0026-895X/01/6006-1189–1194$3.00
MOLECULAR PHARMACOLOGY
Vol. 60, No. 6
1257/946626
Mol Pharmacol 60:1189–1194, 2001
Printed in U.S.A.

ACCELERATED COMMUNICATION

Identification of toxicologically predictive gene sets using cDNA microarrays

RUSSELL S. THOMAS, DAVID R. RANK, SHARRON G. PENN, GINA M. ZASTROW, KEVIN R. HAYES, KALYAN PANDE, EDWARD GLOVER, TOMI SILANDER, MARK W. CRAVEN, JANARDAN K. REDDY, STEVAN B. JOVANOVICH, and CHRISTOPHER A. BRADFIELD

McArdle Laboratory for Cancer Research (R.S.T., G.M.Z., K.R.H., K.P., E.G., C.A.B.) and Department of Biostatistics and Medical Informatics (M.W.C.), University of Wisconsin Medical School, Madison, Wisconsin; Aeomica, Sunnyvale, California (D.R.R., S.G.P.); Department of Computer Science, University of Helsinki, Teollisuuskatu, Finland (T.S.); Department of Pathology, Northwestern University Medical School, Chicago, Illinois (J.K.R.); and Advanced Research Team, Molecular Dynamics Inc., Sunnyvale, California (S.B.J.)

Received July 31, 2001; accepted August 30, 2001

This paper is available online at http://molpharm.aspetjournals.org

ABSTRACT

We have developed an approach to classify toxicants based upon their influence on profiles of mRNA transcripts. Changes in liver gene expression were examined after exposure of mice to 24 model treatments that fall into five well-studied toxicological categories: peroxisome proliferators, aryl hydrocarbon receptor agonists, noncoplanar polychlorinated biphenyls, inflammatory agents, and hypoxia-inducing agents. Analysis of 1200 transcripts using both a correlation-based approach and a probabilistic approach resulted in a classification accuracy of between 50 and 70%. However, with the use of a forward parameter selection scheme, a diagnostic set of 12 transcripts was identified that provided an estimated 100% predictive accuracy based on leave-one-out cross-validation. Expansion of this approach to additional chemicals of regulatory concern could serve as an important screening step in a new era of toxicological testing.

Toxicologists employ a battery of tests to identify chemicals with potential for human toxicity or that might cause environmental harm. According to the United States National Toxicology Program (NTP), a thorough analysis of each chemical requires $2 to 4 million and several years to complete (National Toxicology Program, 1996). Because of the cost- and labor-intensive nature of these studies, the number of chemicals currently tested by the NTP stands at 505 in long-term studies, 66 in short-term tests, and one subchronic study (http://ntp-server.niehs.nih.gov/). Given that there are approximately 70,000 chemicals in commerce today (National Toxicology Program, 1996), it is clearly impossible to apply current testing methods to all chemicals of concern. It is apparent that alternative testing approaches must be developed if science is going to maintain a significant role in environmental and public health policy.

This work was supported by the Burroughs Wellcome Foundation, the National Institutes of Health (Grants ES05703, T32-CA09681, CA07175, GM23750, and HG01775), and a postdoctoral fellowship cosponsored by the Society of Toxicology and the Colgate-Palmolive Corporation.

The development of a screen that would allow prioritization of untested chemicals based upon their toxic potential would have a significant impact on how efficiently we evaluate both synthetic and naturally occurring compounds. One approach for predicting toxic potential is to classify chemicals based upon their capacity to alter transcriptional programs in a manner that is similar to known toxicants (Nuwaysir et al., 1999). Test chemicals that induce transcriptional responses in a manner similar to those induced by a known poison could then be classified as harboring toxic potential and examined carefully by more thorough toxicological means. This approach has two underlying assumptions: 1) that we have enough scientific information to allow proper classification of prototype toxicants and 2) that most if not all toxic chemical exposures will alter gene expression at some level. In support of this second assumption, signal transduction pathways that culminate in a transcriptional response mediate the toxicity of many chemicals. In addition, toxicity is commonly manifested as inflammation, proliferation, apo-

ABBREVIATIONS: AHR, aryl hydrocarbon receptor; PCB, polychlorinated biphenyl; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; BNF, β-naphthoflavone; cipro, ciprofibrate; IL-6, interleukin-6; LPS, lipopolysaccharide; PCB-153, 2,2′,4,4′,5,5′-hexachlorobiphenyl; phenobarb, phenobarbital; phenylhyrzn, phenylhydrazine; TNFα, tumor necrosis factor α.

ptosis, necrosis, and/or cellular differentiation. All of these toxic endpoints are intimately linked to specific alterations in gene expression.

