

Toxicogenomics

Toxicology has come a long way since Paracelsus, a scientist during the late Middle Ages, first uttered the phrase “the dose makes the poison.”

With these words, Paracelsus unveiled the experimental basis of toxicology, a science that has recently attained a level of sophistication that early scientists could have hardly imagined.

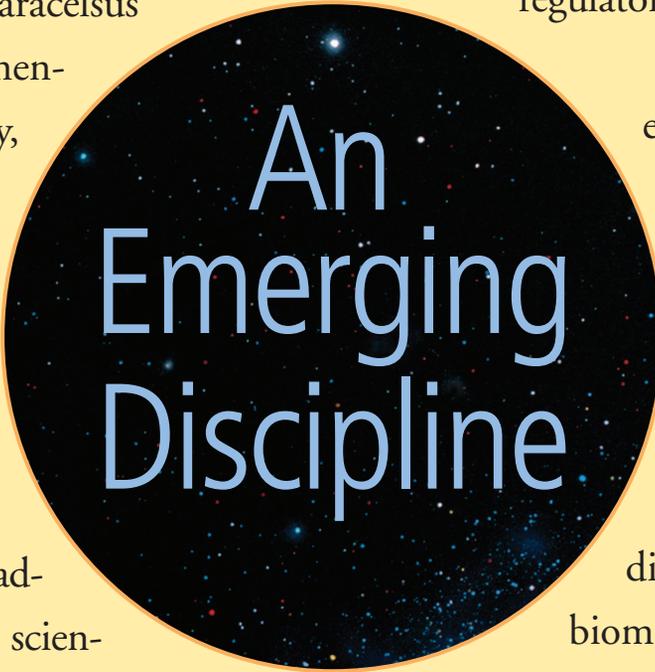
Thanks to rapid advances in technology, scientists are now exploring the complex circuitry of genes and proteins that modulate toxic responses. Until recently, the genes that make up these circuits could be studied only in limited numbers. But

This preview article will also appear in the inaugural issue of *EHP Toxicogenomics Edition*.

new genomic tools are making it possible to study how chemicals affect the expression of thousands of genes either simultaneously or sequentially along regulatory pathways.

The resulting gene expression patterns, or profiles—in addition to the cellular networks they give rise to—are the hallmarks of toxicogenomics. This emerging discipline provides new biomarkers of exposure and

effects, as well as fresh opportunities for preventing environmentally related diseases. “Toxicogenomics opens countless doors to our understanding of how cells and organisms respond to chemical agents,” says Leona Samson, a professor of toxicology at



An
Emerging
Discipline



the Massachusetts Institute of Technology. “We’re seeing totally unexpected cellular processes that turn out to be toxicologically meaningful. There’s so much to learn in terms of how it all fits together.”

A Basis in Genomics

As the name implies, toxicogenomics has its roots in genomics, the study of gene function. Among the greatest genomic achievements is the nearly complete decoding of the human genome. With this achievement, scientists are describing the ordered sequence of genes in human DNA. In recent years, the DNA of many other organisms has also been sequenced, providing scientists with species-specific blueprints to the molecular machinery of life. But as any expert will acknowledge, gene sequences are only the master template for this machinery; in and of themselves, they provide little information about the way living systems function. To understand how genomes govern living processes, scientists must study the biochemical networks they generate.

Scientists engaged in toxicogenomic studies apply genomic tools to learn how these networks modulate a cell’s response to drug or chemical exposures. Microarrays, for example, identify changes in gene expression associated with specific types of toxic responses. Proteomics—a field with many implications beyond toxicogenomics—characterizes the milieu of proteins appearing in a cell after a given exposure. And with metabolomics, scientists look at broad patterns of chemical metabolites in exposed tissues. The gene, protein, and metabolite profiles that make up toxicogenomic data each reflect key steps leading from exposure to disease.

