

TOXICOGENOMICS AND SYSTEMS TOXICOLOGY: AIMS AND PROSPECTS

Michael D. Waters and Jennifer M. Fostel

Abstract | Toxicogenomics combines transcript, protein and metabolite profiling with conventional toxicology to investigate the interaction between genes and environmental stress in disease causation. The patterns of altered molecular expression that are caused by specific exposures or disease outcomes have revealed how several toxicants act and cause disease. Despite these success stories, the field faces noteworthy challenges in discriminating the molecular basis of toxicity. We argue that toxicology is gradually evolving into a systems toxicology that will eventually allow us to describe all the toxicological interactions that occur within a living system under stress and use our knowledge of toxicogenomic responses in one species to predict the modes-of-action of similar agents in other species.

NECROSIS

The localized death of living cells.

MODE-OF-ACTION

The sequence of events from the absorption of a compound into an organism to a toxic outcome or death.

PROTEIN CHIP

A genomic set of proteins that are arrayed on a solid surface without denaturation.

National Center for Toxicogenomics, National Institute of Environmental Health Sciences, PO Box 12233, MD F1-05, 111 Alexander Drive, Research Triangle Park, North Carolina 27709-2233, USA.

Correspondence to M.D.W. e-mail: waters2@niehs.nih.gov

doi:10.1038/nrg1493

The ability to discern the mechanisms of toxicity that are related to health issues is an important challenge facing scientists, public-health decision-makers and regulatory authorities, whose aim is to protect humans and the environment from exposures to hazardous drugs, chemicals and environmental stressors (such as global warming or non-ionizing radiation). Also, the problems of identifying environmental factors involved in the aetiology of human disease and of performing safety and risk assessments for drugs and chemicals have long been formidable issues.

Toxicology — the study of poisons — is focused on the substances and exposures that cause adverse effects in living organisms. A vital part of this study is the empirical and contextual characterization of adverse effects at the level of the organism, the tissue, the cell and intracellular molecular systems. Therefore, studies in toxicology measure the effects of an agent on an organism's food consumption and digestion, on its body and organ weight, on microscopic histopathology, and on cell viability, immortalization, NECROSIS and apoptosis¹.

The rapid accumulation of genomic-sequence data and associated gene and protein annotation has catalysed the application of gene-expression analysis to understanding the MODES-OF-ACTION of chemicals and other environmental stressors on biological systems

(FIG. 1). These developments have facilitated the emergence of the field of toxicogenomics, which aims to study the response of a whole genome to toxicants or environmental stressors^{2–13}. The related field of toxicoproteomics^{14–16} is similarly defined with respect to the proteome — the protein subset of the genome. For example, global technologies such as cDNA and oligonucleotide microarrays, PROTEIN CHIPS and NUCLEAR MAGNETIC RESONANCE (NMR)-based molecular profiling, in corresponding order, can simultaneously measure the expression of numerous genes, proteins and metabolites, therefore providing the potential to accelerate the discovery of toxicant pathways, modes-of-action, and specific chemical and drug targets. So, toxicogenomics combines toxicology with genetics, global '-omics' technologies (BOX 1) and appropriate pharmacological and toxicological models (FIG. 1) to provide a comprehensive view of the function of the genetic and biochemical machinery of the cell.

This review explores the new field of toxicogenomics, delineates some of its research approaches and success stories, and describes the challenges it faces. It discusses how integrating data that is derived from TRANSCRIPTOMICS, PROTEOMICS and METABONOMICS studies can contribute to the development of a toxicogenomics KNOWLEDGEBASE (FIG. 2; BOX 1) and to the evolution of SYSTEMS TOXICOLOGY as it relates to molecular-expression profiling. In many

NUCLEAR MAGNETIC RESONANCE

An analytical chemistry technique that is used to study molecular structure and dynamics; it explores spectral differences that are caused by the differential alignment of atomic spins in the presence of a strong magnetic field.

TRANSCRIPTOMICS

Techniques that measure the full complement of activated genes, mRNAs or transcripts in a particular tissue at a particular time, typically through the use of cDNA or oligonucleotide microarrays.

PROTEOMICS

A collection of techniques used to measure the structural and functional properties of proteins through the use of 2-dimensional gel electrophoresis or liquid chromatography; typically followed by protein identification using some form of mass spectrometry.

METABONOMICS

Techniques that detect changes in the concentration of low-molecular-weight metabolites present in a cell or organism at a given time (the metabolome) by using nuclear magnetic resonance or mass spectrometry coupled to gas or liquid chromatography.

KNOWLEDGEBASE

An archival and computational system that uses data, information and knowledge captured from experts to carry out tasks that create new information and new understanding.

SYSTEMS TOXICOLOGY

The study of the perturbation of biological systems by chemicals and stressors, monitoring changes in molecular expression and conventional toxicological parameters, and iteratively integrating response data to describe the functioning organism.

SYSTEMS BIOLOGY

The integrated study of biological systems (cells, tissues, organs or entire organisms) at the molecular level. It involves perturbing systems, monitoring molecular expression, integrating response data and modelling the molecular structure and network function of the system.

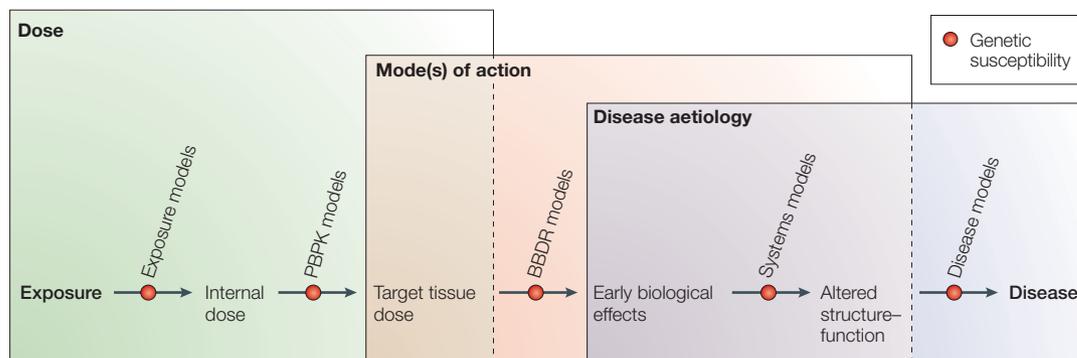


Figure 1 | The role of genetic susceptibility and computational models on the continuum from exposure to disease outcome. The sequence of events between initial exposure to a toxicant and final disease outcome are shown from left to right. Note that genetic susceptibility (red dot) influences every level of toxicological analysis. After exposure, the ADME (absorption, distribution, metabolism and excretion) systems of the body control local concentrations of a chemical stressor in various body compartments. This is affected by genetics through the involvement of specific alleles encoding various transporters and xenobiotic-metabolizing enzymes among others. Mathematical models such as exposure models, physiologically-based pharmacokinetic (PBPK) and biologically-based dose response (BBDR) models can be used to approximate these processes. PBPK models are a set of differential equations structured to provide a time course of a chemical's mass-balance disposition (wherein all inputs, outputs and changes in total mass of the chemical are accounted for) in pre-selected anatomical compartments. BBDR models are dose-response models that are based on underlying biological processes. Once the target tissue is exposed to a local stressor, the cells respond and adapt, or undergo a toxic response; this process can be modelled with systems toxicology approaches. Finally, the disease outcome itself can be mimicked by genetic or chemically induced models of particular diseases; for instance, in the Zucker rat model of diabetes or the streptozotocin-treated rat model. The coloured boxes show the type of toxicologically-relevant information that can be obtained from each set of model.

ways, current gene, protein and metabolite-expression profiles are simple 'snapshots'. By contrast, systems toxicology, as with SYSTEMS BIOLOGY^{17,18}, attempts to define the interactions of all of the elements in a given biological system, under stress or toxicant perturbation, to achieve a mechanistic understanding of the toxicological response.

Toxicogenomics: aims and methods

Toxicogenomics has three principal goals: to understand the relationship between environmental stress and human disease susceptibility (FIG. 1); to identify useful BIOMARKERS of disease and exposure to toxic substances; and to elucidate the molecular mechanisms of toxicity.

