September 5, 2000

The National Center for Toxicogenomics

Table of Contents

Preface

Toxicogenomics is a new scientific field that elucidates how the entire genome is involved in biological responses of organisms exposed to environmental toxicants/stressors. Toxicogenomics combines information from studies of genomic-scale mRNA profiling (by microarray analysis), cell-wide or tissue-wide protein profiling (proteomics), genetic susceptibility, and computational models to understand the roles of gene-environment interactions in disease. The National Center for Toxicogenomics (NCT) is a coordinated, multi- disciplinary research program of the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH). The goal of this document is to specify the mission, goals and operational scheme of the NCT.

Overview

Background

Astounding progress has been made in sequencing and characterizing the human genome. This progress has been so fast that a complete first draft of the human genome sequence was announced in June 2000, far sooner than expected by many experts in the field. The human genome project and other genome sequencing projects have accelerated progress in many important scientific areas. In addition, important new scientific methods based on advances in genomics have recently emerged. In particular, vastly powerful new technologies have been developed for expression profiling of mRNAs and proteins. As a result of these advances, important scientific questions that have long been intractable to toxicology and environmental health are now open to investigation.

Toxicologists can utilize these new methods to obtain a more fundamental understanding of chemical- and drug-induced disease processes. cDNA microarray and proteomics technologies assess changes in gene expression on a genome-wide basis, providing a "global" perspective about how an organism responds to a specific stress, drug, or toxicant. This information can define cellular networks of response genes, identify target molecules of toxicity, provide future biomarkers and alternative testing procedures, and identify individuals with increased susceptibility to environmental agents and/or drugs. These are but a few of the difficult issues which these new tools will help to resolve.

The field of Toxicology is extremely diverse and its research community is crowded with competing and collaborating participants. The stakeholders include members of the pharmaceutical, chemical and consumer product industry, academic scientists, regulatory agencies and federal research organizations, in addition to the private citizens. Toxicogenomics is a new scientific field that elucidates the genomic response of organisms exposed to environmental toxicants. It is clear that advances in Toxicogenomics will be more rapid and efficient if these stakeholders join forces and work together in a coordinated manner.

The National Institute of Environmental Health Sciences (NIEHS) is uniquely positioned to coordinate an effort directed toward centralizing activities in the field of toxicogenomics and allowing free distribution of information associated with investigations centered on discovery of genome-environment interactions. Thus, NIEHS has established **The National Center for Toxicogenomics** (NCT) to coordinate an international research effort to develop the field of Toxicogenomics. The NCT will provide a unified strategy, a public database, and develop the informatics infrastructure to promote the development of the

field of Toxicogenomics. NIEHS will pay special attention to Toxicogenomics as applied to the prevention of environmentally-related diseases. The NCT will work to allow all partners in this unprecedented enterprise to share equally in its benefits and achievements.

Mission and Goal

The goal of the NCT is to use the methodologies and information of genomics science to significantly improve our understanding of basic biological responses to environmental stressors/toxicants. The specific goals of the NCT are to: 1) facilitate further development of gene expression methodology; 2) create a public database relating environmental stressors to biological responses; 3) collect information relating environmental exposures to disease; 4) develop an improved paradigm for use of computational mathematics for understanding responses to environmental stressors and 5) identify biomarkers of disease or exposure to enhance environmental health. Thus, the mission of the NCT is to catalyze the application of "Toxicogenomics" in our quest to improve human health.

Impact

Toxicogenomics could potentially have a revolutionary impact on environmental health, drug safety and risk assessment. However, the scientific problems to be solved in these areas are longstanding and extremely complex, and therefore progress will take significant time and effort. The following text describes three major scientific problems that toxicogenomics is expected to help resolve, how it will contribute to solving these problems, and what the impact is expected to be.

- **Understand biological responses to environmental stressors and identify agents that are a significant risk to human health**

A majority of toxicology research has focused on improving hazard assessment and drug and chemical safety assessment. This research has relied on *in vitro* assays and animal model systems (*i.e.,* rodent bioassays). However, there is significant uncertainty about how to extrapolate these data to humans in order to predict effects on human health. The *in vitro* systems have limitations and idiosyncrasies, and the inbred character of rodent species may also skew results so that they reflect the model as much or more than they reflect the agent being tested. Interspecies extrapolation has therefore been dealt with on a theoretical basis using a relatively small number of genes that are known to confer important allelic variations (*i.e.,* drug and chemical metabolizing enzymes). These studies remain inadequate and do not explain many known differences between different species.

Toxicogenomics is a powerful tool for improving human risk assessment because it will measure specific changes in gene expression in humans and

other species that are exposed to drugs or other agents. Careful data analysis could identify similar patterns in different species, leading to a "signature" for a given pathway of toxicity. Once signatures are identified using large scale, global microarray analysis, it will then be possible to develop smaller, multi-chemical and multi-pathway arrays that can be used to assess the potential toxicity of chemicals in a rapid, prospective manner. This would result in better interspecies extrapolation, greater confidence in animal models, reduction in the number of animals needed for testing, faster testing, and most importantly, insights into pathways of toxicity and their mechanisms.

- **Improve exposure assessment**

Toxicant specific mRNA expression "signatures" will be the basis of new tools for human exposure assessment. Using these signatures, it may be possible to identify the agent and dose to which individuals or populations have been exposed. Eventually this approach will help develop protein biomarkers of exposure and effect. Protein biomarkers of effect may also be important for early detection of environmentally-induced disease. Surveillance programs could be implemented in humans and animals in areas where exposure and/or contamination are suspected. Ideally, exposures will be discovered prior to manifestation of physiological/pathological symptoms.

- **Identify susceptibility factors that influence an individual's response to environmental agents**

It is clear that individuals differ in their susceptibility to environmentally-related disease. Toxicogenomic studies can define gene expression patterns in non-susceptible and susceptible individuals who have suffered adverse responses to drugs or chemicals. This work may lead to highly targeted arrays that can be used to predict which individuals are likely to be adversely affected by different agents. These studies will also help elucidate the mechanisms of susceptibility to different agents, which will lead to safer and more effective therapies or drugs.