By comparing chemically altered genomic profiles to those obtained from nontreated controls, scientists can identify unique biomarkers for disease processes. To illustrate, consider microarray data showing that a chemical activates the genes coding for cytochrome P450, a metabolic enzyme found in liver cells. If the chemical also causes liver damage, scientists might hypothesize that activated P450 is involved in the toxic response mechanism. The gene expression profile for activated P450 is therefore a potential biomarker for chemically induced injury to the liver. Confirming it as such would require additional study. But even in the absence of more complete mechanistic data, putative toxicogenomic biomarkers can be useful diagnostic tools for predicting toxicity, says William Pennie, director of molecular and investigative toxicology at Pfizer in Groton, Connecticut.

“It’s like guilt by association,” he explains. “In isolation, you may not know how each biomarker relates to the disease process. But over time you might find that some are consistently associated with certain toxic end points.

Then the biomarkers become increasingly predictive for these end points.” According to Pennie, pharmaceutical and biotechnology companies are exploring how toxicogenomic biomarkers can be applied during new product development. Specifically, preclinical screening models will look for biomarkers to guide the development of safer drugs and chemicals, he says. Biomarkers that reliably predict certain toxic effects may preclude the need for additional testing, he adds. Therefore, toxicogenomic tools have the potential to reduce the number of animals used in research.

Currently, DNA-based microarrays, or “gene chips,” so dominate the field that toxicogenomics is sometimes viewed erroneously as being concerned solely with gene expression. “This mainly reflects the advanced state of DNA-based microarray technology,” says Kenneth Ramos, a professor of toxicology at Texas A&M University and the editor of *EHP Toxicogenomics Edition*, a new quarterly edition of *EHP* debuting in 2003. Microarrays have emerged as widely available high-throughput tools capable of producing global arrays that identify all the genes expressed in a given sample. But few, if any, laboratories have the ability to assess all the proteins in a biological sample—although new proteomic tools, including surface-enhanced laser desorption/ionization–time of flight mass spectroscopy and protein arrays, should facilitate this capability within a few years.

In fact, it is likely that the most effective diagnostic biomarkers will be proteins, says Cynthia Afshari, associate director of toxicology at Amgen in Thousand Oaks, California, and former codirector of the NIEHS Microarray Center. This is in part because proteins are more stable and easier to collect in a clinical setting than the RNA samples used to elucidate gene expression, she says. Furthermore, expression profiles sometimes will not reflect important functional states that arise from protein changes in the cell.

Over time, proteomics and metabolomics data will become increasingly valuable for toxicogenomic research, experts say. Ramos says the NIEHS’s National Center for Toxicogenomics (NCT), which was established in December 2000 to combine toxicogenomic information with the knowledge emerging from classical toxicology, has embraced these technologies readily. “To achieve the goal of a fully integrated view of toxic responses, we need all this information,” he says. Indeed, by combining all the toxicogenomic biomarkers, scientists will construct the intricate biochemical pathways of toxicity.

An even clearer picture of the human response will emerge from linking toxicogenomic investigations to studies of interindividual variation. Scientists are now cataloging the gene variants that increase or decrease human

susceptibility to chemically induced diseases. Among these variants are simple DNA sequence variations called single nucleotide polymorphisms (SNPs). SNPs can influence toxic responses in a number of ways. For instance, a SNP might inhibit the formation of an enzyme involved in chemical detoxification. Individuals with such a SNP are therefore more vulnerable to chemicals that the enzyme would normally render harmless.

But integrating SNP data with toxicogenomics is a long way off, says NCT deputy director James Selkirk. “When the SNPs are identified in known genes, we can look to see how they affect specific genomic pathways,” he says. But many of the SNPs identified thus far are present in genes for which the function is still unknown.

Enhancement through Microarrays

Experts generally agree that microarrays enabled the rise of toxicogenomics in the late 1990s. Michael Waters, assistant director for database development at the NCT, calls microarrays “the pivotal technology.”