To test the hypothesis that toxicants can be classified according to their influence on global gene expression profiles, we employed cDNA microarray technology (Schena et al., 1995; Golub et al., 1999) and attempted to classify 24 prototype chemical treatments that fall into five well-characterized toxicological classes. Three of the classes include environmental pollutants that are targets of regulatory concern [i.e., peroxisome proliferators, AHR agonists, and noncoplanar PCBs (Schmidt and Bradfield, 1996; Carpenter, 1998; Vanden Heuvel, 1999)]. The remaining two classes consist of agents that stimulate other common toxic endpoints, such as treatments that induce the inflammatory response and treatments that stimulate the hypoxia signal transduction pathway.

Materials and Methods

Animals and Treatment. All animals treated were male C57BL/6J mice. The treatment, dose, vehicle, and time at sacrifice were selected from the literature and are shown in Table 1. All treatments were compared with their respective vehicle controls at the corresponding time points. A summary of these agents and their general toxicological category is provided in Table 2.

RNA Isolation. Total RNA from cells and frozen liver tissue was isolated using the RNeasy system (QIAGEN, Valencia, CA). Poly-A selected RNA was purified using the Oligotex kit (QIAGEN) and checked using gel electrophoresis and/or spectrophotometric measurements. mRNA from the livers of three or more mice were pooled for analysis on the microarrays.

cDNA Microarray Construction and Analysis. Approximately 1200 minimally redundant cDNAs from internal expressed sequence tag projects (http://edge.oncology.wisc.edu/) and the public expressed sequence tag effort were identified and the glycerol stocks were rearrayed into a separate set of clones. An aliquot from these clone sets was amplified by polymerase chain reaction and used to construct custom cDNA microarrays using methods described previously (Worley et al., 2000). Each cDNA clone was spotted four times on each slide for replicate analysis. Labeled cDNA probe was produced from 1 μg of poly-A RNA by incorporation of Cy3- or Cy5-dCTP (Amersham Pharmacia Biotech, Piscataway, NJ) during a standard reverse transcriptase reaction as described previously (Penn et al., 2000). The slides were scanned using a microarray scanner (Molecular Dynamics, Sunnyvale, CA) and the fluorescence data were analyzed using ArrayVision software package (Imaging Reseach, St. Catharines, ON, Canada). For each hybridization, the four spots corresponding to each cDNA were averaged and normalized to an internal suite of 15 housekeeping transcripts coding for ribosomal proteins. Normalization was carried out using methods described previously (Chen et al., 1997).

To eliminate any dye bias, all samples were analyzed at least twice, with one experiment using Cy3 to label the control mRNA and Cy5 to label the treated mRNA and, in the replicate experiment, Cy5 to label the control mRNA and Cy3 mRNA to label the treated mRNA. The results from the replicate hybridizations were averaged. Significance thresholds based on 99% confidence intervals were also calculated for each treatment using the variability in the normalized housekeeping ratios (Chen et al., 1997). The average upper and lower significance thresholds for all of the treatments in the study were 1.4-fold and −1.4-fold, respectively. As a verification of the quality of the data from the microarray experiments, a replicate set of hybridizations were performed of control mRNA compared against the same sample of control mRNA (i.e., homotypic control versus control hybridization). The average ratio from the homotypic hybridizations was 1.01 with a 99% confidence interval of 1.30 to −1.27.

Data Reduction. Before statistical analysis, the data set was screened for transcripts that did not respond to any of the treatments used in the study and would not contribute significantly to the classification. A threshold of 2-fold change in gene expression was used as the cut-off value. The 2-fold cut-off value was slightly more conservative than the confidence limit calculated using techniques published previously (i.e., 1.4-fold) (Chen et al., 1997) and is similar to a standard threshold level used in other studies (e.g., Schena et al., 1996). Using this threshold value, we determined that approximately 500 of the 1200 transcripts changed significantly in response to at least one treatment.