The NCT presents an unprecedented opportunity to have a dramatic impact on environmental health and chemical and drug safety. The rewards to the people of the United States, and of the world, will ultimately be in the three areas outlined above: human risk assessment, human exposure assessment, and identification of individuals with increased susceptibility to environmental exposure. This information will dramatically change our understanding of human disease risk and will provide new opportunities for the nation (and individuals) to protect human health and prevent disease.

Organization

The NCT is an integrated Intramural and Extramural research program of NIEHS (*Chart 1*). The NCT Director reports to the Director, NIEHS, and as the Head of the Intramural Microarray Center, to the Scientific Director, Division of Intramural Research (DIR), NIEHS. Advisory oversight is provided by the NCT Steering Committee, the NIEHS Board of Scientific Counselors and the National Advisory Environmental Health Sciences Council. The NCT Steering Committee will be composed of academic, private sector and government (regulatory and research) scientists with experience and interest in gene expression technology, environmental health and toxicology. In addition to the NCT Director, management and coordination support will be provided by the DERT Toxicogenomics Coordinator and by the NCT Coordination Panel, comprised of the NIEHS Deputy Director and scientists drawn from the Division of Intramural Research (DIR) and the Division of Extramural Research and Training (DERT). Coordination with the Environmental Toxicology Program (ETP) will be achieved through the Toxicogenomics Interest Group and leadership interactions.

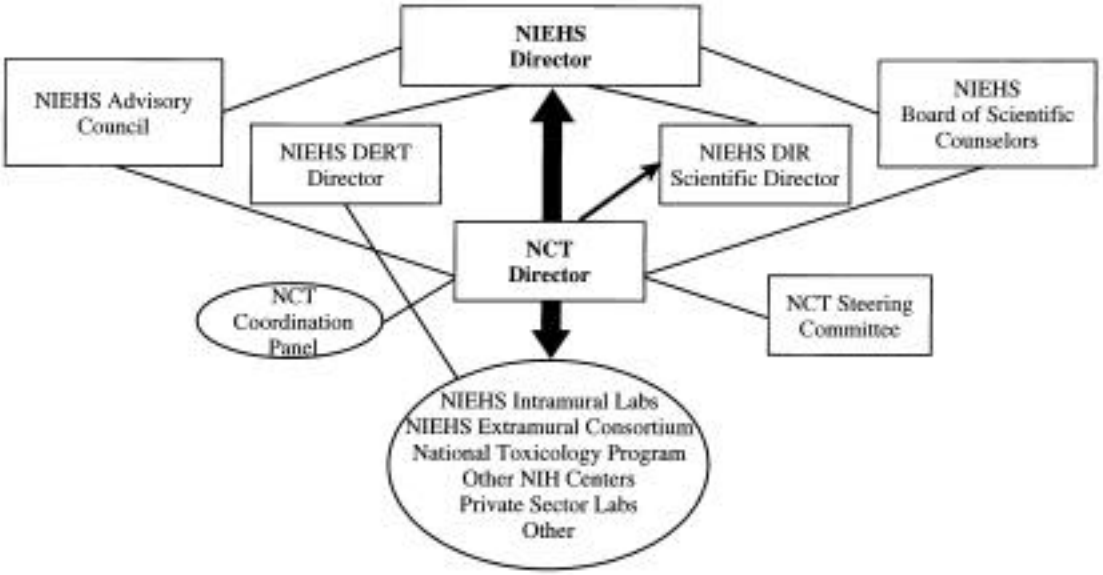
The NCT Director is Dr. Raymond W. Tennant, NIEHS. The NCT management is provided by the NCT Director, NCT Deputy Director, Dr. James Selkirk, and the DERT Coordinator, Dr. Ben van Houten. Administrative support consists of an administrative assistant, secretary, contract coordinator, and information specialist (*Chart 2*). Additional extramural staff (program administrator or program analyst) may be required when all components of the Center are in place.

The operational units of the NCT are the NIEHS Microarray Center, NIEHS Intramural laboratories, The Extramural Toxicogenomics Research Consortium, and other grantees and collaborators. Intramural laboratories provide basic and mission sensitive research on gene expression methodologies, toxicology and environmental health. The Intramural Microarray Center also provides resources and support for Intramural investigators. A Toxicoinformatics Group develops the NCT's National Toxicoinformatics Database and provides liaison and coordination with other genomic-related databases that are developing nationally and internationally.

Contributing efforts by NIEHS Toxicogenomics Research Consortium, (TRC) Extramural Centers and grantees will be developed within the NIEHS DERT and through contracts and cooperative agreements in the private sector *(Chart 3*). The Academic Research Members of the TRC will be developed from organizations at the forefront of functional genomics that are interested in partnering with NIEHS in toxicogenomics. Other components of the NCT will be developed through partnerships and CRADAs with nonprofit and private sector organizations such as pharmaceutical and consumer product companies.