Microarrays are silicon or glass chips coated with thousands of discrete spots of nucleic acids called probes. Each probe corresponds to a specific gene. To use the technology, RNA from a biological sample is “reverse transcribed,” or used to produce an identical copy of the gene that would normally be produced in a living cell. This complementary DNA, or cDNA, is labeled with a radioactive or fluorescent tag and bound to the chip. Ideally, each molecule in the labeled cDNA binds (or “hybridizes”) to its complementary probe on the array. Using quantitative imaging techniques that read the radioactivity or fluorescence, scientists assess cDNA hybridization and thereby identify the genes expressed in the original biological sample and their relative levels of expression.

In toxicogenomic studies, the cDNA is obtained from drug- or chemical-exposed animal tissues or cells in culture. At first, the microarrays used in these studies were small and contained genes for specific pathways such as chemical metabolism or DNA repair, Afshari says. These early microarrays and subsequent refinements have enabled scientists to pursue hypothesis-driven studies focused mainly on obtaining additional information about a particular toxicity mechanism.

As microarrays evolved, the number of genes that could be placed on a chip grew rapidly. Today, investigators can buy chips containing tens of thousands of genes or even whole genomes for certain species. The Santa Clara, California–based firm Affymetrix produces several chips that represent the entire mouse genome. According to Waters, the availability of these larger “discovery” chips has allowed scientists to focus more broadly on gene function

and the identification of novel pathways. Scientists who use the larger chips are able to find entirely new gene expression patterns and associated mechanisms resulting from toxic exposures. "Discovery studies are very important, because we want to develop a more global understanding of toxic mechanisms, and the discovery mode allows us to do that more completely and accurately," Waters says.

Gene Expression and Toxicity

Carl Barrett, scientific director of the Center for Cancer Research at the National Cancer Institute, says the ultimate goal of toxicogenomics is to link chemically induced gene expression patterns to either detrimental, harmless, or even protective effects. "One should not assume the patterns are linked to the cause of toxicity," he cautions. "The patterns could be associated with responses that aren't even toxicologically relevant."

One way to assess the toxicological significance of the patterns is to "phenotypically anchor" them to standard toxicological indices, such as clinical chemistry or tissue pathology. Such experiments are currently a major activity at the NCT. Phenotypic anchoring, says Selkirk, is a technique that couples the unique gene expression patterns induced by chemical exposures to visible evidence of harm. In this fashion, the gene expression patterns provide chemical-specific signatures for toxicological pathways and effects.

Ideally, the signatures should relate to expression profiles obtained at multiple dose levels, Selkirk says. Low-dose signatures, for example, can be correlated with small, ultra-structural changes in cells or tissues, which are observable only with electron microscopy. Explains Waters, "If we can identify the gene signatures that precede an obvious toxic outcome, then we can use the signatures as diagnostic tools. Once we have an understanding of the relationship between gene changes and toxic effects, we can link them to reversible or irreversible damage."

In what NCT director Raymond Tennant calls "an extremely important advance for the field," a series of recent studies have confirmed that gene expression patterns can provide reproducible signatures for toxic mechanisms. In one study, published in the June 2002 issue of *Toxicological Sciences*, Hisham Hamadeh, previously at the NIEHS and now at Amgen, found that compounds acting through a common mechanism (peroxisome proliferation, a cellular process related to oxidative stress) have similar and collectively distinct gene expression profiles. This means that compounds with similar mechanisms can be grouped by a common gene signature, which can then be used to predict the chemical class of an unknown compound.

Linking signature profiles with toxic effects is the goal of ongoing experiments at the NCT.

Some of these experiments are focusing on acetaminophen, a highly studied compound with enormous public exposure. According to Selkirk, NCT scientists are using genomic and proteomic tools to identify the biochemical pathways corresponding to therapeutic and toxic doses. "We want to study both pathways so we know where and when they diverge, and how toxicity is manifested," Selkirk says. So far, the exposures being considered in all these studies are acute. This is mainly a function of practicality—acute exposures produce a series of discrete cellular events that scientists can study to build confidence in toxicogenomic techniques.

