*, 1998). This paucity of metabolic capability almost certainly affects the transcriptional responses observed in HepG2 cells and in other more metabolically competent systems. It should be possible to gauge the likely impact of this issue once a sufficient number of studies are published. In the same way it will be possible to compare the variability and magnitude of re-

sponse patterns in different test systems. Additionally it may be worthwhile to monitor baseline variability using a reference set of mRNAs. Baseline variability may be an important consideration in selection of a test system, especially for methodologies using Cy3/Cy5 fluorescence, in which results are characteristically expressed relative to an untreated control.

Current efforts in this laboratory are underway to determine whether other biological systems may yield more reproducible and robust transcriptional profiles than the HepG2 model system. For instance, studying the effects of toxicants in primary rat hepatocytes will allow comparison with effects in rat liver following treatments *in vivo*. These types of studies will be crucial in determining whether transcriptional profiling in model systems is relevant to *in vivo* toxicity testing and thus useful for streamlining drug discovery. The predictive power of the rather limited gene sets identified in the present studies suggests that gene-expression-based approaches will continue to gain acceptance and application as powerful new tools for toxicity testing.

ACKNOWLEDGMENT

The support and guidance of Dr. William J. Powers is gratefully acknowledged.

APPENDIX

Genes Assessed by Microarray in the Present Study

11-beta hydroxysteroid dehydrogenase type II
12-lipoxygenase
60S ribosomal protein L6
6-O-methylguanine-DNA methyltransferase
Activating transcription factor 3
Activating transcription factor 4
Activating transcription factor 4
Activin receptor type II
Adenine nucleotide translocator 1
Adhesion protein (SQM1)
Alcohol dehydrogenase 2
Alcohol dehydrogenase 4
Aldehyde dehydrogenase 1
Alpha 1-antitrypsin
Alpha-1 acid glycoprotein
Alpha-1 antichymotrypsin
Alpha-2-macroglobulin
Alpha-catenin
Alpha-tubulin
Annexin V
Apolipoprotein A1
Apolipoprotein AII
Apolipoprotein CIII
Aspartate aminotransferase, mitochondrial
Ataxia telangeictasia
ATP-dependent helicase II (70kDa)
ATP-dependent helicase II (Ku80)
BAG-1
BAK
Bax (alpha)
Bcl-xL
Beta-actin
Bilirubin UDP-glucuronosyltransferase isozyme 1
Biliverdin reductase
Branched chain acyl-CoA oxidase
BRCA1
C4b-binding protein
C5a anaphylatoxin receptor
c-abl
Calbindin-D (28kDa)
Calcineurin-B
Calnexin
Calreticulin
Carnitine palmitoyl-CoA transferase
Caspase 1
Caspase 3
Caspase 4
Caspase 6
Caspase 7
Caspase 8
Catalase
Catechol-O-methyltransferase
Cathepsin L
c-erbA-2
c-erbB-2
Ceruloplasmin
c-fms
c-fos
Checkpoint kinase-1
c-H-ras
c-jun
Clusterin
c-myc
Colony-stimulating factor-1
Complement component C3
Connexin-32
Connexin-40
Corticosteroid binding globulin
Creatine kinase b
Cyclin D1
Cyclin D3
Cyclin dependent kinase 2
Cyclin dependent kinase 4
Cyclin G
Cyclin-dependent kinase 4 inhibitor B (P15)
Cyclin-dependent kinase 4 inhibitor B (P16)
Cyclin-dependent kinase 4 inhibitor P27kip1
Cystic fibrosis transmembrane conductance regulator
Cytochrome c oxidase subunit IV
Cytochrome P450 1A1
Cytochrome P450 1B1
Cytochrome P450 2A3
Cytochrome P450 4A
Cytochrome P450 4A1
Damage-specific DNA binding protein p48 subunit
Death receptor 5 (DR5)
Defender against cell death-1
Dihydrofolate reductase
Disulfide isomerase related protein (ERp72)
DNA binding protein inhibitor ID2
DNA dependent helicase
DNA dependent protein kinase
DNA ligase I
DNA ligase IV
DNA mismatch repair protein (MLH1)
DNA mismatch repair protein (PMS2)
DNA mismatch repair protein (PMS-6)
DNA mismatch repair/binding protein (MSH3)
DNA polymerase alpha
DNA polymerase beta
DNA repair and recombination homologue (RAD 52)
DNA repair protein (RAD 50)
DNA repair protein XP-D
DNA repair protein XP-D
DNA replication factor C (36kDa)
DNA topoisomerase I
DNA topoisomerase II
DRA
Dynein light chain 1
E2F-1
Early growth regulated protein 1
E-cadherin
Endothelin converting enzyme
Enolase alpha
Enoyl CoA hydratase
Epoxide hydrolase
ERCC 1 (excision repair protein)
ERCC 3 (DNA repair helicase II)
ERCC 5 (excision repair protein)
ERCC 6 (excision repair protein)
ERK1
Erythropoietin receptor
Estrogen receptor
Farnesol receptor
Fas antigen
FEN-1 (endonuclease)
Ferritin H-chain
FIC1
Flavin containing monooxygenase 1
FosB
Fra-1
Fyn proto-oncogene
Gadd153
Gadd45

