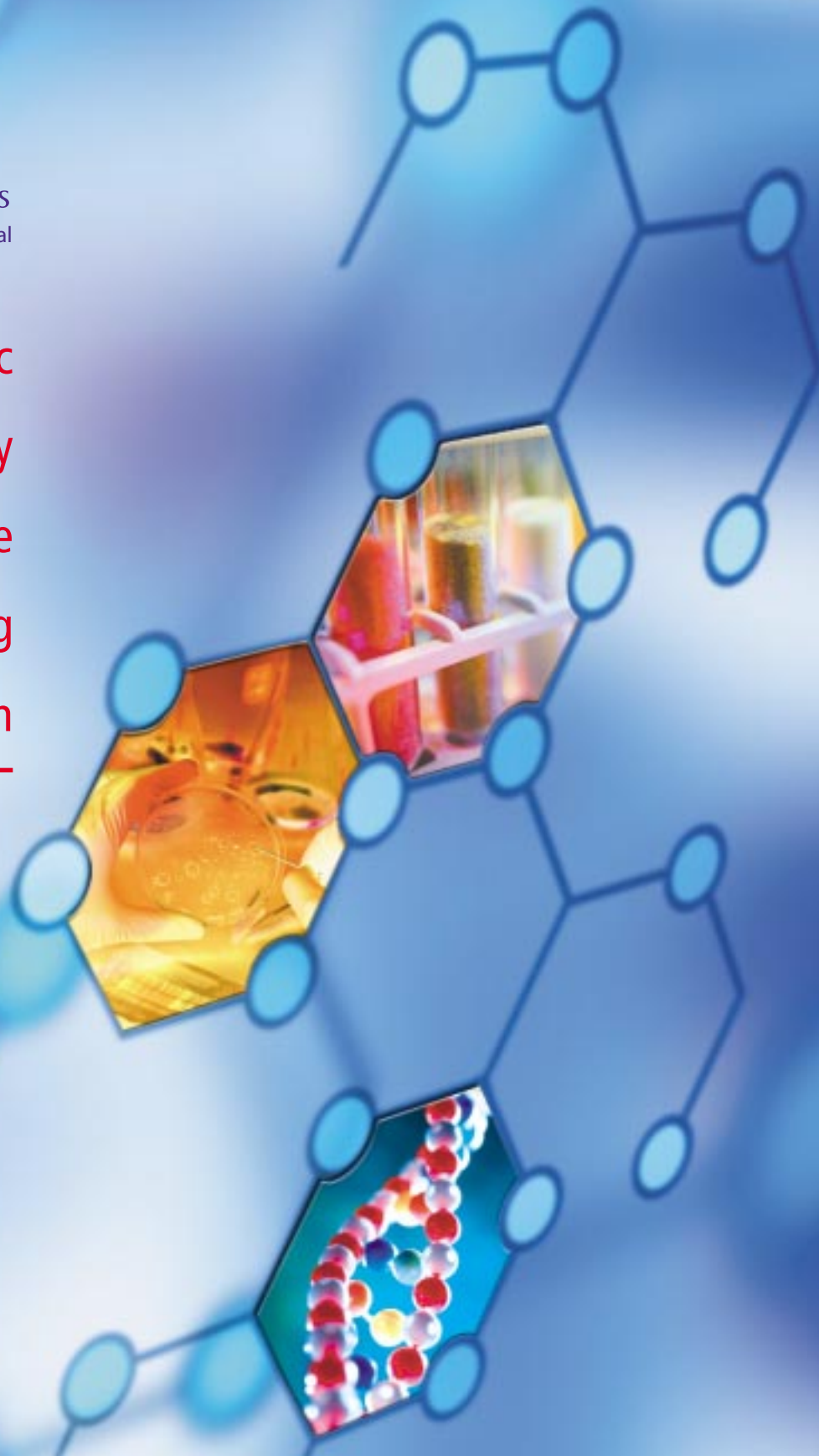
NCT

National Center
for Toxicogenomics

National Institute of Environmental Health Sciences

National Institutes of Health

Using Global Genomic Expression Technology to Create a Knowledge Base for Protecting Human Health



NCT

National Center
for Toxicogenomics

National Institute of Environmental Health Sciences

National Institutes of Health

Using Global Genomic Expression Technology to Create a Knowledge Base for Protecting Human Health

The latter part of the 20th century was marked by a series of rapid advances in genomics-based biotechnology. One such advance was the emergence of the field of toxicogenomics — the collection, interpretation and storage of information about gene and protein activity in order to identify toxic substances in the environment and those populations at the greatest risk of environmental diseases. The National Center for Toxicogenomics (NCT) was created to promote the evolution of gene and protein expression technologies and their use to understand adverse environmental effects on human health. To this end, the NCT will use toxicogenomics methods to measure gene and protein expression, applying the collective effort of many individuals and institutions to develop and populate a knowledge base on Chemical Effects in Biological Systems (CEBS). We envision that the information in the NCT knowledge base will be integrated with other toxicological, genomic and biological information to better understand environmental health risks. Thus, the promise of toxicogenomics is to build an unprecedented body of knowledge that can be used to guide future research, improve environmental health and aid in regulatory decision-making.

The promise of toxicogenomics is no smaller than the challenges with which it confronts us. Toxicogenomics will require the amassing and assimilation of volumes of toxicology data never before attempted. Large numbers of drugs, chemicals and environmental agents must be classified and diverse species with unique biological properties must be catalogued. Highly complex biochemical variation must be described as a function of time, dose and developmental stage. Human disease susceptibilities determined by multiple genetic traits must be defined. All these factors are critical to understanding the consequences of toxicological effects on human health and the environment. In addition, the challenge of managing, analyzing, integrating and storing massive amounts of data demands unprecedented computational resources and novel technological tools. The NCT has amassed powerful resources and is working to meet these challenges in a deliberate, incremental manner.

This brochure provides an overview of the interests and activities of the NCT. Our relatively young program was formally established in September 2000 and has grown tremendously in its first years of existence. Since the program goals are far reaching, we will continue to develop additional mechanisms and partnerships to aid in establishing the CEBS knowledge base. We are eager to collaborate with scientists in academic or private institutions and other governmental organizations. We welcome advice and participation from all individuals who are interested in this enterprise.

Kenneth Olden
Director
National Institute of Environmental Health Sciences (NIEHS)
National Institutes of Health (NIH)

Samuel H. Wilson
Deputy Director
National Institute of Environmental Health Sciences (NIEHS)
National Institutes of Health (NIH)

Raymond W. Tennant
Director
National Center for Toxicogenomics
NIEHS, NIH

Using Global Genomic Expression Technology to Create a Knowledge Base for Protecting Human Health

Contents

Executive Summary

The last two decades of biological research have witnessed dramatic progress toward understanding the molecular basis of life. The technological advancements needed to solve the structure of the human genome have established an infrastructure of new analytical technologies and high throughput methods that have enabled molecular biological databases to grow at rates never seen before. In September 2000 the National Institute of Environmental Health Sciences created the National Center for Toxicogenomics (NCT), whose mission is to coordinate a nationwide research effort for the development of a toxicogenomics knowledge base.

There are five goals of the National Center for Toxicogenomics:

- To facilitate the application of gene and protein expression technology;
- To understand the relationship between environmental exposures and human disease susceptibility;
- To identify useful biomarkers of disease and exposure to toxic substances;
- To improve computational methods for understanding the biological consequences of exposure and responses to exposure; and
- To create a public database of environmental effects of toxic substances in biological systems.