Unfortunately, most human exposures aren't so simple. People are generally exposed to many compounds simultaneously, often on a chronic or intermittent basis. Eventually, Tennant says, toxicogenomics will have to address more realistic exposure scenarios. These pathways are much more complex, he admits. Any evaluation of chronic exposures must contend with the added dimensions of time, adaptive response, and cellular repair. "The signals for each of these processes are masked in the complexity of the response," Tennant says. "With repeat dosing, it all becomes much more intricate."

Knowledge and Standardization

Evaluating chronic exposures merely adds to the already overwhelming challenge of managing toxicogenomic data. Just a few years ago, genes involved in a given mechanism were studied one at a time. But with the advent of microarrays, researchers now study gene pathways by the hundreds, or even thousands, for any single exposure. Add multiple doses in varying tissues and species, not to mention proteomic and metabolomic parameters, and the volumes of data generated quickly overpower most analytical capabilities.

Data analysis and management for genomics are the realm of the associated field of bioinformatics, which applies computational tools toward the understanding of biology. The goal of bioinformatics, says Srinivasa Nagalla, director of the Center for Biomarker Discovery at Oregon Health & Science University, is to codify toxicogenomic information in ways that facilitate "data mining," meaning the quick extraction of relevant parameters stored in a database. This capability will usher in a new era of *in silico* toxicology, Nagalla says, in which scientists use computer searching to screen biomarkers against signature pathways. "Ultimately, for any given tissue, you want to easily identify expressed genes, the degree to which they are expressed, and how those genes are linked to each other [in pathways and networks]," Nagalla says. "It comes down to mathematical equations for predicting gene expression

as a consequence of chemical exposure."

Recently, the NCT laid the groundwork for a repository of toxicogenomic data that will be housed at the NIEHS and made available to scientists all over the world via the Internet. The Chemical Effects in Biological Systems (CEBS) database is being described as a "knowledge base" combining toxicogenomic biomarkers with chemical effects data. Eventually, CEBS data sets will be searchable by compound, structure, toxicity and pathology end point, gene, gene group, pathway, and polymorphism, each as a function of dose, time, species, and target tissue. Similar to the way in which GenBank databases are queried for genome sequences, researchers will query CEBS to obtain information on genes and associated toxicity pathways. If, for example, all that is known about a newly discovered compound is its chemical structure, scientists could query CEBS using this information to obtain pathway and toxicity data associated with other, similar compounds. The CEBS output could provide a rough screen for potential effects and new directions for future research on the mystery compound. According to Waters, a prototype version of CEBS, running on an Oracle database backbone, will be online by late 2003.

But building such a database over time is no easy task—it requires cooperation and data sharing by researchers from many scientific disciplines, from pathology to mathematics, and a mutually agreeable format for linking toxicogenomic data. According to Waters, CEBS will soon begin accepting data from the Toxicogenomics Research Consortium (TRC), a group of five academic research centers, plus the NIEHS Microarray Center, that is funded by the NIEHS [see "Toxicogenomics Research Consortium Sails into Uncharted Waters," p. A744 this issue]. Waters stresses that at some point in the future, when minimal data standards are universally accepted, CEBS will also accept data from many other sources, including industry and other academic and private research centers.

A parallel consortium to the TRC that also aims to participate in CEBS is being coordinated by the Health and Environmental Sciences Institute of the International Life Sciences Institute (ILSI), an independent research organization based in Washington, D.C. The ILSI Technical Committee on Application of Genomics in Mechanism-Based Risk Assessment, which includes scientists from industry, regulatory agencies, and academia, is working on approaches to incorporate toxicogenomic data into safety assessment for new drugs and chemicals. Pennie, who chairs the group, emphasizes that the generated information "is going to be publicly available and of real interest and usefulness to both specialists and general scientists."

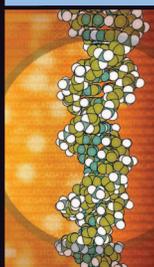
On “-Omics” — A Glossary



Bioinformatics: The science of managing and analyzing biological data using advanced computing techniques. Especially important in analyzing genomic research data.



Data mining: The extraction of relevant parameters stored in a database.