APPENDIX—*Continued*

Gamma-glutamyl hydrolase precursor
Gamma-glutamyl transpeptidase
Glucose-6-phosphate dehydrogenase
Glucose-regulated protein 78
Glucose-regulated protein 94
glucosylceramide synthase
Glutathione peroxidase
Glutathione reductase
Glutathione S-transferase alpha subunit
Glutathione S-transferase theta-1
Glutathione synthetase
Glyceraldehyde 3-phosphate dehydrogenase
GOS24 (NUP475)
Growth-arrested-specific protein 1
Growth-arrested-specific protein 3
GT mismatch binding protein
Heat shock protein 70
Heat shock protein 90
Heme oxygenase-1
Hepatic lipase
Hepatocyte growth factor
Hepatocyte growth factor receptor
Hepatocyte nuclear factor 4
Histone 2A
Histone 2B
HMG CoA reductase
Hypoxanthine-guanine phosphoribosyltransferase
Hypoxia-inducible factor 1 alpha
ID-1
IkB-a
Inhibitor of apoptosis protein 1
Insulin-like growth factor binding protein 1
Insulin-like growth factor I
Integrin beta1
Intercellular adhesion molecule-1
Intercellular adhesion molecule-3
Interferon inducible protein 15
Interleukin-1 alpha
Interleukin-1 beta
Interleukin-13
Interleukin-6
Interleukin-8
JNK1 stress activated protein kinase
Ku autoimmune antigen gene (p80)
Lactoferrin
Leukemia inhibitory factor (LIF)
Lipopolysaccharide binding protein
Lipoprotein lipase precursor
Liver fatty acid binding protein
L-myc
Luteinizing hormone
Lysyl oxidase
Macrophage inflammatory protein-2 alpha
Mannose receptor
Matrix metalloproteinase-2
MDM-2
MET proto-oncogene
Mitogen activated protein kinase (P38)
Mitogen inducible gene-2
Monoamine oxidase A
Monoamine oxidase B
Monocyte chemotactic protein-1
Multidrug resistant protein-1
Multidrug resistant protein-3
MutL homologue (MLH1)
MutS homologue (MSH2)
Myelin basic protein
Myeloid cell differentiation protein-1
NADPH quinone oxidoreductase-1 (DT-diaphorase)
NF-kappaB (p65)
Nitric oxide synthase-1, inducible
Nucleic acid binding protein
Nucleoside diphosphate kinase beta isoform
Octamer binding protein 1
Organic anion transporter 1
Ornithine decarboxylase
Osteopontin
OX40 ligand
Oxygen regulated protein 150
p53
PAPS synthetase
P-cadherin
Peroxisomal 3-ketoacyl-CoA thiolase 2
Peroxisomal acyl-CoA oxidase
Peroxisomal enoyl-CoA hydratase
Peroxisomal fatty acyl-CoA oxidase
Peroxisome assembly factor 1
Peroxisome biogenesis disorder protein-1
Peroxisome proliferator activated receptor alpha
Phenol sulfotransferase
Phenylalanine hydroxylase
Phosphoenolpyruvate carboxykinase
Phosphoglyceride kinase
Phospholipase A2
Plasminogen activator inhibitor 2
Platelet derived growth factor B
Platelet/endothelial cell adhesion molecule-1
Poly(ADP-ribose) polymerase
Prohibitin
Proliferating cell nuclear antigen gene
Prostaglandin H synthase
Protein disulfide isomerase (PDI)
Protein kinase C alpha
Protein-tyrosine phosphatase
RAD
RAD 51 homologue
RANTES
Ref-1
Replication factor C, 40-kDa subunit (A1)
Replication protein A (70 kDa subunit)
Retinoblastoma
Retinoic acid receptor gamma-1
Retinoid×receptor alpha
Ribonucleotide reductase M1 subunit
Ribosomal protein L13A
Ribosomal protein S9
RNA-dependent helicase
Serum amyloid A1
Serum amyloid A2-alpha
Silencer of death domains (SODD)
Spermidine/spermine N1-acetyltransferase (SSAT)
STAT 3
Stromelysin-1
Superoxide dismutase Cu/Zn
Superoxide dismutase Mn
Survivin
Tau protein
T-cell cyclophilin
Thioredoxin
Thymidine kinase
Thymidylate synthase
Thymosin beta-10
Tissue inhibitor of metalloproteinases-1
Tissue transglutaminase
TNF receptor-1 associated protein (TRADD)
Transcription factor IID
Transferrin
Transferrin receptor
Transthyretin
Tryptophanyl-tRNA synthetase
Type 1 interstitial collagenase
Tyrosine aminotransferase
Ubiquitin conjugating enzyme (RAD 6 homologue)
Ubiquitin-homology domain protein PIC1
UDP-glucuronosyltransferase 2B
Uncoupling protein 2
Urokinase plasminogen activator receptor
UV excision repair protein RAD 23 (XP-C)
Vascular cell adhesion molecule 1 (VCAM-1)
Vascular endothelial growth factor receptor 1 (flt-1)
Very long-chain acyl-CoA dehydrogenase
Vimentin
Waf1
XRCC1 (DNA repair protein)
Zinc-finger protein-37