A compendium of gene expression data enhanced by complete proteomic analysis will enable investigators to probe the complexities of the mechanisms of normal genetic and metabolic pathways, and subsequently, to learn how disease occurs when these pathways malfunction. When combined with information on gene/protein groups, functional pathways and networks, and human genetic polymorphisms, these data will confer new knowledge of gene-environment interactions and human health risks.

Introduction

Toxicologists and environmental health scientists have studied the effects of the environment on human health for several decades. Many adverse environmental effects have been identified and important progress has been made in preventing exposure to harmful agents such as γ-radiation, UV-light, lead, pesticides and dioxins. Toxicological research has attempted to develop an efficient, cost-effective, comprehensive strategy for predicting and preventing toxic responses in humans. However, progress toward this goal has been proportionate to the existing technologies and level of scientific knowledge. The field of toxicology could not have risen to this challenge using only the less efficient technologies of the past several decades.

The availability of new technologies has positioned toxicologists to conduct inquiries on an unprecedented global genomic scale. In the post-genomic era, researchers can now use DNA microarrays and proteomics tools to focus their analyses on the basic questions of how normal cells and tissues become diseased. The high sensitivity, rapid throughput technologies that have arisen as part of the Human Genome Project have vastly extended the scope and range of the toxicological sciences and have brought both new opportunities and new challenges to toxicologists and environmental health scientists.

One challenge is to use the human genome sequence as a first step to understand the genetic and biological basis of complex biological traits and diseases such as cancer, diabetes, Alzheimer's disease and Parkinson's disease. Another challenge is to utilize the increased volume of toxicological data to construct genetic and biochemical pathways that will explain the mechanism of toxic responses. Advances in combinatorial chemistry and molecular biology have dramatically accelerated the rate of drug discovery and availability, and the rate at which populations are exposed to new drugs. Such advances intensify the burden of exposure in the population, making it critical that we rapidly increase our understanding of the consequences of such exposure.

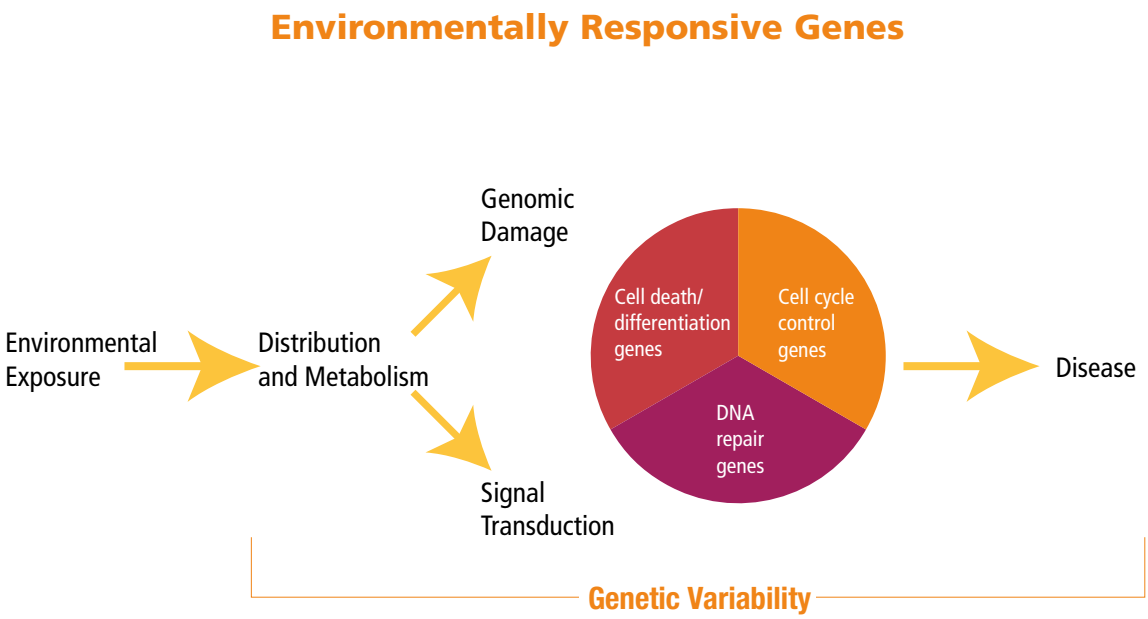
It is now possible for toxicologists to obtain a more fundamental understanding of chemical- and drug-induced disease processes, to develop new ways of evaluating and predicting toxicity, to identify new biomarkers of exposure and toxicity, and to identify individuals who may be at increased risk of adverse effects. These are but a few of the difficult research problems that can now be subject to new inquiry. Obviously, the field is extremely diverse, and reaping the advances in genomics will be much more productive if the constituencies of the pharmaceutical, chemical and consumer product industries, academic scientists, regulatory agencies and Federal research organizations are well coordinated.

The NIEHS is uniquely positioned for the National Center for Toxicogenomics to provide leadership for the development of a unified strategy for toxicogenomics studies, and a public knowledge base with the informatics infrastructure to allow all partners in this

unprecedented enterprise to share equally in its benefits and products. By providing a focus for technological coordination and basic research, a centralized public knowledge base and a center for coordination among all of the partners in the pharmaceutical and chemical industries, the NCT will facilitate this diverse national effort. The NCT will achieve not only economies of time, cost and effort, but will help to ensure the successful development of a broad scientific consensus on the application of toxicogenomics to the improvement of human health.

Moving Beyond Classical Toxicology

The primary goal of toxicological investigations is to assess the effects of environmental exposure on the health status of an exposed animal or human. Toxicologists traditionally measure physical parameters (e.g., body weight, organ weight, blood pressure, activity level), histological features of tissue samples (histopathology), or blood chemistry indicators to identify adverse effects of exposure. Such data are useful diagnostic indicators, but they are less useful for understanding the molecular mechanisms of toxicity or to gain insight into the pathogenesis and progression of disease. In addition, traditional toxicological methods are usually insufficiently sensitive to detect low-level toxicity or early pre-clinical stages of disease. Tools that evaluate toxicity and disease pathology at the mechanistic level are essential in order to understand and prevent adverse environmental exposure.



Environmental toxicogenomics is a new approach to environmental toxicology. Environmental toxicogenomics allows us to identify and characterize genomic signatures of environmental toxicants as gene and protein expression profiles. A major application of gene expression profiling is to understand human genetic variability and susceptibility to disease.

Discovery vs. Hypothesis-Driven Science

Many scientific projects are initiated to test a hypothesis or a model that arises during analysis of previous experimental data. Predictions are made based on the hypothesis, and experiments are carried out that can produce evidence to support or refute the hypothesis. While this approach is valid and important, it can preclude or inhibit discovery of variables, factors and mechanisms that are unknown to the researchers who design the study. An alternate approach is to nonselectively gather information about a particular biological function or system; the results are then analyzed with the hope that significant characteristics will emerge during the analysis to provide insight into the mechanism and function of the system. This approach is known as discovery science, and it contrasts significantly with hypothesis-driven science. Many pharmacogenomics experiments are fundamentally discovery science aimed at defining the pharmacological properties of a new drug.

