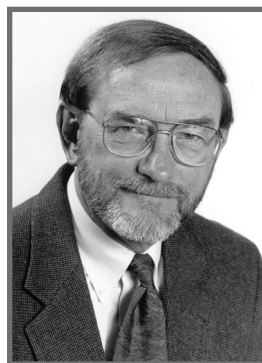*Toxicology will progressively develop from predominantly individual chemical studies into a knowledge-based science in which experimental data are compiled and computational and informatics tools will play a significant role in deriving a new understanding of toxicant-related disease.*

The National Center for Toxicogenomics: Using New Technologies to Inform Mechanistic Toxicology

The science of toxicology has evolved from the empirical codification of dose-related effects to studies directed toward understanding the mechanisms by which individual agents cause their effects in humans. Due to technical limitations, this evolution has been relatively slow, being accomplished one chemical or one effect at a time. To prospectively use the understanding gained on the mode of action of a single chemical, it is also necessary to know about structurally and functionally related chemicals and their time- and dose-dependent biological effects. In addition to chemicals and drugs, there are a plethora of environmental factors and stressors, such as ultraviolet and ionizing radiation, biological agents, and dietary and lifestyle components, that can contribute to the development of disease. The effects of all of these agents must be characterized to a progressively greater depth for us to understand the biochemical and genetic complexity of the cells in which adverse effects are manifested. In this view, toxicology will progressively develop from predominantly individual chemical studies into a knowledge-based science in which experimental data are compiled and computational and informatics tools will play a significant role in deriving a new understanding of toxicant-related disease (*1*).

The application of gene expression technology to understand the actions of chemicals and other environmental stressors on biological systems has been catalyzed by the rapid development of genome-based technology (*2–4*). The capacity to array large numbers of individual gene fragments on small matrices that can be hybridized to mRNA or cDNA has made it possible to synchronously assess the variety of effects that specific chemicals can cause, both good and bad. These technologic advances have led to the development of the field of toxicogenomics, which proposes to apply both mRNA and protein expression technologies to study chemical effects in biological systems (*5–11*). In recognition of the unique scientific opportunities afforded by this approach, the National Institute of Environmental Health Sciences (NIEHS) has created the National Center for Toxicogenomics (NCT). This center's mission is to promote the evolution and coordinated use of gene expression technologies and to apply them to the assessment of toxicologic effects in humans. The primary goal is to provide a worldwide reference system of genome-wide gene expression data and to develop a knowledge base of chemical effects in biological systems. Such a knowledge base will also, as a secondary goal, provide a profound understanding of the mechanisms by which stressor-induced injury occurs.

The NCT was formally established in September 2000 and is working to implement a strategy through which its mission can be achieved. There is implicit recognition that the goals are long-range and that substantial time and effort will be required to develop a truly informative knowledge base. Due to the magnitude and complexity of the science underlying these goals, a central theme of the NCT is the formation of national and international consortia of universities, other federal research and regulatory agencies, and private sector organizations. Some researchers (*10,12*) have expressed concern that the capacity to rapidly obtain large amounts of data on chemical effects using these technologies could result in inappropriate decisions about the potential for chemical-induced adverse effects. However, collective efforts such as those proposed in the NCT partnerships will do much to help develop scientific consensus on the appropriate uses of gene expression data.

Practical matters will dominate in the early stages of the NCT. Importantly, development of a reference database will require some effort at achieving a consensus on content and data quality standards. In addition, it is highly desirable that the data be preserved in a primary form so as to permit reanalysis as bioinformatics tools evolve and improve. We must approach all of these efforts in an incremental fashion, recognizing that in the face of rapid technologic change, it is impossible to anticipate all of the opportunities and problems that can and will develop.

Within this incremental approach several steps can be clearly defined. The first step is to test the hypothesis that signature profiles of individual chemicals, drugs, and other stressors can be defined. It is also necessary to test the hypothesis that specific toxicities will carry signature profiles and that these profiles can be recognized within certain dose-and-time parameters.

The NCT is conducting a series of "proof-of-principle" experiments that are designed to establish signature profiles and to link the patterns of altered gene expression to specific parameters of well-defined, conventional indices of toxicity. This "phenotypic anchoring" of gene expression data to conventional toxic effects is necessary to clearly demarcate pharmacologic or incidental effects from those changes either associated with or causal of adverse effects. A learning set of data on both pharmacologic and toxic gene expression profiles would then allow for distinguishing and predicting adverse effects for other well-defined compounds that could be tested under code. These data will help to establish the use of empirical gene expression profiles for toxicologic characterization, particularly chronic toxicity, which has been and will continue to be a focus of the NIEHS.

