0026-895X/05/6704-1360–1368$20.00
MOLECULAR PHARMACOLOGY
Copyright © 2005 The American Society for Pharmacology and Experimental Therapeutics
Mol Pharmacol 67:1360–1368, 2005

Vol. 67, No. 4
9175/1198128
Printed in U.S.A.

EDGE: A Centralized Resource for the Comparison, Analysis, and Distribution of Toxicogenomic Information

Kevin R. Hayes, Aaron L. Vollrath, Gina M. Zastrow, Brian J. McMillan, Mark Craven, Stevan Jovanovich, David R. Rank, Sharon Penn, Jacqueline A. Walisser, Janardan K. Reddy, Russell S. Thomas, and Christopher A. Bradfield

McArdle Laboratory for Cancer Research, University of Wisconsin-Madison Medical School, Madison, Wisconsin (K.R.H., A.L.V., G.M.Z., B.J.M., J.A.W., R.S.T., C.A.B.); Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin (M.C.); Amersham Biosciences Inc., Sunnyvale, California (S.J., D.R.R., S.P.); Department of Pathology, Northwestern University, The Feinberg School of Medicine, Chicago, Illinois (J.K.R.); and CIIT Centers for Health Research, Research Triangle Park, North Carolina (R.S.T.)

Received November 12, 2004; accepted January 6, 2005

ABSTRACT

Transcriptional profiling via microarrays holds great promise for toxicant classification and hazard prediction. Unfortunately, the use of different microarray platforms, protocols, and informatics often hinders the meaningful comparison of transcriptional profiling data across laboratories. One solution to this problem is to provide a low-cost and centralized resource that enables researchers to share toxicogenomic data that has been generated on a common platform. In an effort to create such a resource, we developed a standardized set of microarray reagents and reproducible protocols to simplify the analysis of liver gene expression in the mouse model. This resource, referred to as EDGE, was then used to generate a training set of 117 publicly accessible transcriptional profiles that can be accessed at http://edge.oncology.wisc.edu/. The Web-accessible database was also linked to an informatics suite that allows on-line clustering and K-means analyses as well as Boolean and sequence-based searches of the data. We propose that EDGE can serve as a prototype resource for the sharing of toxicogenomics information and be used to develop algorithms for efficient chemical classification and hazard prediction.

More than 80,000 chemicals are in commercial use in the United States. This number, plus the addition of approximately 2000 new chemicals each year, makes it impossible to properly assess the toxicity of each compound in a timely manner (http://ntp-server.niehs.nih.gov). It has been proposed that the toxicity of chemicals may be predicted based upon how they influence patterns of global gene expression and how these patterns correlate with profiles induced by prototype chemicals (Amin et al., 2002; Thomas et al., 2002). Preliminary attempts at toxicant classification have been successful using approaches such as hierarchical clustering, linear discriminant analysis, and Bayesian statistics. For small data sets, such approaches have yielded predictive accuracies approaching 90 to 100% (Thomas et al., 2001; Hamadeh et al., 2002a,b; Kramer et al., 2004). The idea that transcriptional profiles will soon become a standard component of toxicology assessment is supported by the observation that the United States Food and Drug Administration recently released a "Guidance to Industry" designed to facilitate the use of such data in the regulatory process (http://www.fda.gov/cder/guidance/5900dft.pdf). This document suggests that future regulatory decisions on drugs and other chemicals are likely to be influenced by the transcriptional profiles they elicit.

Given that regulatory toxicology is largely a descriptive science dependent on reproducibility of observations, many problems must be solved before we can take full advantage of the information found in a chemically induced transcriptional profile. Issues related to data accessibility and standardization are among the most important factors that inhibit the comparison of gene expression data across laboratories (Waring et al., 2004). For example, microarray signals can be dependent on the sequence and size of the target DNAs used in their generation (Kuo et al., 2002).

This work was supported by National Institutes of Health Grants R37-E5005303, P30-CA014520, R01-E5012752, and T32-CA009135. R.T. was a recipient of a Colgate Palmolive/Society of Toxicology postdoctoral fellowship.
Article, publication date, and citation information can be found at http://molpharm.aspetjournals.org.
doi:10.1124/mol.104.009175.

ABBREVIATIONS: EST, expressed sequence tag; PCR, polymerase chain reaction; DMSO, dimethyl sulfoxide; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; LPS, lipopolysaccharide; PPARα, peroxisome proliferator-activated receptor-α; Wy-14,643, pirinixic acid, 4-chloro-6-(2,3-xylidino)-2-pyrimidinyl)thioacetic acid; CAR, constitutively active receptor.

Moreover, experimental results can be influenced by the different protocols for hybridization and analysis used across laboratories (Waring et al., 2004). In theory, interlaboratory variation can be reduced through the application of careful calibration steps and standardization of protocols, but such an effort may be difficult to implement globally (Baker et al., 2004; Chu et al., 2004; Waring et al., 2004).

In an effort to address these issues, we have developed a transcriptional profiling resource for the toxicologist that we refer to as EDGE. This resource is a public database that allows two modes of researcher interaction. First, investigators can query a large and growing database of toxicant-induced transcriptional profiles that have been generated using a common microarray protocol and platform. To aid in this effort, the EDGE resource provides a series of informatics tools that allow data organization and hypothesis testing using a Web interface (http://edge.oncology.wisc.edu/edge.php). Second, the EDGE resource accepts RNA sample submissions from laboratories interested in developing novel toxicant-induced transcriptional profiles for inclusion in the open access database. Directed analysis such as this allows investigators to rapidly compare their chemical responses of interest with a large database of profiles that have been generated on a common microarray platform and consistent protocols.

Materials and Methods

Animals and Treatments. All animal experiments and treatments were reviewed and approved by the Research Animal Resource Center at the University of Wisconsin or by the appropriate oversight department of submitting institutions. Institutions, strains, sex, ages, treatment vehicle, circadian time, and route of administration are recorded for all experiments and can be found for each experiment at the EDGE Web site.