Organizational and Functional Structure of NCT

NCT Organizational Structure



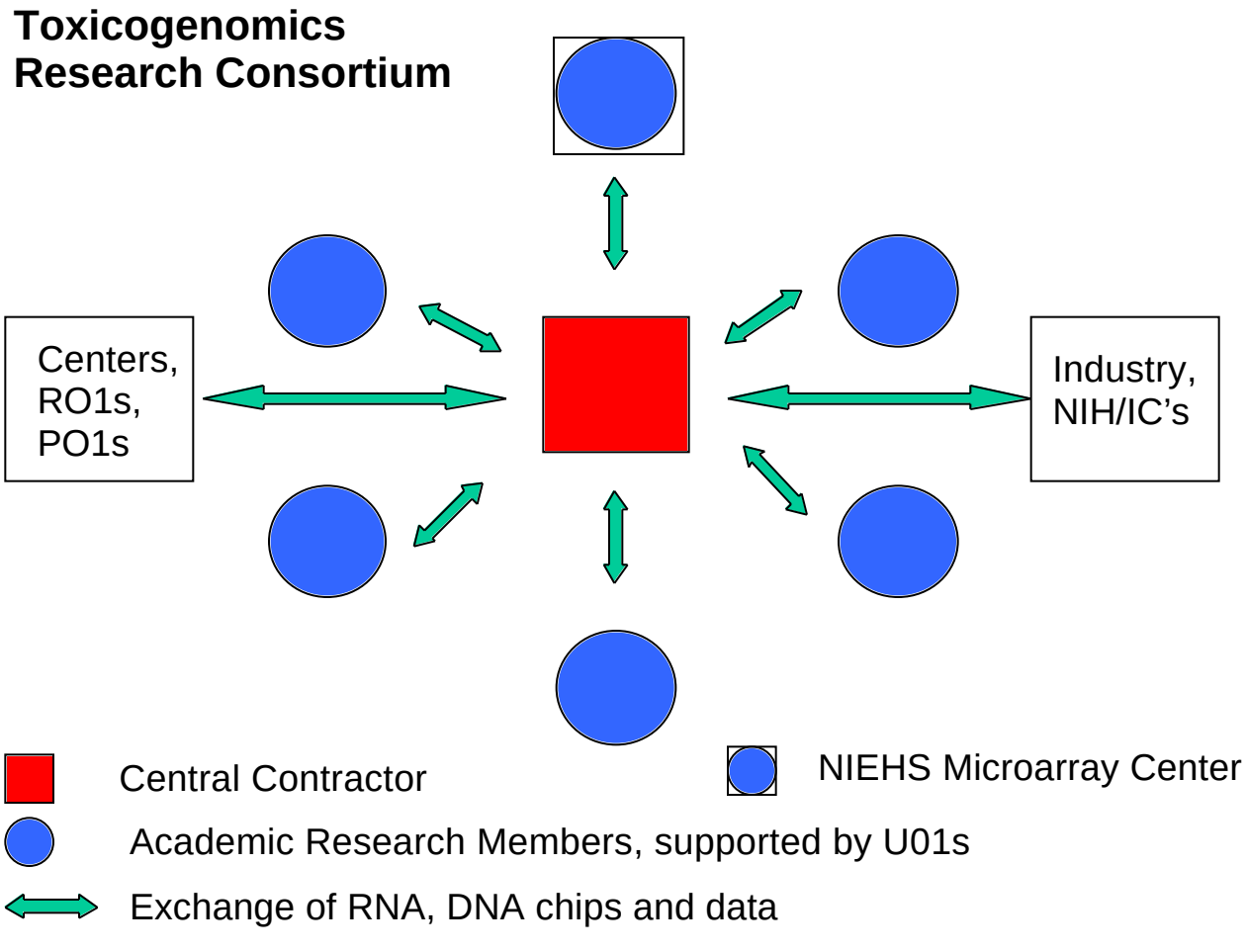
NCT Functional Structure

Leadership:	NIEHS Director, DERT Director, DIR Scientific Director, NCT Director
Oversight:	Advisory Council, Steering Committee, Board of Scientific Counselors
Management:	NCT Director, DERT Director, Coordination Panel

Operational Components of the NCT

NCT Units	NIEHS Units	Non-NIEHS Units
DIR Labs LECM LMC DIR Microarray Center: Microarray Group Proteomics Group Toxicology & Pathology Unit Toxicoinformatics Unit DERT Toxicogenomics Research Consortium	National Toxicology Program NIH Centers Academic Research Members NIEHS Centers Grantees	Private Industry Academic Labs Federal Agency Partners Private Industry Contractor

Chart 3. Toxicogenomics Research Consortium



Strategic Plan

There are five interrelated components of the NCT that must be synchronously developed. The following strategic plan describes each of these components and the phases for their development.

Component 1. Genomic-Scale mRNA Profiling by Microarray Analysis and Protein-Profiling by 2D-gel Analysis

The first component of the NCT is the actual technology and interpretation of expression profiling, including **microarray** and **proteomics**. The NCT will foster development and refinement of these techniques. While significant progress has already been made, it is expected that the methodologies will continue to develop and become more efficient and informative. It will be valuable to all researchers in the field, if methods development is coordinated by the NCT.

1A1. NIEHS Toxicogenomics Research Consortium (TRC) (DERT) Extramural Coordinator:
Bennett Van Houten

Mission:

The Toxicogenomics Research Consortium (TRC) which will accomplish three main goals: 1) enhancement of research in the broad area of environmental stress responses using microarray gene expression profiling; 2) the construction of a robust relational database; and 3) development of standards and practices which will allow proper data generation and management for inclusion in a public database. The TRC will consist of five-six academic research groups, the NIEHS Microarray Center and a contractor for bioinformatics. The contractor will provide expertise in analyzing RNA profiles and will support building a relational database.

The goal of the TRC is to examine the effects of environmental stresses, including; chemicals, physical agents, and physiological stresses, such as heat and osmotic shock in several organisms to examine trans-species effects. It is envisioned that each Academic Research Group will have expertise in micro-array gene expression profiling and is currently performing experiments with human or rodent cells in culture. Additional experience with animal care and extracting RNA from animal tissue is also desirable. Each academic group should have expertise in one or more of the following areas: development, cell cycle regulation, signal transduction, metabolism of xenobiotics, and cellular responses to injury or stress responses.

Gene Expression Profile Database:

It is imperative that there be a single public repository of information on micro-array experiments providing equal access for all researchers. To this end, NIEHS through the TRC will create a relational database for the collection and processing of the large amounts of data produced by gene expression profiling experiments. This database will allow researchers to keep informed on specific experiments including treatments, cell lines, and rodent strains so that unwanted duplication can be avoided. Finally, the real power of gene expression profiling is the ability to compare one data set with another through robust bioinformatics. This will be accomplished through a central contractor (Chart 3). These microarray data will be merged with the Chemical Exposure in Biological Systems (CEBS) database.