A typical toxicogenomics study might involve an animal experiment with three treatment groups: high-dose and low-dose treatment groups and a vehicle control group that has received only the solvent used with the test agent. These groups will be observed at two or three points in time, with three to five animal subjects per group. In this respect, a toxicogenomics investigation resembles a simple, acute-toxicity study. The two approaches differ in the scope of the response they aim to detect, and in the methods used. The highest-dose regimen is intended to produce an overtly toxic response that can be detected in a toxicogenomics study using the global-measurement techniques that are described below (see also BOX 1).

In a typical toxicogenomics experiment, lists of significantly differentially expressed genes are created for each biological sample¹⁹. Alternatively, profile-analysis methods can be applied to dose-related and time-course studies¹⁹ to identify genes and gene profiles of interest. Then, with the aid of the relevant knowledge

that is systematically extracted and assembled²⁰ through literature mining, comparative analysis and iterative biological modelling of molecular-expression datasets, it is possible to differentiate the adaptive responses of biological systems from those changes (or biomarkers) that are associated with or precedent to clinical or visible adverse effects. Over the past 5 years, the field of toxicogenomics has validated the concept of gene-expression profiles as 'signatures' of toxicant classes, disease subtypes or other biological endpoints. These signatures have effectively directed the analytical search for predictive biomarkers of toxicant effects and contributed to the understanding of the dynamic alterations in molecular mechanisms that are associated with toxic and adaptive responses.

The experimental work involved in a toxicogenomics study and the amount of gene-expression data generated is vast. Even to examine, per dose-time group, one tissue per animal requires 18–45 microarrays (more if replicates are used) and the attendant measurement of as many as 20,000 or more transcripts per array. Also, each animal will typically have treatment-associated data on total body and organ-weight measurements, clinical chemistry measurements (often up to 25 parameters) and microscopic histopathology findings for several tissues¹. The careful collection, management and integration of these data, in the context of the experimental protocol, is essential for interpreting toxicological outcomes. So, all data must be recorded in terms of dose, time and severity of the toxicological and/or histopathological phenotype(s). The compilation of such experimental data, together with TOXICOINFORMATICS tools and computational modelling, will be important in deriving a new understanding of toxicant-related disease⁷.

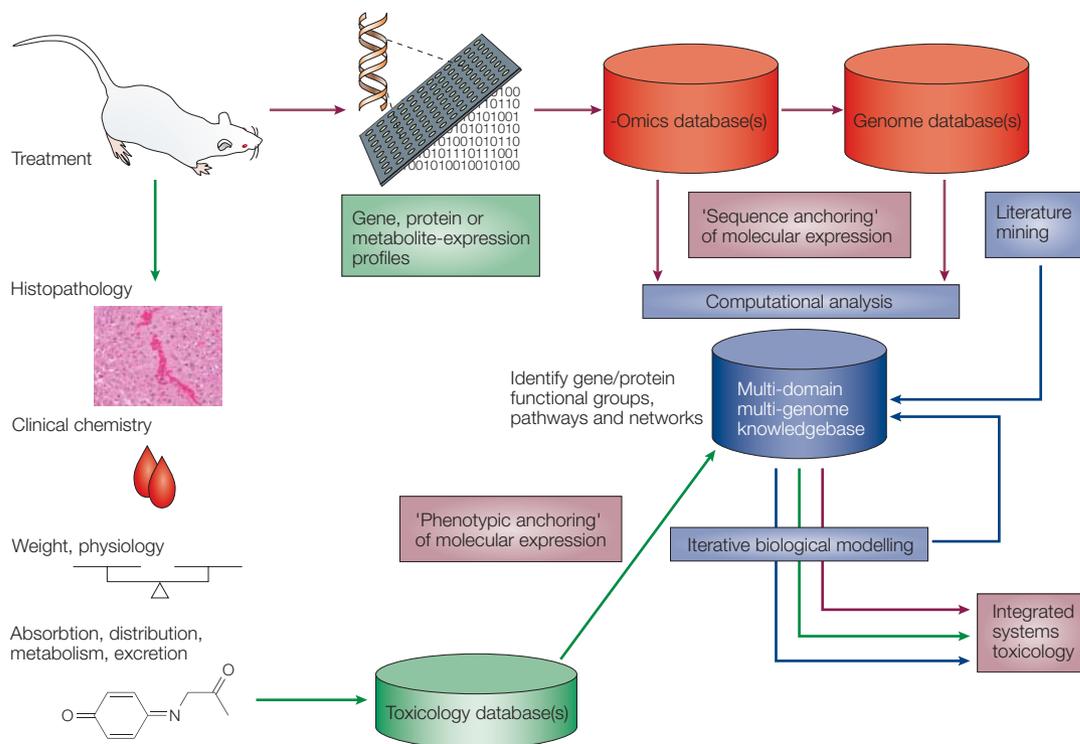


Figure 2 | A framework for systems toxicology. This figure indicates the paths from the initial observation (rat in upper left) to an integrated toxicogenomics knowledgebase (blue cylinder), and so to systems toxicology (bottom right). The ‘-omics’ data stream is shown by the clockwise path from rat to knowledgebase; and the ‘traditional’ toxicology approach is shown in the anti-clockwise path. The knowledgebase will integrate both data streams, along with literature-based knowledge; and by virtue of iterative modelling, will lead to a systems toxicology understanding. The framework involves ‘phenotypic anchoring’ (to toxicological endpoints and study design information) and ‘sequence anchoring’ (to genomes) of multi-domain molecular-expression datasets in the context of conventional indices of toxicology, and the iterative biological modelling of the resulting data.

Toxicogenomics is beginning to integrate the multiple data streams derived from transcriptomics, proteomics and metabolomics with traditional toxicological and histopathological endpoint-evaluation (FIG. 2). This integration has the potential to synergize our understanding of the relationship between toxicological outcomes and molecular genetics. Furthermore, toxicology and toxicogenomics are progressively developing from studies done predominantly on individual chemicals and stressors into a knowledge-based science¹¹. However, the evolution of a truly ‘predictive toxicology’ — in which the knowledge of toxicogenomic responses of a prototypic agent in one species and strain is used to predict the mode-of-action of a similar agent in a related strain or another species — will require the results of numerous toxicogenomics investigations across genotypes and species to be assimilated into a multi-domain, multi-genome knowledgebase (FIG. 2). This knowledgebase must be searchable by chemical formula or stressor type, by gene, protein or metabolite-molecular signature, or by phenotypic outcome, among other entities, to find results that are analogous to those observed with a newly tested agent. Toxicology will then have become an INFORMATION SCIENCE, and public health and risk assessment will be the beneficiaries.

The evolution of the field of toxicogenomics

Toxicogenomics has evolved from early gene-expression studies — which described the response of a biological system to a particular toxicant or panel of reference agents — towards more mature investigations that integrate several -omics domains with toxicology and pathology data (BOX 2). Exposure and outcome-specific patterns of gene, protein and metabolite profiles have been used to identify molecular changes that can be used as biomarkers of toxicity^{8,21–27} and can provide insights into mechanisms of toxicity^{28–38} and disease causation^{39–43}. Crucial to this evolution were extensive genome sequencing and annotation efforts, which are still ongoing^{44,45}, and the ability to describe response profiles in genetically and toxicologically important species, such as mouse, rat, dog and human. Another important contribution to toxicogenomics has been the formation of collaborative research consortia^{46–48} that bring together scientists from regulatory agencies, industrial laboratories, academia and governmental organizations to identify and address important issues for the field.

Profiles of response to toxicants. Nuwaysir *et al.* popularized the term ‘toxicogenomics’ to describe the use of microarrays to measure the responses of toxicologically relevant genes, and to identify selective, sensitive

BIOMARKER
A pharmacological or physiological measurement that is used to predict a toxic event in an animal.

TOXICOINFORMATICS
The description of a toxicological stress and the annotation of the dose-dependent molecular responses that are elicited over time.

INFORMATION SCIENCE
The systematic study and analysis of the sources, development, collection, organization, dissemination, evaluation, use and management of information in all its forms, including the media (formal and informal) and technology used in its communication.