Additional screening was performed to identify a subset of transcripts with a stable temporal expression and provide computational savings by eliminating variables that would have little predictive value. For those treatments with time course profiles, all time-points were collapsed into a single average value representing the average change in that transcript over time. After collapsing the data, only transcripts that showed a change greater than 2-fold in more than one treatment were selected for further analysis. After this data reduction, only 74 transcripts remained. It should be noted that after collapsing the data to identify the stable transcripts, the individual time points were analyzed separately in the classification analysis. Finally, the gene expression values were discretized such that transcripts up-regulated greater than 2-fold were given a value of one, transcripts down-regulated greater than 2-fold were given a value of minus one, and transcripts with less than a 2-fold change were given a value of zero. This data was considered the initial training set for

TABLE 1
Treatment, dose, vehicle, and time of sacrifice

Treatment	Dose[a]	Vehicle		Time of Sacrifice
		Substance	Amount	
Aroclor 1260	200 mg/kg	Corn oil	8 ml/kg	48 h
BNF	50 mg/kg	Corn oil	10 ml/kg	24 h
Cipro	250 mg/kg gavage	DMSO	0.2 ml	2, 4, and 8 h
Cobalt chloride	60 mg/kg	0.9% Saline	10 ml/kg	48 h
TCDD	10 μg/kg	DMSO	100 μl/kg	6, 12, 24, 48, 96, 192, 288, 384, and 480 h
IL-6	25 μg/kg	PBS	10 ml/kg	6 h
LPS	1 mg/kg	PBS	10 ml/kg	6, 12, and 24 h
PCB-153	80 mg/kg	Corn oil	8 ml/kg	48 h
Phenobarb	100 mg/kg/day (3 days)	0.9% Saline	10 ml/kg	72 h
Phenylhyrzn	100 mg/kg	PBS	10 ml/kg	48 h
TNFα	50 μg/kg	PBS	10 ml/kg	6 h
Wy-16,463	0.125% w/w	Powdered chow		2 weeks

DMSO, dimethyl sulfoxide; PBS, phosphate-buffered saline.
[a] Dose was administered intraperitoneally unless noted otherwise.

use in the development of a model to allow organization of future test chemicals into the five classifications defined in Table 2.

Statistical Classification Analysis. The Naïve Bayesian classification used in this article is based on previous work by Kontkanen et al. (1998). Briefly, the predictor variables $X_1,\ldots,X_k$ are assumed to be independent of each other when conditioned on the class variable C. Our model M is constructed by the joint probability distribution for a data vector (x, c) = ($X_1 = x_1, \ldots, X_k = x_k$, C = c) and can be written as follows:

$$P(x,c) = P(C = c) \prod_{i=1}^{k} P(X_i = x_i | C = c) \quad (1)$$

Before incorporating any data, we also assume that all classes are equally probable (i.e., the probability that a chemical will belong to a certain class is the same for each class) and that within each class, the gene expression values of each transcript are also equally probable. Given these assumptions, we can use Bayesian probability theory to calculate the conditional predictive distribution for the class c given x and the data set D by:

$$P(c|x,D) = \frac{P(c,x|D)}{P(x|D)} \quad (2)$$

where the numerator is calculated as:

$$P(c,x|D) = \frac{t_c + 1}{N_t + NC} \prod_{i=1}^{k} \frac{f_{cxi} + 1}{F_c + V_{xi}} \quad (3)$$

where t_c is the number of treatments in class c, N_t is the total number of treatments, NC is the total number of classes, f_{cxi} is the number of cases in class c having a value equal to x_i, F_c is the number of treatments in class c, and V_{xi} is the number of possible values of x_i. The denominator is the same for all c and is calculated as:

$$P(x|D) = \sum_{c'} P(c',x|D) \quad (4)$$

The result of this conditional predictive distribution is then used to classify the data vector. For the parameter selection, a modification of forward parameter selection process outlined in Huberty (1994) was employed. Specifically, an iterative process was used in which transcripts were run individually using the Naïve Bayes model and the transcript with the best internal classification rate and highest confidence (represented by the sum of all probabilities for correctly classified treatments) was selected. In the subsequent round, the selected transcript was fixed and the remaining parameters were added individually to find which transcript, along with the first selected transcript, produced the highest internal classification rate and confidence. This process was repeated until all 74 transcripts were added to the model in the order of their internal classification rate. This type of selection ranks the transcripts in the order of their estimated predictive value and adds them sequentially to the model. It should be noted that the forward selection approach does not necessarily yield the best set or even the smallest set. In addition, genes with similar expression profiles may also be added to the set if they significantly increase the accuracy or confidence. The approach is simply a heuristic to look for a diagnostic set of predictors with a high accuracy. A flow chart describing the data reduction and classification analysis is provided in Fig. 1.

