Application of Genomics to Toxicology Research

Russell S. Thomas,[1,2] David R. Rank,[2] Sharron G. Penn,[2] Gina M. Zastrow,[1] Kevin R. Hayes,[1] Tianhua Hu,[2] Kalyan Pande,[1] Mark Lewis,[2] Stevan B. Jovanovich,[2] and Christopher A. Bradfield[1]

[1]McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, Madison, Wisconsin, USA; [2]Amersham Biosciences,

Traditional models of toxicity have relied on dissecting chemical action into pharmacokinetic and pharmacodynamic processes. However, the integration of genomic information with toxicology will enhance our basic understanding of these processes and significantly change the way we apply toxicological information to risk assessment and regulatory problems. In this article, we summarize the application of gene expression information and polymorphism discovery to four areas in toxicology: toxicity testing, cross-species extrapolation, understanding mechanism of action, and susceptibility. ***Key words:*** **gene expression, genomics, microarrays, polymorphisms, SNP, species extrapolation, susceptibility, toxicogenomics, toxicology.** ***Environ Health Perspect*** **110(suppl 6):919–923 (2002).**

http://ehpnet1.niehs.nih.gov/docs/2002/suppl-6/919-923author name/abstract.html

In the science of toxicology, the fundamental goal is to understand the effects of both single chemicals and mixtures of chemicals on biological systems and, in doing so, to allow the assessment of human health risks associated with exposure. To accomplish this task, toxicologists have modeled the toxic response as consisting of both pharmacokinetic and pharmacodynamic elements (Figure 1). Pharmacokinetics is a description of the distribution, metabolism, and excretion of a chemical, including its metabolites. Pharmacokinetics can be used to predict the ultimate dose of the toxic moiety at the site of action. In contrast, pharmacodynamics is a description of how tissues respond to the chemical once it gets there. These responses can include cell death, adaptation, differentiation, or proliferation and can include any manifestation of the toxic response in the whole organism. However, toxicology is at the beginning of a transition that is being driven by an explosion in the amount of genomic sequence information available and the fast-paced development of technologies to exploit its use. As a result, the science of toxicology is compelled to reanalyze its traditional models and incorporate this new knowledge.

Although the traditional model of toxicity has been used primarily in a descriptive manner by linking chemical doses at the site of action with tissue pathology, system-level toxicity, and overt mortality, the model is still valid and useful as a framework for understanding chemical toxicity. To keep pace with the current progress in biological research, the foundation of this model has shifted from descriptive observations to a mechanistic understanding of toxicity at the molecular level. Genomic information is central to this type of molecular understanding and underlies both the pharmacokinetic as well as the pharmacodynamic aspects of the model. For example, the genomic sequence (i.e., genotype) of an individual can significantly affect the pharmacokinetics for a particular chemical, thereby determining the individual's susceptibility to toxicity via changes in the target tissue dose (*1,2*). In addition, gene expression can change rapidly and dramatically in response to chemical exposure, and these changes are responsible for many of the pharmacodynamic effects.

Like biology, the foundation of toxicology is predicated on the sequence of our genome, and it also underlies many of our current conceptual models of how chemicals produce toxicity. Therefore, the integration of genomics into toxicological research is essential as we search to understand how various chemicals and the corresponding mixtures act in the human body and to develop better tools to assess the risks associated with exposure. With the sequencing of the human genome nearly completed (*3,4*) and significant progress on the mouse and rat models, the integration of genomics into toxicology on even a greater scale can become reality.

Application of Genomics to Toxicity Testing

In our current regulatory and public heath environment, chemicals that are thought to have the potential for a significant level of human exposure and to pose potential heath risk are selected to undergo subsequent testing for toxicity and carcinogenic potency. The traditional method of evaluating carcinogenic activity and chronic toxicity of a specific chemical has been the two-year animal bioassay. The experimental design for these studies involves animals of both sexes and usually from one strain of rat and/or mice. Several dosage levels are chosen, with approximately 50 animals per sex per dose level. The high dose corresponds to the maximum tolerated dose (MTD) as determined in short-term or subchronic toxicity studies and can be based on a variety of end points, including target organ pathology, changes in body weight, and adverse clinical signs. The intermediate dose selection is usually a function of the high dose (e.g., half the MTD), and the spacing between doses is often determined by the confidence in the prediction of the end point used to derive the high dose selection (*5*). The experimental protocol includes detailed records on animal health, vital statistics, and pathology. Because of the cost- and labor-intensive nature of these studies, each bioassay costs between $2 and $4 million and takes several years to complete (*6*). According to the U.S. National Toxicology Program (NTP), the number of chemicals currently tested by the NTP stands at 505 in long-term studies, 66 in short-term tests, and only a single subchronic study (*7*). Given that there are between 70,000 to 85,000 chemicals in commerce today (*8,9*), it is clearly impossible to apply current testing methodologies to all chemicals of concern, let alone the corresponding mixtures (Figure 2). It is apparent that alternative testing approaches must be developed if science is to maintain a significant role in environmental and public health policy.