APPENDIX—*Continued*

Compounds Analyzed by Microarray in the Present Study

1-chloro-2-nitrobenzene	clozapine	methotrexate
2,4-dinitrophenol	colchicine	methyl methanesulfonate
2-acetylaminofluorene	corticosterone	mitomycin C
2-azido-2-deoxycytidine	cycloheximide	mitoxantrone
4-acetamidofluorene	cyclophosphamide w/o S9	naloxone
5-azacytidine	cyclophosphamide with S9	naproxen
5-chlorouracil	cytosine arabinoside	nicotine
5-fluorouracil	dacarbazine	nitrofurantoin
6-mercaptopurine	dexamethasone	N-nitroso-N-ethylurea
acetaminophen	diethylhexylpthalate	N-nitroso-N-methylurea
acetylsalicylic acid	diethylstilbestrol	oligomycin
acridine	diflunisal	o-toluidine
actinomycin	digitoxin	paclitaxel
aminopterin	dimethylhydrazine	phorbol ester (PMA)
aminotriazole	DMSO	prednisone
antimycin A	doxorubicin	proflavin
antipyrine	erythromycin	progesterone
busulfan	ethyl methanesulfonate	puromycin
caffeine	etoposide	rifampicin
camptothecin	fenofibrate	sodium azide
carbamazepine	flufenamic acid	streptozotocin
carboplatin	guanine	tacrine
carmustine	hydroxyurea	tamoxifen
chlorambucil	indomethacin	thioguanine
chloroquine	iodoacetamide	transplatin
cimetidine	isonicotinic acid	triethylenemelamine
cisplatin	mechlorethamine	triethylenethiophosphoramide
clenbuterol	media	verapamil
clofibrate	melatonin	

REFERENCES

Adams, J. M., and Cory, S. (1998). The Bcl-2 protein family: Arbiters of cell survival. *Science* **281,** 1322–1326.