Impact on Toxicology:

DNA micro-array technology will have profound effects on toxicology assays. Rather than a lengthy two-year bioassay for testing new compounds for their carcinogenicity, the use of a unknown toxicant can be quickly compared to a library of known responses to specific toxicants to determine what genes are turned on/off following exposure to a toxic agents. Some key questions that will be addressed are:

- Determine whether specific toxicants have unique signatures.
- Determine whether different cells in different tissues show profoundly different responses.
- Determine whether different animal species will show overlapping patterns of gene responses.
- Determine whether the specific toxicant signature is altered depending on the stage in the developmental process or as a result of other health conditions.
- Determine whether the responses to complex mixtures are more easily addressed by looking at global response through gene expression profiling.
- Determine whether polymorphisms leading to increased susceptibility will be detected by careful analysis of changes in expression profiles to pathology and disease.
- Determine whether gene expression assays will help to demonstrate the effects of chronic low doses of environmental pollutants.

DNA microarray technology will have a major impact on human health and medicine. Here the main research challenges will be to determine whether there are gene expression changes associated with specific low dose or complex exposures to chemical mixtures and whether people with polymorphisms in specific genes show different gene expression profiles. A second area of potential benefit to human health is to determine whether there are discrete gene expression changes associated with toxicant exposure

leading to dysfunction and disease. Finally, DNA microarray technology will become a major tool in medicine, for both diagnosis/prognosis of specific diseases to examination of drug sensitivities and effectiveness. As in any public health initiative, the impact of gene expression profiling must be carefully examined for the ethical, legal, societal issues.

Implementation Strategy:

The TRC will start initially by the interaction of the NIEHS Microarray Center and a central contractor who will provide RNA analysis and bioinformatics support. Through a series of validation experiments specific practice standards, which will enable cross-platform analyses, will be developed. During the next phase three Academic Research Members will be added through a cooperative agreement mechanism who will participate in the consortium. During future years two-to-three more Academic Research Members will be added. These Academic Research Members will have the dual function of performing cutting-edge research employing microarray gene expression profiling and also perform consortium toxicology experiments for cross-platform validation. The central contractor will develop a relational database which will allow evaluation of microarray experiments and establish specific practice standards and cross-platform algorithims. Once having established necessary protocols for cross-platform validation, we envision the use of the database by Extramural grantees of NIEHS, grantees from other Institutes, and the private sector, who will be able to contribute and query the database. This database will be merged with the Chemical Effects in Biological Systems (CEBS) database, one of the main products of the NCT.

Future initiatives from DERT will support targeted solicitation of grant applications using microarray gene expression profiling and proteomics in the elucidation of environmentally induced diseases.

1A2. NIEHS Microarray Center (NMC) (DIR)

Directors:
Richard S. Paules Toxicological Gene Expression Studies (Rodent studies)
Cynthia A. Afshari Basic Research Studies (Human and yeast studies)

Mission:

The mission of the NIEHS Microarray Center (NMC) is to develop a research program in the Division of Intramural Research at the NIEHS that utilizes the advances in genomics to address critical threats to human health as a consequence of environmental exposures, specifically incorporating genome-wide gene expression analysis using microarray technology, with the goal of reducing or eliminating these threats to human health. In order to accomplish this mission, the goals of the NMC include: **1)** the pursuit of mechanistic

investigations of the impact of exposures on gene expression and alterations in critical cellular regulatory pathways; **2)** the utilization of global gene expression changes following environmental stresses to identify specific changes, using bioinformatics tools, that are tightly linked with a deleterious human health outcome as patterns or biomarkers of a particular disease or toxic outcome; and **3)** the continual development of new and better approaches to address genome-wide changes in gene expression. One consequence of this research program will be the accumulation of global gene expression information that will contribute to the population of an environmental health and toxicogenomics gene expression database. Such a public database will allow scientists to better understand the mechanism of action of environmental agents on human health and better evaluate the consequences associated with such exposures prior to the appearance of pathological presentation. This information may allow for the intervention and prevention of disease outcomes. The NMC will have a leadership role in developing sound scientific data and the best scientific approaches to help the entire scientific community to make rapid advances in utilizing genomic information to benefit human health.

cDNA microarray technology has the promise of allowing the recent advances in the field of genomics to increase our understanding of the action of toxic or potentially toxic environmental insults. This will be accomplished by application into two general and intertwined areas of research, mechanistic-based studies and predictive, or "toxicant-response signature" defining studies. Studies that allow for detailed analysis of molecular mechanisms and gene expression networks associated with a compound's toxicity across experimental animal models to humans should increase our ability to extrapolate potential dangers to human health.

NMC Initiated Studies:

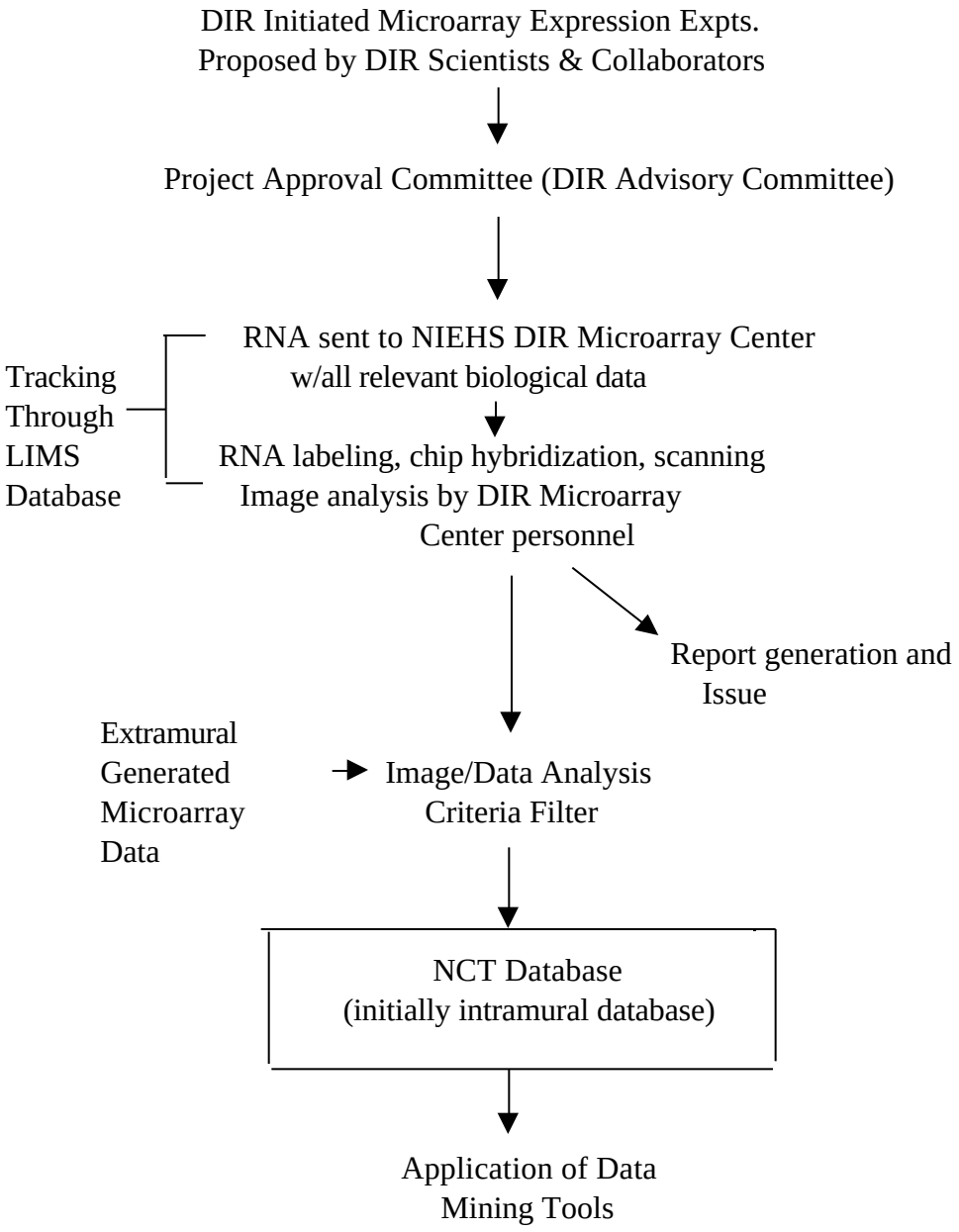
The goal of the NMC is to design and perform critical experiments that utilize global gene expression analysis using microarray technology to provide critical mechanistic information that provides insight into biological consequences of exposure as well as signature patterns of gene expression changes linked with biological/toxicological endpoints. These efforts will entail a significant percentage of the effort of the NMC.

Collaborator Proposed Projects:

The NIEHS cDNA Microarray Center (NMC) was initially established as a collaborative center to provide investigators within the NIEHS intramural and extramural communities access to the emerging technology of cDNA microarrays. While the NMC will continue to provide the opportunity for investigators to perform collaborative projects with the Center, projects will be prioritized and selected on the basis of the criteria outlined in the guidelines for

Chart 4.

Microarray Flowchart



submission of project proposals found on the NMC website, and for contribution to the scientific goals of the NMC. A small committee of NIEHS scientists with knowledge and expertise in genomics and with representation from each of the Division of Intramural Research Programs, will be established by appointment of the Scientific Director, NIEHS, to review proposals for collaborative studies on a quarterly basis. Criteria for evaluation of projects will continue to be based on soundness of scientific hypothesis, relevance to NIEHS mission, and expertise of the proposing laboratory to validate data.

Continued Technology Development/Evaluation:

A portion of the effort of the NMC will be reserved for and devoted to ongoing technology development which is critical in this very rapidly developing area of science. Alternate RNA isolation protocols, labeling protocols, array formats, instrumentation, and analysis tools are under constant development by academic and commercial investigators and need to be evaluated if they will potentially impact the cost (dollars and time), sensitivity and reproducibility of current protocols implemented in the NMC.

1B. NCT Proteomics (DIR)

NIEHS Proteomics Working Group:

Kenneth B. Tomer Mass Spectrometry, Laboratory of Structural Biology
B. Alex Merrick Protein Separations, Laboratory of Molecular Carcinogenesis

Mission:

The mission of the intramural Proteomics program is to study global changes in protein expression that reflect responses to environmental toxicants/stressors. The goals of the Proteomics Group within the NCT are: 1) to identify protein biomarkers; 2) biochemical pathways; 3) critical protein-protein interactions; 4) post-translational events that are altered by exposure or environmental diseases; and 5) to develop and foster technical innovations for global protein analysis. These goals will be accomplished initially by an **Intramural Proteomics Group** that will establish a research program and well as collaborate with other DIR groups and by **Extramural contracts, grants and collaborations.** Data from Proteomics studies are envisioned to be publicly accessible in the CEBS database.

Overview:

The technology for Proteomics, the volume and type of data output as well as the number of analyses possible are different from genomic expression technologies. The differences warrant a brief explanation to understand what the current capabilities are and the potential contribution of Proteomics to Toxicogenomics.

Proteomics examines the entire range of gene expression of a specific cell or organism at the protein level. While the Human Genome Project will ultimately define the number of human genes, the number of proteins is expected to be much higher because post-translational processes may alter each gene product and give rise to several structurally distinct proteins. Defining the exact structure and function of each protein species will be important for understanding environmentally-induced disease and in finding reliable biomarkers of chemical exposure.