The development of an efficient screening process that would allow prioritization of untested chemicals and mixtures based on their toxic potential would significantly affect how efficiently we evaluate both synthetic and naturally occurring compounds. One approach for predicting toxic potential is to classify chemicals based on their capacity to alter transcriptional programs in a manner similar to that of known toxicants or chemicals already tested in the two-year bioassay. Test chemicals that induce transcriptional responses in a manner similar to those induced by an established toxicant could then be classified as harboring toxic potential. These chemicals could be examined carefully by more thorough toxicological means and

This article is part of the monograph *Application of Technology to Chemical Mixture Research.*

Address correspondence to R.S. Thomas, Kalypsys, Inc., 11099 North Torrey Pines Rd., La Jolla, CA 92037 USA. Telephone: (858) 754-3316. Fax: (858) 754-3301. E-mail: rthomas@kalypsys.com

This work was supported in part by the Burroughs Wellcome Foundation, the National Institutes of Health (grants ES05703, T32CA09681, CA07175, GM23750), and a postdoctoral fellowship cosponsored by the Society of Toxicology and the Colgate-Palmolive Corporation.

Received 18 December 2002; accepted 24 June 2002.

subjected to interim regulations until they are proven safe. By taking this approach, one would have to make one overriding assumption—that most, if not all, toxic chemical exposures will alter gene expression at some level. In support of this assumption, toxicity by its very nature results from some form of cellular dysfunction or cell death that will result in changes in gene expression at some level. These changes can be associated either with the root cause of the toxicity or downstream of the initial event. Nonetheless, the resulting overall pattern of gene expression changes can act as a diagnostic "fingerprint" for that chemical to match with known toxicants or chemicals already tested in the two-year bioassay.

The widespread application of microarray technology to toxicity testing would require the integration of sophisticated statistical tools into the analysis. Specifically, the use of statistical classification techniques would be necessary to build predictive models for the subsequent classification of unknown/untested chemicals. These statistical techniques range from linear and nonlinear discriminant analysis (*10*) to Bayesian classification (*11*), nearest-neighbor approaches (*12*), and neural networks (*13*). With any of these techniques, a training data set is needed in which the model is trained on data where the answer is already known. For example, a potentially beneficial data set could come from a subset of the compounds already bioassayed by the NTP. These compounds could be fed to small numbers of rodents in short-term studies and the changes in gene expression evaluated using microarray technology. Provided that the subset of chemicals contains chemicals that show both positive and negative results in the bioassay, a classification model could be built to predict the response with subsequent untested chemicals. The accuracy of these predictions could be evaluated using additional cross-validation techniques (*14*) and the results used to make critical regulatory decisions about the safety of the analyzed compounds.

The application of microarray analysis and statistical classification tools to the type of chronic toxicity study outlined above has the potential to be extremely useful from both a scientific and an economic perspective. This type of study would allow the construction and validation of a statistical model based on gene expression patterns that would predict the long-term toxicity of an unknown chemical based on short-term rodent studies using only a small number of animals. The potential savings in cost and time could be immense, and the current gap between the number of tested and untested chemicals could be reduced in the interest of public health. However, there are significant hurdles to overcome before this type of analysis will be useful on a large scale. First, multiple factors converge to ultimately influence the manifestation of toxicity and the associated gene expression patterns. Among these factors are time, dose, route of administration, age of the animal, and sex. Fully characterizing the influence of all of these variables on transcript profiles with even a small number of treatments would require considerable resources (e.g., 500 treatments × 3 time points per treatment × 3 doses per time point × 3 routes per dose × 3 ages per treatment × 2 sexes × 3 tissues = 243,000 microarray studies).