Box 1 | Descriptions of selected '-omics' technologies

The terms transcriptomics, proteomics and metabonomics or METABOLOMICS refer to highly parallel, analytical technologies in which simultaneous measurements are made of expressed genes, proteins or metabolites. These technologies are used to ascertain the function of the genome. Toxicogenomics makes use of all of these FUNCTIONAL GENOMICS technologies in the study of toxicology. The terms toxicoproteomics and toxicometabolomics are sometimes used in a technology-centric sense to discuss the response of the proteome or metabolome to toxicants.

Transcriptomics — cDNA microarray hybridization and analysis

Early gene-expression profiling experiments that were carried out for toxicogenomics studies used cDNA microarrays⁴⁹. Although this cDNA technology is rapidly being supplanted by synthetic-oligonucleotide — short and long — microarrays the technological concepts underlying the two approaches are mostly analogous: cDNAs are derived from sequence-verified clones representing the 3' ends of the genes, which are either spotted onto glass slides using a robotic arrayer or synthesized *in situ*. Each RNA sample is labelled with dye-conjugated dUTP (deoxyuridine triphosphate) by reverse transcription from an oligo-dT (deoxythymine) primer. The fluorescently labelled cDNAs are then hybridized to the microarray and the microarray is scanned using laser excitation of the fluorophores¹⁹. Raw pixel intensity-images that are derived from the scanner are analysed to locate targets on the array, measure local background for each target and subtract it from the target intensity value.

Proteomics

An established proteomics strategy⁹⁰ uses global protein-stratification systems, such as PAGE, followed by protein identification through mass spectrometry. Two-dimensional PAGE separation, by charge and by mass, can resolve thousands of proteins to near homogeneity. This separation is a necessary prerequisite to enzymatic digestion and mass-spectrometry identification, which requires unique peptide-fingerprint masses or amino-acid sequence tags. Where proteins are separated by liquid chromatography instead of PAGE, a new and promising platform that involves multidimensional liquid chromatography can be used to fractionate and reduce the complexity of the protein mixture before peptide sequencing by mass spectrometry or TANDEM MASS SPECTROMETRY. This approach is being augmented by SELDI (surface enhanced laser desorption/ionization) time-of-flight mass spectrometry; a method that results in the isolation of tens-of-thousands to hundreds-of-thousands of low-molecular-weight fragments that represent a proteome.

Metabolomics and metabonomics

Quantitative analytical methods have been developed to identify metabolites in pathways or classes of compounds. This collective directed approach has been called metabolite profiling or metabolomics. Semi-quantitative, nuclear-magnetic-resonance (NMR) based metabolic fingerprinting has also been applied to high-abundance metabolites and has been termed 'metabonomics'⁹¹. Peaks detected in NMR spectra carry information regarding the structure of the metabolites, whereas peaks detected by mass spectrometry have associated molecular weights. In addition, specific mass-spectrometry methods can be established to fragment the parent molecule, allowing metabolites to be identified through investigation of fragmentation patterns.

METABOLOMICS

The directed use of quantitative analytical methods for analysing the entire metabolic content of a cell or organism at a given time (the metabolome).

FUNCTIONAL GENOMICS

The development and application of global (genome-wide or system-wide) experimental approaches to assess gene function by making use of information and reagents provided by physical mapping and sequencing of genomes.

TANDEM MASS SPECTROMETRY

The use of two mass spectrometers in series to detect and identify substances on the basis of mass and charge.

SYNOVIOCYTES

Cells believed to be responsible for the production of synovial-fluid components in joints, for absorption from the joint cavity, and for blood/synovial fluid exchanges.

CHONDROCYTES

Cartilage cells that produce the structural components of cartilage.

biomarkers of toxicity⁴⁹. The first published toxicogenomics study compared the gene-expression profiles of human cells responding to the inflammatory agent lipopolysaccharide (LPS) with those responding to mitogenic activation by phorbol myristate acetate (PMA)⁵⁰. RNA samples that were isolated at various times after exposure showed the expected increases in cytokine, chemokine and matrix metalloproteinase transcripts. Similar expression profiles were seen in SYNOVIOCYTES and CHONDROCYTES from a patient with rheumatoid arthritis, confirming the ability of the system to mimic the biological changes that occur during inflammatory disease. Subsequent studies extended this type of observation to other tissues and for a wide range of toxicants, enabling the association of specific molecular profiles with specific toxicities.

Phenotypic anchoring. Conventional toxicology has used surrogate markers that are correlated with toxic responses to monitor adverse outcomes in inaccessible tissues⁵¹. For example, the liver enzymes alkaline phosphatase (ALT) and aspartate aminotransferase (AST) are released after hepatic damage has occurred, and concentrations of these enzymes that are found in serum correlate with histopathological changes in the liver^{51,52}. These serum

enzyme markers, in conjunction with histopathology, facilitate the 'phenotypic anchoring' of molecular-expression data^{7,11,53}. Phenotypic anchoring is the process of determining the relationship between a particular expression profile and the pharmacological or toxicological phenotype of the organism for a particular exposure or dose and at a particular time⁷. The dose and time alone are often insufficient to define the toxicity experienced by an individual animal, so another measure of toxicity is needed for the full interpretation of the data obtained during a toxicogenomics study. Conversely, the phenotype alone might be insufficient to anchor the molecular profile, because an elevated value for ALT in serum can be observed both before peak toxicity (as it rises) and after peak toxicity (as it returns to baseline). Therefore, anchoring the molecular-expression profile in phenotype, dose and time helps to define the sequence of key molecular events in the mode-of-action of a toxicant.

Phenotypic anchoring can also be used in conjunction with lower doses of the toxicant to classify agents and to explore the mechanisms of toxicity that occur before histopathological changes are seen. For example, transcriptional changes that occur after both low- and high-dose exposures of acetaminophen were identified,

Box 2 | **The scope and evolution of toxicogenomics****Toxicogenomics tools and model systems**

Toxicogenomics began with 'toxicology-specific' cDNA microarrays designed to measure the amounts of acute phase and xenobiotic-metabolizing enzymes such as cytochrome P450s^{49,99}. These were superseded as commercial platforms were developed for toxicologically important species such as rat. The armamentarium of pre-clinical gene-expression platforms was completed with the canine microarray⁴⁵. It is now possible to use commercial oligonucleotide microarrays to measure expression responses in species ranging from nematodes (*Caenorhabditis elegans*) to frogs (*Xenopus laevis*), to zebrafish (*Danio rerio*) to rodents (rat and mouse) and non-human primates to man. Toxicogenomics tools for sentinel aquatic species have been developed as well⁶³. Later experiments began to focus on more challenging subjects such as subcellular organelles¹⁰⁰, non-standard tissue such as saliva¹⁰¹, less well-characterized species¹⁰², genetic models of diseases¹⁰³ and integration of data from different '-omics' disciplines^{29,37,42,63,64,104}. Additionally, comprehensive studies of yeast have become increasingly important^{83,84,105}.

Some tissues used in toxicogenomics studies

Most toxicogenomics studies so far have involved hepatotoxicants^{8,19,21–27,30,32,33,35–38,40,42,43,106}, as the liver is the primary source of xenobiotic metabolism and detoxification and because liver injury is the principal reason for withdrawal of new drugs from the market¹⁰⁷. However, toxicogenomics studies have also addressed, for example, nephrotoxicity^{27,28,34}, neurotoxicity^{108,109} and reproductive toxicity³¹, as well as lung toxicity^{22,39}, skin toxicity¹¹⁰ and cardiotoxicity¹¹¹.

Phenotypic anchoring

Phenotypic anchoring relates expression profiles to specific adverse effects defined by conventional measures of toxicity such as histopathology or clinical chemistry^{7,11,53}. Experiments have been designed to correlate expression patterns with disease pathologies such as necrosis, apoptosis, fibrosis or inflammation^{19,21,39,45,112}. Also, phenotypic anchoring can be used to provide the biological context for toxicogenomics observations made at subtoxic doses^{24,36}.