Separation and identification of all cellular proteins in a single procedure is a daunting task. It is the combination of 2D gel separation and mass spectrometry that forms the current basis for Proteomics at most research centers. Unlike cDNA microarray technologies where thousands of known genes or EST's can be hybridized with sample cDNA, the identity of most proteins appearing on 2D gel maps are initially unknown. In order to identify individual proteins separated on 2D gels, they are excised, proteolytically digested, and the fragments are measured with high accuracy by mass spectrometry (matrix assisted laser desorption ionization-mass spectroscropy or MALDI-MS) to form a searchable protein fingerprint in protein or genomic databases for identification. With further analysis, *de novo* amino acid sequence can be generated by tandem MS (MS/MS) analysis of individual digestion fragments for entry into genomic databases for identification. Because Proteomics is an open system in generating primary data, it may also serve as a tool of discovery for new gene products in any tissue or species. Finally, different types of MS techniques are capable of determining various types of post-translational modifications which may be altered by hazardous agent exposure with important functional consequences for the protein, tissue and organism. It should also be mentioned that other global protein analysis technologies, in addition to the 2D gel - MS paradigm, are under active experimental and commercial development. As such technological developments become available, the Proteomics program at NIEHS should explore and adapt new technologies that achieve greater data output, higher throughput and improved accuracy for global protein analysis that will best advance the Toxicogenomics program.

Proteomics Initiated Studies

A major goal of the Proteomics field is to ultimately identify and structurally characterize every protein expressed in human cells, tissues and organs as well as rodent tissues. Several commercial entities are interested in pursuing this goal. However, from a research standpoint, a *first* practical strategy will be to focus initial efforts in Proteomics upon that critical group of proteins which are differentially expressed from control among experimental test samples. Therefore, the more crucial goal in Proteomics research is to link experimental manipulations or biological effects to a group differentially expressed proteins.

Once differentially expressed proteins are located by 2D gel analysis, they can be targeted for quantitation and identification.

The initial information output of Proteomics is not expected to be as large as gene expression technologies because *de novo* analysis is usually performed for each protein. Once proteins are identified on 2D patterns, this information will be annotated in 2D protein maps and electronically curated so it can be readily referenced for future Toxicogenomics studies. Over time, the cumulative annotations of identified proteins stored 2D protein maps will allow Proteomics information volume to grow and be more comparable to cDNA microarray platforms.

NIEHS Collaborative Projects

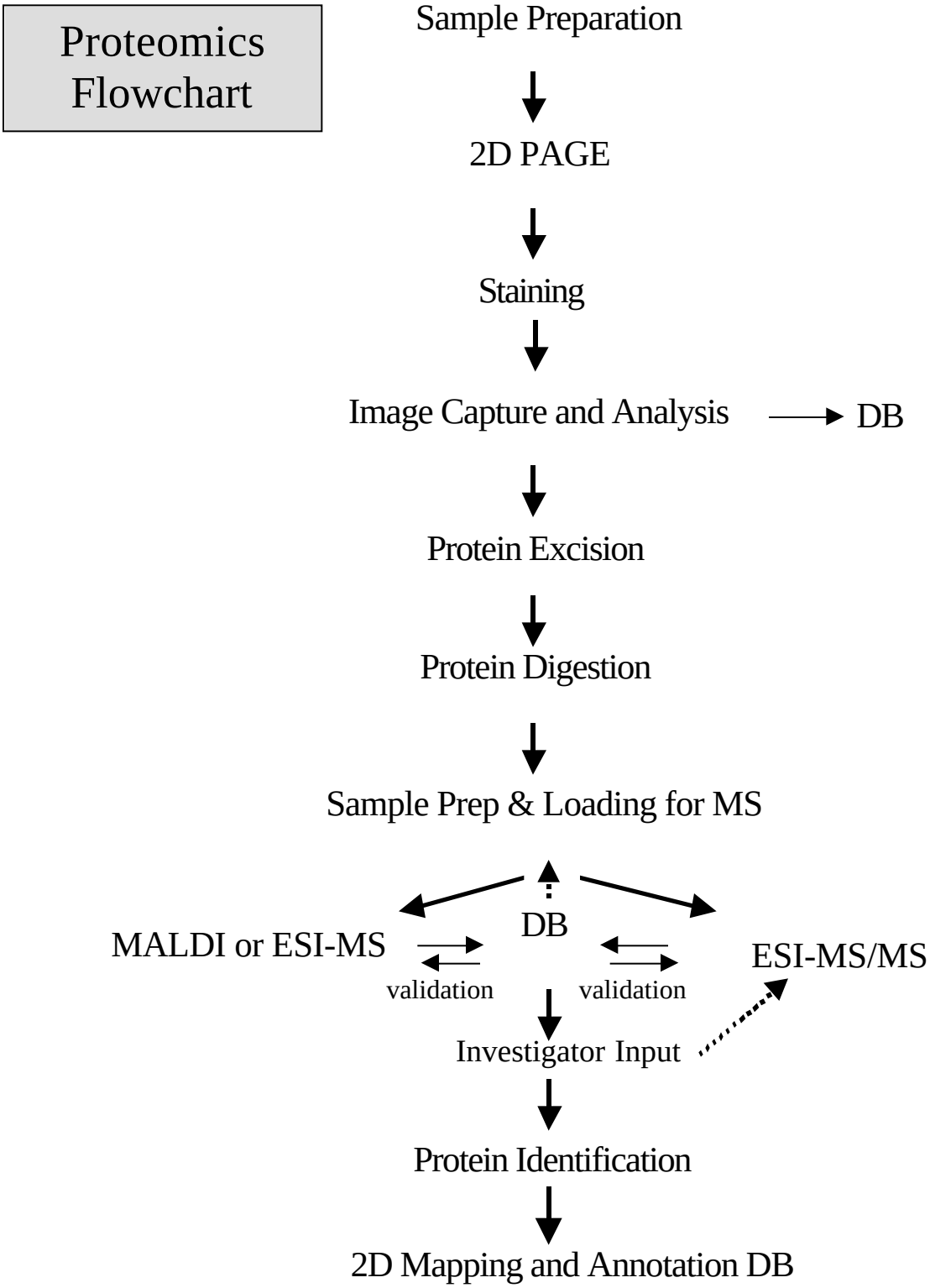
A substantial effort of the intramural Proteomics group will be spent in collaborating on projects with the NIEHS Microarray Center (NMC). This is expected to be a dynamic relationship in which Proteomics data can also influence NMC research activities. The Proteomics Group will, beyond separation and analysis, bring forth functional significance to identified proteins and answer biological questions posed by toxicant exposure. The ability to select appropriate biochemical methods, and the planning of additional experiments after initial protein identification will help address toxicological questions being considered by the NCT.