Conceptually, some of these variables could be standardized in the animal studies and, in principle, the classification models incorporate a number of these variables because the models statistically relate gene expression measurements performed in a short-term study to the results from the two-year bioassay. Whether the classification models will be robust enough to allow accurate predictions despite these factors remains to be seen. Second, the training set must be carefully chosen to incorporate a broad spectrum of toxicants that act via different mechanisms. The ability of the statistical model to predict long-term toxicity of a chemical that differs significantly from the training set will be limited. Therefore, the larger the public database of chemically induced gene expression and the more toxicological categories added to the model, the more predictive the models will become. Although some progress is being made toward classifying chemical toxicants based on gene expression patterns (*11,15,16*), many of these hurdles still exist. However, toxicogenomics will continue to advance, and the era of the resource-intensive animal bioassay will undoubtedly end.

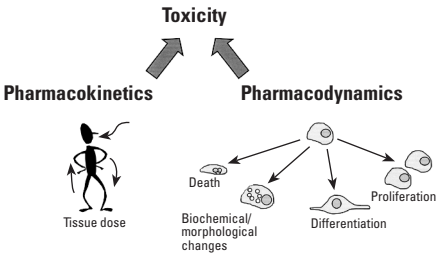

Figure 1. The traditional model of toxicity encompassing a number of quantitative and qualitative characteristics that fall into two categories—pharmacokinetics and pharmacodynamics. Pharmacokinetics is described as what the body does to the chemical via fate and distribution. Pharmacodynamics is described as what the chemical does to the body and how the body responds. These responses can include cell death, adaptation, differentiation, or proliferation that lead to the manifestation of the toxic response in the whole organism. Both pharmacokinetics and pharmacodynamics are integrated to link the dose of the chemical at the site of action with the response of the tissue, thereby providing a basis for estimating the risk of human exposure.

Application of Genomics to Species Extrapolation

It is well established that most known human chemical carcinogens are also carcinogenic in at least one species of laboratory animal. Whether the reverse is true cannot, and probably will never, be established with any degree of confidence. However, the assumption that it is true provides part of the foundation for toxicological testing in animal systems and is the basis for one of the underlying principles in toxicology—that experimental results in animal models, given certain restrictions, are generally applicable to humans. Regulatory agencies, academic researchers, pharmaceutical companies, and chemical companies all rely heavily on results from animal studies for establishing health and safety guidelines, assessing toxicity, and evaluating the potential efficacy of new drugs. For example, animal studies are commonly used at the U.S. Environmental Protection Agency (U.S. EPA) in both the hazard identification and response assessment phases of the risk assessment process. During hazard evaluation, the U.S. EPA evaluates a chemical's inherent toxicity using laboratory animal models (i.e., the type and extent of harmful effects), and the response assessment includes additional dose–response and chronic exposures to find the no-observed-effect level (NOEL) and calculate a reference dose (RfD). Although additional safety factors are built into calculating the RfD, the basis for its calculation relies on results from the animal studies.

Despite the heavy reliance on animal models in toxicological research, most toxicologists recognize that significant quantitative and qualitative differences exist between humans and the animal models used in toxicological research. For example, there are significant differences in the responses of rodents and humans to peroxisome proliferators (*17*), and the lethal dose for 2,3,7,8-tetrachlorodibenzo-*p*-dioxin varies by more

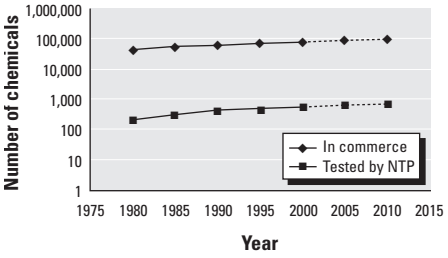

Figure 2. A comparison between the estimated number of chemicals in commerce and the number of chemicals tested in chronic animal bioassays by the NTP. Estimates of the number of chemicals in commerce were derived from previously published reports (*8*), and the number of chemicals tested was obtained from the NTP website (*7*). Based on the current rate of testing, it is apparent that the number of chemicals tested will never approach the number of chemicals in commerce.

than 1,000-fold between species (*18*). With these examples and many others, it is evident that the identification of toxicological species differences and their implementation in the regulatory arena are essential from both a health and economic standpoint. To do this, a clear understanding of both the pharmacokinetic and pharmacodynamic differences between species is necessary. For the pharmacokinetics, studies of the fate and distribution of the chemical in each species can be performed, and the application of biologically based models to this type of data has proven useful for species extrapolations (*19*). For the pharmacodynamics, the identification can be more challenging, particularly without a molecular understanding of the chemical's action. It is here that the application of genomics can identify cross-species differences at the molecular level.