Some classes of toxicants characterized

Studies have examined responses to toxicants with established mechanisms of toxicity^{21,26,27,32,33,35,43,113}, environmental toxicants^{40,63,110,114} or exposure to suprapharmacological concentrations of drugs^{22,24,29,30,36,37,42,106,111}.

Examples of toxicant or stressor mechanisms

Acetaminophen^{24,29,37,106,115} (see also BOX 4); oestrogenic agents^{31,116}; oxidant stress^{105,117}; and peroxisome proliferators^{8,25,27,33,35}. Peroxisome proliferators are compounds that induce increased numbers of peroxisomes — single-membrane cytoplasmic organelles that metabolize long-chain fatty acids.

Importance of reporting husbandry and other technical details

Expression profiles are altered by experimental conditions including the harvesting method, the *in vitro* culture method, the vehicle used to deliver an agent, the time of day of sacrifice and the diet. Up to 9% of the transcripts in mouse liver fluctuated with circadian cycling¹¹⁸. These included genes that control glucose metabolism and vesicle trafficking or cytoskeleton, as might be anticipated from changes in the diet of animals during the day and night. In addition, however, transcript levels of Cyp17 and Cyp2a4, which are important for steroid synthesis, and Cyp2e1, which is important for detoxification of xenobiotics, also fluctuated. These changes might be expected to affect the response to test agents, and reflect a requirement to report the time of day of dosing and sacrifice, along with the diet, vehicle, and harvest and culture methods, when summarizing or publishing results of toxicogenomics studies.

Commercial database resources for toxicogenomics profiles

Toxicogenomics studies for the purpose of developing commercial databases have been done by both GeneLogic and Iconix (<http://www.genelogic.com>; <http://www.iconixpharm.com>). These companies have each gathered data from several hundred samples produced from short-term exposures of agents at pharmacological and toxicological dose levels. Customers of both companies can access the respective databases to classify the mode-of-action of novel agents of interest.

Integration of toxicogenomics efforts

This should occur through the International Life Sciences Institute (ILSI) Committee on the Application of Toxicogenomics to Risk Assessment^{46,47}, the Toxicogenomics Research Consortium (TRC) and the Consortium on Metabonomics and Toxicology (COMET)⁴⁸. Through such consortia, the technical factors affecting data can be identified and overcome, approaches to data analysis and interpretation can be agreed on, and high-quality public datasets can be prepared. The field of toxicoproteomics is currently not represented by a consortium, although the ILSI Genomics Committee and the TRC are working toxicogenomics consortia in transcriptomics and COMET is a working toxicogenomics consortium in metabonomics.

Integration of data domains

Integration of data can provide a more complete picture of the expression profiles that are associated with a particular treatment, shedding light not only on what the cell is planning (transcriptomics), but what occurred in the proteome and metabolome^{29,37,63,64,119}.

which indicates that biological responses can be detected using transcriptome measurements before histopathological changes are easily detected²⁴. Additionally, phenotypic anchoring can help to explain a toxicant's mechanism of action. For example, the

transcriptional responses in a rat model to superpharmaceutical doses of WAY-144122 (a negative regulator of insulin) were observed before histopathological changes were seen in either the liver or ovaries, and reflected different mechanisms of toxicity in the two organs³⁶.

Biomarkers. Some toxicities lack conventional biomarkers, which leads to increased risk in clinical trials and motivates the search for new pre-clinical biomarkers to support drug development. A class of LEAD COMPOUNDS identified in a discovery programme based on γ -secretase inhibition as therapy for Alzheimer disease also have been found to have an undesirable effect of inhibiting cleavage of the *Hes1* gene-product by Notch1; a process that is important for the differentiation of intestinal epithelial cells. Through the use of gene-expression profiling and subsequent protein analysis, Searfoss *et al.*⁵⁴ identified adipsin as a biomarker for this toxicity.

Carcinogenic potential is conventionally measured using a 2-year study, incurring notable expense in both animal and human resources. It is therefore of great interest to identify biomarkers of carcinogenicity that can be detected in acute, short-term studies, and efforts towards achieving this have been reported^{19,23,41–43,55}. Biomarkers with clinical relevance have also been found using toxicogenomics approaches. For example, Petricoin *et al.*⁵⁶ found a set of protein markers that distinguished patients with high concentrations of prostate-specific antigen (PSA) — a clinical marker correlated with prostate cancer — from those with low PSA concentrations who were therefore presumed to be healthy. In addition, the marker set also correctly predicted the 71% of patients with intermediate PSA concentrations.

Consortia. The issues facing toxicogenomics are larger than can be solved by scientists independently, and the rapid advancement of the field requires common efforts towards data collection and comparison. Three main collaborative research consortia have been formed, principally to standardize measurements and to guide the interpretation of toxicogenomics experiments. These groups of scientists — formed from industry-based, government and academic laboratories, as well as from regulatory agencies — were organized by research institutions to address a relevant scientific question.

The ILSI (International Life Sciences Institute) Health and Environmental Services Institute Genomics Committee — the first of these groups — began its work in 1999 and reported its main findings in 2004. These findings included the mechanisms of toxicity of several agents (the hepatotoxicants clofibrate and methapyrilene⁴⁶, and the nephrotoxicants cisplatin⁵⁷, gentamicin and puromycin⁵⁸), the successful applications of toxicogenomics to genotoxicity⁵⁵ and the establishment of a collaboration with the EMBL-European Bioinformatics Institute to develop a database¹. The second group, the Toxicogenomics Research Consortium of the NIEHS (National Institute of Environmental Health Services) National Center for Toxicogenomics (NCT), is engaged in a project to standardize toxicogenomics investigations and to analyse environmental stress responses. In 2003, the Consortium for Metabonomics Technology (COMET) — the third group — reported its interim progress towards producing a metabonomics database containing studies of

80 agents⁴⁸. Member laboratories reported data that were free of inter-laboratory bias, indicating that the COMET standardized method was robust, and that findings obtained in different laboratories could be subjected to LONGITUDINAL DATA MINING for patterns that are associated with various toxicity endpoints.

The ILSI Genomics Committee also found that microarray results from different laboratories and different platforms were comparable in the identification of a common biological-response profile, although the responses of individual genes contributing to the pattern differed between platforms^{57,59–61}. This, together with the metabonomics reproducibility reported by COMET, is a vitally important finding that supports the use of public toxicogenomics databases for meaningful meta-analysis of results obtained in different laboratories. Although some researchers^{5,62} are concerned that the capacity to assemble data on drug and toxicant effects using these technologies could result in inappropriate safety and risk decisions, collective efforts such as these will do much to help develop scientific consensus on the appropriate uses of gene-expression data.

Integration of data

A key objective in toxicogenomics is to integrate data from different studies and analytical platforms to produce a richer and biologically more refined understanding of the toxicological response of a cell, organ or organism (BOX 3). For example, one goal would be to describe the interplay between protein function and gene expression, or between the activity of certain metabolizing enzymes and the excretion into serum or urine of populations of small metabolites. The integration of data from different domains — such as proteomics and transcriptomics^{37,63,64}, or transcriptomics and metabonomics²⁹ — has been reported. In these experiments, tissue samples that were derived from the same individual animals or from comparably treated animals were analysed in parallel using different technologies. However, the data from different studies were integrated only after a shortlist of differentially responsive transcripts or protein spots had been derived.

The experience gained from integrating global-proteomics or metabonomics data, such as spot intensities from 2D gels or metabonomics fingerprint data from NMR, tells us that cluster or PRINCIPAL-COMPONENT ANALYSIS can be done to derive global signatures of molecular expression in much the same way as in transcriptomics analyses. If biological samples segregate into unique clusters that show similar expression characteristics, further efforts can be made to discern the new proteins or metabolites that are expressed in these samples. Further steps can also be taken to evaluate these proteins or metabolites as potential biomarkers and as a means to determine the underlying toxicological response.

Although software is plentiful for managing expression-profiling data at the laboratory level, there is a compelling need for public databases that combine profile data with associated biological, chemical and toxicological endpoints¹. Comparisons of gene, protein and

LEAD COMPOUNDS

Chemicals or drugs that show promise for commercialization.