To estimate the predictive accuracy of this approach, the process of parameter selection was integrated with leave-one-out cross-validation in which one of the treatments is removed from the analysis, and the model is constructed and then used to predict the left-out treatment. The predictive accuracy was then assessed after each parameter was added and the number of genes in the 'diagnostic set' was chosen based on the peak predictive accuracy and confidence of the model. The final 'diagnostic set' was derived by following the same procedure on the complete data set (i.e., no treatment left out). It should be noted that in choosing the number of genes to include in the 'diagnostic set', we had the benefit of understanding how the estimated accuracy changed with the addition or removal of genes from this list. In a real setting, however, this information would be absent and the decision on the number of genes to include may be more difficult. Further research in this area is needed.

Results and Discussion

To allow the accurate classification of large numbers of toxicants based on gene expression, several important factors were considered in the experimental design. First, exposures were performed in inbred mice in an effort to minimize the influence of genetic polymorphism on transcriptional responses to toxicants. Second, gene expression monitoring was focused on the liver, because this organ is often the first significant site of chemical exposure and exhibits a wide array of pathological and adaptive responses to a broad spectrum of toxicants. Third, gene expression was characterized using mRNA obtained from whole liver after an acute exposure to a test chemical because the whole-organ response better represents the full range of transcriptional changes that can result from toxicity. In contrast to cell culture studies, this in vivo approach is particularly important when multiple cell-types and paracrine signaling are required for the complete toxic response (e.g., inflammation). Finally, in an effort to understand and adjust for the influence of temporal and sample acquisition variables, multiple time-points were analyzed for several of the treatments.

To represent the transcriptional response as a whole, a two-dimensional hierarchical clustering method was employed (Eisen et al., 1998) and the relationships between treatments based on gene expression profiles are highlighted in the adjacent dendrogram (Fig. 2). A visual inspection of the treatment-dendrogram indicates that individual chemicals, with a few exceptions, generally fall into their anticipated toxicological classes. Specifically, the hypoxia treatments (i.e., phenylhydrazine and cobalt) are intermixed with the noncoplanar PCBs showing some similarity among the gene expression profiles for these treatments (Fig. 2). Application of a more formal nearest-neighbor classification anal-

TABLE 2
General toxicological classes and corresponding treatments classified in this study using gene expression profiles from cDNA microarrays

Noncoplanar-PCBs	Peroxisome Proliferators	Inflammatory	Hypoxia	AHR Agonists
PCB-153	Cipro[a]	TNF-α	Cobalt	TCDD[a]
Aroclor-1260[b]	Wy-16,463	LPS[a]	Phenylhrzn	BNF
Phenobarb		IL-6		

[a] Multiple time points investigated.
[b] Aroclor-1260 is a PCB mixture that contains primarily non-coplanar, phenobarbital-like PCBs (Harris et al., 1993; Ngui and Bandiera, 1999).

ysis using a similar correlation based metric indicated that 7 of 24 treatments were closer to treatments outside of their class as opposed to within (data not shown).

The imperfect classification scheme attained by the initial clustering analysis on the transcriptional response across the whole microarray was not surprising. Although our classification scheme assumes that chemicals in each toxicant class act through a common mechanism, individual members of each class may stimulate unique pathways that may be secondary or unrelated to toxicity. For example, although considerable pharmacological and genetic evidence indicates that halogenated-dioxins lead to toxicity through their binding to the Ah-receptor, the agonist BNF is also known to stimulate the antioxidant response pathway (Poland and Glover, 1980; Radjendirane and Jaiswal, 1999). Similarly, although agents that stimulate the transcriptional response to hypoxia act through up-regulation of HIF1α, some of these agents have also been shown to induce the acute phase response (Wenger et al., 1995). The resulting pattern of expression across the whole microarray reflects small, chemical-specific differences at the molecular level that result in a virtual 'fingerprint' of expression for individual compounds. Therefore, that subset of transcripts that allows classification of these treatments must be defined according to our understanding of the primary toxic mechanism. In other words, predictor variables (in our case transcripts) must be screened for their ability to discriminate between groups, and a subset of these variables must be derived that allows accurate predictions.

