Chemical Research in Toxicology

MARCH 2005
VOLUME 18, NUMBER 3

Review

Advances in Toxicogenomics

Kevin R. Hayes and Christopher A. Bradfield*

The McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, Madison, Wisconsin, 53706

Received November 30, 2004

Contents

1. Introduction

Toxicogenomics is an emerging scientific discipline that combines the tools of traditional toxicology with those of genomics, bioinformatics, and high throughput experimentation (*1*). The underlying premise of toxicogenomics is that a global assessment of the biology of chemical exposure will lead to a more thorough understanding of how toxicants act. It is the promise of toxicogenomics that such information will lead to better judgments regarding human health risks and improved pipelines for the next generation of therapeutics.

Toxicogenomics had its origins in the technologies related to transcriptional profiling. To date, dozens of manuscripts have appeared in the peer-reviewed literature that employ global transcriptional profiling methods to link toxic sequelae with changes in specific batteries of mRNAs. These data sets have the potential to provide insights into toxic mechanism and to identify sensitive biomarkers of exposure. These early experiments also

* To whom correspondence should be addressed. Tel: 608-262-2024. Fax: 608-262-2824. E-mail: Bradfield@oncology.wisc.edu.

10.1021/tx0496690 CCC: $30.25
Published on Web 03/21/2005

define the most prominent obstacle to toxicogenomics. That is, how do we gain insight from descriptive data sets that are of a volume and a type that are unprecedented in modern toxicology?

The methods of toxicogenomics are varied. In the most common approach, the chemical response is described by profiling the corresponding changes in cellular mRNAs, proteins, or metabolites (*1*−*3*). In the complementary approach, the toxic pathway is interrogated by employing high throughput functional screens to assess the influence of sets of transcripts, proteins, or small molecules on a given biological process (*4*). Taken together, these avenues allow the toxicologist to investigate the relationship of thousands of gene products or small molecules to toxic mechanism.

In this review, we will summarize the current state of the art in toxicogenomics, with particular emphasis placed on the application and interpretation of transcriptional profiling. The focus of this review was the result of the rapid acceptance of this tool in modern toxicology. The fact that the volume of such data is increasing in the scientific literature has required the toxicology community to understand and interpret such data. Another reason to review transcriptional profiling is related to the importance of these prototype data sets as model systems that will be used to move the science of toxicology forward. In this regard, although transcriptional profiling data can be viewed as an important example of the power of toxicogenomics, the tools being developed to interpret these data are setting the stage for how data from parallel approaches will be analyzed in the future.

2. Microarray Studies

Central to the field of toxicogenomics has been the use of DNA microarray technology. Microarray technology can be defined as the use of high-density DNA probe sets to simultaneously assay the transcriptional states of hundreds to thousands of genes (*5*). These transcriptional states are usually recorded as a relative value where the experiment measures the differential response of each gene between control and experimental conditions. Differentially expressed genes are those that are found to be different, based upon some statistical or often informal assessment.

2.1. Basics of Microarray Analysis

The term "microarray" encompasses a wide variety of strategies and technologies. In general, the term refers to an array of hybridization targets that can be simultaneously assayed (*5*). With respect to transcriptional profiling, these targets are made from DNA. The two main approaches used for transcriptional profiling are based on either single or double "fluor" protocols. In the single fluor protocol, the control and experimental samples are hybridized against separate microarrays. The data are then related, with expression ratios calculated from the values of the two microarrays. The alternative is the use of two fluors with different excitation and emission spectra, where both the control and the experimental samples are hybridized against the same microarray (*5*). The expression ratio is then calculated from the data for the two samples at the same location. In experiments using the double fluor protocols, it is common that hybridizations are performed in both color directions. For example, the fluor Cy3 is used to label the control and fluor Cy5 is used to label the experimental sample in the "forward" direction. In the "reverse" direction, Cy5 is used to label the control and Cy3 is used to label the experimental sample (*6*−*8*).

Data collected from a microarray experiment are commonly presented through the generation of a "heat map". In such a display, the experiment is presented as a two-dimensional grid with the experimental condition as one axis and the transcript identity as the other axis. Each coordinate on the grid is then color shaded according to the change in gene expression of that particular treatment. In the most common convention, transcripts that are up-regulated are colored red and transcripts that are down-regulated are colored in green (Figure 1a).

Global expression profiles are often organized using an unsupervised approach known as "clustering". For the purpose of this review, the complete set of transcriptional changes occurring in response to an experimental condition is referred to as a "profile". In a method known as hierarchical clustering, treatments are "clustered" based on the similarity of changes in target expression. Alternatively, targets may be clustered based on their expression changes across treatments. A statistical metric, usually a correlation coefficient or Euclidian distance, is then calculated to rank the similarity between each profile. A hierarchical tree is then created based upon these metrics until a single dendogram is formed (Figure 1a) (*9*). Distances in branches of the resulting tree correlate with similarity, with short branches signifying close relationships. The underlying data are then plotted as a heat map next to the dendogram to produce a full visualization of the data.

One related form of clustering that has been used extensively in toxicogenomics is *K* means. In *K* means analysis, targets are again grouped based on similarity of expression, except that the number of clusters is defined a priori (*10*). That is, transcripts that are to be clustered are forced into a preset number of cluster groups. Targets are then reassigned to different cluster groups until the distance metric is minimized within a given cluster group and maximized between the cluster groups (Figure 1b).

A third form of clustering that is commonly used is principle component analysis (PCA) (*11*). PCA collapses multiple values into "eigenvectors" generated from matrices of gene expression data that retain much of the data variability. These eigenvectors can be comprised of any number of individual target data points. The eigenvectors can then be plotted to give relative locations of profiles. Again, the distance between profiles in this space directly correlates with similarity. In this manner, a gene expression experiment that may contain hundreds of vectors can be reduced to three dimensions (Figure 1c).

2.2. Design of Microarray Experiments

Experimental design affects the information that can be gained from a microarray. Design issues affect all microarray experiments but are especially critical for two color hybridizations. This section deals exclusively with those issues.

The most common designs used in transcriptional profiling experiments are referred to as "direct", "reference", and "loop" (*6*−*8*). In the direct design, the control sample is hybridized directly against the experimental