LONGITUDINAL DATA MINING

The process of locating previously unknown patterns and relationships within data that result from multiple observations of a population of genes, animals or patients.

PRINCIPAL-COMPONENT ANALYSIS

A statistical method that seeks to reduce the dimensionality of a data set by projecting the data onto new axes that align with the variability in the data.

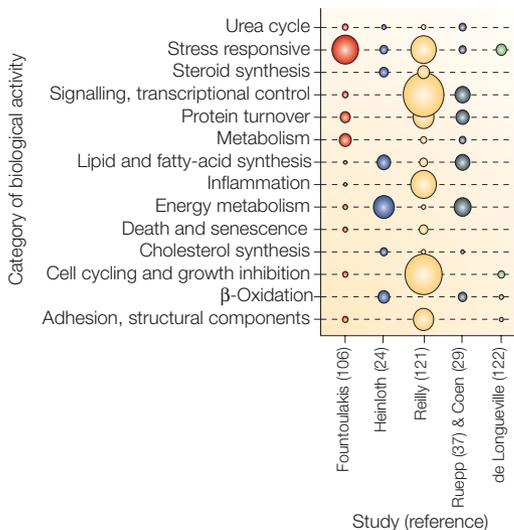
Box 3 | **Integration of acetaminophen toxicogenomics profiles**

Acetaminophen (APAP; paracetamol) overdose is a leading cause of hospitalization for acute liver failure in the United States, and its mechanism of toxicity is well-characterized¹²⁰. Data from six toxicogenomics studies are compared as an example of the power of integrating data derived under different conditions and for different purposes.

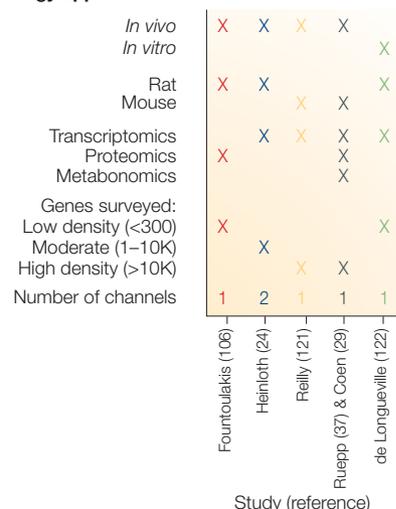
In the figure, genes with altered expression in any of the studies were organized into functional categories. Circles in part **a** represent the number of genes in a given category of biological activity identified in the referenced study. The size of a circle is proportional to the number of expressed genes (smallest circles, one gene; largest, 23 genes). The total number of genes represented is 228. The matrix in part **b** compares the referenced studies by experimental subject and the technology used. Fountoulakis *et al.*¹⁰⁶ carried out a proteomics study of C57BL/6 mouse liver after exposure to either APAP or its non-toxic isomer, AMAP. The work was based on a database of 256 identified proteins. Changes in histopathology and amounts of 35 proteins were observed at 8 hours post-exposure. The proteins that were identified were those expected on the basis of knowledge of the mechanism (free-radical-based protein adducts). AMAP-related changes were generally a subset of those seen after treatment with APAP. Heinloth *et al.*²⁴ used a high-density (6,000 genes) two-colour cDNA array to observe expression changes in F-344/N rats exposed to one of three APAP doses, at pre-toxicity (6 hours), peak toxicity (24 hours) or recovery (48 hours), as judged by conventional histopathology. Even at low doses, cellular energy loss and occasional mitochondrial damage was observed. Reilly *et al.*¹²¹ used high-density DNA single-channel oligonucleotide arrays (>11,000 genes) and RNA from livers of C57Bl/6 X 129/Ola hybrid mice 6 hours after APAP exposure. Significant alterations in nearly 100 genes from 7 principal categories of biological activity were detected. Reupp *et al.*³⁷ exposed CD1 mice to subtoxic and toxic doses of APAP, sampled at 15 minutes to 4 hours, and then did microarray (mouse ToxBlots of 450 genes) and 2D-PAGE proteomics characterizations of liver (mitochondrial subfractions). They reported changes in glutathione S-transferase, inflammatory-signalling molecules and in mitochondrial proteins within 15 minutes of exposure, before changes in transcript levels were observed. Coen *et al.*²⁹ carried out the metabolomics characterization of liver and plasma from AP1 mice exposed to APAP (from 15 minutes to 4 hours), and then integrated the findings with parallel-microarray analysis (high-density oligonucleotide array) of liver. They reported an increased rate of hepatic glycolysis and changes in lipid and energy metabolism. A study by de Longueville *et al.*¹²² used low-density arrays (59 genes) to identify changes in expression seen in rat hepatocytes *in vitro* after exposure to a range of toxicants, including APAP.

metabolite data in public databases will be valuable for promoting a global understanding of how biological systems function and respond to environmental stressors^{48,65}. As these repositories are developed, experiments will be deposited from disparate sources, using different experimental designs, but targeting the same toxicity endpoint or a similar class of toxicant. In these cases, it will be important that the databases integrate data from related studies before data mining occurs. To maximize the value of deposited datasets, the repositories must also be able to integrate data from different technological domains (see BOXES 1,3). Furthermore, a standard representation of data types in each domain is

a Relative numbers of genes expressed



b Studies compared by experimental subject and technology applied



a pre-requisite for efficient and accurate storage, access, analysis, comparison and data exchange. International standards that encompass technological and biological domains are under development by the Microarray Gene Expression Data (MGED) Society, or Reporting Structure for Biological Investigations (RSBI), Working Group (BOX 4). Furthermore, members of regulatory bodies are working with scientists from industry, academic and government laboratories participating in the ILSI Genomics Committee and **Clinical Data Interchange Standards Consortium** to develop standards for the exchange, analysis and interpretation of transcriptomics data.

Box 4 | **Databases and standards for exchange of data****Databases**

Public databases allow the scientific community to publish, share and compare the data obtained from toxicology and toxicogenomics experiments. They are a resource for data mining, and for the discovery of novel genes or proteins through their co-expression with known molecules. They also help to identify and minimize the use of experimental practices that introduce undesirable variability into toxicogenomics datasets.

Guidelines

Public data-repositories promote international database and data-exchange standards^{92–95} through guidelines developed by specific regulatory agencies. For example, the Clinical Data Interchange Standards Consortium (CDISC) develops guidelines for the electronic submission of clinical data, whereas the Standards for Exchange of Nonclinical Data (SEND) Consortium addresses the submission of toxicology study data. Minimum Information About a Microarray Experiment (MIAME) guidelines⁹⁶ specify sufficient and structured information that needs to be recorded to correctly interpret and replicate microarray experiments or to retrieve and analyse the data from a public microarray database (such as ArrayExpress (Europe)⁹⁴, GEO (US)⁹² or CIBEX (Japan)⁹⁷). Similar guidelines that describe what information should be included in a published set of toxicogenomics data are under development by the Microarray Gene Expression Data (MGED) Society. The MGED Toxicogenomics Working Group has recently broadened its scope to include environmental genomics and NUTRIGENOMICS and has changed its name to Reporting Structure for Biological Investigations (RSBI). The RSBI has proposed a tiered checklist to describe a biological investigation; such a checklist should enhance harmonization of related disciplines and reduce overlap in data collection. Tier I includes checklists (such as the ‘investigation design description’) and describes the study design — including the role of each experimental subject, the treatment given and the times of tests or other events in the study. Tier II comprises checklists for studies of toxicology (MINTox), environmental genomics (MIAME/Env) and nutrigenomics (MIAME/Nut). Tier III includes technology modules for transcriptomics (MIAME), proteomics (MIAPE^{98,126}) and metabonomics (SMRS). Modules in Tier II and III aim to collect the data necessary to interpret an investigation in a specific domain. This structure is analogous to the Object Model for Functional Genomics¹²⁷.

NUTRIGENOMICS

The study of the nutritional environment and related cellular or genetic processes at the level of the genome.

PHYSIOLOGICALLY-BASED PHARMACOKINETIC MODELLING

Involves deriving a set of mathematical (differential) equations that are structured to provide a time course for a chemical's mass–balance disposition (wherein all inputs, outputs and changes in total mass of the chemical are accounted for) in preselected anatomical compartments.