As a method of identifying a diagnostic set of predictor transcripts, we applied a probabilistic approach based upon Bayesian statistics (Duda and Hart, 1973; Kontkanen et al., 1998). Here, gene expression values were discretized and a standard forward parameter selection algorithm was employed to select predictor variables to be added to the model (Huberty, 1994). This type of selection ranks the transcripts in the order of their estimated predictive value and adds them sequentially to the model. For example, in the first round of selection, all transcripts were run individually using

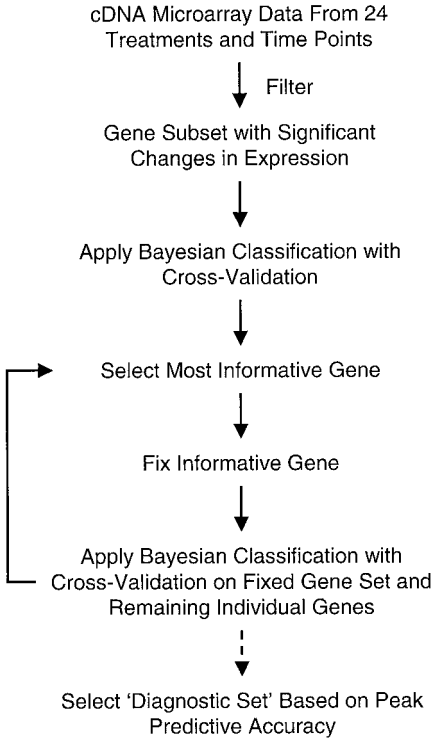

Fig. 1. A flow-chart illustrating the method for the data reduction and classification analysis leading to the diagnostic set of transcripts.