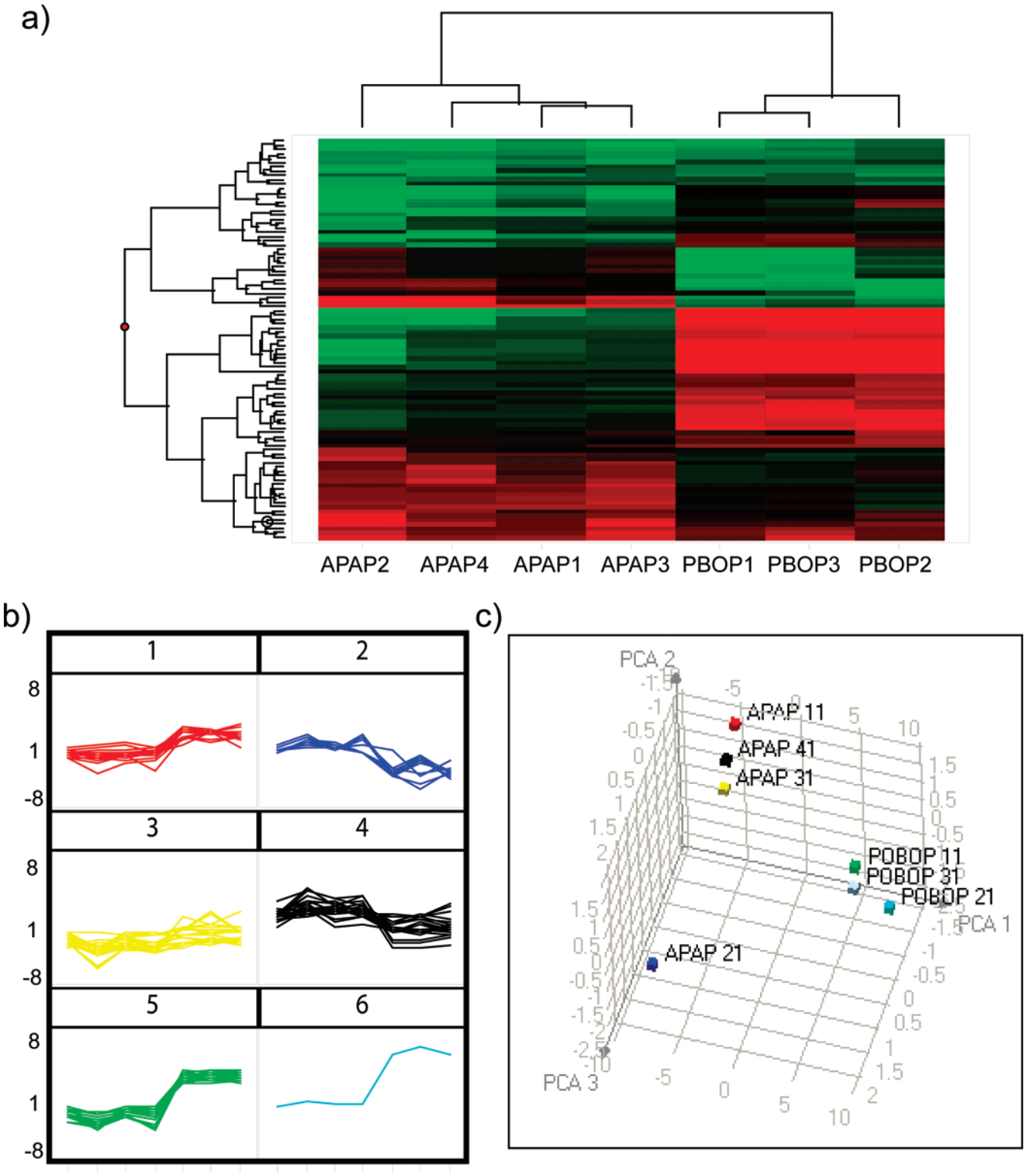

Figure 1. Depiction of various unsupervised clustering methods of four acetaminophen and three TCPOBOP dosed liver samples using 92 targets that have a minimum 2-fold change in at least two experiments. (A) Hierarchical cluster of the 92 targets. The colorimetric display clearly shows the differential expression of the targets and the directions of the regulation. Also note that both targets and treatments have been shown, displaying the relationships between both targets and treatments. (B) *K* means cluster of the 92 targets after defining six clusters. (C) PCA of the same seven treatments using the 92 targets. Experiments have been plotted into 3D space using the three eigenvectors from PCA. The plot shows the great differences (distances) between the two treatments and that one acetaminophen treatment (APAP2) differs from the other three.

sample with different color fluors for each sample (Figure 2a,b). The reference design employs the hybridization of both the control and the experimental groups against a common reference RNA sample (Figure 2c). In the loop design, biological replicates from multiple groups are sequentially hybridized against one another until all of the samples have been connected (Figure 2d). For all methods, biological variance may be estimated by replication of several biologically independent samples (Figure 2b,c).

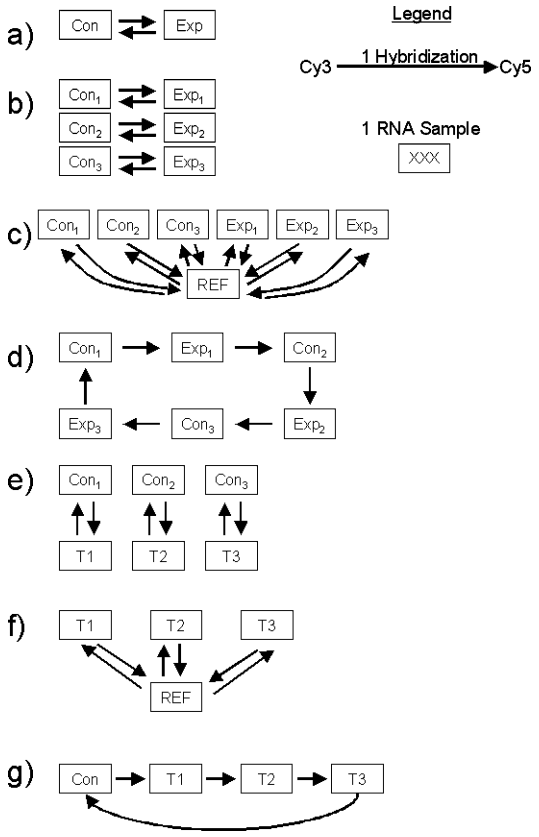

Figure 2. Various experimental designs commonly used in microarray experiments. In each panel, a box depicts a RNA sample, with each arrow corresponding to a microarray hybridization. The RNA sample at the head of each arrow is labeled with Cy3, and the RNA sample at the tail of each arrow is labeled with Cy5. (A) A direct hybridization using "dye swap". (B) A direct dye swap design with biological replication. (C) A reference hybridization design wherein both control and experimental samples are hybridized against a common reference RNA. This reference RNA need not have relevance to the experiment. (D) A loop hybridization design where each sample is connected to successive samples, until all samples are related, closing the loop. (E) A time series design where each time point is hybridized to a control sample, increasing the precision at determining which targets are regulated differently from the control state. (F) A time series design where a reference sample is hybridized against all time points, increasing the relative accuracy between time points. (G) A time series loop design that increases the precision in determining how targets change over time.

The question being asked influences the choice of direct, reference, or loop design (*6*−*8*). In a time course experiment, the identification of differentially expressed targets at a particular time is often the objective. When the experiment attempts to identify a list of targets changing at a particular time, a direct strategy that compares each time point to a time-matched control provides the most sensitivity (Figure 2e). When the experiment is an attempt to understand how targets change over time, then a loop strategy comparing each time point against the preceding and subsequent points is a more appropriate choice (Figure 2g). This design provides less sensitivity with respect to changes but provides better information on the influence of time on a particular target. When both the identity of targets and the temporal nature of their changes are important, then the reference design might be of greatest value (Figure 2f). This design provides the magnitude of the change and the temporal relationship but at a cost of statistical power for both.

3. Transcriptional Profiling of the Chemical Response

One area of great promise for the field of toxicogenomics is its use in providing more efficient identification of hazardous chemicals. If the concept of classification by transcriptional profiling can be developed, the profiles elicited by an unknown compound can be compared to those induced by known toxicants. When similarities to a known class of toxicants are found, this information may be used to prioritize those chemicals that require further toxicological study. Similarly, it may quickly identify those compounds that can be expected to pose a particular human or environmental risk by comparison with known contaminants.