The development of a system that would allow the assessment of all molecular differences between species after chemical exposure would have a significant impact on how we evaluate the pharmacodynamic aspects of cross-species extrapolations. Although assessing all molecular differences is nearly impossible, one approach for assessing these differences is to measure transcriptional alterations in orthologous sets of genes (i.e., genes in different species that evolved by speciation from a common ancestral gene; Figure 3). Test chemicals that induce different transcriptional responses between species could then be classified as harboring potential pharmacodynamic differences. These chemicals could be examined carefully on a mechanistic basis and subjected to additional regulatory review. By taking this approach, one would have to make two assumptions. First, for many chemicals, *in vitro* results from human cell types would have to be similar to those obtained *in vivo* because of obvious limitations on human studies. Second, conserved changes in gene expression equate to conserved pharmacodynamic end points. For the first point, arguments for and against extrapolating from *in vitro* results to *in vivo* predictions are important but will not be addressed here. In support of the second assumption, evolution and selection have maintained the structure and function of many biochemical pathways over time, resulting in the conservation of many important processes (*20*). Although the function of some genes may have changed over time, similar changes in orthologous sets of genes are likely to produce the same physiological end point whether it is inflammation, proliferation, apoptosis, necrosis, or cellular differentiation.

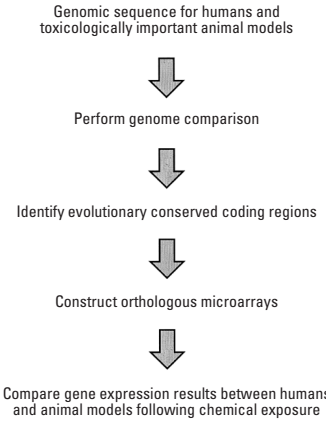

Figure 3. A flow chart describing the construction of orthologous microarrays to address critical species extrapolation issues in toxicology. From the top of the flow chart, the available genome sequences of both humans and the important rodent models are used to identify evolutionary conserved regions. These regions are screened for coding potential and arrayed on glass slides. Measurements of gene expression in both organisms after chemical insult will allow the evaluation of conserved pharmacodynamic end points.

To develop a system that can exhaustively measure alterations in orthologous genes, the majority of the genomic sequence of both species must be known and the sequences compared in order to identify evolutionarily conserved coding regions. For toxicology, this would mean the completion of the mouse (currently at 96% in both draft and finished sequence), rat (currently at 64% in both draft and finished sequence), and human genomes (currently at 98% in both draft and finished sequence). Previous studies have already compared large segments (>30 kb) of human and rodent sequences, demonstrating that coding domains are generally well conserved, whereas noncoding regions exhibit variable levels of conservation (*21*). After completion of the genomes, the sequences could be compared using various sequence alignment algorithms and the results screened for coding potential (i.e., containing an open reading frame over the length of the conserved sequence). The resulting list would contain orthologous sets of putative exons that could be used to create parallel rodent and human microarrays.

The application of genomics to species extrapolation issues in the form of cross-species sequence comparisons and orthologous microarray analysis has the potential to provide critical molecular data on the ways different species react to toxicants. This type of analysis would allow the integration of both pharmacokinetic information and pharmacodynamic information at a basic level and significantly improve the uncertainty associated with conventional risk assessments. Fundamentally, this issue is at the root of both biological and toxicological research. It is also tied to large amounts of money in terms of research dollars and Superfund-related cleanup costs, and it forms the basis for the protection of human health from chemical exposure. As a result, the use of genomic information of this type should be leveraged in toxicology to its fullest extent.

Application of Genomics to Understanding Mechanism

Although related to chemical classification and cross-species extrapolation, the identification of toxic mechanism based on gene expression is a unique problem with unique challenges. The primary difficulties are reliably detecting subtle changes in gene expression (i.e., <2-fold) that may be biologically relevant, assigning functional significance to any observed alterations, and separating downstream transcriptional changes from the causative changes. These problems are not just related to toxicology but also include the microarray field as a whole.