It is clear that discovery science and hypothesis-driven science are complementary approaches that can advance the field of toxicology. Discovery science is a high-throughput method that can rapidly screen many genes for potential involvement in a biological or toxicological process. Hypothesis-driven science is, in contrast, a low-throughput method. Toxicogenomics, as practiced by the NCT, will employ both approaches. Global expression profiling using microarray technology or proteomics can be used to generate specific hypotheses on the mechanism of a drug or toxicant as well as to rapidly survey known genes that may be involved in that or other mechanisms. Furthermore, hypotheses may be novel and may involve previously uncharacterized genes. Toxicogenomics methods thus may help formulate novel hypotheses that may lead to important breakthroughs in toxicology and environmental health.

Program Goals

In September 2000 the National Institute of Environmental Health Sciences (NIEHS) created the National Center for Toxicogenomics (NCT) to foster development in this burgeoning field. The mission of The National Center for Toxicogenomics (NCT) is to coordinate a nationwide research effort that will facilitate application of the new genomics and proteomics sciences to the study of the biological response to environmental agents and stressors and prevention of environmental diseases. The NCT will serve as the focus for research efforts and as a coordinator of national and international research through the development of a toxicogenomics knowledge base on Chemical Effects in Biological Systems (CEBS). This Center will seek to promote new understanding of the mechanisms of biological responses to environmental stressors, including toxic injury, and to identify biomarkers of exposure and disease that can be used to improve and protect human health.

The NCT aims to use and promote toxicogenomics as a means to guide federal agencies and legislators in developing guidelines and laws that regulate the levels of various chemicals in the environment. This knowledge will reduce the likelihood of needless and expensive over-regulation as well as potentially dangerous under-regulation of environmental toxicants, as new guidelines will be based on defensible scientific standards and information rather than on estimates and educated opinions. The NCT is currently doing the groundbreaking research needed to develop and demonstrate the techniques that will become tomorrow's routine testing methods in toxicogenomics.

In the near future, toxicogenomics will likely have a significant impact on three key areas of human health: risk assessment, exposure assessment, and understanding human susceptibility to disease. Toxicogenomics and initiatives such as the NCT bring the promise of understanding complete biochemical mechanisms of toxicity. Many challenges remain to be solved before the potential of toxicogenomics is fulfilled. For example, it is critical that toxicogenomics technologies and other approaches to determine an individual's or group's susceptibility be implemented in an ethically responsible manner that preserves and protects the legal and social concerns of all individuals. Toxicogenomics has the potential to increase dramatically our understanding of the complex biochemistries that can lead to toxicological hazards and to uncover potentially dangerous situations that can be addressed with preventive strategies. This knowledge will enhance our ability to improve and protect human health. As toxicogenomics matures in the beginning of the 21st century, the NCT will be at the forefront of this important research frontier.


1
Facilitate the application of gene and protein expression technology.
2
Understand the relationship between environmental exposures and human disease susceptibility.
3
Identify useful biomarkers of disease and exposure to toxic substances.
4
Improve computational methods for understanding the biological consequences of exposure and responses to exposure.
5
Create a public database of environmental effects of toxic substances in biological systems.
The Five Goals of the National Center for Toxicogenomics (NCT)

Technologies

DNA microarray technology enables the simultaneous measurement of transcription of thousands of genes, using microchips containing thousands of probes of complementary DNA (cDNA) immobilized in a predefined array. The development of this revolutionary technology requires the capacity to synthesize and print large numbers of DNA probes that correspond to thousands of different genes; it also depends on the availability of large amounts of DNA sequence information in publicly accessible databases. Researchers are further developing microarray technology to study disease endpoints, toxic responses, drug-responsiveness, mechanisms of action, regulatory networks and more.

Proteomics, the global analysis of protein expression, is a technology that developed in parallel with microarray-based gene expression profiling. The emergence of proteomics depends on the development of highly sensitive and powerful mass spectrometry methods as well as the availability of publicly accessible protein sequence databases. The use of mass spectrometry has enabled the generation of highly accurate measurements of the molecular masses of peptide fingerprints as well as amino acid sequence information for peptide fragments from proteins. Gene expression provides the template for protein molecules that are responsible for functional activities within the cell. It is critically important that we understand the normal gene sequence as well as translated protein structures and levels of activity. Using protein sequence databases, a protein "fingerprint" can unequivocally identify a protein species, even when it is present in small quantities in a complex mixture.

Together, **Microarray** and **Proteomics** technologies have made it possible to examine the biological effects of a host of distinct toxicants and to observe multiple biological endpoints with efficient high-throughput methods (i.e., computational systems with high output). These combined technologies have led to the emergence of the new field of toxicogenomics.

Toxicogenomics applies global expression profiling, including microarray and proteomics, to study the relationship between exposure and disease and to understand gene-environment interactions and their impact on human health. Toxicogenomics thus seeks to create a comprehensive knowledge base for a comprehensive understanding of environmental effects in complex biological systems.

The Chemical Effects in Biological Systems (CEBS) Knowledge Base will support research that promotes mechanistic understanding of environmentally induced toxicity and disease. Such mechanistic knowledge will make predictive toxicology possible and improve exposure assessment and risk assessment. The ultimate goal of the NCT, therefore, is to create a knowledge base that allows environmental health scientists and practitioners to understand and prevent adverse environmental exposure in the 21st century.

The NCT is a multilevel, multicomponent program that encompasses resources in the NIEHS intramural and extramural communities, including academic and private sector organizations. The NCT is also involved in relationships and partnerships with other federal research and regulatory organizations. The combined assets of these resources will be leveraged to achieve the goals of the NCT.

The Main Components of the NCT

Intramural Programs

- NIEHS Microaray Center (NMC)
- NCT "Tox/Path Team"
- NIEHS Proteomics Program
- NIEHS intramural laboratories

Extramural Programs

- Toxicogenomics Research Consortium (TRC)
- DERT Working Group
- Proteomics

Shared Programs

- Microarray and proteomics resource contracts
- Database resource contract
- Intramural and extramural grantees and other research groups which interact with the NCT through partnerships, cooperative research and development agreements (CRADAs), and other mechanisms

New NCT components which may be added include:

- Proteomics Research Consortium
- Additional database resources

NIEHS Microarray Center Staff Members

Intramural Programs

The NIEHS Microarray Center

The NMC supports the NCT mission by developing and applying microarray technology to study environmental effects on human health. Microarray studies are carried out primarily in collaboration with NIEHS intramural labs.

The goals of the NMC are:

- To identify toxicant-specific patterns of gene expression;
- To elucidate molecular mechanisms of environmental agents;
- To develop gene expression-based biomarkers of human exposure;
- To evaluate how well environmental effects of toxicants extrapolate from one species to another;
- To study the toxicological effects of chemical mixtures;
- To evaluate dose response curves and high or low dose-specific toxicological responses; and
- To develop a public database of microarray expression profiles.

The NMC operates a state-of-the-art microarray facility including the following equipment and technology:

- A high speed, XYZ robotic arrayer that can print up to 96 slides at one time by transferring cDNA clones from 96- or 384-microcell plates. The arrayer is equipped with a pre-spotting station, plate stacker, pin wash-dry station, and temperature and air quality controls. High-density high quality arrays can be printed with up to 25,000 elements.
- Three microarray scanners, two GenePix 4000A/B scanners and the Agilent scanner. These scanners are customized to read fluorescent signals from the microarray chips and transfer the data directly into a computer for analysis.