Expressed Sequence Tag Libraries. Poly-A mRNA was isolated from total RNA via the Oligotex kit (QIAGEN, Valencia, CA). Construction of expressed sequence tag (EST) libraries was typically performed by using the cDNA Timesaver kit (Amersham Biosciences Inc., Sunnyvale, CA). In brief, 5 μg of poly-A was used to generate first strand cDNA with an anchored poly-T oligonucleotide [dT(18)-N] that also contained a NotI-specific restriction site. First strand cDNA was then purified via a phenol/chloroform extraction. After second strand synthesis, cDNA was again purified using phenol/chloroform extraction. Ligation of adaptors containing an EcoRI site was then performed and the cDNA was digested with both NotI and EcoRI. The cDNA underwent size selection to give fragments with a minimum size of 400 base pairs. The resultant cDNA was repurified as described above and then directionally cloned into the NotI and EcoRI sites of pGEM11zf- (Promega, Madison, WI)

Full-Length Libraries. Five micrograms of total RNA was dephosphorylated for 1 h using calf intestinal phosphatase (Stratagene, La Jolla, CA). The reaction was then cleaned up by a phenol/chloroform extraction followed by ethanol precipitation. The mRNA cap was then removed from the RNA using tobacco acid pyrophosphatase (Epicenter Technologies, Madison, WI). After phenol/chloroform extraction and ethanol precipitation, a custom RNA oligonucleotide sequence was then ligated to the mRNA using RNA ligase. A reverse transcription reaction was performed using Powerscript reverse transcriptase (BD Biosciences, San Jose, CA). The cDNA was amplified using the polymerase chain reaction (PCR), and products were checked on an agarose gel for amplification success and quality. The PCR products were purified using a PCR purification kit (QIAGEN) and digested with AscI and PacI enzymes corresponding to sites incorporated during PCR. The PCI-neo vector (Promega) was digested with AscI and PacI restriction enzymes to create complementary ends for directional cloning. The library was then ligated into PCI-Neo, transformed into DH5a, and plated on agar plates with kanamycin selection.

cDNA Microarray Construction. For the liver microarray, clones were selected from a minimally redundant set of liver-derived EST and full-length cDNAs to create a minimally redundant set of clones. This was accomplished by screening them against the RefSeq database of known gene products (Wheeler et al., 2004). This set of clones was then fortified with cDNAs representing toxicologically relevant genes such as drug-metabolizing enzymes (e.g., cytochromes P450), inflammatory responsive genes (e.g., serum amyloid a1), and genes that we have previously identified as useful in classification of chemicals (Thomas et al., 2001). All clones were then sequence verified from both 3′ and 5′ directions. These cDNAs were then amplified through PCR and purified using the Millipore UltraScreen system (Millipore Corporation, Billerica, MA). The cDNAs were then dried and resuspended in a 50/50 H_2O/DMSO solution. Six replicates of each cDNA were printed on each array using a Microgrid II (Genomic Solutions, Ann Arbor, MA) at a pitch of 180 μm for a total of 27,648 features per chip. The slides were then baked at 80°C for 3 h and stored in a desiccator until their use.

cDNA Annotation. The EDGE microarray was annotated based on the sequence verifications of the PCR products. These sequence verifications were then compared against the RefSeq database, as described above. All clones that had a verifiable hit against a RefSeq record were then annotated with the name given by RefSeq. All transcripts in the text are called by the official RefSeq abbreviation. For clones that had high-quality sequence information but did not record a hit to RefSeq, sequence was compared against both the Unigene and GenBank databases. These clones were then hand-annotated based on the homology of these clones to the databases. Hand-annotated clones are identifiable by all capitalized letters being used.

RNA Isolation for Microarray Studies. For RNA samples generated in-house, and the majority of submitted samples, total RNA was prepared using the RNA Protect system (QIAGEN). Quality and quantity of RNAs were determined on an Agilent 2100 Bioanalyzer using the RNA Nano Labchip (Agilent, Palo Alto, CA). Typical RNA samples are at a concentration more than 1 μg/μl, and the 280/260 absorbance ratio is between 1.7 and 2.0.

cDNA Microarray Hybridization. When the amount of total RNA exceeded 20 μg, labeling and hybridizations were carried out according to the Genisphere cDNA50 kit (Genisphere, Hatfield, PA). The Genisphere cDNA350 kit was used when total RNA was less than 20 μg. In brief, RNA is reverse-transcribed into cDNA using specific primers that contain a sequence complementary to an oligonucleotide attached to a dendrimer containing Cy3 or Cy5 dye molecules. The labeled cDNA from both the treatment group and from a pool of its corresponding control are then mixed and hybridized to the liver microarray for 18 h. The arrays are then washed according to the manufacturers' specifications, and a second hybridization is performed using DNA complementary to the capture sequences that is covalently linked to a dendrimer containing 50 or 350 dye molecules (cDNA50 or cDNA350 kit, respectively). All experiments were performed using a "dye flop" experimental design where replicate hybridizations of the same sample and control are performed in each dye direction. That is, Cy3-labeled RNA from control sample is hybridized against Cy5-labeled RNA from the treated sample in dye direction 1, and Cy5-labeled control RNA is hybridized against Cy3-labeled treated RNA in dye direction 2.

cDNA Microarray Image Acquisition and Analysis. After hybridization, slides are scanned on a DNA Microarray Scanner, model G2565BA (Agilent, Palo Alto, CA). Images are processed, and the data are extracted using Agilent Feature Extractor software. Further data processing is then performed by custom Perl scripts and the data are then imported into MySQL databases. Normalization was performed by the LOWESS method in the Feature Extrac-

tor software. Data from the six cDNA replicate spots are then averaged after removing the maximum and minimum Cy5/Cy3 spot ratios to give a single value, and the values from the dye flop are then averaged.

Contributed RNA Samples. EDGE accepts samples from interested laboratories for inclusion in the data set. The above-mentioned protocols are provided as recommended procedures, but samples are not always subject to the same RNA preparation protocols. For inclusion in the EDGE database, all submitted samples are required to disclose the methods used to generate the sample [i.e., organism, number of biological samples represented in the RNA, animal strain, genetic variation (such as knockouts, transgenics, or induced mutations), animal age, sex, sample tissue, treatment, vehicle, dose, route of administration, RNA isolation method, time of dose, time of harvest, time elapsed, and proof of animal usage approval]. The quality and concentration of submitted RNA is analyzed as described above.