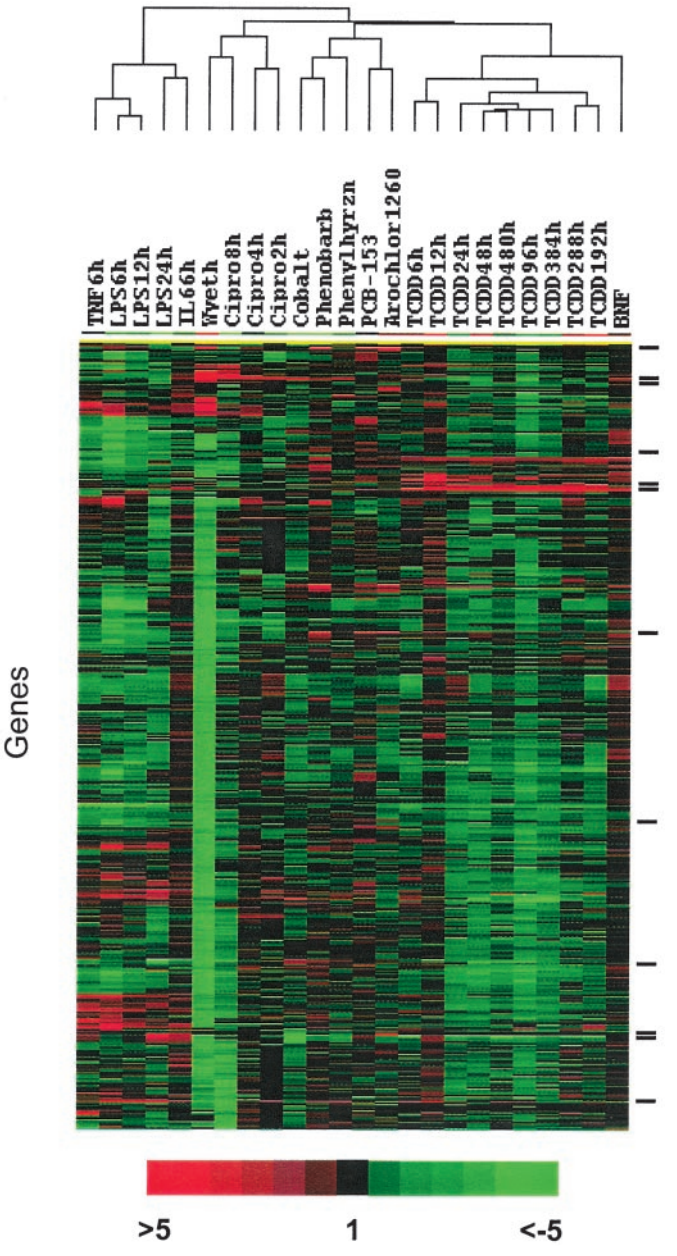

Fig. 2. Variation in expression of approximately 500 transcripts in 24 experimental treatments. The data are presented after two-dimensional hierarchical clustering, which organizes transcripts and treatments on the basis of similarity. Each row represents a single transcript and each column an experimental treatment. For each treatment, the ratio of the expression of each transcript to the expression of the transcript in control samples is represented by the color of the corresponding cell in the graph. Green represents down-regulation of the transcript, black means no change, and red represents up-regulation of the transcript. Color saturation reflects the magnitude of the change. Clustering was performed using complete linkage clustering with a centered correlation coefficient as the similarity metric (Eisen et al., 1998). A dendrogram representing similarities between experimental treatments is shown at the top and the approximate locations of the genes in the final diagnostic set are noted with bars to the right of the figure. Wyeth, Wy-16,463.

the Naïve Bayes model, and the transcript with the best internal classification rate and highest confidence (represented by the probabilities for all correctly classified treatments) was selected. The internal classification rate for a given set of predictor variables is measured based on the classification rate within the training set. In the next round, the selected transcript was fixed and the remaining parameters were added individually to find which transcript, along with the first selected transcript, produced the highest internal classification rate and confidence. This process was repeated until all transcripts were added to the model in the order of their internal classification rate. To estimate the predictive accuracy of this approach, the process was integrated with leave-one-out cross-validation, in which one of the treatments is removed from the analysis, then the model is constructed and used to predict the left-out treatment.

The results of this analysis show the predictive accuracy as a function of the number of transcripts added to the model during the forward parameter selection process (Fig. 3). Based on this analysis, we found that the predictive accuracy and confidence of the model began to level out after the addition of approximately a dozen transcripts and even began to drop soon thereafter. Consequently, a set of 12 transcripts was considered 'diagnostic' for the classification of treatments into the toxicological classes investigated in this study. The final 'diagnostic set' was derived by following the same procedure on the complete data set (i.e., no treatment left out). The transcriptional profile of the 12 diagnostic transcripts, and their order of their addition to the model, is shown in Fig. 4.

The forward selection and cross-validation process identified several properties in our toxicological profiles. First, a 'diagnostic set' of 12 transcripts was identified that allows an estimated 100% predictive accuracy for the toxicological classes chosen for this study. Interestingly, the transcripts in this diagnostic set are a mix of transcripts previously known to be altered by these treatments and relatively uncharacterized transcripts with respect to toxicant regulation. For example, CYP1A2 and CYP4A10 are known to be up-regulated by TCDD and peroxisome proliferators, respectively (Bell et al., 1993; Schmidt and Bradfield, 1996). In contrast, for IL-18 and betaine homocysteine methyltransferase, we have little information regarding their response after toxicant exposure. Overall, about half of the changes in the optimal set were described previously at some level in the literature. Second, the forward selection approach also explains the previously described 50 to 70% predictive accuracy attained when using the whole data set (500 transcripts), in that soon after the addition of the diagnostic set transcripts, the estimated predictive accuracy begins to decline significantly. In other words, the further addition of transcripts beyond the diagnostic set begins to split out treatments and the 'individuality' of the treatment profiles begins to take over.

Despite the apparent success of our study in establishing the potential for classification of toxic chemicals according to diagnostic sets of genes, the success should be framed in a couple ways. It should be noted that the exposures chosen in this study represent single acute doses and may not reflect the gene expression changes at lower, environmentally relevant doses. In an organism, multiple factors converge to ultimately influence the manifestation of toxicity and the associated gene expression patterns. Among these factors are time, dose, route of administration, age of the animal, and sex. Characterizing the influence of all of these variables on transcript profiles with even a small number of treatments would require considerable resources (e.g., 12 treatments $\times$ 3 time points per treatment $\times$ 3 doses per time point $\times$ 3 routes per dose = 324 microarray studies). Although in this study we chose to primarily address the factor of time, additional experimentation was also performed to look at how different doses of TCDD affected classification. In these preliminary dose-response studies, our statistical model proved to be resistant to the variation in TCDD doses with correct classification at doses as low as 0.05 μg/kg and as high as 100 μg/kg (data not shown). Arguably, other chemicals may not be as easy to classify over such a large dose range and additional studies will be needed to address this issue and other factors that may affect the predictive accuracy of the model. To understand how our statistical model performed with a treatment that was not in our five toxicological categories, results from arsenic-treated mice were also analyzed. Interestingly, the model did not classify the arsenic treatment with any