Tox/Path Team Staff Members

NCT Tox/Path Team

The NCT Tox/Path Team is composed of NCT personnel who support the NCT mission by designing and performing NCT research activities within the NIEHS Division of Intramural Research. The Tox/Path team also is developing proteomics methods and applications.

The goals of the Tox/Path Team are to:

- Design and conduct mechanistic toxicology studies;
- Define gene expression signatures of toxicological endpoints and environmentally-induced disease;
- Define surrogate biomarkers for use in predictive hazard assessment and risk assessment; and
- Evaluate the dose-, time- and tissue-dependence of gene expression profiles.

Structure

Proteomics Program Staff Members

Proteomics Program

The NCT is developing proteomics research capabilities by collaborating with the NIEHS Proteomics Group, a program within the NIEHS Division of Intramural Research, and by implementing extramural proteomics activities.

The mission of the NCT Proteomics Program is to use proteomics technology to identify key proteins and pathways in response to environmental toxicants and stressors.

The goals of the Proteomics Program are to:

- Analyze biochemical pathways that contribute to the response to environmental agents;
- Study protein-protein interactions and post-translational events that are affected by environmental exposure;
- Promote technical innovations and improvements in proteomics technology; and
- Develop protein biomarkers of environmental exposure.

The Proteomics Group is applying mass spectrometry-based techniques to protein identification, differential quantitation, and identification of post-translational modifications resulting from exposure to environmental toxicants and stressors.

Extramural Programs

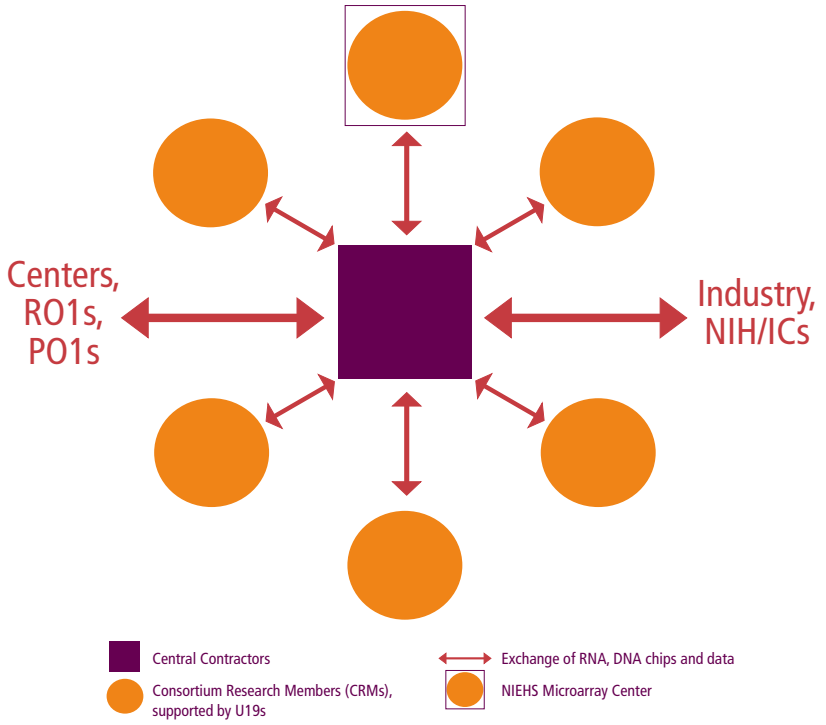
DERT Toxicogenomics Research Consortium (TRC)

Division of Extramural Research and Training (DERT) Working Group Members

The TRC is the major extramural mechanism by which the NCT is currently applying microarray technology. The TRC was formally established in November 2001 when the NIEHS Division of Extramural Research and Training (DERT) awarded grants totalling $37 million over five years to five institutions to participate in the TRC as Cooperative Research Members (CRMs). Five institutions and the NIEHS Intramural Microarray Center (NMC) form the core of this Consortium. The Cooperative Research Members (CRMs) provide specialized expertise in gene expression profiling, bioinformatics and proteomics and will conduct both cooperative and independent research on different aspects of toxicogenomics.

Structure

Toxicogenomics Research Consortium (TRC)



The initial goal of the TRC is to conduct a series of cooperative gene expression experiments using shared and complementary microarray platforms within the CRMs. The collaborative experiments will be used to develop standard operating procedures and quality control standards for gene expression experiments and to develop technology standards and bioinformatics tools for data comparison across the CRMs. This will be a unique contribution to the field of toxicogenomics that can best be achieved through cooperative efforts of the Consortium members. This will also lay the foundation for additional toxicogy experiments conducted by the TRC.

The TRC also interacts with other academic, industry and government research entities. The TRC member institutions (CRMs) and their areas of expertise are:

University of North Carolina at Chapel Hill

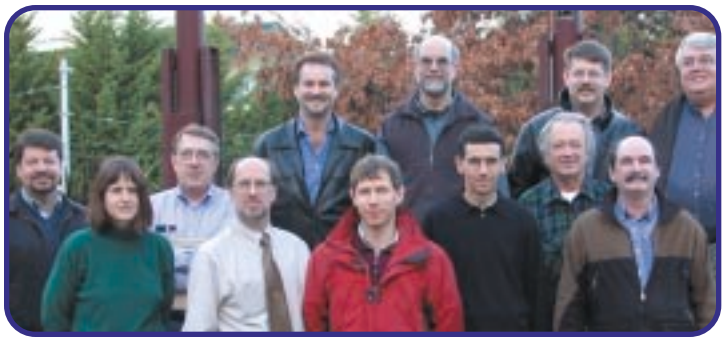
(Director, William Kaufmann)

Expertise: susceptibility factors in the genomic response to toxicants

Fred Hutchinson Cancer Research Center
Seattle, Washington

(Director, Helmut Zarbl)

Expertise: gene expression profiling in transgenic rodents and human cell lines exposed to environmental toxicants



Oregon Health and Science University
Portland, Oregon

(Director, Peter Spencer)

Expertise: cell-specific injury in the central nervous system and mechanisms of neurotoxicants

Duke University, Durham, North Carolina

(Director, David Schwartz)

Expertise: gene expression profiling to explore environmental stresses on human health

Academic Members of the TRC

Massachusetts Institute of Technology
Cambridge, Massachusetts

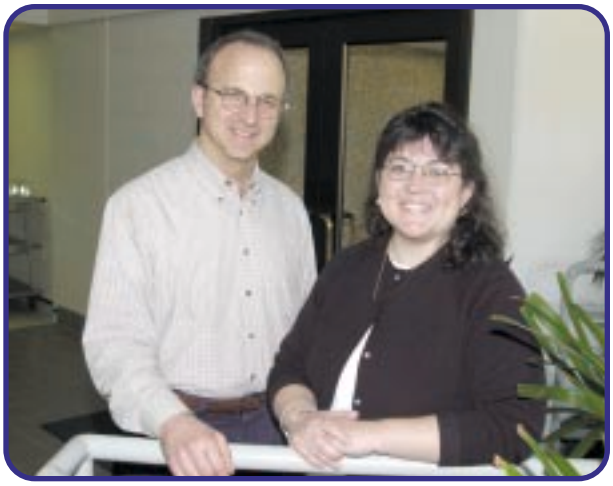
(Director, Leona Samson)

Expertise: effects of environmental alkylating agents on gene expression and human health

The NIEHS Microarray Center

(Director, Richard Paules and former Co-Director, Cynthia Afshari)

The NMC supports the NCT mission by developing and applying microarray technology to study environmental effects on human health.