Data Access and Assurance. At the request of the submitter, data generated from EDGE can be embargoed from public access for a period of 4 months from the time of data generation. During this time, submitters have exclusive access to the data for comparison with all publicly accessible data. After this time, the data enters the public domain and is available to all users. Our rules of data access are superseded by rules imposed by scientific publications. In this regard, data that must be submitted to either GEO or ArrayExpress will be available on these databases as well as on EDGE. Deposition of EDGE data into GEO or ArrayExpress is simplified for the user, because all microarray experiments adhere to the minimum information about a microarray experiment guidelines (Brazma et al., 2001).

Results and Discussion

Strategy

Our objective was to develop a resource for the toxicology community that would allow comparison of transcriptional profiles that occur in response to chemical, physiological, environmental, and genetic stimuli. We proposed that for such a resource to be useful, it should meet the following criteria. First, it should serve to minimize interlaboratory differences, allow easy accessibility to data, and adhere to principles of information sharing. For this resource to serve a broad spectrum of interests, it should allow for comparisons between large numbers of chemically induced profiles in addition to comparison with profiles corresponding to specific physiological and pathological states.

Our solution is to perform the analysis of microarray experiments using a standardized toxicogenomics platform. To this end, we first developed a cDNA-based microarray that is enriched for responsive targets of hepatotoxicants in the mouse model. We then developed a pipeline for the rapid and economical generation of transcriptional profiles for a set of prototype chemicals and pathological states. Last, we implemented an on-line database replete with informatics that allows researchers to query and interpret large numbers of transcriptional profiles. We propose that the economy gained by this large-scale centralized effort will encourage the efficient use of cDNA microarray technology by the toxicology community and allow for the more efficient development of algorithms useful in chemical classification and hazard prediction.

Choice of cDNA Microarray Platform. In this first phase of development, we chose to construct a custom cDNA microarray platform that was enriched for transcripts expressed in toxicant exposed liver. Our decision not to use a commercial array solution was based on issues of cost and open data access. With respect to cost, we estimate that our array platform has a per experiment cost of approximately one fifth of that required by most commercial solutions. With respect to open access data and protocols, we intentionally steered away from any platform that relies upon proprietary technologies or that uses quality control information or bioinformatic approaches that are not equally available to all within the scientific community. This decision is related to our intent to keep all aspects of EDGE-derived data and analysis as transparent as possible to the scientific community.

Our decision to generate a novel cDNA clone set for the microarray platform was based on a number of factors. First, we found that many of the available curated cDNA sets were highly redundant and did not contain clones commonly expressed in toxicant-exposed target tissues, specifically liver. We predicted that custom generation and sequencing of cDNAs would provide an important resource that would complement currently available cDNA sets. Finally, the ability to confirm and readily distribute clones allows researchers access to these reagents once microarrays point them to a need for confirmation or functional studies.

To maximize efficiency of this platform for hepatotoxicity studies, cDNAs were initially derived from libraries of hepatic clones generated in-house. To increase the representation of toxicant-specific cDNAs, libraries were constructed from livers

TABLE 1

Distribution of liver cDNAs from control and toxicant-induced libraries

Libraries were constructed from mRNA isolated from livers under the specified conditions. Primers specific for the poly-adenylation site were used to anchor the libraries to the 3′ end of the transcript. Also listed is the percentage of gene discovery and the number and percentage of genes that were exclusive to that library.

Library	No. Sequenced	% Discovery	No. Unique	% Unique
Untreated libraries				
C57Bl/6J (6 weeks)	865	37.5	29	3.3
E13.5	1048	49.0	0	0.0
E17.5	2422	33.6	163	6.7
AHR$^{-/-}$	1033	54.3	123	11.9
Subtotal	5368	43.6	315	5.5
Chemically treated libraries				
TCDD	8772	33.5	1679	19.1
Control	1305	37.7	110	8.4
Phenobarbital	945	54.5	142	15.0
Cobalt	1677	44.7	317	18.9
Subtotal	12,699	42.6	2248	15.4
Total	18,067	22.8	2563	14.2

AHR, aryl hydrocarbon receptor.

that were treated with various chemicals or that came from different developmental or pathological states. By increasing the frequency of chemically induced cDNAs in our clone set, we predicted that the applicability of the arrays to a broad spectrum of chemical stimuli would be increased. The toxicants used to create unique cDNA libraries include the aryl hydrocarbon receptor agonist 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), cobalt chloride, and phenobarbital (Denison and Whitlock, 1995; Gu et al., 2000). Physiological states that were used to build liver libraries also included mouse livers from both embryonic days 13.5 and 17.5, as well as livers from $Ahr^{-/-}$ mice that exhibit altered hepatic blood flow (Lahvis et al., 2000).

To date, this effort has led to the cloning and characterization of 18,067 ESTs from eight libraries (Table 1). To analyze these ESTs, we first compared them with the RefSeq database to obtain identity. When RefSeq did not provide identity, we turned to UniGene and finally GenBank. These analyses reveal that the 18,067 ESTs fall into 4123 gene identities. Representative clones were selected and rearrayed. This set was supplemented with 485 cDNA clones historically linked to toxicant mechanism to form the base clone set used in microarray construction. Resequencing of this base clone set via both 5′ and 3′ ends resulted in approximately 3700 unambiguous identities representing 2364 different gene products. That is, our base clone set contains approximately 2400 cDNA clones representing products of unique loci. An additional 1300 cDNA clones provide redundant representation, and 900 clones have yet to be unambiguously confirmed (listed as "verification in progress"). Confirmation of these remaining 900 clones is ongoing, and the database will be updated in real time to reflect this new information as it is generated.

For all analyses, the presented data are derived from the 2364 products of unique genes. In cases where multiple clones for the same gene existed, an average was taken of all available data. The analyses presented below also exclude data from all clones whose identity remains unverified.