To reliably detect subtle changes in gene expression, technology and associated statistical methods must continue to advance in order to evaluate these potentially important transcriptional changes. However, there are obvious limits to what this or any other technology can detect in biological samples, and assessing potential changes that are less than the natural biological variability may be impossible. One potential solution to this problem is to separate the affected cell population from the surrounding normal tissue using microdissection (*22*). Using this technique, transcriptional alterations can be identified that would have normally gone undetected when analyzing the tissue as a whole.

To assign functional significance to observed changes in gene expression, various statistical methods have been used together with multiple experimental conditions in order to group genes according to common expression patterns. These methods include self-organizing maps (*23*), support vector machines (*24*), *k*-means (*25*), and hierarchical clustering techniques (*26*). The underlying assumption of the clustering approach is that genes changing in a coordinate fashion are functionally related, and if these changes correspond to the onset of toxicity, they may be mechanistically involved. Although there are many assumptions and data gaps in this type of analysis, genes with similar function do tend to be coordinately expressed. However, the link to toxic mechanism is much more tenuous, and it has not yet proved to be the proverbial silver bullet for understanding the mechanism of action for a particular treatment.

In less complex organisms, other approaches have proven useful for understanding and assigning significance to specific gene expression patterns. For example, the analysis of gene expression changes in yeast deletion mutants together with chemical treatments has allowed the identification of previously unidentified targets for commonly

used drugs (*27*). Although the experiment with the yeast deletion mutants was relatively ideal because of the compound chosen and the complexity of the model system, the application of these types of techniques toward the understanding of toxic mechanisms in mammalian systems shows promise. A recent estimate of the number of genes in the human genome is approximately 35,000 (*3,4*), and the rodent has even less. As a result, the systematic screening of various gene subsets could potentially be a reality provided that there is access to large numbers of *in vivo* or *in vitro* deletion mutants or an equivalent knockdown technology. A few emerging techniques show promise in this area. In particular, application of small interfering RNAs (siRNAs) in mammalian cell culture has the potential to provide a knockdown system on a genomewide scale (*28*).

Despite the potential of these types of experiments, they are not currently feasible for many toxicological microarray investigations. For some investigators, classic protein synthesis inhibitors have been used to separate primary and secondary effects (*29*). However, despite the apparent separation of these effects, the authors were still unable to arrive at a mechanism for the chemical and concluded that the biological mechanism was more complicated than previously imagined. For many treatments in our laboratory, a similar conclusion was reached, and it has become apparent that our understanding of what a particular gene expression change means biologically and how these patterns relate to a phenotype is limited at best. Thus, the application of arrays for understanding the mechanism of chemical toxicity may not yet be achievable given our current set of experimental tools and our present understanding of the biology involved.

Although the potential ability to measure global patterns of gene expression and instantly understand how an unknown chemical produces toxicity is an exciting possibility, this problem may prove to be the most difficult to solve in the near future because it relies on an intricate knowledge of the biological system being altered, and alterations at the transcriptional level may account for only a subset of the toxic responses. However, the field of genomics is progressing fast, and we are accumulating knowledge at an ever-increasing pace. Databases of gene expression are appearing, and studies of well-characterized toxicants are adding to our understanding. As a result, this ability to make these comparisons may become reality as our knowledge base continues to grow.

Application of Genomics to Susceptibility Prediction

The conceptual model of toxicology outlined in the introductory paragraphs views the etiology of the toxic response resulting from chemical exposure as a complex mix of pharmacokinetic and pharmacodynamic factors with an underlying genetic component. Part of this genetic component is the variability of the individual organism at the nucleotide level of the gene. Variations in a predominant allele are usually referred to as genetic polymorphisms, a term used to describe variation at an incidence of >1%. Of these changes, single nucleotide polymorphisms (SNPs) are the most common between individuals and are thus the primary affective agent in phenotypic variation. The frequency of polymorphism in humans is approximately one SNP per 1,000 bases of DNA, and currently more than 1.42 million SNPs have been identified across the human genome by the SNP consortium alone (*30*).