Afshari, C. A., Nuwaysir, E. F., and Barret, J. C. (1999). Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. *Cancer Res.* **59,** 4759–4760.

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J. Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403,** 503–511.

Aubrecht, J., Narla, R. K., Ghosh, P., Stanek, J., and Uckun, F. M. (1999). Molecular genotoxicity profiles of apoptosis-inducing vanadocene complexes. *Toxicol. Appl. Pharmacol.* **154,** 228–235.

Ben-Dor, A., Shamir, R. and Yakhini, Z. (1999). Clustering gene expression patterns. *J. Comp. Biol.* **6,** 281–297.

Braxton, S., and Bedilion, T. (1998). The integration of microarray information in the drug development process. *Curr. Opin. Biotech.* **9,** 643–649.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. (1996). Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614.

DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686.

Drews, J. (2000). Drug discovery: A historical perspective. *Science* **287,** 1960–1964.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 14863–14868.

Farr, S., and Dunn, R. T., II (1999). Concise review: Gene expression applied to toxicology. *Toxicol. Sci.* **50,** 1–9.

National Institutes of Health.

Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C., Trent, J. M., Staudt, L. M., Hudson, J., Jr., Boguski, M. S., Lashkiri, D., Shalon, D., Botstein, D., and Brown, P. O. (1999). The transcriptional program in the response of human fibroblasts to serum. *Science* **283,** 83–87.

Jover, R., Bort, R., Gomez-Lechon, M. J., and Castell, J. V. (1998). Re-expression of C/EBPα induces CYP2B6, CYP2C9 and CYP2D6 genes in HepG2 cells. *FEBS Lett.* **431,** 227–230.

Liu, H., Miller, E., van de Water, B., and Stevens, J. L. (1998). Endoplasmic reticulum stress proteins block oxidant-induced Ca2+-increases and cell death. *J. Biol. Chem.* **273,** 12858–12862.

Marshall, E. (1998). NIH to produce a "working draft" of the genome by 2001. *Science* **281**, 1774–1775.

Marton, M. J., Derisi, J. L., Bennett, H. A., Iyer, V. R., Meyer, M. R., Roberts, C. J., Stoughton, R., Burchard, J, Slade, D., Dai, H., Bassett, D. E., Jr., Hartwell, L. H., Brown, P. O., and Friend, S. H. (1998) Drug target

validation and identification of secondary drug effects using DNA microarrays. *Nat. Med.* **4,** 1293–1301.

Muller, M., Strand, S., Hug, H., Heinemann, E.-M., Walczak, H., Hofmann, W. J., Stremmel, W., Krammer, P. H., and Galle, P. R. (1997). Drug-induced apoptosis in hepatoma cells is mediated by the CD95 (Apo-1/Fas) receptor/ligand system and involves activation of wild-type p53. *J. Clin. Invest.* **99,** 403–413.

National Institutes of Health, human genome sequencing. Availabile at http://www.ncbi.nim.nih.gov/Web/Genbank. Accessed May 10, 2000.

Nehme, A., Baskaran, R., Aebi, S., Fink, D., Nebel, S., Cenni, B., Wang, J. Y., Howell, S. B., and Christen, R. D. (1997). Differential induction of c-jun NH_2-terminal kinase and c-abl kinase in DNA mismatch repair-proficient and -deficient cells exposed to cisplatin. *Cancer Res.* **57,** 3253–3257.

Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C., and Afshari, C. A. (1999). Microarrays and toxicology: The advent of toxicogenomics. *Mol. Carcinog.* **24**, 153–159.