Microarray Data Analysis

Linearity and Quality. Given the limitations of our current microarray printing technology, we chose to develop an

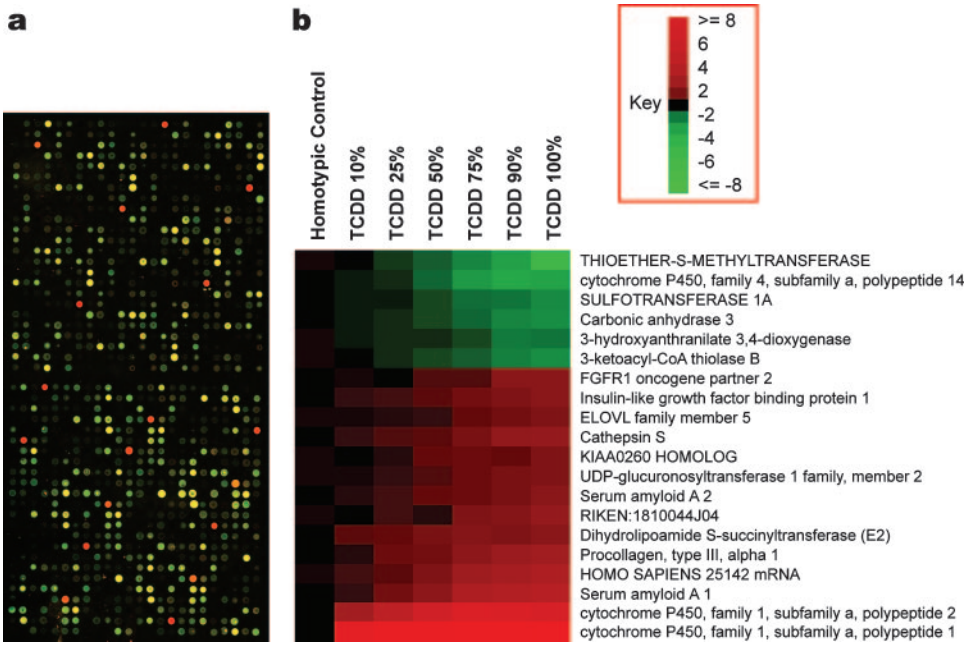

Fig. 1. Experiments showing qualitative and quantitative performance of the EDGE microarray. a, image of approximately 4% of a microarray hybridized with RNAs corresponding to Cy3-labeled vehicle-treated control (green) and Cy5 labeled TCDD sample (red). b, heat Map of 21 clones showing linear range of observed responses. RNA from the liver of a single animal treated with 64 μg of TCDD for 96 h was diluted with control RNA from eight pooled vehicle controls at the indicated dilutions. Heat map contains 21 clones that responded more than 3-fold in pure treated RNA. Linear range for cytochrome P450s 1a1 and 1a2 holds true but is not evident because of color saturation. Color key displayed corresponds to the same color scheme used in all figures.

TABLE 2

Analysis of variance on EDGE microarray experiments

Coefficients of variance were determined scrutinizing spot replicates, clone replicates, and biological replicates on the EDGE platform. Three biological replicates of C57Bl/6J mice dosed with lipopolysaccharide for 24 h were analyzed. Only genes whose fold change was greater than 2-fold in one animal and had multiple clones corresponding to that gene were analyzed. Coefficients of variance were then determined for the spot replicates, clone replicates, and biological replicates.

Gene Symbol	Refseq	No. of Clones	Average Fold Change	Coefficient of Variance		
				Spot	Clone	Biological
Saa2	NM_011314.1	6	13.06	6.7	19.5	10.7
Apcs	NM_011318.1	3	7.67	6.1	19.5	13.4
Orm1	NM_008768.1	4	7.05	16.1	10.7	12.4
Hpxn	NM_017371.1	4	4.16	7.8	11.0	10.8
Fgl1	NM_145594.1	5	3.62	28.5	27.0	21.8
Saa4	NM_011316.2	4	2.89	12.4	9.8	21.7
Cp	NM_007752.1	3	2.17	7.2	5.0	9.0
1100001G20Rik	NM_183249.1	3	1.98	9.3	10.1	10.2
Trf	NM_133977.1	5	1.82	2.3	9.1	7.9
S100a9	NM_009114.1	3	1.66	9.0	19.4	16.0
Cyp1a2	NM_009993.2	7	−2.02	6.7	13.5	9.9
Cyp2c50	NM_134144.1	3	−2.17	2.6	17.2	6.7
Cyp2c37	NM_010001.1	4	−2.31	4.0	15.8	9.2

array with 27,648 features. This density allowed for six replicate spots of each cDNA. Typical hybridization experiments displayed uniform spot morphology, allowing easy spot by spot delineation and fluorescence quantitation (Fig. 1a). In an effort to understand the linear range afforded by the EDGE microarray, we performed an experiment where control RNA was mixed with increasing proportions of RNA from TCDD-exposed livers (Fig. 1b). For the 21 transcripts that exhibited the greatest magnitude of up or down-regulation, we observed a linear relationship between the estimates of transcript quantity and proportion of induced sample. These results suggest the array data are linear over several orders of magnitude. For the most up-regulated gene, the linear relationship holds over a range of 44-fold.

Variability in microarray experiments can be of either biological or technical nature. To address the relative contributions of technical and biological variance to observed results, we performed a careful analysis of variance in response to a single dose of lipopolysaccharide (LPS) across three individual biological replicates. Targets were filtered for those genes that responded greater than 2-fold response, up or down (Table 2). Coefficients of variance for both replicate spots for a single clone and across redundant clones representing the same target gene ranged between 5 and 30% (Table 2).

Prototype Chemicals

To develop a foundation for future classification studies, we generated a preliminary cluster analysis of transcriptional profiles for a number of prototype chemicals and phys-

TABLE 3

Treatments used in generation gene expression profiles

Chemicals and physiological conditions that were used in generating the first phase of experiments for the EDGE database. Experiments listed are performed on C57Bl/6 male mice. The exception is the male liver RNA compared with female RNA. More detailed information regarding experimental conditions is available on the EDGE Web site at edge.oncology.wisc.edu