Although the majority of these polymorphisms have no consequences, a small subset can have dramatic effects on gene function, leading to a variety of phenotypic responses. As a result, the identification and characterization of these polymorphisms have become important in both the pharmaceutical industry and environmental toxicology. For example, in the pharmaceutical industry, a significant number of approved drugs have serious side effects not detected before approval and lead to approximately 100,000 deaths per year in the United States (*31*). These undiscovered side effects are due primarily to the heterogeneity of the human population and the relatively small number of patients in clinical trials, which therefore do not identify small, susceptible populations. From a public health perspective, the 100,000 deaths per year is roughly equivalent to the number of automobile fatalities in the United States in 1999 (*32*). Financially, if the pharmaceutical company has to pull the drug, it may experience losses as high as $500 million, not including potential lawsuits.

The identification and characterization of toxicologically relevant SNPs are difficult problems for a number of reasons, including their low relative frequency, problems in relating polymorphisms to function, and the sheer number of potential genes directly or indirectly involved in a particular toxicological phenotype. In general, polymorphisms that alter gene function do so in several ways. First, the change produces an amino acid substitution that alters protein stability, function, or activity. Second, the variation occurs in the regulatory region, which alters the rate of transcription. Third, the change produces a premature termination of the protein. Fourth, the variant produces changes in RNA stability or splicing. Relating these changes to function and human health consequences is the next big step and usually includes molecular epidemiologic methods, functional *in vitro* assays, and animal models (*33*).

The molecular epidemiologic approaches are difficult because of issues such as dose reconstruction and confounding variables such as concurrent exposure to other chemicals (i.e., chemical mixtures). The *in vitro* studies can also be difficult to interpret because of the relevance of *in vitro* results for *in vivo* end points. However, combining the two approaches provides both a mechanistic characterization of the polymorphism and relevance to the *in vivo* condition that will have a higher probability of success. In addition, identification of polymorphisms in model organisms will also prove to be important to the identification of relevant human SNPs. There are a significant number of chemicals in the environment that we know little or nothing about, and using the differential sensitivities of the various mouse strains will help map out important loci related to chemical toxicity as well as chemical carcinogenesis. Despite the inherent difficulties with each approach, SNPs located in toxicologically pertinent genes will most likely be uncovered through highly directed sequencing efforts focusing on these allelic differences (*34*). This type of sequencing effort, together with epidemiologic and functional *in vitro* analysis, is currently being undertaken at the U.S. National Institutes of Environmental Health Sciences with the Environmental Genome Project (*35*).

Understanding human susceptibility to exposure is at the heart of all toxicology. Clinical, environmental, and industrial toxicologists all deal with susceptible subpopulations, and the resulting variation in toxic responses can be large enough to warrant pulling a drug off the market or to lead to unpredictable debilitating diseases such as Gulf War syndrome. It is only by identifying these subpopulations that we can educate them on the various risks they may face. Although this also carries with it tremendous ethical implications, this application of genomics to toxicology has the potential to personalize the risk assessment process and tailor it for the genotype of the individual.

Conclusions

Although the study of genes and their impact on toxicity has been around for a number of years, the recent progress in genome sequencing and the ability to simultaneously monitor the expression of thousands of genes have moved genomics to the forefront of toxicology and spurred the coining of the term toxicogenomics. The near completion of the human genome and the progress on the rodent sequencing mark a new era not just in biology but in toxicology as well. However, the availability of the sequence information and the

development of the new genomic tools are only part of the effort. The integration of this information and application to problems such as toxicity testing, cross-species extrapolation, and susceptibility is the goal of the future and where some of the true benefits of these advancements lie. This will be true for both single chemical exposures and chemical mixtures that may have additive, greater than additive, or antagonistic effects on the assessed molecular end points.