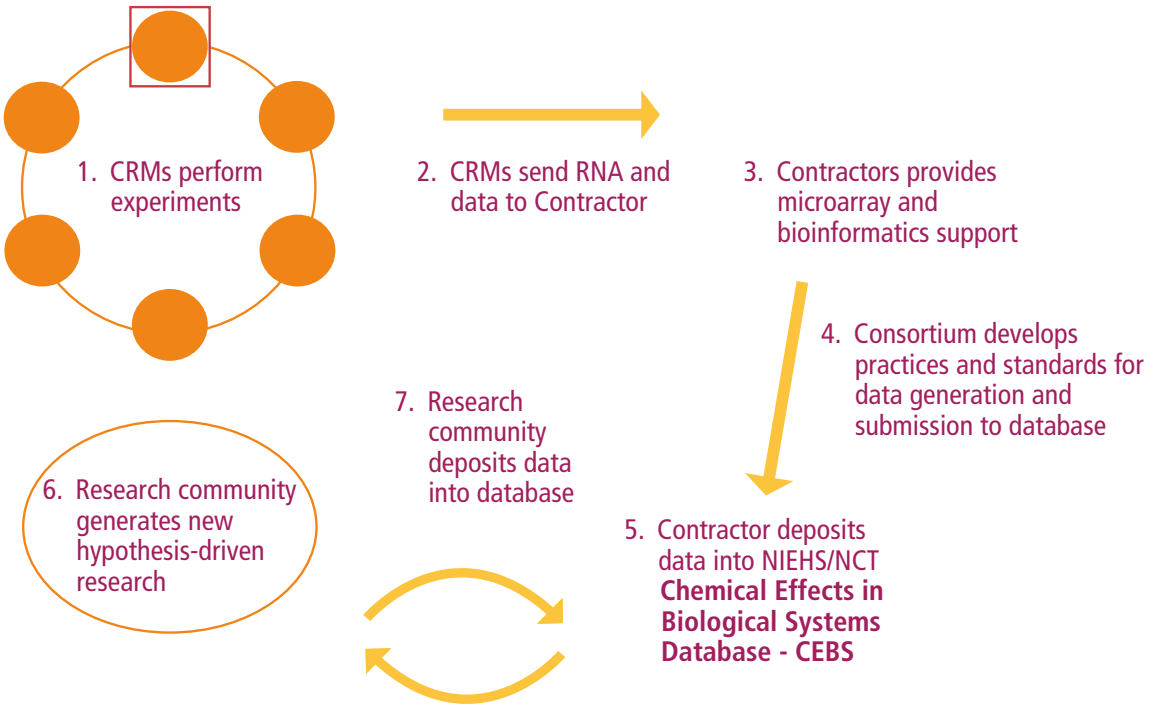


The Goals of the TRC

- Identify and characterize sources of variation in gene expression experiments and establish standard protocols ("best practices") for gene expression experiments and bioinformatics standards and tools;
- Evaluate toxicant-specific patterns of gene expression and define trans-species comparisons for responses to toxicant-induced stress;
- Integrate gene expression profiling data with other sources of data on proteomics, metabonomics and phenotypic anchoring;
- Elucidate molecular mechanisms of cellular responses to environmental agents;
- Develop gene expression-based biomarkers of human exposure;
- Study the toxicological effects of chemical mixtures;
- Evaluate toxicant dose-response behavior and high/low dose-specific responses; and
- Contribute gene expression and proteomics data to the CEBS.

Toxicogenomics Research Consortium (TRC) and Resource Contracts

• Information Flow •



DERT Working Group

The DERT Working Group was established as a forum for the exchange of information and integration of DERT programs related to "-omics" research. Members of the DERT Working Group include program representatives from the TRC; the Comparative Mouse Genomics Centers Consortium (CMGCC); Proteomics, Metabonomics, and Small Business Innovation Research programs; as well as grants management.

Proteomics

The DERT has initiated programs in proteomics that will support the NCT. A recent Request for Applications (RFA) was announced to provide funding in basic research related to proteomics.

The Goals of the RFA Solicitation

- To encourage the use and development of innovative proteomic technologies to study environmentally-responsive protein networks;
- To study the relationship between changes in gene expression as a consequence of exposure and protein abundance using in vivo and in vitro models;
- To address the dynamics of protein networks with the cell by hypothesis-driven research that focuses on understanding the expression, modifications and interactions of proteins on a global scale; and
- To provide bioinformatics tools to analyze data from proteomics research.

Proteomics Research Consortium

The NIEHS is currently using proteomics to complement gene expression microarray studies and is working to develop this technology in collaboration with intramural and extramural collaborators. Independent research initiatives in proteomics are anticipated in the future, including a Proteomics Research Consortium.

Ethical, Legal and Social Implications (ELSI)

The completion of the human genome sequence and the emergence of toxicogenomics have many implications of an ethical, legal and social nature. The ELSI research community is just beginning to attempt to address and resolve the questions raised by recent technological developments. It is apparent that ELSI issues could play a major role in determining the nature of the impact of toxicogenomics on society in the next several years.

Many toxicogenomics studies focus on understanding inherited and acquired susceptibility to disease. Individual susceptibility is determined by many variables including current exposure, past exposure, health status and genetic traits. Such variables are often unknown or difficult to measure with precision. It is expected that toxicogenomics will provide new and much more precise information about gene/environment interactions and other factors that cause disease susceptibility. In addition, toxicogenomics will undoubtedly be used to identify individuals or population subgroups that carry specific disease susceptibility alleles. Public health policy will face the challenge of learning how to use and prevent misuse of this information.

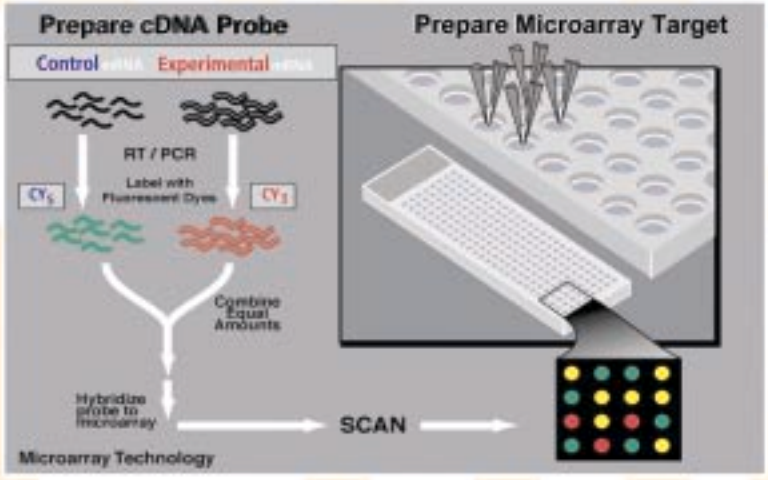
Bioethicists can and should play an important role in resolving ELSI issues that arise as toxicogenomics methods are used more widely and begin to impact the general public. In some areas, bioethicists have developed guidelines for ethical human genetic research. For example, in studies involving human subjects, the following questions must be asked:

- What genetic research should be done on which population?
- What do individuals need to know before choosing to participate as research subjects?
- How should the privacy of participants be protected?
- When and in what manner should research findings be disclosed to participants?