Recent studies using a relatively small learning set (*13–15*) and RNAs that are tested in blind studies have shown that it is possible to identify signature expressed gene patterns (*13*). This is an extremely important advance for the field and serves as the first major validation of the hypothesis that signature gene arrays can be defined and reproduced. Because these studies were conducted on acutely exposed animals, the array patterns appear to be representative of the pharmacologic activity of the chemicals. One group within the NCT

has begun to conduct further proof of principal experiments designed to distinguish between the pharmacologic and toxicologic effects of chemicals, and to develop a learning set of responses that are linked to conventional phenotypic parameters of toxicity (i.e., hepatomegaly, hepatocellular necrosis, inflammation, etc.) (16). These studies will take us one step closer to being able to address an issue that is of prime importance to the National Toxicology Program: the use of toxicogenomic approaches for understanding the biochemical processes associated with chronic chemical exposures. This is a particularly difficult problem but an important component of our strategy. To address this issue, it is important to determine whether or not serum/blood cells can be used as an alternative to specific target organ tissue, that is, can an informative subset of the pharmacologic and toxic parameters of acute chemical exposure seen in target tissues also be seen in blood. We are now testing this hypothesis, and if blood components can be used as a surrogate for tissue-specific chemical effects, this will open the door for comparative studies with exposed human populations.

One goal of the NCT is to define pathways and gene interactions through which chemical or environmental stressor effects are mediated; the secondary goal of developing a knowledge base, including empirical signature profiles, is equally important. One model for achieving both goals was initiated by the development of a Toxicogenomics Research Consortium that will work under a National Institutes of Health cooperative agreement. The announcement for development of this grant-supported consortium was divided into two components: an independent component comprising individual research projects within the framework of a program project grant, and a dependent component in which members of the consortium will collaborate in the development of studies to bring definition to toxicogenomics. In the current state of gene expression technology, there are different methodologies for arraying genes and for assessing mRNA expression, and different informatics tools that are being applied to the management and analysis of such data. To develop a substantial reference database, it will be necessary to collectively define quality criteria for submission of data to the common Chemical Effects in Biological Systems (CEBS) database and to have the capacity to store primary array data for subsequent reanalysis.

There is another parallel consortium that is dealing with the same platform and informatics issues and will complement the work of the NCT. The Health and Environmental Sciences Institute of the International Life Sciences Institute (ILSI) is coordinating the efforts of approximately thirty pharmaceutical companies in a worldwide effort to evaluate the harmonization of gene expression data and analysis. The ILSI Genomics Project (17) is focusing on three categories of toxicants: genotoxicants, hepatotoxicants, and nephrotoxicants. One strategy that the ILSI consortium is applying involves in-life studies conducted at specific laboratories where animals are dosed, tissues are taken for histopathology, and RNA is extracted and distributed to participating laboratories for microarray analysis using methods chosen by the respective participating laboratories. This type of collaboration will go a long way toward minimizing problems associated with RNA extraction and quality and in providing a basis for useful comparisons of expression data from various microarray platforms. The acquisition of tissue samples for histopathology will make it possible to characterize the type of toxicity associated with the chemical exposures at the time tissues are sampled for mRNA analysis so phenotypic anchoring of the microarray results will be possible in relationship to chemical pathology.

It is very likely that time- and dose-dependent microarray data will be reflective of phases of chemical activity. Initial responses of organisms or tissue to chemical exposure within 24–48 hr at doses that are not acutely toxic may provide data on specific genes involved in the pharmacologic action of the drug or chemical. As exposure to the stressor or agent is increased in time or dose, toxicity or cellular injury will become progressively obvious and various adaptive functions will be expressed. The use of microarrays should thus provide the opportunity to search for signature pathways of toxic injury. Such data will allow insight into the mode or mechanism of toxic injury and will also provide a means of distinguishing array patterns indicative of the adverse effect of the agent. If array data can be "phenotypically anchored" to conventional indices of toxicity (histopathology, clinical chemistry, etc.) it will be possible to search for evidence of injury prior to its clinical or pathologic manifestation. This approach could lead to development of early biomarkers of toxic injury, and it may also help to resolve issues related to interspecies extrapolation and variation in susceptibility across individuals.