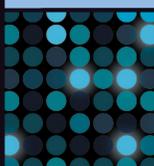
Genomics: The study of genes and their function. Genomics differs from genetics in that genetics looks at single genes, one at a time, whereas genomics looks at all genes, over time, to determine how they interact and influence biological pathways, networks, and physiology.



Metabolomics: The study of the total metabolite pool (the metabolome) through a wide spectrum of technologic methods including liquid chromatography–mass spectrometry, gas chromatography–mass spectrometry, and nuclear magnetic resonance.



Metabonomics: The study of the total metabolite pool (the metabolome) specifically through nuclear magnetic resonance profiling.



Microarray: Sets of miniaturized chemical reaction areas that may also be used to test DNA fragments, antibodies, or proteins.

Pharmacogenomics: The analysis of the effect of genomics—in particular, genetic variation (polymorphisms)—on drug response.

Pharmacogenetics: A subset of pharmacogenomics encompassing the study of genetic variation underlying differential response to drugs, particularly genes involved in drug metabolism. Often used to refer specifically to tests that predict drug response.



Proteomics: The study of the full set of proteins (the proteome) encoded by a genome.



Single nucleotide polymorphism (SNP): A position in the genome where some individuals have one DNA base (e.g., A) and others have a different base (e.g., C). SNPs and point mutations are structurally identical, differing only in their frequency. Variations that occur in 1% or less of a population are considered point mutations, and those occurring in more than 1% are considered SNPs.



Toxicogenomics: The study of how genomes respond to environmental stressors or toxicants. Combines genome-wide mRNA expression profiling with protein expression patterns, using bioinformatics to understand the role of gene–environment interactions in disease and dysfunction.



Transcriptomics: The generation of messenger RNA expression profiles.



Clockwise from top left: EyeWire; EyeWire; Digital Vision; EyeWire; U.S. Department of Energy Human Genome Program; U.S. Department of Energy Human Genome Program; EyeWire; EyeWire

The ILSI committee has already completed the first phase of its research—profiling toxicogenomic parameters for chemicals with hepatotoxic, nephrotoxic, and genotoxic mechanisms. “The idea is to make the data complementary with CEBS and available in [a form that can be accessed by others],” Pennie says. “The NCT will have continual access, and I hope the data sharing will be reciprocal.” According to Pennie, these data are also being prepared for entry into a database managed by the European Bioinformatics Institute, a multinational research organization based in England.

The prospect of sharing data across databases and research centers raises a highly challenging issue: the standardization of experimental protocols and data formats. Therefore, a system for ensuring uniform data quality is of key importance, experts say. However, as in any field, toxicogenomic experiments are highly variable in terms of approach, outcome, and interpretation.

One line of reasoning suggests that scientists should apply standardized methods to minimize this variability. According to Pennie, the ILSI committee tackled this question as one of its first priorities. But after a period of internal debate, he says, committee members concluded that standardized approaches wouldn't yet be possible to adopt, in part because they felt researchers shouldn't be constrained in method development. “The techniques are still evolving at a rapid pace,” he explains. “Furthermore, we want to understand what the sources of variability are.”

With respect to standardization, both Pennie and Waters (whose NCT duties include project officership of the CEBS database) say they will lean heavily on the recommendations of the Microarray Gene Expression Data Society, an international organization working to facilitate the sharing of microarray data. This group has developed a core set of guidelines that is loosely termed Minimum Information About a Microarray Experiment, or MIAME. According to the society's website, these guidelines, which apply to microarray gene expression research, will “assist in the development of microarray data repositories and data analysis tools.” The guidelines provide a uniform nomenclature for describing toxicogenomic experiments and data. Says Waters, “With regards to CEBS, we will follow the recommendations coming from that society in addition to the guidance we obtain from the Toxicogenomics Research Consortium and the NIEHS Microarray Center.”