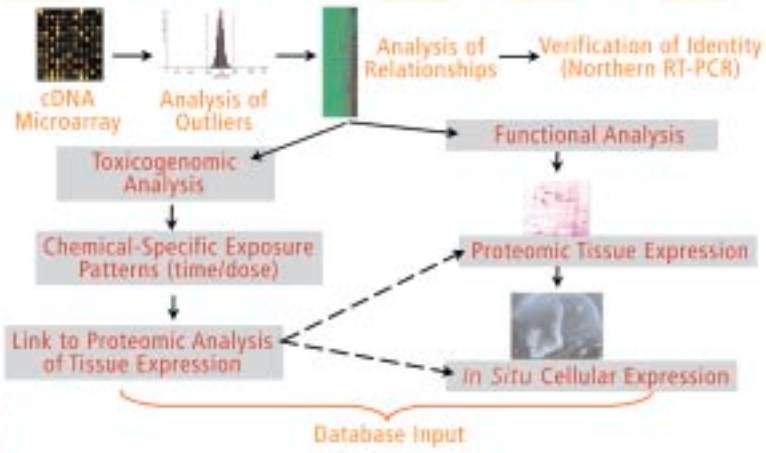
The NCT includes bioethicists on its staff and promotes research and education in ELSI. The NCT has established ELSI as a high priority in its program and is firmly committed to applying toxicogenomics methods in an ethically responsible manner.

The Goals of the NCT

Goal 1
Facilitate the application of gene and protein expression technology

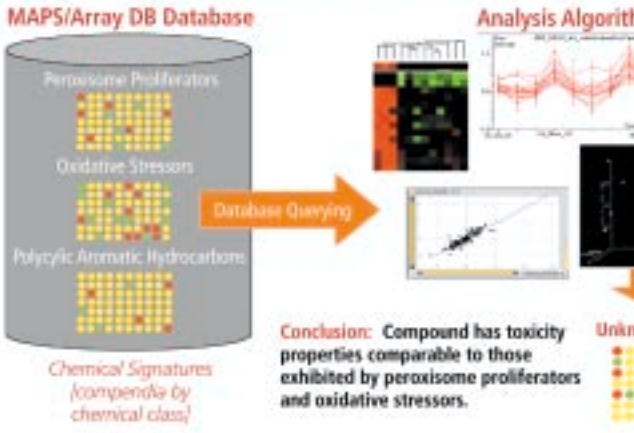


Integration of Gene/Protein Expression Data

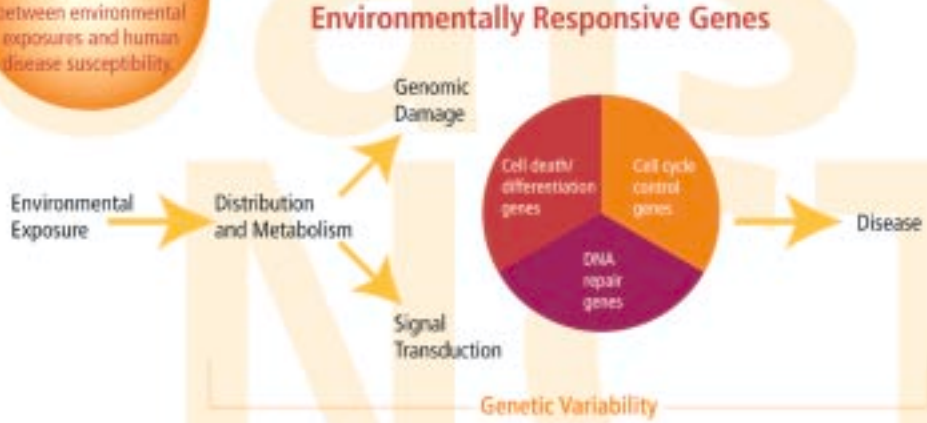


Goal 5
Create a public database of environmental effects of toxic substances in biological systems.

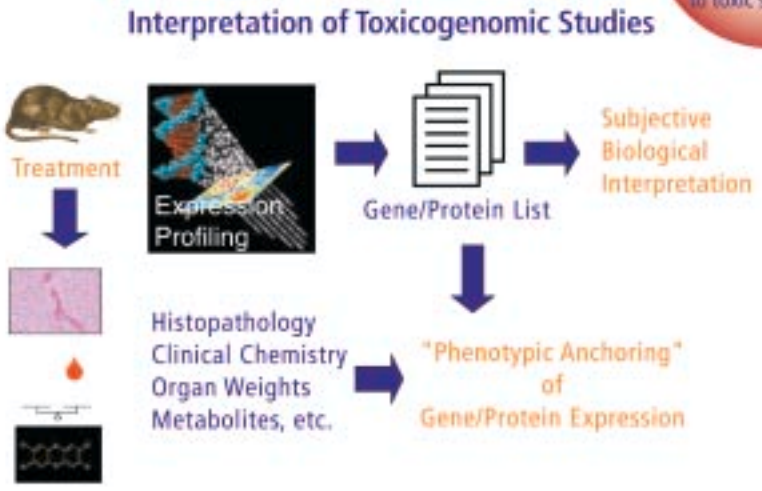
Database-Enabled Prediction M




Goal 2
Understand the relationship between environmental exposures and human disease susceptibility.


Environmentally Responsive Genes
Genomic Damage
Environmental Exposure
Distribution and Metabolism
Cell death/ differentiation genes
Cell cycle control genes
DNA repair genes
Disease
Signal Transduction
Genetic Variability


Goal 3
Identify useful biomarkers of disease and exposure to toxic substances.


Interpretation of Toxicogenomic Studies
Treatment
Expression Profiling
Gene/Protein List
Subjective Biological Interpretation
Histopathology
Clinical Chemistry
Organ Weights
Metabolites, etc.
"Phenotypic Anchoring" of Gene/Protein Expression


odel
nms
own Toxicant


Goal 4
Improve computational methods for understanding the biological consequences of exposure and responses to exposure.

The Goals of the NCT

Global expression profiling by microarray and proteomics methods provides the means to analyze many genes and proteins simultaneously. Standard methods for measuring gene and protein expression, including the Northern blot and the Western blot, are labor-intensive and can be used only for a limited number of samples. The new technologies enable the comparison of the expression profiles associated with many toxicants, tissues or cell types. Because microarray chips can assess the expression of uncharacterized genes and expressed sequence tags, it is also possible to identify novel genes that play roles in toxicity or disease.

The power and potential of these new toxicogenomics methods are revolutionizing the field of toxicology and environmental health. Studies of gene expression can provide mechanistic insight into toxic responses and disease pathologies. Changes in gene expression as a result of chemical exposure can indicate activation of a signal transduction pathway that may be a significant step in disrupting a metabolic process. Such information can be combined with other biological and chemical information to explain the mechanism underlying a toxic response. It is anticipated that our understanding of mechanisms of toxicity will improve as these new high-throughput methods are used more extensively and as the database of toxicogenomics information is developed more fully.