Another critical component of a toxicogenomics strategy is the analysis of global alterations of protein expression (18). Although mRNA analysis is a potentially powerful tool for recognizing chemical-induced effects, analysis of protein sequence, structure, and modification provides advantages more clearly reflecting the actual current state of activity of the cell or tissue. Promising new methods in proteomics are emerging, including the capacity to profile proteins with surface-enhanced laser desorption mass spectrometry (19–23) and antibody arrays (24,25). Correlations between changes in mRNA and protein levels may offer insights into the function of genes or serve as a guide in the search for protein biomarkers of chemical exposure and predictive toxicity.

Another technologic innovation, called metabonomics (26), involves the application of nuclear magnetic resonance spectroscopy to characterize tissue-wide patterns of chemical metabolites. In conjunction with the use of microarray and proteomic characterization, metabonomics will be a useful adjunct to define mechanisms of injury.

Although genome-wide alterations in either mRNA, proteins, or metabolites in tissue extracts may be useful in identifying signature gene changes, a critical step in verifying that the gene product(s) plays a role in a toxic process requires localization of the target genes and its products to specific cell types. This requires the use of *in situ* hybridization, immunohistochemistry, laser capture microdissection, and other techniques to identify the cells expressing the gene(s). Other techniques, such as Northern or Western blotting or real-time polymerase chain reaction, are used to verify the expressed gene or to selectively analyze its expression over time or dose parameters. It will also become more important to analyze expression in specific cell populations in order to profile the alterations in gene expression involved in chronic chemical exposure that lead to tumor development. The capability to focus on limited cell populations is dependent upon cell separation methods that will minimize the opportunity for cells to alter the patterns of genes expressed *in situ*. Methods that prolong the isolation and separation of target cells will induce adaptive responses in the cells that are not related to chemical exposure. It is also dependent on high fidelity linear amplification of mRNA, the use of array platforms that require minimal amounts of cDNA, or proteomic methods that are highly sensitive.

At the present state of development of the field of toxicogenomics, the major advances in understanding toxic effects will still be made one chemical, agent, or mechanism at a time. However, the promise of this new technology is such that it can be used to generate data on large numbers of chemicals and exposure conditions and to develop an unprecedented knowledge base that can be used to guide future research, improve environmental health, and aid in regulatory decisions. Development of the knowledge base must proceed incrementally and requires the collective efforts of many individuals and institutions. From the results of individual mRNA arrays and proteomic or metabonomic analyses it is difficult to discern the implications of all of the expression changes observed.

However, as the database expands to include structurally or functionally related agents and as gene identity, functional genomics, and annotation progresses, it will be possible to search in a comprehensive way for common, critical, or causal changes. As it becomes possible to create pathway maps of common cellular processes, it will be possible to map partial genome arrays to pathways and to link such changes to known phenotypic markers of toxicity. The proposed databases and relational linkages must grow incrementally, and developers and users must have the patience and dedication to remain on course. Such incremental growth will eventually become exponential growth, and the field of toxicology will be profoundly changed. Given the vast numbers and diversity of drugs, chemicals, and environmental stressors, the diversity of species in which they act, the time and dose factors that are critical to the induction of beneficial and adverse effects, and the diversity of phenotypic consequences of exposures, it is only through the development of a rich knowledge base and its availability to all of the scientific community that toxicology and environmental health can rapidly advance. Concomitant with development of the data/knowledge base must be the evolution of informatics (computational and statistical) and data mining tools (query algorithms, relational interfaces, etc.) and the individuals trained to apply them (*27–30*).

The NCT has committed itself to the national effort to develop a CEBS knowledge base as a long-range goal of its strategy. The magnitude of the effort required to populate the databases that will comprise the knowledge base requires a collective will and collaborative efforts. We will continue to develop additional partnerships with scientists in academia, the private sector, and other governmental organizations to create a public knowledge base that will be a lasting resource for the scientific community. The efforts of the NCT can be followed on the NCT Web site (*31*).