The NIEHS Proteomics Group will provide the NCT with intramural research capabilities for global analysis of proteins, protein identification and protein characterization. Opportunities for investigator initiated projects will be prioritized and selected on the basis of contribution to the NCT and NIEHS mission. NCT management and scientists with knowledge and expertise in Proteomics will review and prioritize collaborative studies based on soundness of scientific hypothesis, relevance to NIEHS mission and expertise of the proposing laboratory to validate data.

Continued Technology Development/Evaluation:

A portion of the Proteomics Group's efforts will be apportioned for technology development of rapid, global analysis of proteins and for characterization of specific post-translational modifications that bear particular importance to biomarker identification. Rapidly developing technological advances in protein separations, image analysis, mass spectrometry of biomolecules and protein bioinformatics by academic and commercial investigators need to be continually monitored and evaluated for possible implementation within the Proteomics Group.

Chart 5.


Proteomics Flowchart
Sample Preparation
2D PAGE
Staining
Image Capture and Analysis
DB
Protein Excision
Protein Digestion
Sample Prep & Loading for MS
DB
MALDI or ESI-MS
ESI-MS/MS
validation
validation
Investigator Input
Protein Identification
2D Mapping and Annotation DB

Component 2. Linking Biological Responses to Environmental Stressors

The application of the microarray and proteomics technologies outlined in Component 1 will be in the development of understanding of the link between biological responses and environmental stress. Gene Expression data will be collected on the cellular responses to exposure to a wide variety of toxicants. To facilitate database development, initial data collected by the NCT will be derived from a very limited number of platforms, and verification will ensure that results are consistent across laboratories. Initial studies will include toxicants whose mechanism of toxicity is at least partly understood in well-defined model systems. Hypothesis-based studies will be carried out to search for toxicant-specific trans-species patterns of altered gene expression, indicative of tissue-specific acute or chronic toxicity. Ultimately, signature patterns of gene expression will be identified and applied to hazard assessment, biomarker development, and analysis of genetic susceptibility.

Component 3. NCT Public Access Database On Chemical Effects in Biological Systems (CEBS)

It is imperative that there be a public repository of toxicogenomics information, which offers access to all researchers. NIEHS will create such a repository for data that will be generated in part from toxicological-based experiments involving microarray and proteomics technologies described previously. To establish the database, NCT supported research will need to conform to a specific reporting standard. Furthermore, due to the large expense of microarray and proteomics experiments (both in cost and time), the database will be used to minimize duplication of effort in these studies. Finally, the standards and conformity of the database will be essential as a basis for later data analyses. For this and many other reasons, the NCT will maintain specific criteria for reporting to the NCT public access database. It is expected that a large number of NIEHS grantees and collaborators will contribute to and use this database.

Component 4. Toxicoinformatics

Toxicoinformatics is the fourth component of the NTC. Toxicoinformatics are essential computational tools for the analysis of time- and dose-dependent changes in patterns of toxicant-induced gene expression. Development of these capabilities will require a database that compiles high quality data from diverse sources involving different gene expression platforms, assay methods, validation results and diverse drugs, chemicals or environmental agents. Complete analysis will require linkage between and among additional databases that provide most current sequence identity, can annotate gene identity and function, chemical structure, toxicity, pathology, pharmacokinetics/ biodistribution, genotoxicity, etc. Relational searching of these databases will

be essential. It is intended that the contents of these databases eventually will be freely accessible to the scientific community. In addition, when feasible, analytical tools will also be linked to the datamining and database. Partnership with other NIH units and private sector units is anticipated.

Component 5. Biomarkers

Biomarkers are the fifth component of the NCT. The NCT will begin to assemble a comprehensive database of biomarkers. Protein biomarkers will be derived from Proteomics and other ongoing studies. Although the thrust will be with human cells, it is anticipated 2D-gel data from other species will also be incorporated into the program. Since 2D-gel patterns of any specific cell or tissue have unique marker proteins in their pattern, and there is a world-wide effort to produce web-based master 2D-gel images of both normal and abnormal tissues, it is anticipated the NCT effort will incorporate that knowledge into the database, as well as contribute to this growing international tissue and cell profiling effort.

Biomarkers will play an important role in early detection of environmentally-induced disease, since routine surveillance programs in both people and animals in suspected environmentally hazardous areas could be instituted. Also, chemicals routinely used in the home or workplace that are suspected hazards can be tested for early and subtle changes in the protein pattern of body fluids and tissues to determine which biochemical pathways are altered.

Also planned is a Website that will be designed for facile data search by an assembly of software routines for comparative biomarker searching. It is apparent that there will be a large number of significant biomarkers found with each chemical treatment in the arrays of several thousand genes and proteins, and it is anticipated these changes will represent toxicity to numerous biochemical pathways. Therefore, we can expect complex patterns of overlap with different chemical treatments that should lend greater knowledge to the mechanism of activation, and detoxification of all chemicals, both toxic and beneficial that confront the cell.

Once a database is established, effects of chemicals of unknown toxicity can be assayed for their pattern of protein biomarkers. Beneficial chemicals such as medicines as well as toxic materials can be studied for the biochemical pathways associated with drug efficacy, or those where toxicity has occurred. Once a sufficiently large database is assembled, it is assumed that it will be possible to predict toxic effects of chemical classes. Cell and tissue biomarkers found by microarray and 2D-gel analysis from potential consumer or industrial use could be matched against profiles already in the database for toxicity prediction and as an indication for further testing. This knowledge will also contribute to validating surrogate models used in toxicological testing, since the

enzymatic pathways of activation and detoxification can be compared to human processes. The ultimate goal is to build a source of biomarker information as a central focus for public and private institutions, to be utilized for both basic and applied research. Our approach will be to partner with academia and industry for simultaneous growth of information on many individual cells and tissues and produce the voluminous data needed to populate this database in a reasonable timeframe.