Initial studies indicate that classes of toxicants and toxic responses can be recognized as gene expression signatures using microarray technology. The NCT is performing experiments to validate the concept of gene expression profiles as "signatures" of toxicant classes, disease subtypes or other biological endpoints. Such experiments will correlate gene expression profiles with other well-defined parameters, including toxicant class, chemical structure, pathological or physiological response, or other validated indices of toxicity. For example, experiments can be designed to correlate gene expression patterns with liver pathologies such as hepatomegaly (enlarged liver), hepatocellular necrosis (death of liver cells or tissues) or inflammation. It is also possible to look for correlative patterns -- for example in enzyme levels -- in liver and other tissues or cells such as blood. Changes in serum enzymes provide diagnostic markers of organ function that are commonly used in medicine and in toxicology. Such "phenotypic anchoring" of gene expression data using conventional indices will distinguish the toxicological signal from other gene expression changes that may be unrelated to toxicity, such as the pharmacological or therapeutic effects of a compound.

A large body of data is needed to develop the capacity to interpret and define toxicological gene expression signatures. Once sufficient data have been accumulated

and assimilated, the predictive value of the database can be tested. Ultimately, the NCT will develop the capacity to use gene expression signatures to facilitate toxicological characterization of toxicants and their biological effects. The analysis of these expression profiles for different chemicals from different classes over dose and time is used to identify genes and protein expression profiles that are consistently and mechanistically linked to specific exposures and disease outcomes. Further, these approaches will eventually have a strong impact on hazard assessment, risk assessment and their ability to protect human health.

The Goals of the NCT

Rodents have been the models of choice for toxicology studies in the past. Rodent model systems have biochemical processes that closely resemble those of humans. Such models have been used to generate an important set of toxicological data that has been essential in establishing safe levels of exposure in the human population. Human epidemiological data often are used in combination with toxicological data from rodent models to guide regulatory decisions. However, many questions remain about how to extrapolate from the rodent dose-response curve in order to predict the equivalent dose in humans, especially at low doses. Surrogate measures made in model systems in vitro or in tissues (such as blood) apart from those normally observed for toxic endpoints can be very useful in resolving such questions. Thus, there is keen interest in using toxicogenomics methods as a means to compare and validate dose-dependent effects in animal and human cells or in expendable tissues.

Field and clinical research applications of toxicogenomics methods are anticipated. It is well known that a Single Nucleotide Polymorphism (SNP), a single base change in the message of a gene, can cause a protein to malfunction. Preliminary data indicate that gene expression profiles will be useful as diagnostic tools for identifying early stages of cancer or other pathologies. If this approach enables earlier detection of disease than is currently possible through other approaches, it could allow earlier initiation of therapeutic interventions. Additionally, gene expression profiling may become an important tool for predicting therapeutic outcome, and may be particularly useful in addressing the significant variability that has been observed in how well patients respond to different types of drug therapy. Such patterns of variability have been studied using expression profiling; in some cases, expression signatures have been associated with individuals who are responders or nonresponders for a particular type of drug therapy. Once this kind of result is validated, it may be possible to use expression profiling to optimize the therapeutic regimen for individual patients, thus increasing the chance of a good treatment outcome.

The Goals of the NCT

Biomarkers of disease or toxicity are one of the most important and powerful tools that can be used to achieve the NCT's ultimate goal of improving environmental health and reducing the incidence and impact of environmentally related diseases. A biomarker is a molecular indicator of a specific biological property and can be used as an early warning of pathology to follow. Biomarkers can be indicators of exposure, effect or susceptibility. Each potential biomarker must be carefully validated and tested for specificity, sensitivity, reliability and predictive value. Microarray and proteomics methods can be used to focus attention on a specific diagnostic transcript or protein that has potential for development as a biomarker of disease or toxicity. Diagnostic biomarkers may be identified and developed from microarray and proteomics data, which could also accelerate disease detection and facilitate timely initiation of therapy. Subsequent testing and validation studies are then required to determine if the biomarker is useful. While the NCT considers biomarker development an important part of its mission, it is likely that biomarker development will proceed relative to the volume of toxicogenomics data accumulation in the NCT databases. Ultimately, it is anticipated that biomarkers will be recognized as key factors in a sequence of events that will help to define the way in which a specific chemical or environmental exposure causes disease. In other words, toxicogenomics may help to delineate the mode of action of agents as an important step in assessing potential risk of human exposure.

As the field of toxicogenomics evolves, toxicogenomics databases will support predictive toxicology and hazard assessment. This will help scientists predict the toxicologic impact of suspected toxicants and calculate how much of a hazard these toxicants actually represent to human and environmental health. As phenotypic anchoring is developed and expression profiles are indexed and compared in order to discern diagnostic "signatures," it will become increasingly possible to characterize an unknown biological or physical sample by comparing its gene expression profile to profiles in the database. However, this predictive capability will be realized only after the NCT databases mature and are populated with many datasets.

Toxicogenomics will support new basic research initiatives in computational and mechanistic toxicology and environmental health. Many of these initiatives may not have been possible without toxicogenomics technology or insights gained from toxicogenomics data. Toxicogenomics will provide the means to explore toxicological

mechanisms at a new level of intensity and detail by investigating dose response, kinetics, species and tissue specificity, genetic interactions and polymorphism, and multiple pathways that contribute to toxicity or pathology. Basic research into mechanisms of toxicity and disease may also facilitate and streamline the process of drug and chemical development. Computational methods undoubtedly will be used in conjunction with toxicogenomics data to model critical networks within the cell that enable it to respond to environmental stressors, toxicants and drugs.

The NCT is currently developing the initial version of the Chemical Effects in Biological Systems (CEBS) database. The TRC and other NCT components are generating large amounts of microarray data, and these data will be analyzed, annotated, and stored within the CEBS database. As query algorithms are refined and proteomics develops further with the expansion of NCT programs, the CEBS database will evolve into the CEBS knowledge base. The CEBS knowledge base will gradually become a larger component of NCT that assimilates data from the TRC, other NCT activities, NIEHS programs such as the Environmental Genome Project, and other NIH programs.

The goals of the CEBS knowledge base are to:

- Create a reference toxicogenomic information system of studies on environmental chemicals/stressors and their effects.
- Develop relational and descriptive compendia on toxicologically important genes, groups of genes, SNPs, mutants, and their functional phenotypes that are relevant to human health and environmental disease.
- Create a toxicogenomics knowledge base to support hypothesis-driven research.

CEBS Knowledge Technology

The NCT is working to help the field of environmental health research evolve into a knowledge-based science in which experimental data are compiled. Computational and informatics tools will play a significant role in improving our understanding of toxicant-related disease. CEBS will be created as a high quality, publicly accessible relational database that is compatible with standard laboratory output platforms. Database development will be integrated with strategic toxicogenomics experimental design and conduct. Standardized procedures, protocols, data formats, and assessment methods will be used to ensure that data meet a uniform high level of quality. Raw data sets from NCT experiments will be available in their entirety. Relational and descriptive compendia will be included on toxicologically important genes, groups of