PHARMACODYNAMIC MODELLING

Involves the development of a mathematical description of a toxicological or disease outcome after therapy.

TARGET TISSUE

The tissue or tissues that are damaged as a result of exposure to a toxicant or stressor.

REAL-TIME PCR

A process that allows the amount of PCR product to be quantified during each cycle of a PCR reaction. The product concentration, as a function of cycle number, provides a good estimation of the relative quantity of the mRNA being tested.

A proposal has been made to extend toxicogenomics and combine it with computational approaches such as PHYSIOLOGICALLY-BASED PHARMACOKINETIC (PBPK) and PHARMACODYNAMIC (PD) MODELLING¹¹. PBPK modelling can be used to derive quantitative estimates of the dose of the test agent or its metabolites that are present in the TARGET TISSUE at any time after treatment, thereby allowing molecular-expression profiles to be anchored to internal dose, as well as to the time of exposure and to the toxicant-induced phenotype. Relationships between gene, protein and metabolite expression can then be described both as a function of the applied dose of an agent and the ensuing kinetic and dynamic dose-response behaviours that occur in various tissue compartments. Such models also must take into account the fact that the transcriptome, proteome and metabolome are themselves dynamic systems, and are therefore subject to significant environmental influences, such as time of day and diet^{66–68}.

Despite the numerous successes of toxicogenomics in the context of toxicology, a poorly addressed but confounding issue that is pertinent to drug safety and human risk assessment is the effect of the individual genetic background on the response of an animal or human patient. The PharmGKB pharmacogenetics knowledgebase⁶⁹ catalogues the relationship between different human genetic backgrounds and susceptibility to drug therapy. In addition, the NIEHS Environmental Genome Project⁹ is identifying SNPs in genes that are important in environmental disease, detoxification and repair. Linking toxicogenomics knowledgebases with those containing information about SNPs and human susceptibility will gradually lead to a more complete picture of the relevance of the responses and genotypes of surrogate animal species to human risk assessment.

Challenges and technical considerations

Predicting potential human health risks from chemical stressors raises three general challenges: the diverse properties of thousands of chemicals and other stressors that are present in the environment; the time and dose parameters that define the relationship between exposure to a chemical and disease; and the genetic and experiential diversity of human populations and of organisms used as surrogates to determine the adverse effects of a toxicant. FIGURE 1 illustrates the effect of genetic susceptibility on the continuum from toxic exposure to disease outcome. Knowledge of this continuum, and the role that genetics has in it, can help us to understand environmentally induced diseases, assess risk and make public-health decisions. Associated with these challenges are others of a more technical nature; these pertain specifically to toxicogenomics studies and are described below.

Technical issues. Although genome-wide alterations in mRNA, protein or metabolite concentrations in tissue extracts are clearly useful in identifying ‘signature’ gene changes, verifying that one or more gene products are involved in a toxic process depends on knowing the specific cell types in which the target-gene transcripts and products are located. Northern or western blotting, or REAL-TIME PCR, are typically used to verify the expression profile of a gene or to selectively analyse its expression as a function of toxicant dose or time of exposure. *In situ* hybridization, immunohistochemistry and other techniques can be used to identify the specific cell types that express the gene(s).

The ability to focus molecular-expression analysis on only a limited number of cell types depends on cell-separation methods that minimize the opportunity

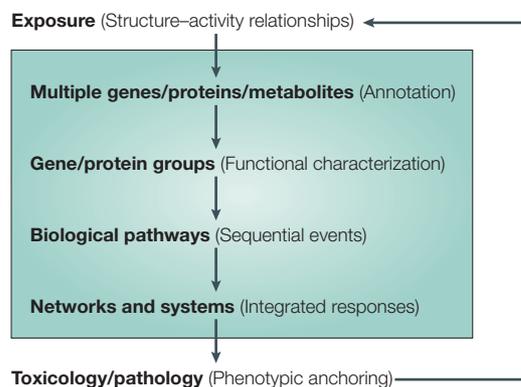


Figure 3 | Bioinformatics challenges and biological complexity. The focus of bioinformatics (in brackets) in interpreting molecular-expression data depends on the level of biological complexity (bold) — shown here progressing from genes/proteins/metabolites to networks and systems. For toxicology and/or pathology, the focus is on phenotypic anchors — observed biological responses that can be related to the chemical structure of the test agent or exposure.

for other cell types to contribute to gene expression. Even the most carefully gathered biological samples contain many cell types, especially if the sample is from inflamed or necrotic tissue. More homogeneous samples are provided by laser capture microdissection (LCM), a method that isolates individual cells or sections of tissue from a fixed sample^{70–74}. The use of LCM minimizes contributions by non-target cell populations in comparisons of diseased and normal tissue, but also introduces handling and preparation steps that can affect detection accuracy.

Concomitant with new technology that selectively samples cell populations must come the ability to reliably detect signals from increasingly smaller samples. For example, it will frequently be necessary to amplify mRNA from the biological sample used for transcriptomics analysis. The need to detect weak signals or small but biologically important changes in expression levels remains, as toxicologists explore the initial steps in biological-signalling cascades and compensatory processes. At present, cDNA microarray hybridization can detect strong signals within a mixed cell population in samples that are diluted by up to 20-fold⁷⁵. Therefore, this technology can probably detect a strong signal from a population comprising 5–10% of the total tissue, but might miss more subtle changes associated with signalling or other initial responses to a stressor. With LCM, a relatively pure cell population could be sampled, so the technology would be expected to detect much more subtle changes; for instance, responses seen only in a subpopulation, or asynchronous responses occurring in 10% of the cells at the time of sampling. The ultimate goal would be the ability to quantify genomic changes that occur in a single cell.

Although mRNA analysis is a powerful tool for recognizing toxicant-induced effects, analysis of protein structure and modification and, more importantly, of global protein-expression provides distinct advantages for understanding the functional state of the cell or tissue. Promising new methods are emerging including the capacity to profile proteins with antibody arrays⁷⁶ and surface-enhanced laser-desorption mass spectrometry^{77,78} (BOX 1). Alterations in patterns of mRNA and protein

Box 5 | The Chemical Effects in Biological Systems (CEBS) knowledgebase

To promote a systems biology approach to understanding the biological effects of environmental chemicals and stressors, the CEBS knowledgebase is being developed to house data from many complex data streams in a way that will allow extensive and complex queries from users. Unified data representation will occur through a systems-biology object model (a system for managing diverse -omics and toxicology/pathology data formats) that incorporates current standards for data capture and exchange (CEBS SysBio-OM)¹²³. Data streams will include gene expression, protein expression, interaction and changes in low-molecular-weight metabolite levels on agents studied, in addition to associated toxicology, histopathology and pertinent literature⁸⁸.

The conceptual design framework for CEBS (FIG. 4) is based on functional genomics approaches that have been used successfully for analysing yeast gene-expression datasets^{10,11,84}. Because CEBS will contain data on molecular expression, and associated chemical/stressor-induced effects in several species (for example, from yeast to humans), it will be possible to derive functional pathway and network information on the basis of cross-species homology. Genomic homology can be used within a knowledgebase such as CEBS to gain new understanding in toxicology, as well as in basic biology and genetics.

CEBS will index and sequence-align to the respective genomes all datasets known to the knowledgebase. Therefore, changes or differences in the expression patterns of entire genomes at the levels of mRNA, protein and metabolism can be determined. It will be possible to query CEBS globally; that is, to 'BLAST'¹²⁴ the knowledgebase with a profile of interest and have it return information on similar profiles observed under defined experimental conditions of dose, time and phenotype. CEBS will provide dynamic links to relevant sites such as genome browsers, animal-model databases, genetic quantitative trait loci and SNP susceptibility data, and physiologically-based pharmacokinetic and biologically-based dose response modelling. Using search routines optimized for PARSING known gene/protein groups onto toxicologically relevant pathways and networks, CEBS will automatically survey the literature and integrate this new knowledge with existing knowledgebase annotations. The current status of the CEBS infrastructure and that of other toxicogenomics databases is described in a recent review¹. These repositories offer the regulatory community reference resources for comparison with toxicogenomics data submitted in the compound registration-process¹²⁵. Progress in the development of CEBS can be monitored at <http://cebs.niehs.nih.gov>.