The prospect of transcriptional profiling is based on the assumption that chemicals with similar toxicology may also invoke similar patterns of transcriptional response. Early studies using prototype chemicals support this idea (*12*−*15*). The use of microarray-based profiling in response to well-characterized compounds has confirmed known transcriptional targets, provided preliminary evidence of distinctive signatures of these compounds, and added to our knowledge of novel chemically induced targets (*16*).

3.1. Mechanism of Chemical Action

There are several mechanisms by which chemicals invoke similar transcriptional profiles in vivo. The first is that toxic chemicals often induce cognate adaptive metabolic pathways. This adaptation to xenobiotic exposure is exemplified by the rapid metabolic response to polycyclic aromatic hydrocarbons and phenobarbital (*17*, *18*). The second general mechanism underlying similarity in transcriptional response is that gene expression often changes as the consequence of pathological events (*19*). For example, inflammation can influence transcriptional profiling either by direct up-regulation of cytokine responsive genes or through granulocyte recruitment (*20*). A third example of how chemicals might influence the transcriptional profile of a tissue is from the appearance of differentiated cell types. An example of this phenomenon is the rapid differentiation of epithelial cell types in response to pollutants such as 2,3,7,8-*p*-tetrachlorodibenzodioxin (*21*).

3.2. Chemical Classification

The idea that chemicals can be classified based upon their influence on global transcription has considerable support from the related field of molecular oncology. In this arena, early studies have demonstrated successes from using transcriptional profiles to differentiate tumor types. Examples include the use of microarrays to discriminate between soft tissue tumors, which can be difficult to histologically diagnose, the discrimination between primary tumors and metastases, and the correlation between target expression and efficacy of a large battery of chemotherapeutic agents (*22*−*25*). Further studies have advanced this field by correlating gene expression patterns with clinical outcomes (*26*). These results suggest that pathological states can be categorized based on transcriptional profiling in addition to classical histological methods.

A number of recent publications support the potential of transcriptional profiling in toxicology. In one method,

Table 1. Listing of Web Accessible Genomic Databases with Appropriate URLs

resource	URL	type
Tox-MIAME Express	http://www.ebi.ac.uk/tox-miamexpress/	protocol
CEBS	http://cebs.niehs.nih.gov/microarray/index.jsp	database
ArrayExpress	http://www.ebi.ac.uk/arrayexpress/	database
GEO	http://cebs.niehs.nih.gov/microarray/index.jsp	database
SYMATLAS	http://symatlas.gnf.org/SymAtlas/	database
NCT	http://www.niehs.nih.gov/nct/home.htm	consortium
TRC	http://www.niehs.nih.gov/dert/trc/home.htm	consortium
EDGE	http://edge.oncology.wisc.edu/	database
PharmGKB	http://www.pharmgkb.org/index.jsp	database
dbZach	http://dbzach.fst.msu.edu/	database
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/	database
CTD	http://ctd.mdibl.org/	database

chemicals are classified using an approach based on Bayesian probability (*27*). In an early example, expression profiles are generated from 24 different chemicals and then assigned to five toxicological classes based upon known mechanisms of action. Each DNA target is then tested for its potential to classify the treatments correctly, and the best overall classifier is selected. Additional targets are then added to the model in a sequential manner. Cross-validation was accomplished via a leave-one-out-across method for each addition. A "diagnostic set" was then selected from the minimum number of targets that produced the maximum classification accuracy.

Classification of chemicals has also been accomplished by linear discriminant analysis and genetic algorithm/*K*-nearest neighbors (GA/KNN) (*28*). In this example, a training set of transcriptional profiles was generated from four chemicals, three representing one class of toxicant and one representing another. The GA/KNN method was then used to select informative genes that were able to discriminate the classes. Twenty-two samples of chemicals were then profiled and tested for membership in these two classes.

A related study examined the correlation between global gene expression changes and carcinogenic potential, providing an example of how classification can be accomplished based upon pathological knowledge (*29*). In this study, changes in gene expression for nine chemicals at multiple doses were correlated with their known carcinogenic potency. Several targets were highly correlated with carcinogenic potential in a five day assay. These targets include cytochrome P450 oxidoreductase and "transforming growth factor-β stimulated clone 22". The ability to test compounds for carcinogenic potential in a short-term assay could have a significant impact on the development and regulatory decisions for pharmaceutics and industrial chemicals.

While these studies have shown great promise, many problems still remain. With over 70 000 chemicals in commerce today, a robust method that can incorporate and accurately predict toxicity must first be validated on much larger sets. In this regard, it remains to be seen if any of these algorithms are able to accurately predict toxicity as the number of classes and chemicals within each class increase. Future algorithms may take into account extra knowledge about the function and relation of targets in classifier sets to help refine accuracy as sets grow.

4. Toxicogenomic Databases

The potential for microarray data to provide insights into chemical classification and toxic mechanism is highly dependent upon methods for the sharing of transcriptional profiles across laboratories. Although not specific to toxicology, general repositories of transcriptional profiles are now emerging on-line (*30*, *31*). The impact of such databases on toxicology will be related to how efficiently laboratories can access and interpret this growing volume of data, how these different data sets can be related to one another, and how well these data can be linked to additional toxicological information.

The ability to understand microarray results is dependent on the technical information that is supplied with the data. A clear report of the methods used to analyze the data is crucial in the interpretation of related conclusions. To this end, the Microarray Gene Expression Data Society proposed the "Minimum Information About a Microarray Experiment" (MIAME) guidelines for reporting and publication of microarray experiments (*32*). These guidelines suggest the reporting of all raw microarray data and mathematical transformations. Numerous journals now require adherence to these guidelines for publications involving microarray experiments (*33*, *34*).

Comparison of gene expression data across laboratories has been problematic (*35*). Standardization of protocols and use of common platforms are one way to reduce interlaboratory variation (*36*). Unfortunately, such standardization is not easily achieved due to the large number of platforms and protocols available to the scientific community (*35*, *37*−*40*). Another factor that inhibits comparisons across laboratories is related to the different informatic strategies that have been used to annotate targets (*41*−*43*). The differing nomenclature that can result from these strategies can confound comparisons. This problem may be reduced, as genomes are more carefully sequenced and curated. Taken in sum, these findings highlight the difficulties in comparing any single profiling experiment with the existing data in the toxicogenomics literature.

5. Basic Resources for Transcriptional Profiling

The development of open access genomic databases is beginning to provide the toxicologist with the resources to better interpret profiling experiments. General resources with relevance to toxicology are described below.

5.1. Gene Expression Omnibus (GEO)

The GEO at the National Center for Biotechnology Information has become the largest microarray repository in the world (Table 1) (*31*). The GEO serves as a central warehouse for over 18 000 microarray experiments from

numerous branches of biology. Adherence to MIAME guidelines allows the user to interpret the quality and analysis of the data. Although only a small fraction of the database, data for many toxicology experiments are available. In addition to its repository, the GEO is adding visualization tools, such as hierarchical clustering for data within a series. These visualization tools allow user interaction but do not currently allow analyses of data subsets or data outside of the series. Although this resource does not address cross-laboratory variation, it serves as a powerful tool for data mining and hypothesis generation.