Chemical	Route	Dose	Time	Mechanism
Acetominaphen	p.o.	400 mg/kg	1 day	Electrophile
TCDD	p.o.	10 μg/kg	2 days	AHR agonist
3-Methylcholanthrene	i.p.	100 mg/kg	2 days	AHR agonist
β-Napthoflavone	p.o.	100 mg/kg	2 days	AHR agonist
Aroclor 1260	i.p.	200 mg/kg	2 days	AHR agonist
Phenobarbital	p.o.	100 mg/kg/day	2 days	CAR agonist
TCPOBOP	i.p.	3 mg/kg/day	2 days	CAR agonist
Circadian rhythm	None	None	0, 4, 8, 12, 16, 24 h	Circadian
Ketoconazole	i.p.	250 mg/kg	1, 4 days	P450 inducer
Cobalt	p.o.	60 mg/kg	2 days	Heme disrupter
Phenylhydrazine	p.o.	100 mg/kg	2 days	Heme disrupter
DEHP	Chow	2%	2 weeks	PPAR agonist
Ciprofibrate	Chow	0.1%	2 weeks	PPAR agonist
Wy-14,643	Chow	0.1%	2 weeks	PPAR agonist
Corn oil	p.o.	1 μl/g	2 days	Vehicle
Corn oil	i.p.	1 μl/g	2 days	Vehicle
DMSO	i.p.	1 μl/g	2 days	Vehicle
Saline	p.o.	1 μl/g	2 days	Vehicle
Lipopolysaccharide	i.p.	1 mg/kg	6, 12, 24, 48 h	Inflammatory agent
Interleukin-1β	i.p.	25 μg/kg	6 h	Inflammatory response
Interleukin-6	i.p.	25 μg/kg	6 h	Inflammatory response
Tumor necrosis factor-α	i.p.	5 g/kg	6, 12 h	Inflammatory response
Fasting	None	None	1, 2 days	Metabolism
Sex	None	None	None	Sex

AHR, aryl hydrocarbon receptor; DEHP, diethyl-hexyl-phthalate; TCPOBOP, 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene.

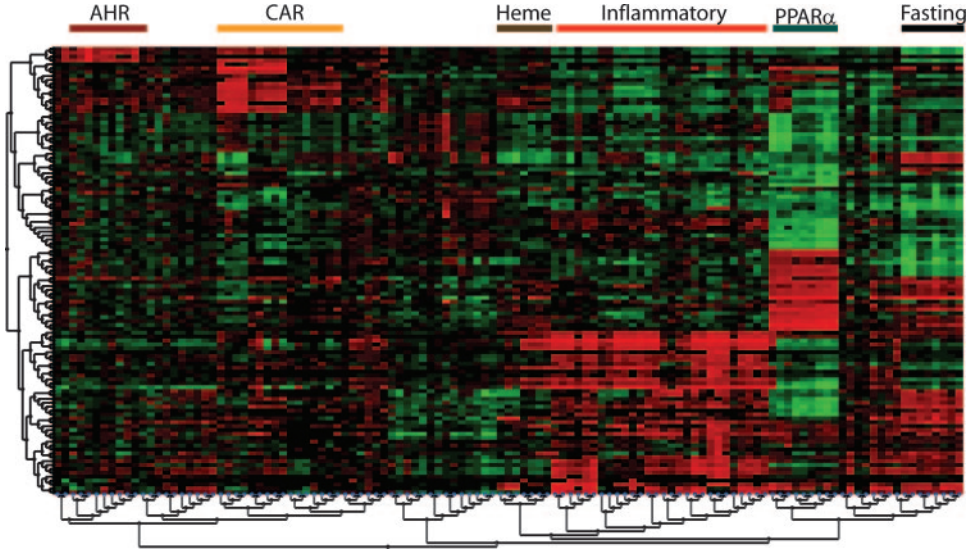

Fig. 2. Hierarchical cluster of all 117 treatments currently available to the public on EDGE[2]. The 115 targets respond with at least a 5-fold change in expression to at least one of the treatments. Clustering of these responses exhibits large batteries of genes that are highly responsive to such insults as inflammatory agents, PPARα agonists, agonists of the aryl hydrocarbon receptor and constitutively active receptor, heme function disruptors, and fasting. Clustering of the treatments also shows the utility of this approach in grouping treatments by mechanism of action, because inflammatory agents group closely together.

iological states (Table 3). In many cases, multiple doses and time points were selected to provide examples of how these variables influence qualitative and quantitative aspects of the response. Given that biological replicates were often performed, these data sets also provide an example of the biological variance observed in global gene expression studies. Included in the chemical response set were inflammatory cytokines, aryl hydrocarbon receptor agonists, peroxisome proliferators, and activators of the constitutively activated receptor. These initial treatments are meant to serve as a scaffold to which future treatments can be compared. So far, there are 117 treatments publicly available in the EDGE database and more than 400 additional treatments will become public within 4 months of this writing. All graphical results from these following sections can be generated using the informatics available at http://edge.oncology.wisc.edu/.

Inflammatory Agents. The response of liver to inflammatory agents or cytokines has been well studied (Monshouwer and Hoebe, 2003). Given that many hepatotoxic agents can lead to liver inflammation, we placed a priority on the characterization of the battery of transcripts that respond to LPS, tumor necrosis factor-α, interleukin-1β, and interleukin-6. The largest magnitude of response came in a gene battery of transcripts that correspond to what are commonly referred to as "acute phase" proteins (Ramadori and Armbrust, 2001; Streetz et al., 2001). In the EDGE array, this battery of cytokine-responsive genes includes the known acute phase genes: *Saa1–4*, *Orm1*, *Hp*, *Hpxn*, *Fn1*, and *Serpina 1–6*. The transcriptional changes in response to these inflammatory treatments provide a clear transcriptional signature of liver inflammation (Fig. 2).

Aryl Hydrocarbon Receptor Agonists. Given their central role in molecular toxicology, we have also provided transcriptional profiles for a number of aryl hydrocarbon receptor agonists. These chemicals serve as prototypes for a wide range of environmental pollutants that include polychlorinated biphenyls and dioxins. In this regard, we used various doses of 3-methylcholanthrene and β-naphthoflavone. Consistent with what is reported in the literature, we observed marked induction of transcripts for known xenobiotic-metabolizing enzymes such as Cyp1a1, Cyp1a2, Gstm1, and Ugt1a (Whitlock et al., 1997). In addition, we have identified several