Raymond W. Tennant
National Center for Toxicogenomics
National Institute of Environmental Health Sciences
Research Triangle Park, North Carolina
E-mail: tennant@niehs.nih.gov

REFERENCES AND NOTES

1. Iannaccone PM. Toxicogenomics: "the call of the wild chip" [Editorial]. Environ Health Perspect 109:A8–A11 (2001).
2. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470 (1995).
3. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14:1675–1680 (1996).
4. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37 (1999).
5. Afshari CA, Nuwaysir EF, Barrett JC. Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. Cancer Res 59:4759–4760 (1999).
6. Olden K, Guthrie J. Genomics: implications for toxicology. Mutat Res 473:3–10 (2001).
7. Hamadeh HK, Bushel P, Nuwaysir E, Paules R, Barrett JC, Afshari. CA. cDNA microarray technology: merging toxicology and genomics. Comments Toxicol (in press).
8. Hamadeh HK, Bushel P, Paules R, Afshari CA. Discovery in toxicology: mediation by gene expression array technology. J Biochem Mol Toxicol 15:231–242 (2001).
9. Lobenhofer E, Bushel P, Afshari CA, Hamadeh, HK. Progress in the application of DNA microarrays. Environ Health Perspect 109:881–891 (2001).
10. Fielden MR, Zacharewski TR. Challenges and limitations of gene expression profiling in mechanistic and predictive toxicology. Toxicol Sci 60:6–10 (2001).
11. Hamadeh H, Afshari CA. Gene chips and functional genomics. Am Sci 88:508–515 (2000).
12. Smith LL. Key challenges for toxicologists in the 21st century. Trends Pharmacol Sci 22:281–285 (2001).
13. Hamadeh HK, Bushel P, Jayadev S, Martin K, DiSorbo O, Sieber S, Bennett L, Tennant RW, Stoll R, Barrett JC, et al. Unpublished data.
14. Bartosiewicz MJ, Jenkins D, Penn S, Emery J, Buckpitt A. Unique gene expression patterns in liver and kidney associated with exposure to chemical toxicants. J Pharmacol Exp Ther 297:895–905 (2001).
15. Bulera SJ, Eddy SM, Ferguson E, Jatkoe TA, Reindel JF, Bleavins MR, De La Iglesia FA. RNA expression in the early characterization of hepatotoxicants in Wistar rats by high-density DNA microarrays. Hepatology 33:1239–1258 (2001).
16. Bessems JG, Vermeulen NP. Paracetamol (acetaminophen)-induced toxicity: molecular and biochemical mechanisms, analogues and protective approaches. Crit Rev Toxicol 31:55–138 (2001).
17. ILSI Health and Environmental Sciences Institute. Application of Genomics and Proteomics in Mechanism-Based Risk Assessment Subcommittee. Available: http://hesi.ilsi.org/activities/actslist.cfm?pubentityid=8&pubactivityid=29 [cited 29 November 2001].
18. Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology 14:61–65 (1996).
19. Rubin RB, Merchant M. A rapid protein profiling system that speeds study of cancer and other diseases. Am Clin Lab 19:28–29 (2000).
20. Merchant M, Weinberger SR. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis 21:1164–1177 (2000).
21. Li X, Mohan S, Gu W, Miyakoshi N, Baylink DJ. Differential protein profile in the ear-punched tissue of regeneration and non-regeneration strains of mice: a novel approach to explore the candidate genes for soft-tissue regeneration. Biochim Biophys Acta 1524:102–109 (2000).
22. Borrebaeck CA. Antibodies in diagnostics - from immunoassays to protein chips. Immunol Today 21:379–382 (2000).
23. Liotta L, Petricoin E. Molecular profiling of human cancer. Nat Rev Genet 1:48–56 (2000).
24. Huang RP. Detection of multiple proteins in an antibody-based protein microarray system. J Immunol Methods 255:1–13 (2001).
25. Lueking A, Horn M, Eickhoff H, Bussow K, Lehrach H, Walter G. Protein microarrays for gene expression and antibody screening. Anal Biochem 270:103–111 (1999).
26. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29:1181–1189 (1999).
27. 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).
28. Ermolaeva O, Rastogi M, Pruitt KD, Schuler GD, Bittner ML, Chen Y, Simon R, Meltzer P, Trent JM, Boguski MS. Data management and analysis for gene expression arrays. Nat Genet 20:19–23 (1998).
29. Tamayo P, Slonim D, Mesirov 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).
30. Quackenbush J. Computational analysis of microarray data. Nat Rev Genet 2:418–427 (2001).
31. National Center for Toxicogenomics. Available: http://www.niehs.nih.gov/nct/home.htm [cited 29 November 2001].