5.2. ArrayExpress

ArrayExpress is a large public data repository based at the European Bioinformatics Institute (EBI) (*30*). Containing data on more than 5000 microarray hybridizations, ArrayExpress serves as the second largest microarray data repository in the world. The EBI has worked with the MGED society to create an additional set of controlled vocabularies to describe toxicogenomic experiments, MIAME/Tox (Table 1). It has also developed a database to house these experiments, Tox-MIAMExpress (Table 1). Tox-MIAMExpress works by incorporating descriptive data about toxicology experiments (i.e., dose, route of administration, treatment time, etc.) using the controlled vocabulary into the existing ArrayExpress database schema. This new subsystem of the ArrayExpress represents one of the first attempts to create a public repository for toxicogenomic research.

5.3. Symatlas

Symatlas is a specialized database that is maintained by the Genomics Institute of the Novartis Research Foundation (*44*). This database is a catalog of gene expression in varying tissues and cell culture models (Table 1). Comprised of microarray data for 79 human tissues, 61 mouse tissues, and 43 rat tissues, the expression of approximately 22 000 targets can be interrogated. This resource allows researchers to identify tissues or cell culture models where their gene of interest is expressed. It also allows for screening of potential target tissues based on expression patterns of receptors and effectors. As with many of the databases, Symatlas devotes a good amount of resource toward proper annotation and function. To this end, the targets are continually reanalyzed and updated, linking each to a wide array of outside literature and structural and functional databases.

6. Toxicogenomic Resources

6.1. National Center for Toxicogenomics (NCT)

The mission of the NCT is to coordinate a nationwide research effort to develop the use of genomic tools in toxicology (*45*). Recently, the NCT has developed a multicenter effort known as the Toxicogenomic Research Consortium (TRC) (Table 1). Although the efforts of the NCT and TRC are only now emerging, one result has been the development of a database providing access to recent publications and selected raw data from microarray experiments. It is proposed that the efforts from the NCT and the TRC will be housed in the Chemical Effects in Biological Systems (CEBS) knowledge base (Table 1). Although at the time of this review the CEBS knowledge base was still in development, this resource will eventually relate toxicogenomic data with gene function, polymorphism data, and proteomic information (*46*).

6.2. Environment, Drugs, and Gene Expression (EDGE)

The EDGE database is a publicly accessible microarray resource that is devoted to toxicogenomic research (*47*). The EDGE resource circumvents many of the issues associated with data comparison described above. This is accomplished by the centralization of the microarray experiment performance and the use of a standardized microarray platform. The EDGE represents a database as well as a resource as it accepts samples from outside researchers for analysis. This allows researchers with little microarray experience the ability to compare their data with a large volume of toxicogenomic data under similar conditions. The EDGE also offers a number of powerful informatic tools that allow users to direct their analyses. These tools include hierarchical and *K* means clustering, BLAST searching, rank analysis, and classification tools. These analyses are not prerun but directed by the user so that novel comparisons between any data in the database may be performed (*47*). Although the EDGE resource is currently focused on toxicogenomics of mouse liver, the resource is expanding to other organ systems and has been shown to be useful in profiling samples from rat models.

6.3. Pharmacogenomics Knowledge Base (PharmGKB)

The PharmGKB provides a catalog of the interaction between genes and drugs (Table 1) (*48*). For each known gene−drug interaction, PharmGKB provides a link to the corresponding literature, as well as known human polymorphism for that gene. It also relates the known genetic variation with primary data on metabolism, efficacy, and clinical outcome. In summary, this database provides a powerful tool to help researches understand how genetic constituency might influence the toxic response to xenobiotics.

6.4. Emerging Toxicogenomic Databases

While the databases discussed above are currently accessible, numerous others are under development and will be available soon. Several of these will have direct application to toxicogenomics. In this regard, the dbZach database will house microarray data and analysis tools with particular emphasis on endocrine disruption and testicular toxicity (Table 1). Similarly, the Comparative Toxicology Database (CTD) will provide associations between toxic agents and biological systems (Table 1). Ultimately, the linking of all of these databases should provide the toxicologist with rich online resources to strengthen hypotheses, direct experiments, and understand chemical actions.

7. Comparative and Computational Genomics

In addition to gene profiling, a number of recent advances are providing researchers with new tools that can help provide insight into chemical−biological interac-

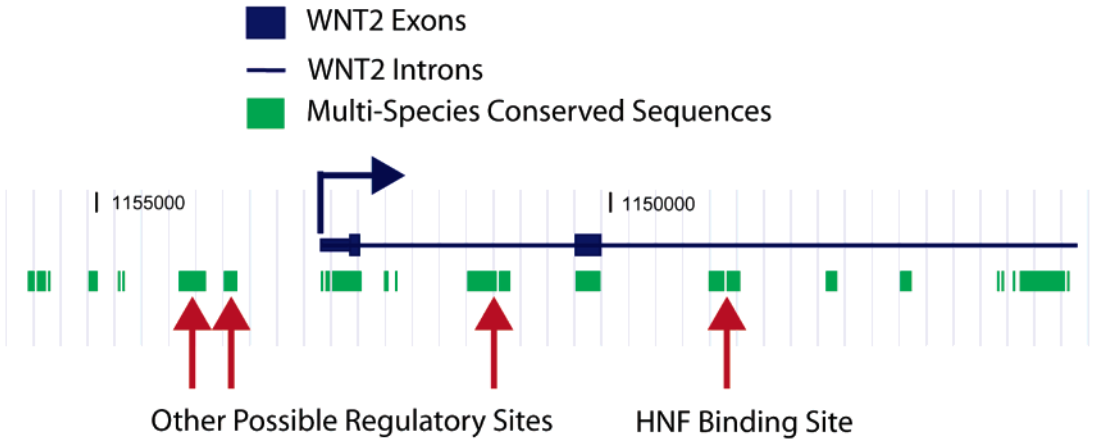

Figure 3. Genomic map of exons one and two of the WNT2 gene with multispecies conserved sequences. This figure illustrates six large regions surrounding the first two exons and the transcriptional start site that show high correlation between distant species. Noted is a site enriched for the hepatic nuclear factor (HNF) binding site, which has long been proposed to regulate WNT2.

tions. Probably most important in this regard was the sequencing of multiple vertebrate genomes (*49*−*53*). A parallel result has been the identification of genes and their regulatory regions by both experimental and computational methods (*49*, *50*). These advances are only beginning to be employed by toxicologists to answer questions regarding toxicant mechanism of action.

7.1. Comparative Genomics

The sequencing of multiple vertebrate genomes has allowed detailed comparison of the sequences that encode genes and their regulatory elements across species. Multispecies comparisons have indicated that about 70% of conserved sequences lie outside of exonic regions (*54*, *55*). Conserved regions within and flanking genes are largely thought to represent conserved regulatory domains that control expression of nearby loci. These comparisons have made it possible to build statistical models for putative transcription factor response elements for any loci with conservation across species (*56*, *57*) (Figure 3). It is proposed that genes that have positionally conserved elements are highly likely to be functional elements (*56*, *57*). Nonconserved elements may also provide insight, as they may explain differential responses across species. Such knowledge could have a great impact on toxicology research and regulatory decisions based on animal models.

7.2. Comparative Genomic Hybridization (CGH)

Another genomic advance has been the development of CGH arrays. These arrays were first used in profiling chromosomal translocations, insertions and deletions in tumors (*58*). More recently, researchers have coupled Chromatin Immunoprecipitation (ChIP) with genomic arrays, or "ChIP chips". The ChIP chips provide an empirical method for discovering promoter occupancy by transcription factors given an experimental state. This approach can complement expression profiling as it may allow a link between promoter occupancy and chemically induced transcriptional response (*59*).