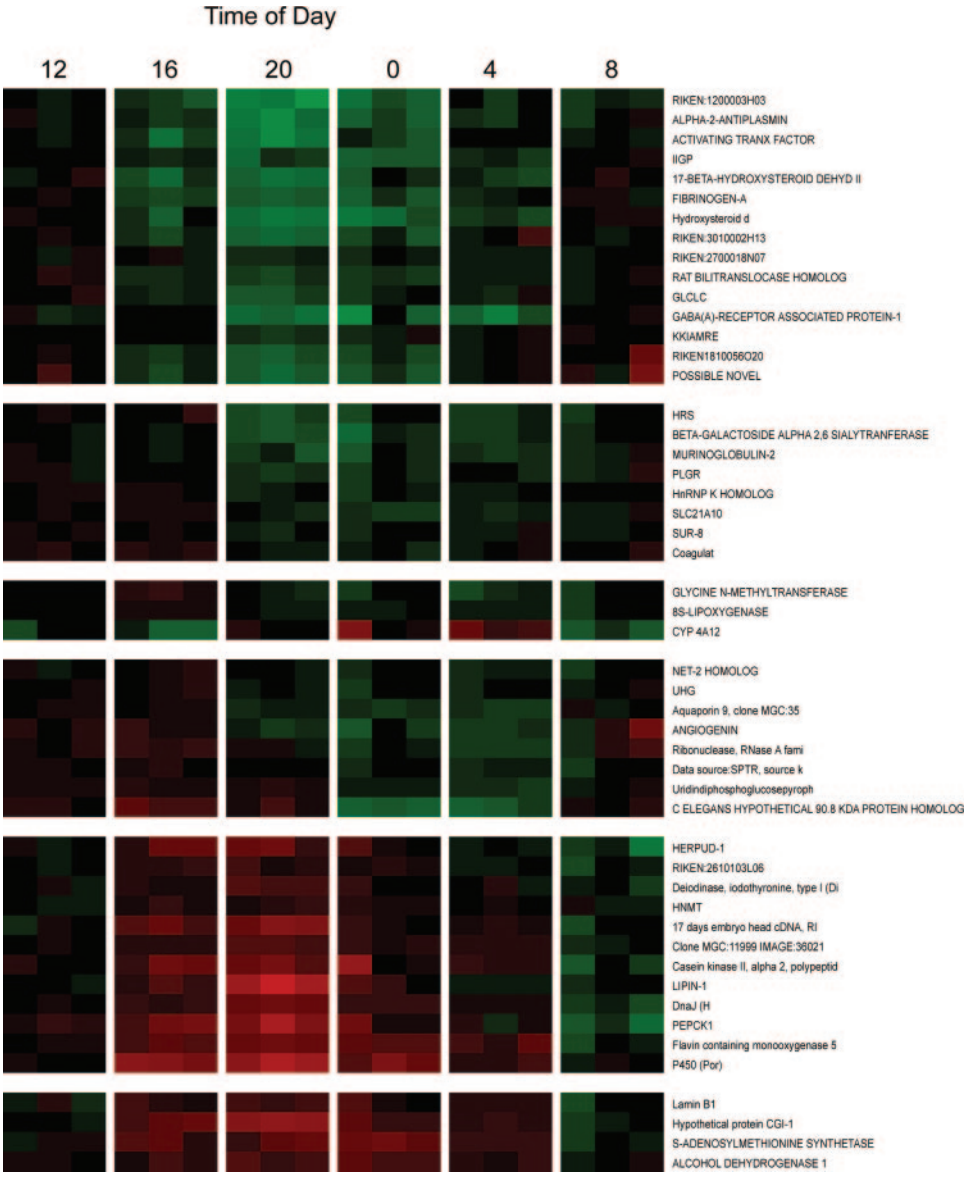

Fig. 3. K-means clustering of hepatic circadian expression of 50 targets whose profiles cycled over a 24-h period with a significance of $p < 0.01$. Significance was determined using the COSOPT program as having a periodicity between 23 and 25 h. All samples were compared with expression in hepatic tissue at noon (time of day, 12).

other aryl hydrocarbon receptor responsive transcripts such as HectD2 and TATIP-30kDa. The transcriptional changes in response to these environmental contaminants also provide a clear transcriptional signature (Fig. 2).

Peroxisome Proliferators. Agonists of the peroxisome proliferator-activated receptor-α (PPARα) are common therapeutics, environmental pollutants, and hepatotoxic agents (Yu et al., 2003). To characterize the hepatic response to this class of chemicals, we profiled the transcriptional changes of mouse liver to diethyl-hexyl-phthalate, ciprofibrate, and Wy-16,463. Data from the EDGE resource demonstrated the induction of the classic PPARα-responsive genes Cyp4a10, Cyp4a14, Acox1, and Ehhadh (Qi et al., 1999). A number of transcripts not previously known to be responsive to peroxisome proliferators were also identified in these array studies. Included in this responsive set are the products of CD36 and Xcr1 loci. These profiles also provide one of the best examples of the breadth and magnitude of hepatic transcriptional changes that can occur after chemical insult (Fig. 2).

CAR Activators. Activators of CAR are well characterized inducers of a broad spectrum of cytochromes P450 and as etiologic agents in hepatomegaly (Masahiko and Honkakoski, 2000). In agreement with this known activity, transcriptional data generated from the powerful CAR stimulators 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene, aroclor-1260, and phenobarbital led to marked up-regulation of Cyp2b10, 2c29, 2c37, and 3a11 (Honkakoski et al., 1996). Also coordinately up-regulated with these genes were other phase II metabolizing enzymes such as Gstm1 and Gsta2. Among genes that were markedly down-regulated were Cyp4a10, Hepcidin, and Bhmt. Like the inflammatory response these responses exhibit an exclusive signature that readily identifies stimulators of the CAR pathway (Fig. 2).

Treatment Vehicles and Routes of Administration. Toxicology experiments may be significantly influenced by choice of vehicle and route of administration. In an effort to define the influence of these variables, we profiled the transcriptional response of corn oil, DMSO, and saline solutions by varying routes of administration. We found that corn oil or DMSO via intraperitoneal injection caused an acute phase response consistent with treatment of inflammatory agents such as LPS (interspersed in inflammatory clade; Fig. 2). To a lesser degree, we found that oral gavage of vehicles had subtle but reproducible effects on gene expression. In this regard, we found that oral administration of saline or corn oil repressed expression of the Saa1–2 by approximately 3-fold compared with untreated animals (data not shown). These experiments illustrate the necessity of appropriate vehicle controls in toxicogenomic studies.

Fig. 4. Time-course data on expression changes of 41 genes stimulated by LPS treatment (1 mg/kg) by at least a 4-fold change over 2 days is displayed by a hierarchical cluster. The cluster shows the transient nature of hepatic cytokine response and the importance of the temporal nature of responses in gene profiling. Most genes have returned to basal levels by 2 days.