The Goals of the NCT

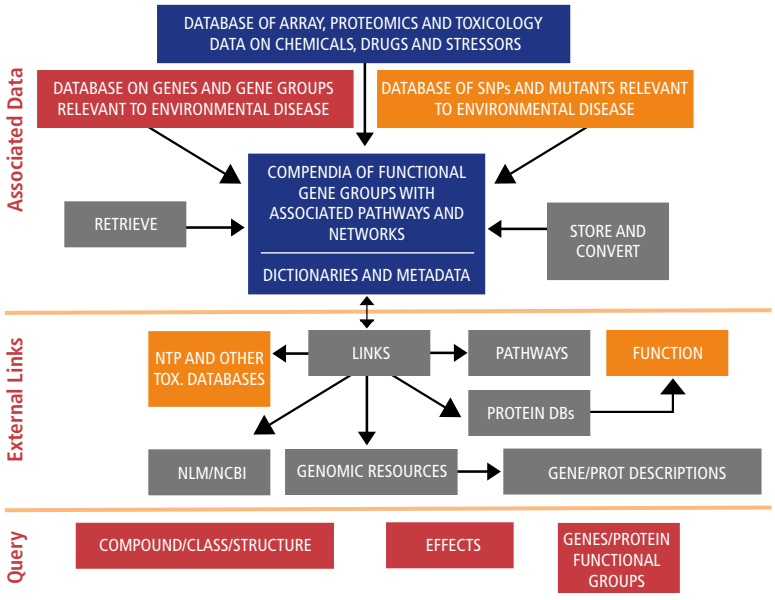
genes, SNPs, and mutants and their functional phenotypes. Information about the biological effects of chemicals and other agents and their mechanism of action will be collected from the literature and stored. CEBS will be fully searchable by compound, structure, toxicity, pathology, gene, gene group, SNP, pathway, and network. Dictionaries and explanatory text will guide researchers in understanding toxicogenomics datasets. CEBS will be linked extensively to other databases and to Web genomics and proteomics resources, providing users the suite of information and tools needed to fully interpret toxicogenomics data.

Future promises of CEBS include:

- Developing large context-annotated datasets that allow precise definition of biological/toxicological pathways and lead to the identification of new biomarkers.
- Linking genomic sequence to expression data to determine those genes that may be responsible for the coordinated regulation of sets of genes.
- Increasing the interpretability and dimensionality of expression data by including data from new types of arrays including protein arrays and SNP arrays.
- Aiding in development of new algorithms and computational tools that allow predictive modeling of gene interactions and networks.

The CEBS knowledge base will support research that promotes mechanistic understanding of environmentally induced toxicity and disease. Such mechanistic knowledge will make predictive toxicology possible and improve exposure assessment and risk assessment. The ultimate goal of the NCT, therefore, is to create a knowledge base that allows environmental health scientists and practitioners to understand and prevent adverse environmental exposure in the 21st century.

CEBS Vision - Bioinformatics to Knowledge



Glossary of Terms

Allele: An alternative form of a gene or any other segment of a chromosome.

Bioinformatics: The analysis of biological information using computers and statistical techniques; the science of developing and utilizing computer databases and algorithms to accelerate and enhance biological research.

Biomarker: A molecular indicator of a specific biological property; a biochemical feature or facet that can be used to measure the progress of disease or the effects of treatment.

Complementary DNA (cDNA): DNA made from a messenger RNA (mRNA) template. The single-stranded form of cDNA is often used as a probe in physical mapping.

DNA (deoxyribonucleic acid): One of two types of molecules that encode genetic information. (The other is RNA. In humans DNA is the genetic material; RNA is transcribed from it. In some other organisms, RNA is the genetic material and, in reverse fashion, the DNA is transcribed from it.)

Expressed sequence tag: A unique stretch of DNA within a coding region of a gene that is useful for identifying full-length genes and serves as a landmark for mapping.

Gene: The basic biological unit of heredity; a segment of deoxyribonucleic acid (DNA) needed to contribute to a function.

Genome: All of the genetic information or hereditary material possessed by an organism; the entire genetic complement of an organism.

Genomics: The study of genes.

Genotype: The genetic composition of an organism or a group of organisms; a group or class of organisms having the same genetic constitution.

In vitro: Literally, "in glass," i.e., in a test tube or in the laboratory; the opposite of in vivo (in a living organism).

In vivo: In a living organism, as opposed to in vitro (in the laboratory).

Knockout: Inactivation of specific genes. Knockouts are often created in laboratory organisms such as yeast or mice so that scientists can study the knockout organism as a model for a particular disease.

Mapping: Charting the location of genes on chromosomes.

Appendix

Mass spectrometry: A method used to determine the masses of atoms or molecules in which an electrical charge is placed on the molecule and the resulting ions are separated by their mass to charge ratio.

Messenger RNA (mRNA): A type of RNA that reflects the exact nucleoside sequence of the genetically active DNA. mRNA carries the "message" of the DNA to the cytoplasm of cells where protein is made in amino acid sequences specified by the mRNA.

Metabonomics: The evaluation of tissues and biological fluids for changes in metabolite levels that result from toxicant-induced exposure.

Microarray: A tool used to sift through and analyze the information contained within a genome. A microarray consists of different nucleic acid probes that are chemically attached to a substrate, which can be a microchip, a glass slide or a microsphere-sized bead.

Northern blot: A technique used to separate and identify pieces of RNA.

Nucleotide: A subunit of DNA or RNA. To form a DNA or RNA molecule, thousands of nucleotides are joined in a long chain.

Phenotype: The observable physical or biochemical traits of an organism, as determined by genetics and the environment; the expression of a given trait based on phenotype; an individual or group of organisms with a particular phenotype.

Polymorphism: The quality or character of occurring in several different forms.

Proteome: All of the proteins produced by a given species, just as the genome is the totality of the genetic information possessed by that species.

Proteomics: The study of the proteome.

RNA (ribonucleic acid): A nucleic acid molecule similar to DNA but containing ribose rather than deoxyribose.

Signal transduction pathway: The course by which a signal from outside a cell is converted to a functional change within the cell.

Single nucleotide polymorphism (SNP): A change in which a single base in the DNA differs from the usual base at that position.

Throughput: Output or production, as of a computer program, over a period of time.

Toxicology: The study of the nature, effects and detection of poisons and the treatment of poisoning.

Toxicogenomics: The collection, interpretation, and storage of information about gene and protein activity in order to identify toxic substances in the environment, and to help treat people at the greatest risk of diseases caused by environmental pollutants or toxicants.

Transgenic: Having genetic material (DNA) from another species. This term can be applied to an organism that has genes from another organism.

Western blot: A technique used to separate and identify proteins

Appendix

The Organization

Organization

The NCT is an integrated Intramural and Extramural research program of the National Institute of Environmental Health Sciences. Advisory oversight is provided by the NCT Scientific Advisory Group, the NIEHS Board of Scientific Counselors and the National Advisory Environmental Health Sciences Council. The NCT Scientific Advisory Group is composed of academic, private sector and government (regulatory and research) scientists with experience and interest in gene expression technology, environmental health and toxicology.

NCT Scientific Advisory Group

J. Michael Bishop, M.D.
Chancellor
University of California San Francisco
San Francisco, California

Philip Leder, M.D.
Chair, Department of Genetics
Harvard Medical School
Boston, Massachusetts

Leroy E. Hood, M.D., Ph.D.
Director and President
Institute for Systems Biology
Seattle, Washington

Gerald N. Wogan, Ph.D.
Professor of Chemistry
Division of Bioengineering and Environmental Health
Massachusetts Institute of Technology
Cambridge, Massachusetts

The NCT Director is Dr. Raymond W. Tennant, NIEHS. The NCT management is provided by the NCT Director, NCT Deputy Director, Dr. James Selkirk, and the DERT Coordinator, Dr. Benjamin van Houten. Dr. Michael Waters is NCT Assistant Director for Database Development.

For more information about the

National Center for Toxicogenomics

Please visit the NCT website at
http://www.niehs.nih.gov/nct

or send email to nct@niehs.nih.gov