7.3. Microarrays and the Genome

Given that the majority of genes have been physically mapped in human and mouse, it is becoming possible to link gene expression data with chromosomal location. Linking data in this manner allows researchers to look for chromosomal "hot spots" that may be under control of a common element. It also has the potential to gauge the availability of genomic DNA to chemically activated transcription factors due to chromatin structure. A promising direction of this research that is only beginning to be realized is the linking of gene expression data with quantitative trait loci data (*60*). This marriage of classic genetics with functional genomics will allow for much quicker identification of important genes, such as those associated with the innate immune response or toxicant susceptibility.

8. Relating Profiling Data to Signal Transduction Pathways

Microarrays have considerable power to reveal global transcriptional changes in response to a chemical insult. This approach has the potential to identify sets of coordinately regulated genes. One approach to making these connections is to first assume that all genes are regulated independently and then look for deviations from this prediction. Genes whose profiles show correlation can then be examined for coordinated regulation and associated biology.

8.1. Pathway Mapping

The process of overlaying microarray data onto biological networks is commonly known as "pathway mapping". One advance in the field of pathway mapping is in the development of gene annotations using a controlled vocabulary. The Gene Ontology Consortium (GO) has provided a framework that has supplied functional annotations to over 14 000 mouse genes, 10 000 rat genes, and 9500 human genes (*61*). These annotations are divided into three subsections: biological process (e.g., toxin metabolism), molecular function (e.g., insulin receptor activity), and cellular component (e.g., Golgi vesicle). By coupling such databases to results from microarray experiments, one can query groups of targets to search for "overrepresented" or "underrepresented" GO terms. In this way, insight into any related functionality of a coregulated cluster can be gained in an unsupervised manner.

Investigators are also taking a direct approach and overlaying microarray results onto information about

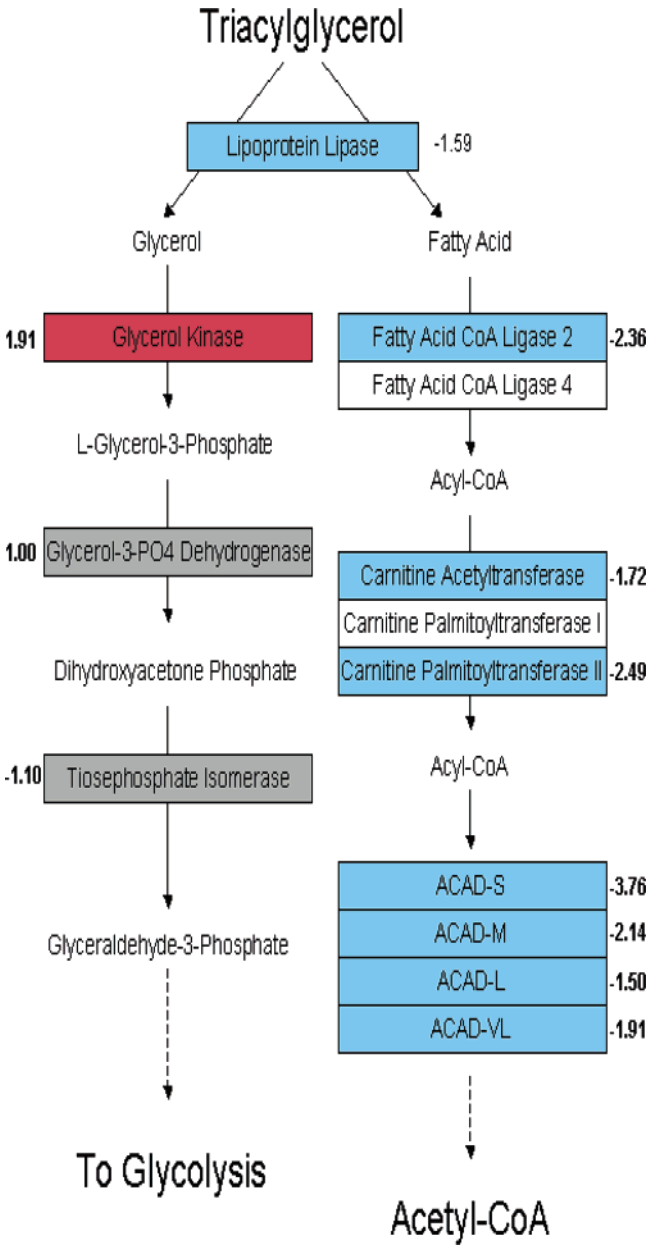

Figure 4. Fatty acid degradation pathway overlaid with microarray results. GenMAPP (*62*) was used to create a map overlaid with results from a mouse model of cardiomyopathy (Conklin, unpublished). The figure illustrates a shift from Acetyl-CoA production toward glycolysis by up-regulation of a rate limiting enzyme (glycerol kinase, shown in red) and the coordinated down-regulation of an entire pathway (shown in blue).

signal transduction or metabolic pathways. Pathways ranging from metabolic routes to signal transduction cascades are annotated into gene networks. These networks can then be overlaid with data from a microarray experiment (*62*, *63*) (Figure 4). Using such an approach, chemically induced shifts in metabolic pathways can be defined. These analyses are highly supervised and are limited by both the scope of the annotations and the number of pathways that have been mapped in the database.

Prior knowledge on the function of genes can also be used to strengthen analyses. By looking at sets of genes known to interact biologically, it has been found that sets of genes that share a pathway or function are changed coordinately and detected with increased statistical sensitivity (*64*). Alternatively, when single target DNAs in the study are examined individually, statistical sensitivity is reduced. Applying this approach to toxicogenomics has the potential to identify pathways that may be minimally affected by a toxicant, but that contributes significantly to pathology.

8.2. Protein Interaction Networks (PINs)

PINs provide a powerful approach to map toxicant-affected pathways. Construction of these networks involves building a database from known gene−gene interactions and using experimental results to identify subnetworks of participant genes. The basis of these networks can be built from any number of interaction databases: protein−protein, transcriptional activation/repression, protein modification/degradation, etc. (*65*). After importation of an experimental data set, a map can be drawn containing clusters of interacting genes. The complexity of these maps is dependent on the number of interactions in the database as well as the size of the experimental sample. Given their smaller genomes, initial databases are biased toward signaling in simpler organisms. Thus, the majority of initial studies have focused on simpler organisms, such as *Saccharomyces cerevisiae* (*4*). Computational methods for searching the known literature and building databases for higher organisms are being developed that allow for analysis of more complicated systems (*66*−*69*).

9. Toxicogenomics in Drug Development

Advances in toxicogenomics will have applications in drug design and development. The use of microarray technology holds promise to speed drug discovery pipelines, increase therapeutic index, and target therapies to an individual's response (*70*, *71*).

This technology is now being used to identify patients in need of treatments and which treatments to prescribe. Recent studies on breast cancer suggest that in many cases gene profiling data can more accurately predict patient outcome than traditional methods (*26*, *72*−*74*). Clinical trials are now underway using data from microarrays to match patients with treatment regimens (*74*). The U.S. Food and Drug Administration is also interested in how this technology is to be incorporated into the clinic (*75*). New rules and guidelines are now being worked out so that a wide range of microarray diagnostic tools may be employed.