Practical Applications

The implications of toxicogenomic research ultimately extend far beyond laboratories and clinics, and reach into public

and environmental policy. Even now, policy experts are grappling with fundamental questions about how toxicogenomic data will transform the process of setting standards for chemical hazards. The risk assessment protocols used to set these standards have always focused on protecting the most sensitive subgroups of the human population. Usually, these subgroups are defined using equations that relate human responses to lowest-observed-effect levels seen in animal experiments. These low-level effects—usually fairly benign responses such as altered lung function or changes in blood chemistry—are thought to be shared by all members of the population. Therefore, the standards are said to protect against “population-level effects.”

Now, toxicogenomics is revising the concept of low-level effects. Once the purview of the pathologist, the limits of sensitivity are increasingly being defined according to genetic susceptibility to toxic exposure. Therefore, sensitive subgroups are transformed from an amorphous entity into a clearly defined genetic subset of individuals within the population.

According to Richard Sharp, an assistant professor of medical ethics at the Center for Medical Ethics and Health Policy of the Baylor College of Medicine, the identification of these subpopulations raises some important questions. Most importantly, who bears the cost of protecting these individuals? This issue will likely first come to bear in the workplace, he says. Once sensitive subgroups are defined, employers may deny jobs to applicants on the basis of genetic screens that show them to be sensitive to workplace hazards. “How much discretion should an employer have to make these kinds of decisions?” Sharp asks. Sharp chairs an NCT working group on ethical, legal, and policy issues that is now exploring this and other issues.

Regulatory agencies are also grappling with toxicogenomics issues. In June 2002, the U.S. Environmental Protection Agency (EPA) issued its first guidelines for using genomics data for the standard-setting process. In the guidelines, the agency opined that toxicogenomics will potentially have an “enormous impact on our ability to assess the risk from exposure to stressors and ultimately to improve our risk assessments.” But the guidelines also make clear that “the relationships between changes in gene expression and adverse effects are unclear at this time and may be difficult to evaluate.” The EPA's current position on the matter is that, although useful, toxicogenomic data alone are insufficient to characterize risk. Therefore, the agency will accept the data but will consider them only on a case-by-case basis.

The U.S. Food and Drug Administration

(FDA) is also closely watching developments in the field. Frank Sistare, director of the Division of Applied Pharmacology Research at the FDA's Center for Drug Evaluation and Research, says the administration has just completed a series of multistakeholder meetings on the use of genomic data for drug evaluations. A report describing the FDA's position on the issue is now being prepared. Echoing the concerns of EPA officials, Sistare emphasizes the difficulty of linking microarray results to adverse effects. With respect to submitting toxicogenomic data for drug approval processes, Sistare says, “It's more of a challenge to industry than it is to us right now. I don't think anyone can provide a clear answer on what every signal means on an experiment conducted with microarrays that query thousands of end points at once. This puts industry in the unenviable position of generating all these data that they can't really explain.”

If data are produced for safety evaluations, they must be submitted to the FDA by law, Sistare adds. Therefore, companies must decide whether the studies are worth the investment, particularly if the resultant data simply immerses them in a drug approval quagmire. Sistare admits the policy might deter drug and biotechnology companies from undertaking toxicogenomic research on developmental compounds. Referring to this particular issue, Sistare adds that viewpoints both within and outside the FDA suggest that companies pursuing toxicogenomic studies for new drug screening (when potential drugs are selected for further research and development) should not be obligated to submit their experimental data. “Either way,” he says, “we don't want to inhibit application of the technology to drug development. We think it's a powerful tool, and all of us need to come to resolution on when the data need to be submitted, how the data should be submitted, and how we are going to use them.”

Clearly, toxicogenomics is a new field brimming with potential benefits. But many challenges remain. Scientists are just beginning to explore the remarkable complexities of cellular response mechanisms. And assembling the pieces of the toxicogenomics puzzle is a challenge that will require ever more sophisticated technology and decades of research. But the environmental health payoff is significant: more effective diagnosis and treatment of environmentally related diseases, expedited evaluations of chemicals and new drugs, and better risk assessment. “We have to understand how all the pathways fit together in a systems biology perspective,” concludes Samson. “That's the biggest hurdle of all.”

Charles W. Schmidt