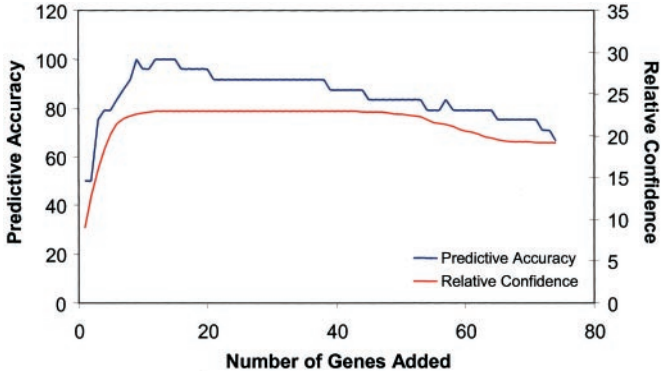

Fig. 3. Estimated predictive accuracy and relative confidence of the classification model with the addition of selected transcripts. The order of transcripts added to the model and the predictive accuracy were performed using a forward selection scheme and leave-one-out cross-validation, respectively (see text for details).

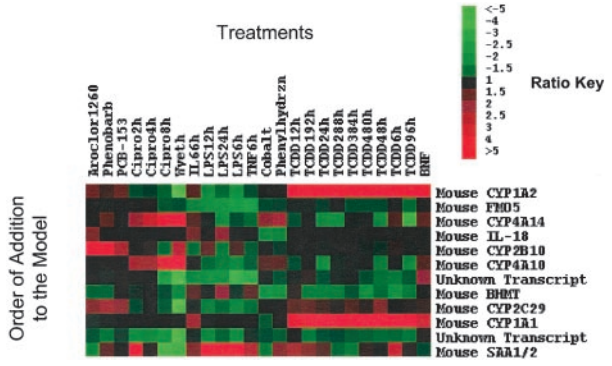

Fig. 4. Variation in expression of the diagnostic set of transcripts used to classify the 24 treatments into the five toxicological classes. Each row represents a single transcript and each column an experimental treatment. The order of the transcripts, from top to bottom, is in the order of addition to the probability model. For each treatment, the expression ratio of each transcript in a treated sample to the expression in the control sample is represented by the color of the corresponding cell in the graph. Green represents down-regulation of the transcript, black means no change, and red represents up-regulation of the transcript. Color saturation reflects the magnitude of the change and can be compared with the ratio key. Wyeth, Wy-16,463.

degree of confidence, with inflammatory as the closest category and AHR agonists as the least similar category (data not shown). This adds additional confidence that accurately predicting toxicological endpoints based on gene expression is achievable.

Several interesting conclusions can be drawn from this work. Most importantly, we have presented evidence that the accurate classification of toxic chemicals according to their transcript expression profiles is possible. This opens the door to a new era of toxicological testing where relatively short and inexpensive studies using transcript expression as an endpoint allow the prioritization of untested chemicals based upon their classification. This would mean significant savings in both animal usage and financial resources and would reduce the disparity between the number of tested and untested chemicals in commerce today. However, this study is just the first step toward this goal. The toxicological categories selected in our study primarily reflect the model compounds that toxicologists have studied extensively over the last decade and represent only a small percentage of the 70,000 chemicals in commerce today.

Obviously, as the public gene expression database grows, more toxicological categories can be added to the model and the more predictive our model and those like it will become. Another conclusion from this work is that large arrays with thousands of transcripts are unnecessary to make these classifications. Although the large arrays are necessary to initially identify the diagnostic gene set, once these 'diagnostic' sets of indicator transcripts are identified, simple measurements of only one or two dozen transcripts may allow the average investigator the ability to make judgments as to the relative toxicity of a particular chemical. Finally, we have purposely chosen to use a robust set of statistical methods and conservative assumptions to develop this predictive set. The discrete nature of the approach does not require an accurate measurement of transcriptional changes beyond the assessment of only significant up- or down-regulation. This should make subsequent evaluations more resistant to inter- and intralaboratory variability such as that observed when switching arrays or performing hybridizations in multiple laboratories.