10. Environmental Regulation

The impact of toxicogenomics on environmental regulatory decision has yet to be determined. While one use of toxicogenomics is to predict the toxicity of environmental and industrial compounds, the details of how this information will be used are as yet unclear. Toxicogenomics data have a great potential to verify toxicities of related compounds that are regulated under one rubric. An example of this application lies in the toxic equivalency factors (TEFs) for dioxins, dibenzofurans, and polychlorinated biphenyls (*76*−*78*). The TEFs for these compounds are constructed from both in vivo and in vitro studies and can be created from a number of end points. The power of toxicogenomics to profile these compounds and compare their relative effects across the genome may aid the choice of biomarkers used in hazard prediction.

11. Conclusions

Toxicogenomics is a field that is emerging from idea to practice. Problems associated with platform and data compatibility, completeness of information, assimilation into usable databases, and statistical power still require considerable attention. Once addressed, toxicogenomics can emerge as a powerful tool for hazard prediction, mechanistic understanding, and drug development. The integration of transcriptional profiling, gene function annotation, statistical methodologies, and informatics will have great utility in the field of toxicogenomics. Powerful tools to accomplish this integration are currently being constructed and tested. These tools, and others in development, will shed light into the mechanisms of the vast numbers of chemicals in commerce today whose toxic properties are poorly understood.

Acknowledgment. We thank Andrew Smith, Tim Gant, Jacqueline Walisser, and Aaron Vollrath for their comments and review of this manuscript. This paper was supported by grants from the National Institutes of Health (R01-ES012752, T32-CA009135, and P30-CA014520).

References

(1) 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.

(2) Griffin, J. L., and Bollard, M. E. (2004) Metabonomics: Its potential as a tool in toxicology for safety assessment and data integration. *Curr. Drug Metab. 5*, 389−398.

(3) Pognan, F. (2004) Genomics, proteomics and metabonomics in toxicology: Hopefully not "fashionomics". *Pharmacogenomics 5*, 879−893.

(4) Yao, G., Craven, M., Drinkwater, N., and Bradfield, C. A. (2004) Interaction networks in yeast define and enumerate the signaling steps of the vertebrate aryl hydrocarbon receptor. *PLoS Biol. 2*, E65.

(5) 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.

(6) Yang, Y. H., and Speed, T. (2002) Design issues for cDNA microarray experiments. *Nat. Rev. Genet. 3*, 579−588.

(7) Kerr, M. K. (2003) Design considerations for efficient and effective microarray studies. *Biometrics 59*, 822−828.

(8) Churchill, G. A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nat. Genet. 32*, 490−495.

(9) 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.

(10) Sherlock, G. (2000) Analysis of large-scale gene expression data. *Curr. Opin. Immunol. 12*, 201−205.

(11) Bleharski, J. R., Li, H., Meinken, C., Graeber, T. G., Ochoa, M. T., Yamamura, M., Burdick, A., Sarno, E. N., Wagner, M., Rollinghoff, M., Rea, T. H., Colonna, M., Stenger, S., Bloom, B. R., Eisenberg, D., and Modlin, R. L. (2003) Use of genetic profiling in leprosy to discriminate clinical forms of the disease. *Science 301*, 1527−1530.

(12) Ueda, A., Hamadeh, H. K., Webb, H. K., Yamamoto, Y., Sueyoshi, T., Afshari, C. A., Lehmann, J. M., and Negishi, M. (2002) Diverse roles of the nuclear orphan receptor CAR in regulating hepatic genes in response to phenobarbital. *Mol. Pharmacol. 61*, 1−6.

(13) Frueh, F. W., Hayashibara, K. C., Brown, P. O., and Whitlock, J. P., Jr. (2001) Use of cDNA microarrays to analyze dioxin-induced changes in human liver gene expression. *Toxicol. Lett. 122*, 189−203.

(14) Wei, Y. D., Tepperman, K., Huang, M. Y., Sartor, M. A., and Puga, A. (2004) Chromium inhibits transcription from polycyclic aromatic hydrocarbon-inducible promoters by blocking the release of histone deacetylase and preventing the binding of p300 to chromatin. *J. Biol. Chem. 279*, 4110−4119.

(15) Kramer, J. A., LeDeaux, J., Butteiger, D., Young, T., Crankshaw, C., Harlow, H., Kier, L., and Bhat, B. G. (2003) Transcription profiling in rat liver in response to dietary docosahexaenoic acid implicates stearoyl-coenzyme a desaturase as a nutritional target for lipid lowering. *J. Nutr. 133*, 57−66.

(16) Cherkaoui-Malki, M., Meyer, K., Cao, W. Q., Latruffe, N., Yeldandi, A. V., Rao, M. S., Bradfield, C. A., and Reddy, J. K. (2001) Identification of novel peroxisome proliferator-activated receptor alpha (PPARalpha) target genes in mouse liver using cDNA microarray analysis. *Gene Expression 9*, 291−304.

(17) Conney, A. H. (1967) Pharmacological implications of microsomal enzyme induction. *Pharmacol. Rev. 19*, 317−366.

(18) Remmer, H., and Merker, H. J. (1963) Drug-induced changes in the liver endoplasmic reticulum: Association with drug-metabolizing enzymes. *Science 142*, 1657−1658.

(19) Fielden, M. R., and Zacharewski, T. R. (2001) Challenges and limitations of gene expression profiling in mechanistic and predictive toxicology. *Toxicol. Sci. 60*, 6−10.

(20) Hoffmann, K. F., McCarty, T. C., Segal, D. H., Chiaramonte, M., Hesse, M., Davis, E. M., Cheever, A. W., Meltzer, P. S., Morse, H. C., 3rd, and Wynn, T. A. (2001) Disease fingerprinting with cDNA microarrays reveals distinct gene expression profiles in lethal type 1 and type 2 cytokine-mediated inflammatory reactions. *FASEB J. 15*, 2545−2547.

(21) Loertscher, J. A., Sattler, C. A., and Allen-Hoffmann, B. L. (2001) 2,3,7,8-Tetrachlorodibenzo-p-dioxin alters the differentiation pattern of human keratinocytes in organotypic culture. *Toxicol. Appl. Pharmacol. 175*, 121−129.

(22) Nielsen, T. O., West, R. B., Linn, S. C., Alter, O., Knowling, M. A., O'Connell, J. X., Zhu, S., Fero, M., Sherlock, G., Pollack, J. R., Brown, P. O., Botstein, D., and van de Rijn, M. (2002) Molecular characterisation of soft tissue tumours: A gene expression study. *Lancet 359*, 1301−1307.

(23) Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. (2003) A molecular signature of metastasis in primary solid tumors. *Nat. Genet. 33*, 49−54.

(24) 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.

(25) 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.

(26) van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature 415*, 530−536.

(27) Thomas, R. S., Rank, D. R., Penn, S. G., Zastrow, G. M., Hayes, K. R., Pande, K., Glover, E., Silander, T., Craven, M. W., Reddy, J. K., Jovanovich, S. B., and Bradfield, C. A. (2001) Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol. Pharmacol. 60*, 1189−1194.

(28) Hamadeh, H. K., Bushel, P. R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J. C., Paules, R. S., Blanchard, K., and Afshari, C. A. (2002) Prediction of compound signature using high-density gene expression profiling. *Toxicol. Sci. 67*, 232−240.