Pennisi, E. (1999). Mouse genome added to sequencing effort. *Science* **286,** 211.

Pennisi, E. (2000a). Human genome: Finally the book of life and instructions for navigating it. *Science* **288,** 2304–2307.

Pennisi, E. (2000b). Mouse sequencers take up the shotgun. *Science* **287,** 1179–1181.

Phase 1 J toxicology database. Phase 1 Molecular Toxicology, Inc. Available at: http://www.phase1tox.com. Accessed May 10, 2000.

Rodi, C. P., Bunch, R. T., Curtiss, S. W., Kier, L. D., Cabonce, M. A., Davila, J. C., Mitchell, M. D., Alden, C. L., and Morris, D. L. (1999). Revolution through genomics in investigative and discovery toxicology. *Toxicol. Pathol.* **27,** 107–110.

Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J. C., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D., and Brown, P. O. (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24,** 227–235.

Sanchez-Perez, I., and Perona, R. (1999). Lack of c-jun activity increases survival to cisplatin. *FEBS Lett.* **453,** 151–158.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene-expression patterns with a complementary DNA microarray. *Science* **270,** 467–470.

Scherf, U., Ross D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., Kohn, K. W., Reinhold, W. C., Myers, T. G., Andrews, D. T., Scudiero, D. A., Eisen, M. B., Sausville, E. A., Pommier, Y., Botstein, D., Brown, P. O., and Weinstein, J. N. (2000). A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* **24,** 236–244.

Shalon, D., Smith, S. J., and Brown, P.O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* **6**, 639–645.

Shivakuvar, C. V., Brown, D. R., Deb, S., and Deb, S. P. (1999). Wild-type human p53 transactivates the human proliferating cell nuclear antigen promoter. *Mol. Cell. Biol.* **15,** 6785–6793.

Simon, J. A., Szankasi, P., Nguyen, D. K., Ludlow, C., Dunstan, H. M., Roberts, C. J., Jensen, E. L., Hartwell, L. H., and Friend, S. H. (2000). Differential toxicities of anticancer agents among DNA repair and checkpoint mutants of *Saccharomyces cerevisiae*. *Cancer Res.* **60,** 328–333.

Spotfire, Inc., toxicant expression profiles. Available at: http:// www.spotfire.com. Accessed May 10, 2000.

Sun, X., Shimizu, H., and Yamamoto, K. (1995). Identification of a novel p53 promoter element involved in genotoxic stress-inducible p53 gene expression. *Mol. Cell. Biol.* **15,** 4489–4496.

Suzuki, A., Tsutomi, Y., Akahane, K., Araki, T., and Miura, M. (1998). Resistance to fas-mediated apoptosis: Activation of caspase 3 is regulated by cell cycle regulator $p21^{waf1}$ and IAP gene family ILP. *Oncogene* **17,** 931–939.

Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. (1998). Shotgun sequencing of the human genome. *Science* **280**, 1540–1542.

Weinstein, J. N., Myers, T. G., O′Connor, P. M., Friend, S. H., Fornace, A. J., Jr, Kohn, K. W., Fojo, T., Bates, S. E., Rubinstein, L. V., Anderson, N. L., Buolamwini, J. K., van Osdol, W. W., Monks, A. P., Scudiero, D. A., Sausville, E. A., Zaharevitz, D. W., Bunow, B., Viswanadhan, V. N., Johnson, G. S., Wittes, R. E., and Paull, K. D. (1997). An information-intensive approach to the molecular pharmacology of cancer. *Science* **275,** 343–349.

Yoshida, H., Haze, K., Yanagi, H., Yura, T., and Mori, K. (1998). Identification of the cis-acting endoplasmic reticulum stress response element responsible for transcriptional induction of mammalian glucose-regulated proteins. *J. Biol. Chem.* **273,** 33741–33749.

Zamble, D. B., Jacks, T., and Lippard, S. J. (1998). p53-Dependent and -independent responses to cisplatin in mouse testicular teratocarcinoma cells. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 6163–6168.