PARSING

The process of determining the syntactic structure of a sentence or string of symbols in a language.

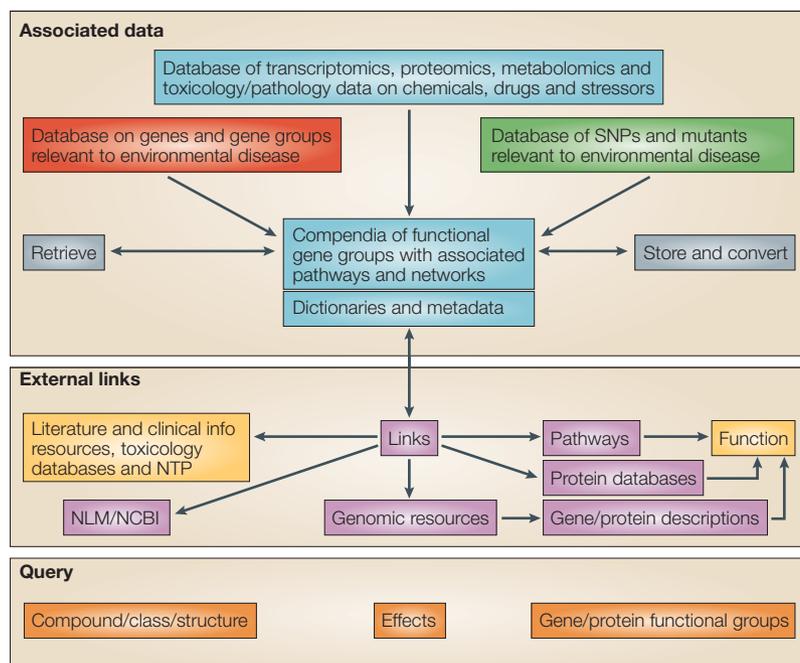


Figure 4 | Conceptual framework for the development of the Chemical Effects in Biological Systems (CEBS) knowledgebase. The CEBS knowledgebase is a cross-species toxicogenomics information system that can be used for reference to chemicals/stressors and their effects. The upper section indicates the data associated in CEBS; the central section indicates the external links from CEBS; and the lower section indicates the sample query types that CEBS will support. The boxes in the upper section include primary data (blue), important genetic loci (red) and genetic markers such as SNPs (green). The tasks that CEBS will carry out are shown in grey boxes. In the central section, the links to databases are shown in purple, and the links to unstructured data are in yellow. NTP, National Toxicology Program; NLM, National Library of Medicine; NCBI, National Center for Biotechnology Information.

expression in accessible tissues such as serum¹⁵ might offer new insights into the function of genes in the context of toxicity and guide the search for protein biomarkers of toxicant exposure or predictive toxicity.

Bioinformatics challenges. Full realization of the potential of molecular profiling in toxicogenomics requires a substantial investment in bioinformatics to extract biological sense from the myriad of interrelated numerical-molecular identifiers and their associated annotations. Advances in bioinformatics and mathematical modelling provide powerful approaches for identifying the patterns of biological response that are imbedded in genomic datasets. However, straightforward interpretation of global molecular-datasets derived from -omics technologies is currently constrained by the 'bioinformatics bottleneck'. Bioinformatics must improve in gene, protein and metabolite identification and annotation to open the field of toxicogenomics to high-throughput applications in drug development and toxicant evaluation. Several useful resources address the annotation problem by linking identifiers used in genomic databases at the National Center for Biotechnology Information (NCBI), the European Molecular Biology Laboratory (EMBL) and

the DNA Data Bank of Japan (DDBJ) to other annotation resources. Crucial to resolving annotation inconsistencies is the knowledge of the sequence of the actual nucleotide or protein that is used to query the genome.

The use of advanced bioinformatics tools to extract information from microarray results⁷⁹ is valuable only if the data used by these tools have a high degree of internal specificity and accuracy¹. Additionally, the interpretation of molecular-expression profiles must emphasize both biological coherence and statistical validity when deriving knowledge from toxicogenomics experiments. This means that once a set of genes with altered expression is identified, their biological functions must be ascertained. Mechanistic interpretation of transcript changes might be impeded by the non-standard or imprecise annotation of a sequence element (that is, a gene). Without appropriate synonyms for gene names, the effectiveness of a literature search might be limited. Differences in annotation within and among different microarray platforms might hamper the comparison of results. Such inconsistency frequently arises from annotation resources that use different lexicons, or from annotation information being compiled at different times.

Further bioinformatics and interpretive challenges arise at many levels of biological organization (FIG. 3). Our current focus and level of understanding of the global molecular-landscape encompasses only the lower levels of complexity (genes/proteins, gene/protein groups, and biological pathways). The resolution of this knowledge might be termed linear toxicoinformatics; that is, the description of environmental stimuli and responses, over dose and time, following a toxicological stress. Toxicologists and risk assessors typically define a sequence of key events and linear modes-of-action for environmental chemicals and drugs^{80–82}. By contrast, the networks and systems level of biological organization might show highly nonlinear cellular-expression state changes in response to environmental stimuli^{83,84}. Therefore, the statistical and bioinformatics-based separation of the complex adaptive, pharmacological and toxicological responses of drugs, chemicals and even dietary constituents will probably be a matter of degree — reflecting the kinetic and dynamic responses of specific tissues to toxicants as directed by the genome, the genetic heritage of the individual and that individual's current and prior exposures.

Systems toxicology

Ideker *et al.*¹⁷ used the phrase 'systems biology' to describe the integrated study of biological systems at the molecular level — involving perturbation of systems, monitoring molecular expression, integrating response data and modelling the systems' molecular structure and network function. Here, we similarly use the phrase 'systems toxicology' to describe the toxicogenomics evaluation of biological systems, involving perturbation by toxicants and stressors, monitoring molecular expression and conventional toxicological parameters, and iteratively integrating response data to model the toxicological system.

Several approaches are being developed to model network behaviour, with different assumptions, data requirements and goals. However, it is not likely that toxicogenomics and systems toxicology models will be assembled exclusively from knowledge of cellular components, without equivalent knowledge of the response of these components to toxicants⁸³. Therefore, the 'stress testing' of the structural biology of the system and the capture of that data in the context of the functioning organism adapting, surviving or succumbing to the stress will be required.

Development of a knowledgebase to accurately reflect network-level molecular expression and to facilitate a systems-level biological interpretation requires a new model of data management, data integration and computational modelling. A knowledgebase that fully embraces systems toxicology (BOX 5; FIG 4) will use precise sequence data to define macromolecules, interaction data based experimentally on co-localization, co-expression and analyses of protein-protein interactions, and functional and phenotypic data that is based on gene knockouts, knockins and RNA-INTERFERENCE studies, in addition to studies of responses to chemical, physical and biological stressors. These data will allow specific molecules to be accurately related to the biological phenomena that reflect both the normal and the stressed cell, tissue, organ or organism. In the best circumstances, a systems toxicology approach will build a toxicogenomics understanding from global molecular-expression changes that are informed by PBPK or PD modelling and BIOLOGICALLY-BASED DOSE-RESPONSE (BBDR) MODELLING. The challenge in constructing a robust systems toxicology knowledgebase is formidable.

The future of toxicogenomics

New toxicogenomics methods have the power and potential to revolutionize toxicology. Technological innovations that are already in use allow RNA profiling of formalin-fixed tissues⁸⁵, potentially making archived tissues from generations of toxicological studies accessible to gene-expression analysis. Methods that can be used to array hundreds of toxicologically relevant protein antibodies, and to profile hundreds of small molecules in high-throughput mode using gas chromatography, liquid chromatography or mass spectrometry are in development.