(29) Kramer, J. A., Curtiss, S. W., Kolaja, K. L., Alden, C. L., Blomme, E. A., Curtiss, W. C., Davila, J. C., Jackson, C. J., and Bunch, R. T. (2004) Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. *Chem. Res. Toxicol. 17*, 463−470.

(30) Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G. G., Oezcimen, A., Rocca-Serra, P., and Sansone, S. A. (2003) ArrayExpress—A public repository for microarray gene expression data at the EBI. *Nucleic Acids Res. 31*, 68−71.

(31) Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res. 30*, 207−210.

(32) Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet. 29*, 365−371.

(33) Bradfield, C. (2004) Genomics and proteomics. *Chem. Res. Toxicol. 17*, 2.
(34) (2002) Microarray standards at last. *Nature 419*, 323.
(35) Kuo, W. P., Jenssen, T. K., Butte, A. J., Ohno-Machado, L., and Kohane, I. S. (2002) Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics 18*, 405−412.
(36) Thompson, K. L., Afshari, C. A., Amin, R. P., Bertram, T. A., Car, B., Cunningham, M., Kind, C., Kramer, J. A., Lawton, M., Mirsky, M., Naciff, J. M., Oreffo, V., Pine, P. S., and Sistare, F. D. (2004) Identification of platform-independent gene expression markers of cisplatin nephrotoxicity. *Environ. Health Perspect. 112*, 488−494.
(37) Mah, N., Thelin, A., Lu, T., Nikolaus, S., Kuhbacher, T., Gurbuz, Y., Eickhoff, H., Kloppel, G., Lehrach, H., Mellgard, B., Costello, C. M., and Schreiber, S. (2004) A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol. Genomics 16*, 361−370.
(38) Ghosh, D., Barette, T. R., Rhodes, D., and Chinnaiyan, A. M. (2003) Statistical issues and methods for meta-analysis of microarray data: A case study in prostate cancer. *Funct. Integr. Genomics 3*, 180−188.
(39) Li, J., Pankratz, M., and Johnson, J. A. (2002) Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. *Toxicol. Sci. 69*, 383−390.
(40) Baker, V. A., Harries, H. M., Waring, J. F., Duggan, C. M., Ni, H. A., Jolly, R. A., Yoon, L. W., De Souza, A. T., Schmid, J. E., Brown, R. H., Ulrich, R. G., and Rockett, J. C. (2004) Clofibrate-induced gene expression changes in rat liver: A cross-laboratory analysis using membrane cDNA arrays. *Environ. Health Perspect. 112*, 428−438.
(41) Irizarry, K., Kustanovich, V., Li, C., Brown, N., Nelson, S., Wong, W., and Lee, C. J. (2000) Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences. *Nat. Genet. 26*, 233−236.
(42) Mattes, W. B. (2004) Annotation and cross-indexing of array elements on multiple platforms. *Environ. Health Perspect. 112*, 506−510.
(43) Marshall, E. (2001) DNA arrays. Affymetrix settles suit, fixes mouse chips. *Science 291*, 2535.
(44) Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R., and Hogenesch, J. B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U.S.A. 101*, 6062−6067.
(45) Tennant, R. W. (2002) The National Center for Toxicogenomics: Using new technologies to inform mechanistic toxicology. *Environ. Health Perspect. 110*, A8−A10.
(46) Waters, M., Boorman, G., Bushel, P., Cunningham, M., Irwin, R., Merrick, A., Olden, K., Paules, R., Selkirk, J., Stasiewicz, S., Weis, B., Van Houten, B., Walker, N., and Tennant, R. (2003) Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *EHP Toxicogenomics 111*, 15−28.
(47) Hayes, K. R., Vollrath, A. L., Zastrow, G. M., B. J., M., Craven, M., Jovanovich, S., Walisser, J. A., Rank, D., Penn, S., Reddy, J. K., Thomas, R., and Bradfield, C. A. (2005) EDGE: A centralized resource for the comparison, analysis and distribution of toxicogenomic information. *Mol. Pharmacol.* Submitted for publication.
(48) Chang, J. T., and Altman, R. B. (2004) Extracting and characterizing gene-drug relationships from the literature. *Pharmacogenetics 14*, 577−586.
(49) Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., Szustakowki, J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., and International Human Genome Sequencing, C. (2001) Initial sequencing and analysis of the human genome. *Nature 409*, 860−921.
(50) Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen,

N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001) The sequence of the human genome. *Science 291*, 1304−1351.

(51) Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S. P., Zdobnov, E. M., Zody, M. C., Lander, E. S., and Mouse Genome Sequencing, C. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature 420*, 520−562.

(52) Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., Scott, G., Steffen, D., Worley, K. C., Burch, P. E., Okwuonu, G., Hines, S., Lewis, L., DeRamo, C., Delgado, O., Dugan-Rocha, S., Miner, G., Morgan, M., Hawes, A., Gill, R., Celera, Holt, R. A., Adams, M. D., Amanatides, P. G., Baden-Tillson, H., Barnstead, M., Chin, S., Evans, C. A., Ferriera, S., Fosler, C., Glodek, A., Gu, Z., Jennings, D., Kraft, C. L., Nguyen, T., Pfannkoch, C. M., Sitter, C., Sutton, G. G., Venter, J. C., Woodage, T., Smith, D., Lee, H. M., Gustafson, E., Cahill, P., Kana, A., Doucette-Stamm, L., Weinstock, K., Fechtel, K., Weiss, R. B., Dunn, D. M., Green, E. D., Blakesley, R. W., Bouffard, G. G., De Jong, P. J., Osoegawa, K., Zhu, B., Marra, M., Schein, J., Bosdet, I., Fjell, C., Jones, S., Krzywinski, M., Mathewson, C., Siddiqui, A., Wye, N., McPherson, J., Zhao, S., Fraser, C. M., Shetty, J., Shatsman, S., Geer, K., Chen, Y., Abramzon, S., Nierman, W. C., Havlak, P. H., Chen, R., Durbin, K. J., Egan, A., Ren, Y., Song, X. Z., Li, B., Liu, Y., Qin, X., Cawley, S., Cooney, A. J., D'Souza, L. M., Martin, K., Wu, J. Q., Gonzalez-Garay, M. L., Jackson, A. R., Kalafus, K. J., McLeod, M. P., Milosavljevic, A., Virk, D., Volkov, A., Wheeler, D. A., Zhang, Z., Bailey, J. A., Eichler, E. E., Tuzun, E., Birney, E., Mongin, E., Ureta-Vidal, A., Woodwark, C., Zdobnov, E., Bork, P., Suyama, M., Torrents, D., Alexandersson, M., Trask, B. J., Young, J. M., Huang, H., Wang, H., Xing, H., Daniels, S., Gietzen, D., Schmidt, J., Stevens, K., Vitt, U., Wingrove, J., Camara, F., Mar Alba, M., Abril, J. F., Guigo, R., Smit, A., Dubchak, I., Rubin, E. M., Couronne, O., Poliakov, A., Hubner, N., Ganten, D., Goesele, C., Hummel, O., Kreitler, T., Lee, Y. A., Monti, J., Schulz, H., Zimdahl, H., Himmelbauer, H., Lehrach, H., Jacob, H. J., Bromberg, S., Gullings-Handley, J., Jensen-Seaman, M. I., Kwitek, A. E., Lazar, J., Pasko, D., Tonellato, P. J., Twigger, S., Ponting, C. P., Duarte, J. M., Rice, S., Goodstadt, L., Beatson, S. A., Emes, R. D., Winter, E. E., Webber, C., Brandt, P., Nyakatura, G., Adetobi, M., Chiaromonte, F., Elnitski, L., Eswara, P., Hardison, R. C., Hou, M., Kolbe, D., Makova, K., Miller, W., Nekrutenko, A., Riemer, C., Schwartz, S., Taylor, J., Yang, S., Zhang, Y., Lindpaintner, K., Andrews, T. D., Caccamo, M., Clamp, M., Clarke, L., Curwen, V., Durbin, R., Eyras, E., Searle, S. M., Cooper, G. M., Batzoglou, S., Brudno, M., Sidow, A., Stone, E. A., Payseur, B. A., Bourque, G., Lopez-Otin, C., Puente, X. S., Chakrabarti, K., Chatterji, S., Dewey, C., Pachter, L., Bray, N., Yap, V. B., Caspi, A., Tesler, G., Pevzner, P. A., Haussler, D., Roskin, K. M., Baertsch, R., Clawson, H., Furey, T. S., Hinrichs, A. S., Karolchik, D., Kent, W. J., Rosenbloom, K. R., Trumbower, H., Weirauch, M., Cooper, D. N., Stenson, P. D., Ma, B., Brent, M., Arumugam, M., Shteynberg, D., Copley, R. R., Taylor, M. S., Riethman, H., Mudunuri, U., Peterson, J., Guyer, M., Felsenfeld, A., Old, S., Mockrin, S., Collins, F., and Rat Genome Sequencing Project, C. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature 428*, 493−521.