Prototype Physiological States

To begin to correlate gene expression patterns with physiological and pathology, we generated profiles for several

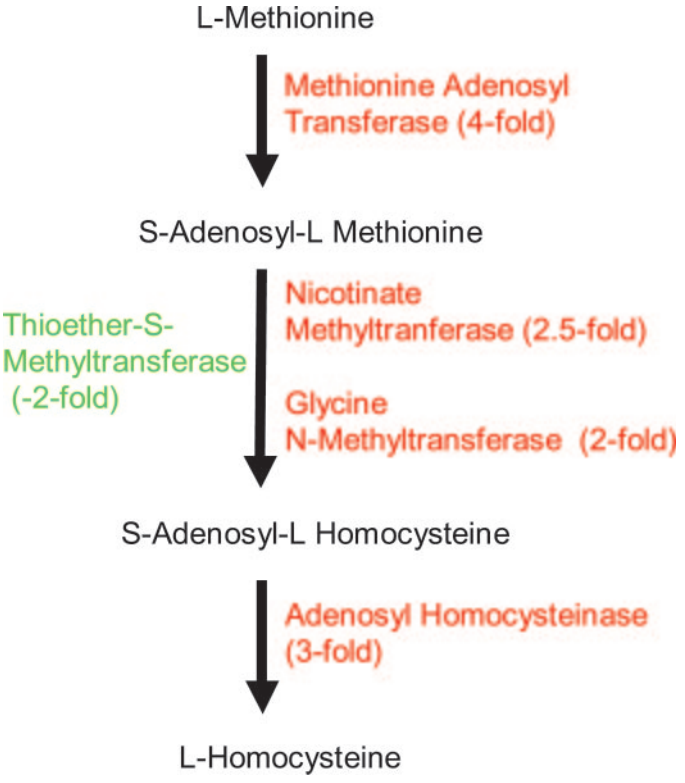

Fig. 5. Activation of metabolic pathway from L-methionine to L-homocysteine activated by fasting is displayed as an enzymatic pathway. Genes shown in red (MHT, GNMT, HNMT, and AHCY) are induced by fasting at both 24 and 48 h. TEMT, shown in green, also plays a role in this pathway, and is repressed during fasting. Fold change of expression in parentheses.

prototype states. These conditions included circadian rhythm, decreased food intake, and sex.

Circadian Rhythm. One potential variable in toxicology experiments is the choice of time of day for administration of the treatment or for measurement of the response. Given that gene expression can cycle over a 24-h period, it seemed important to describe those batteries of genes that correspond to particular circadian times (Panda et al., 2002; Oishi et al., 2003). Using the EDGE platform, we observed many genes important to xenobiotic metabolism and toxicant action that cycle with a circadian period (Fig. 3). For example Cyp2a5 and Cyp4a12 both cycle over a 24-h period with peak and nadir expression differing by approximately 7-fold (Fig. 3). Likewise, genes such as Pepck and G6pc are major components of glucose mobilization and storage pathways and are also regulated in a circadian manner (Fig. 3). Consistent with what is known about energy metabolism, mice do not require much glucose mobilization during the night while they are ingesting new food sources (Rutter et al., 2002). Also during the night, genes associated with insulin resistance and fat storage, such as *Igfals* and *Fabp5* are induced. These data suggest common transcriptional control for a large percentage of hepatic transcripts.

LPS Time Course. Gaining insight into the temporal aspects of the transcriptional response to chemical exposure is important for the classification of unknown toxicants. Therefore, we provided an example of the influence of time after exposure on subsequent expression profiles. In this regard, the transcriptional responses to LPS are evident 6 h after exposure and quickly attenuate by 24 h (Fig. 4). These data illustrate the importance of gathering multiple time points as a method to capture the most diagnostic transcriptional responses. In the case of compounds with unknown inflammatory properties, a time point of 2 days might miss expression patterns associated with cytokine signaling, causing misclassification of that compound.

Fasting. Given that exposure to many toxicants can induce a cachetic response, it seemed important that the EDGE database provide a preliminary characterization of those batteries of genes that respond to decreased food intake. The availability of such information can have two uses. First, observation of such batteries in toxicology studies can warn the investigator that decreased food intake may be an issue. Second, the knowledge that certain transcriptional sets are related to decreased food intake may allow investigators to quickly identify these targets, allowing focus on more primary genomic endpoints.

Learning about Mechanism

Classification. One of the great promises of toxicogenomics is the prediction of toxicity based on gene expression. Several articles have demonstrated that the classification of chemicals based on the expression patterns they induce has considerable potential (Thomas et al., 2001; Hamadeh et al., 2002a,b; Kramer et al., 2004). Methods used for microarray classification studies have varied considerably from both the standpoint of user supervision and the statistical concepts used. Several of these methods are available as part of the analysis suite on the EDGE resource, including unsupervised methods such as hierarchical and K-means clustering.

Of the 2364 genes represented on the current liver array, 115 (~5%) responded with at least a 5-fold change to at least one of the 117 treatments/conditions in prototype set (Fig. 2). Applying a hierarchical clustering algorithm to these data provide support for the idea that microarrays can effectively group chemical treatments based solely on their influence on global transcriptional profiles. For example, a simple nearest neighbor analysis of the cluster analysis reveals that the known peroxisome proliferators diethyl-hexyl-phthalate, ciprofibrate, and Wy-14,643 fall within the same peroxisome proliferator clade. Likewise, of the 21 inflammatory treatments, 19 fall within the same inflammatory clade (Fig. 2).

Pathway Mapping. Integration of knowledge about gene expression with information related to pathway function has the potential to strengthen insights regarding chemical signaling and mechanism of action. To provide an example of this type of analysis, we first identified a set of transcripts dysregulated by fasting. This list was then compared with the Gene Ontology database to identify similarities in function. It is interesting that we found genes important in the biosynthesis of sulfur-containing amino acids to be induced. For example, gene products such as Bhmt, Gnmt, Mat1a, and Ahcy were significantly up-regulated. Mapping of these genes to a metabolic pathway provides a graphical representation of how these genes are important in the metabolism of methionine to homocysteine (Fig. 5). It is noteworthy that other isoschizomeric proteins in the same pathway, such as Temt, are down-regulated in response to fasting. These results show a coordinated shift in the processing of amino acids in fasted animals.