Acknowledgments

We thank Dr. Elaina M. Kenyon of the United States Environemntal Protection Agency for contributing tissue from arsenic-exposed mice.

References

Bell DR, Plant NJ, Rider CG, Na L, Brown S, Ateitalla I, Acharya SK, Davies MH, Elias E, Jenkins NA, et al (1993) Species-specific induction of cytochrome P-450 4A RNAs: PCR cloning of partial guinea-pig, human and mouse CYP4A cDNAs. *Biochem J* **294:**173–180.

Carpenter DO (1998) Polychlorinated biphenyls and human health. *Int J Occup Med Environ Health* **11:**291–303.

Chen Y, Dougherty ER, and Bittner ML (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J Biomed Optics* **24:**364–374.

Duda R and Hart P (1973) *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York.

Eisen MB, Spellman PT, Brown PO, and Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**:14863–14868.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, et al. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science (Wash DC)* **286:**531–537.

Harris M, Zacharewski T, and Safe S (1993) Comparative potencies of Aroclors 1232, 1242, 1248, 1254, and 1260 in male Wistar rats—assessment of the toxic equivalency factor (TEF) approach for polychlorinated biphenyls (PCBs). *Fund Appl Toxicol* **20:**456–463.

Huberty CJ (1994) *Applied Discriminant Analysis.* John Wiley & Sons, New York.

Kontkanen P, Myllymaki P, Silander T, and Tirri H (1998) BAYDA: Software for Bayesian classification and feature selection, in *Proceedings of the Fourth International Conference on Knowledge Discovery & Data Mining*, AAAI Press, Menlo Park, CA.

Ngui JS and Bandiera SM (1999) Induction of hepatic CYP2B is a more sensitive indicator of exposure to Aroclor 1260 than CYP1A in male rats. *Toxicol Appl Pharmacol* **161:**160–170.

National Toxicology Program (1996) *National Toxicology Program: Annual Plan for Fiscal Year 1996*. National Institute of Environmental Health Sciences, Research Triangle Park, NC.

Nuwaysir EF, Bittner M, Trent J, Barrett JC, and Afshari CA (1999) Microarrays and toxicology: The advent of toxicogenomics. *Mol Carcinogenesis* **24:**153–159.

Penn SG, Rank DR, Hanzel DK, and Barker DL (2000) Mining the human genome using microarrays of open reading frames. *Nat Genet* **26:**315–318.

Poland A and Glover E (1980) 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin: segregation of toxicity with the Ah locus. *Mol Pharmacol* **17:**86–94.

Radjendirane V and Jaiswal AK (1999) Antioxidant response element-mediated 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) induction of human NAD(P)H:quinone oxidoreductase 1 gene expression. *Biochem Pharmacol* **58:**1649–1655.

Schena M, Shalon D, Davis RW, and Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (Wash DC)* **270:**467–470.

Schena M, Shalon D, Heller R, Chai A, Brown PO, and Davis RW (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* **93:**10614–10619.

Schmidt JV and Bradfield CA (1996) Ah receptor signaling pathways. *Annu Rev Cell Dev Biol* **12:**55–89.

Vanden Heuvel JP (1999) Peroxisome proliferator-activated receptors (PPARs) and carcinogenesis. *Toxicol Sci* **47:**1–8.

Wenger RH, Rolfs A, Marti HH, Bauer C, and Gassmann M (1995) Hypoxia, a novel inducer of acute phase gene expression in a human hepatoma cell line. *J Biol Chem* **270:**27865–27870.

Worley J, Bechtol K, Penn S, Roach D, Hanzel D, Trounstine M, and Barker D (2000) A systems approach to fabricating and analyzing DNA microarrays, in *Microarray Biochip Technology* (Schena M ed) pp 65–86, Biotechniques Books, Natick, MA.

Address correspondence to: Dr. Christopher A. Bradfield, McArdle Laboratory for Cancer Research, 1400 University Avenue, Madison, WI 53706-1599. E-mail: bradfield@oncology.wisc.edu