References and Notes

1. Ueshima Y, Tsutsumi M, Takase S, Matsuda Y, Kawahara H. Acetaminophen metabolism in patients with different cytochrome P-4502E1 genotypes. Alcohol Clin Exp Res 20:25A–28A (1996).
2. Meyer UA, Skoda RC, Zanger UM. The genetic polymorphism of debrisoquine/sparteine metabolism-molecular mechanisms. Pharmacol Ther 46:297–308 (1990).
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. Science 291:1304–1351 (2001).
4. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 409:860–921 (2001).
5. Bucher JR, Portier CJ, Goodman JI, Faustman EM, Lucier GW. National Toxicology Program studies: principles of dose selection and applications to mechanistic based risk assessment. Fundam Appl Toxicol 31:1–8 (1996).
6. NTP. Annual Plan for Fiscal Year 1996. Research Triangle Park, NC:National Toxicology Program, 1996.
7. National Toxicology Program. Home Page. Available: *http://ntp-server.niehs.nih.gov/* [accessed 8 June 2001].
8. Ziegler J. Health risk assessment research: the OTA report. Environ Health Perspect 101:402–406 (1993).
9. Lucier GW, Schecter A. Human exposure assessment and the National Toxicology Program. Environ Health Perspect 106:623–627 (1998).
10. Huberty CJ. Applied Discriminant Analysis. New York:John Wiley & Sons, 1994.
11. Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Pande K, Glover E, Silander T, Craven MW, Reddy JK, et al. Identification of toxicologically predictive gene sets using cDNA microarrays. Mol Pharmacol 60:1189–1194 (2001).
12. Dasarathy BV. Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. Los Alamitos, CA:IEEE Computer Society Press, 1991.
13. Fausett L. Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. Englewood Cliffs, NJ:Prentice Hall, 1994.
14. Lachenbruch PA. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics 23:639–645 (1967).
15. Waring JF, Jolly RA, Ciurlionis R, Lum PY, Praestgaard JT, Morfitt DC, Buratto B, Roberts C, Schadt E, Ulrich RG. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. Toxicol Appl Pharmacol 175:28–42 (2001).
16. Hamadeh HK, Bushel PR, Jayadev S, DiSorbo O, Bennett L, Li L, Tennant R, Stoll R, Barrett JC, Paules RS, et al. Prediction of compound signature using high density gene expression profiling. Toxicol Sci 67:232–240 (2002).
17. Vanden Heuvel JP. Peroxisome proliferator-activated receptors (PPARs) and carcinogenesis. Toxicol Sci 47:1–8 (1999).
18. Poland A, Knutson JC. 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity. Annu Rev Pharmacol Toxicol 22:517–554 (1982).
19. Yang RSH, Andersen ME. Pharmacokinetics. In: Introduction to Biochemical Toxicology (Hodgson E, Levi PE, eds). Norwalk, CT:Appleton and Lange, 1994;49–73.
20. Schmidt-Nielsen K. Scaling in biology: the consequences of size. J Exp Zool 194:287–307 (1975).
21. Koop BF. Human and rodent DNA sequence comparisons: a mosaic model of genomic evolution. Trends Genet 11:367–371 (1995).
22. Alevizos I, Mahadevappa M, Zhang X, Ohyama H, Kohno Y, Posner M, Gallagher GT, Varvares M, Cohen D, Kim D, et al. Oral cancer *in vivo* gene expression profiling assisted by laser capture microdissection and microarray analysis. Oncogene 20:6196–204 (2001).
23. Tamayo P, Slonim D, Mestrov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 96:2907–2912 (1999).
24. Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Haussler D. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci USA 97:262–267 (2000).
25. Aronow BJ, Toyokawa T, Canning A, Haghighi K, Delling U, Kranias E, Molkentin JD, Dorn GW. Divergent transcriptional responses to independent genetic causes of cardiac hypertrophy. Physiol Genomics 6:19–28 (2001).
26. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863–14868 (1998).
27. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, et al. Functional discovery via a compendium of expression profiles. Cell 102:109–126 (2000).
28. Elbashir SM, Harborth J, Weber K, Tuschl T. Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods 26:199–213 (2002).
29. Puga A, Maier A, Medvedovic M. The transcriptional signature of dioxin in human hepatoma HepG2 cells. Biochem Pharmacol 60:1129–1142 (2000).
30. International SNP Map Working Group. A map of human genome sequence variation containing 1.4 million SNPs. Nature 409:928–933 (2001).
31. Wood AJJ, Stein MC, Woosley R. Making medicines safer—the need for an independent drug safety board. N Engl J Med 339:1851–1854 (1998).
32. U.S. Department of Transportation. Fatality Analysis Reporting System. Available: *http://www-fars.nhtsa.dot.gov/* [accessed 6 June 2001].
33. Guengerich FP. The Environmental Genome Project: functional analysis of polymorphisms. Environ Health Perspect 106:365–368 (1998).
34. Rieder MJ, Taylor SL, Clark AG, Nickerson DA. Sequence variation in the human angiotensin converting enzyme. Nat Genet 22:59–62 (1999).
35. National Institute of Environmental Health Sciences. Environmental Genome Project. Available: *http://www.niehs.nih.gov/envgenom/home.htm* [accessed 6 June 2001].