Conclusions

The EDGE database was developed to enhance sharing of toxicogenomic information. In addition to generating novel transcriptional profiles, the EDGE resource allows investigators to perform in silico experiments where the results of the large-scale effort can be interrogated to ask specific questions about toxicant challenge or physiological condition. Because of this resource, investigators often do not have to establish microarray technology and expertise within their own laboratory.

The EDGE resource currently houses liver transcriptional profiles from 117 different treatments, doses, and time points. Once embargoes have expired, the size of this data set will increase to more than 500 profiles in the next 4 months. In addition, parallel platforms containing at least 1000 unique genes each have been developed for mouse skin, lung, kidney, palate, and tendon. Similar platforms for mouse thymus, heart and ureter, are being developed and should be available within a year.

References

Amin RP, Hamadeh HK, Bushel PR, Bennett L, Afshari CA, and Paules RS (2002) Genomic interrogation of mechanism(s) underlying cellular responses to toxicants. *Toxicology* **181–182:**555–563.

Baker VA, Harries HM, Waring JF, Duggan CM, Ni HA, Jolly RA, Yoon LW, De Souza AT, Schmid JE, Brown RH, et al. (2004) Clofibrate-induced gene expression changes in rat liver: a cross-laboratory analysis using membrane cDNA arrays. *Environ Health Perspect* **112:**428–438.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, et al. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* **29:**365–371.

Chu TM, Deng S, Wolfinger R, Paules RS, and Hamadeh HK (2004) Cross-site comparison of gene expression data reveals high similarity. *Environ Health Perspect* **112:**449–455.

Denison MS and Whitlock JP Jr (1995) Xenobiotic-inducible transcription of cytochrome P450 genes. *J Biol Chem* **270:**18175–18178.

MOL PHARM
MOLECULAR PHARMACOLOGY
aspet

Gu Y-Z, Hogenesch J, and Bradfield C (2000) The PAS superfamily: sensors of environmental and developmental signals. *Annu Rev Pharmacol Toxicol* **40:**519–561.
Hamadeh HK, Bushel PR, Jayadev S, DiSorbo O, Bennett L, Li L, Tennant R, Stoll R, Barrett JC, Paules RS, et al. (2002a) Prediction of compound signature using high density gene expression profiling. *Toxicol Sci* **67:**232–240.
Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, Bennett L, Tennant R, Stoll R, Barrett JC, et al. (2002b) Gene expression analysis reveals chemical-specific profiles. *Toxicol Sci* **67:**219–231.
Honkakoski P, Moore R, Gynther J, and Negishi M (1996) Characterization of phenobarbital-inducible mouse Cyp2b10 gene transcription in primary hepatocytes. *J Biol Chem* **271:**9746–9753.
Kramer JA, Curtiss SW, Kolaja KL, Alden CL, Blomme EA, Curtiss WC, Davila JC, Jackson CJ, and Bunch RT (2004) Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. *Chem Res Toxicol* **17:**463–470.
Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, and Kohane IS (2002) Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* **18:**405–412.
Lahvis G, Lindell S, Thomas R, McCuskey R, Murphy C, Glover E, Bentz M, Southard J, and Bradfield C (2000) Portosystemic shunts and persistent fetal vascular structures in Ah-receptor deficient mice. *Proc Natl Acad Sci USA* **97:**10442–10447.
Masahiko N and Honkakoski P (2000) Induction of drug metabolism by nuclear receptor CAR: molecular mechanisms and implications for drug research. *Eur J Pharm Sci* **11:**259–264.
Monshouwer M and Hoebe KH (2003) Hepatic (dys-)function during inflammation. *Toxicol In Vitro* **17:**681–686.
Oishi K, Miyazaki K, Kadota K, Kikuno R, Nagase T, Atsumi G, Ohkura N, Azama T, Mesaki M, Yukimasa S, et al. (2003) Genome-wide expression analysis of mouse liver reveals CLOCK-regulated circadian output genes. *J Biol Chem* **278:**41519–41527.
Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, Schultz PG, Kay SA, Takahashi JS, and Hogenesch JB (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109:**307–320.
Qi C, Zhu Y, Pan J, Usuda N, Maeda N, Yeldandi AV, Rao MS, Hashimoto T, and Reddy JK (1999) Absence of spontaneous peroxisome proliferation in enoyl-CoA Hydratase/L-3-hydroxyacyl-CoA dehydrogenase-deficient mouse liver. Further support for the role of fatty acyl CoA oxidase in PPARalpha ligand metabolism. *J Biol Chem* **274:**15775–15780.
Ramadori G and Armbrust T (2001) Cytokines in the liver. *Eur J Gastroenterol Hepatol* **13:**777–784.
Rutter J, Reick M, and McKnight SL (2002) Metabolism and the control of circadian rhythms. *Annu Rev Biochem* **71:**307–331.
Streetz KL, Wustefeld T, Klein C, Manns MP, and Trautwein C (2001) Mediators of inflammation and acute phase response in the liver. *Cell Mol Biol (Noisy-le-grand)* **47:**661–673.
Thomas RS, Hayes KR, Zastrow GM, Tran K, Penn SG, Rank D, and Bradfield CA (2002) Application of DNA microarrays for predicting toxicity and evaluating cross-species extrapolation, in *Toxicogenomics* (Inoue T and Pennie W eds) pp 31–38, Springer, Tokyo, Japan.
Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Pande K, Glover E, Silander T, Craven MW, Reddy JK, et al. (2001) Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol Pharmacol* **60:**1189–1194.
Waring JF, Ulrich RG, Flint N, Morfitt D, Kalkuhl A, Staedtler F, Lawton M, Beekman JM, and Suter L (2004) Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyrilene. *Environ Health Perspect* **112:**439–448.
Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, et al. (2004) Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* **32:**D35–D40.
Whitlock JP Jr, Chichester CH, Bedgood RM, Okino ST, Ko HP, Ma Q, Dong L, Li H, and Clarke-Katzenberg R (1997) Induction of drug-metabolizing enzymes by dioxin. *Drug Metab Rev* **29:**1107–1127.
Yu S, Rao S, and Reddy JK (2003) Peroxisome proliferator-activated receptors, fatty acid oxidation, steatohepatitis and hepatocarcinogenesis. *Curr Mol Med* **3:**561–572.

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