Partnerships

One of the primary goals of the NCT is to coordinate the field of toxicogenomics on an international level. Thus, it will be important to establish partnerships with other organizations and centers actively engaged in toxicogenomics research. Such centers are located in private sector companies that are developing methods of gene array analysis, proteomics, and informatics, in academic and government centers conducting basic research, and in industry where there is great interest in applying Toxicogenomics to chemical and drug safety assessment, and to the identification of individuals who are susceptible to adverse reactions. Partnerships will be fostered between all of these stakeholders in the field. The mechanisms for these partnerships will include collaborations, cooperative research and development agreements (CRADAs), Cooperative Agreements and contracts. These partnerships will be formed during the initial phase of the strategic plan. Some partnerships are currently in place, including CRADAs with Boehringer-Ingelheim Pharmaceutical Company, Glaxo Wellcome and Paradigm Genetics, and a collaborative consortium of pharmaceutical companies working under the International Life Sciences Institute (ILSI).

Private Sector - Partnerships

Abstracts for public release derived from CRADAs (Cooperative Research and Development Agreements) between NIEHS and industrial partners:

Boehringer-Ingelheim Pharmaceuticals (BIPI):

NIEHS and BIPI are initiating a Cooperative Research and Development Agreement (CRADA) for the development and utilization of rat and human gene expression arrays for toxicological analysis. Both parties will work together to develop the genomic resources necessary to analyze expression patterns in a variety of tissues after treatment with toxic agents. Information obtained from this effort could be used to develop molecular assays for predictive, high throughput toxicity screens or diagnostic mechanistic information for preclinical and clinical toxicity evaluation studies.

Note: This collaboration also allows a joint collaboration with **Phase-1 Molecular Toxicology, Inc.,** a biotech company directed at analysis of small toxicology arrays.

Paradigm Genetics:

NIEHS and Paradigm Genetics are initiating a Cooperative Research and Development Agreement (CRADA) for the development of yeast expression arrays for toxicological analysis. Both parties will work together to develop the genomic resources and analysis protocols to analyze expression patterns in yeast after treatment with a variety of toxic agents. It is expected that the culmination of this CRADA will result in joint publication of these results in peer-reviewed journals.

Glaxo Wellcome:

The large amount of DNA sequence data, made available as a result of the Human Genome Initiative, has made possible the development of high-density DNA microarrays. Using these arrays, qualitative and quantitative gene expression information for thousands of genes (the transcriptome) within a given cell population or tissue can be obtained. This information will make it possible to identify genes or patterns of gene expression that are associated with specific cellular processes. The goal of the research partnership between Glaxo Wellcome and the NIEHS will be to use a toxicity-specific 2,000 gene microarray (termed ToxChip), developed by the NIEHS, to determine whether transcriptome analysis can be predictive of cellular chemical toxicity, and to develop a database consisting of gene expression information representative of various mechanisms of chemical toxicity. Information obtained from this effort could be used to develop molecular assays for predictive, high throughput toxicity screens, or diagnostic mechanistic information for preclinical and clinical toxicity evaluation studies.

SBIR collaboration with **Gene Trace** on testing of novel, high throughput gene expression technology.

ILSI Initiative:

NIEHS NMC scientists are participating in a joint effort with industrial scientists, primarily from major international pharmaceutical companies, and scientists from other governmental agencies including US FDA, US EPA, NCI/NIH, and NIHS of Japan to evaluate the “Application of genomics to mechanism based risk assessment” under the auspices of the Health and Environmental Sciences Institute (HESI) of the International Life Sciences Institute (ILSI). In the initial effort, global gene expression changes will be analyzed on the same or similar biological samples prepared at various sites under the same protocols and then analyzed on various platforms at various locations. These first efforts will focus

on gene expression changes involved in hepatotoxicity, nephrotoxicity, and genotoxicity, while another subgroup evaluates and puts into place or locates a suitable database with analytical software tools for processing of expression results from this effort.

Public Access Database

Important partnership agreements will be used to develop informatics tools and the **public access (CEBS) database** mentioned above. This database will include the informatics associated with storage and interpretation of microarray data, cross-organization, cross-platform comparisons, and quality assessment of data. The extended database will be developed to allow relational searches that can annotate toxicogenomic data for chemical structure, metabolic pathways, acute and chronic toxicity, tissue specific pathology, human exposure data and other parameters. Ultimately, it is the relational databases that will permit searches to associate specific molecular pathways of gene expression with specific chemical toxicity.

The Environmental Genome Project (EGP)

The primary goal of the Environmental Genome Project (EGP) is the identification of human polymorphisms indicative of specific disease, drug or environmental agent susceptibility. The EGP performs sequence analysis of DNA samples from target and cohort populations, and develops a catalogue of single nucleotide polymorphism (SNP) and other DNA variations in the human genome. The NCT will use gene expression data obtained from non-susceptible and susceptible individuals suffering adverse drug reactions to identify expression markers of susceptibility. It may be possible to associate specific patterns of gene expression or molecular pathways with specific toxic effects. The EGP and the NCT, therefore, have a very complementary nature and the NCT will coordinate its efforts closely with the EGP.

The NIH Biomarkers Program

In collaboration with the FDA, the NIH has initiated a trans-institute Biomarkers Program to identify surrogate biomarkers for preclinical assessment of drug efficacy and safety. The NIEHS has been given a lead role in the development of surrogate biomarkers for drug safety assessment. The NCT will contribute heavily to this effort, because, as explained above, biomarkers can be based on expression profiles that are diagnostic of chemical or drug exposures and/or effects. The NCT will therefore represent a major contribution to the Biomarkers Program.