(53) Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M. D., Roach, J., Oh, T., Ho, I. Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S. F., Clark, M. S., Edwards, Y. J., Doggett, N., Zharkikh, A., Tavtigian, S. V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y. H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D., and Brenner, S. (2002) Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science 297*, 1301−1310.

(54) Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science 304*, 1321−1325.

(55) Margulies, E. H., Blanchette, M., Haussler, D., Green, E. D., and Program, N. C. S. (2003) Identification and characterization of multi-species conserved sequences. *Genome Res. 13*, 2507−2518.

(56) Conkright, M. D., Guzman, E., Flechner, L., Su, A. I., Hogenesch, J. B., and Montminy, M. (2003) Genome-wide analysis of CREB target genes reveals a core promoter requirement for cAMP responsiveness. [(2003) *11* (5), 1417]. *Mol. Cell 11*, 1101−1108.

(57) Iourgenko, V., Zhang, W., Mickanin, C., Daly, I., Jiang, C., Hexham, J. M., Orth, A. P., Miraglia, L., Meltzer, J., Garza, D., Chirn, G. W., McWhinnie, E., Cohen, D., Skelton, J., Terry, R., Yu, Y., Bodian, D., Buxton, F. P., Zhu, J., Song, C., and Labow, M. A. (2003) Identification of a family of cAMP response element-binding protein coactivators by genome-scale functional analysis in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A. 100*, 12147−12152.

(58) Albertson, D. G., and Pinkel, D. (2003) Genomic microarrays in human genetic disease and cancer. *Hum. Mol. Genet. 12*, R145−R152.

(59) Kondo, Y., Shen, L., Yan, P. S., Huang, T. H., and Issa, J. P. (2004) Chromatin immunoprecipitation microarrays for identification of genes silenced by histone H3 lysine 9 methylation. *Proc. Natl. Acad. Sci. U.S.A. 101*, 7398−7403.

(60) Lan, H., Stoehr, J. P., Nadler, S. T., Schueler, K. L., Yandell, B. S., and Attie, A. D. (2003) Dimension reduction for mapping mRNA abundance as quantitative traits. *Genetics 164*, 1607−1614.

(61) Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet. 25*, 25−29.

(62) Doniger, S. W., Salomonis, N., Dahlquist, K. D., Vranizan, K., Lawlor, S. C., and Conklin, B. R. (2003) MAPPFinder: Using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol. 4*, R7.

(63) Dahlquist, K. D., Salomonis, N., Vranizan, K., Lawlor, S. C., and Conklin, B. R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet. 31*, 19−20.

(64) Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., and Groop, L. C. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet. 34*, 267−273.

(65) Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res. 32*, D449−D451.
(66) Chaussabel, D., and Sher, A. (2002) Mining microarray expression data by literature profiling. *Genome Biol. 3*, RESEARCH0055.
(67) Pandey, R., Guru, R. K., and Mount, D. W. (2004) Pathway Miner: Extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data. *Bioinformatics 20*, 2156−2158.
(68) Hu, Y., Hines, L. M., Weng, H., Zuo, D., Rivera, M., Richardson, A., and LaBaer, J. (2003) Analysis of genomic and proteomic data using advanced literature mining. *J. Proteome Res. 2*, 405−412.
(69) Masys, D. R., Welsh, J. B., Lynn Fink, J., Gribskov, M., Klacansky, I., and Corbeil, J. (2001) Use of keyword hierarchies to interpret gene expression patterns. *Bioinformatics 17*, 319−326.
(70) Basik, M., Mousses, S., and Trent, J. (2003) Integration of genomic technologies for accelerated cancer drug development. *Biotechniques 35*, 580−582, 584, 586 passim.
(71) Guerreiro, N., Staedtler, F., Grenet, O., Kehren, J., and Chibout, S. D. (2003) Toxicogenomics in drug development. *Toxicol. Pathol. 31*, 471−479.
(72) Chang, J. C., Wooten, E. C., Tsimelzon, A., Hilsenbeck, S. G., Gutierrez, M. C., Elledge, R., Mohsin, S., Osborne, C. K., Chamness, G. C., Allred, D. C., and O'Connell, P. (2003) Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet 362*, 362−369.
(73) Sotiriou, C., Powles, T. J., Dowsett, M., Jazaeri, A. A., Feldman, A. L., Assersohn, L., Gadisetti, C., Libutti, S. K., and Liu, E. T. (2002) Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res. 4*, R3.
(74) van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med. 347*, 1999−2009.
(75) Hackett, J. L., and Lesko, L. J. (2003) Microarray data—the US FDA, industry and academia. *Nat. Biotechnol. 21*, 742−743.
(76) van den Berg, M., Peterson, R. E., and Schrenk, D. (2000) Human risk assessment and TEFs. *Food Addit. Contam. 17*, 347−358.
(77) Van den Berg, M., Birnbaum, L., Bosveld, A. T., Brunstrom, B., Cook, P., Feeley, M., Giesy, J. P., Hanberg, A., Hasegawa, R., Kennedy, S. W., Kubiak, T., Larsen, J. C., van Leeuwen, F. X., Liem, A. K., Nolt, C., Peterson, R. E., Poellinger, L., Safe, S., Schrenk, D., Tillitt, D., Tysklind, M., Younes, M., Waern, F., and Zacharewski, T. (1998) Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. *Environ. Health Perspect. 106*, 775−792.
(78) Finley, B. L., Connor, K. T., and Scott, P. K. (2003) The use of toxic equivalency factor distributions in probabilistic risk assessments for dioxins, furans, and PCBs. *J. Toxicol. Environ. Health A 66*, 533−550.

TX0496690