Toxicoproteomics research is anticipated to lead to the identification, measurement and evaluation of proteins and other biomarkers that are more accurate, sensitive and specific than those available now, and that might be targeted to particular human genetic subpopulations. Metabonomics research will help to identify alterations in the levels of small endogenous molecules as important changes in a sequence of key metabolic events; such 'metabolite fingerprints' might then help to diagnose and define the ways in which specific chemicals, environmental exposures or stressors cause disease. This, coupled with the ability to detect damage to particular organs by observing alterations in serum and urine components, is expected to lead to the more

sensitive detection of exposure or risk factors⁸⁶. Further considerations in assessing the toxicogenomic response to environmental exposures are the individual genotype, lifestyle, age and exposure history⁶⁷. Toxicogenomics will help to ascertain the degree to which these factors influence the balance between healthy and disease states.

Toxicogenomics will increase the relevance of toxicology through the global observation of genomic responses with therapeutically and environmentally realistic dose-regimens. It will help to delineate the modes-of-action of various classes of agents and the unique genetic attributes of certain species and population subgroups that render them susceptible to toxicants^{10,81}. Studies on strains within a species that are sensitive or resistant to the chemical induction of specific disease phenotypes will be particularly valuable. Extending this thinking to the phylogenetic analysis of both core, conserved biological processes⁸⁷ and to the toxicological responses seen in different species, will provide further comparative insights into genetic susceptibility and probable disease outcomes.

The combined application of the -omics technologies will improve our overall understanding of mechanisms of toxicity and disease aetiology as integrated toxicogenomics databases are developed more fully¹⁰. Data on gene, protein or metabolite changes collected in the context of dose, time, target tissue and phenotypic severity across a range of species — from yeast, to nematode, to man — will provide the comparative information needed to assess the genetic and molecular basis of gene-environment interactions. Toxicology will emerge as an information science that will facilitate scientific discovery across biological species, chemical classes and disease outcomes⁷. Although there are great challenges in developing public toxicogenomic data repositories, the nucleotide-sequence databases — GenBank, EMBL and DDBJ — provide an excellent example of the benefits of sharing data to the wider scientific and medical community.

Concomitant with development of toxicogenomics databases must be the evolution of bioinformatics methods and data-mining tools, and increased numbers of individuals who have been trained to apply them⁷⁹. We believe that a predictive systems toxicology will gradually evolve, aided by knowledge that is systematically generated²⁰ through literature mining^{88,89}, comparative analysis and iterative biological modelling of molecular-expression datasets over time. However, given the vast numbers and diversity of drugs, chemicals and environmental agents, and the diversity of species in which they act, we believe that it is only through the development of a comprehensive and public knowledgebase that toxicology and environmental health can rapidly advance. The ultimate goal of the NCT is to create the Chemical Effects in Biological Systems (CEBS) knowledgebase — a public resource (BOX 5) that will enable health scientists and practitioners to understand and mitigate or prevent adverse environmental exposures and related diseases in the twenty-first century.

RNA INTERFERENCE

An ancient natural antiviral mechanism that directs silencing of gene expression in a sequence-specific manner and can be exploited artificially to inhibit the expression of any gene of interest.

BIOLOGICALLY-BASED DOSE-RESPONSE MODELLING

The science of establishing dose-response models based on underlying biological processes.

1. Mattes, W. B., Pettit, S. D., Sansone, S. A., Bushel, P. R. & Waters, M. D. Database development in toxicogenomics: issues and efforts. *Environ. Health Perspect.* **112**, 495–505 (2004).
This paper describes the current status of development of databases related to toxicogenomics, with an emphasis on those that are public.
2. Aardema, M. J. & MacGregor, J. T. Toxicology and genetic toxicology in the new era of 'toxicogenomics': impact of '-omics' technologies. *Mutat. Res.* **499**, 13–25 (2002).
3. Afshari, C. A. Perspective: microarray technology, seeing more than spots. *Endocrinology* **143**, 1983–1989 (2002).
4. Ulrich, R. & Friend, S. H. Toxicogenomics and drug discovery: will new technologies help us produce better drugs? *Nature Rev. Drug Discov.* **1**, 84–88 (2002).
5. Fielden, M. R. & Zacharewski, T. R. Challenges and limitations of gene expression profiling in mechanistic and predictive toxicology. *Toxicol. Sci.* **60**, 6–10 (2001).
6. Hamadeh, H. K., Amin, R. P., Paules, R. S. & Afshari, C. A. An overview of toxicogenomics. *Curr. Issues Mol. Biol.* **4**, 45–56 (2002).
7. Tennant, R. W. The National Center for Toxicogenomics: using new technologies to inform mechanistic toxicology. *Environ. Health Perspect.* **110**, A8–A10 (2002).
The author describes the field of toxicogenomics and the goals of the NCT, as well as important concepts such as phenotypic anchoring.
8. Thomas, R. S. *et al.* Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol. Pharmacol.* **60**, 1189–1194 (2001).
9. Olden, K. & Guthrie, J. Genomics: implications for toxicology. *Mutat. Res.* **473**, 3–10 (2001).
10. Waters, M. D., Olden, K. & Tennant, R. W. Toxicogenomic approach for assessing toxicant-related disease. *Mutat. Res.* **544**, 415–424 (2003).
11. Waters, M. D. *et al.* Systems toxicology and the chemical effects in biological systems knowledge base. *Environ. Health Perspect.* **111**, 811–824 (2003).
This paper proposes the concept and strategy for building a true toxicogenomics knowledgebase, integrating -omics and conventional data domains, and anchoring data in sequence, phenotype and chemical structure.
12. Lobenhofer, E. K., Bushel, P. R., Afshari, C. A. & Hamadeh, H. K. Progress in the application of DNA microarrays. *Environ. Health Perspect.* **109**, 881–891 (2001).
13. Burchiel, S. W. *et al.* Analysis of genetic and epigenetic mechanisms of toxicity: potential roles of toxicogenomics and proteomics in toxicology. *Toxicol. Sci.* **59**, 193–195 (2001).
14. Merrick, B. A. & Tomer, K. B. Toxicoproteomics: a parallel approach to identifying biomarkers. *Environ. Health Perspect.* **111**, A578–A579 (2003).
15. Petricoin, E. F. *et al.* Toxicoproteomics: serum proteomic pattern diagnostics for early detection of drug induced cardiac toxicities and cardioprotection. *Toxicol. Pathol.* **32** (Suppl. 1), 122–130 (2004).
16. Wilkins, M. R. *et al.* From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (NY)* **14**, 61–65 (1996).
17. Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372 (2001).
18. Nurse, P. Understanding cells. *Nature Biotechnol.* **21**, 883 (2003).
19. Hamadeh, H. K. *et al.* Methapyriene toxicity: anchorage of pathologic observations to gene expression alterations. *Toxicol. Pathol.* **30**, 470–482 (2002).
The authors describe the methods that have been developed to anchor expression changes to pathology phenotypes, to identify expression changes seen before the detection of overt histopathology and to explore the mechanism of toxicity of a reference hepatotoxicant.
20. Zweiger, G. Knowledge discovery in gene-expression-microarray data: mining the information output of the genome. *Trends Biotechnol.* **17**, 429–436 (1999).
21. Waring, J. F. *et al.* Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharmacol.* **175**, 28–42 (2001).
22. Mortuza, G. B., Neville, W. A., Delaney, J., Waterfield, C. J. & Camilleri, P. Characterisation of a potential biomarker of phospholipidosis from amiodarone-treated rats. *Biochim. Biophys. Acta* **1631**, 136–146 (2003).
23. Kramer, J. A. *et al.* Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. *Chem. Res. Toxicol.* **17**, 463–470 (2004).
This paper identifies molecular markers seen after acute exposure to chemical carcinogens from various classes, making possible a short-term test for potential rodent carcinogens.
24. Heinloth, A. N. *et al.* Gene expression profiling of rat livers reveals indicators of potential adverse effects. *Toxicol. Sci.* **80**, 193–202 (2004).
25. Hamadeh, H. K. *et al.* Prediction of compound signature using high density gene expression profiling. *Toxicol. Sci.* **67**, 232–240 (2002).
26. Bulera, S. J. *et al.* RNA expression in the early characterization of hepatotoxicants in Wistar rats by high-density DNA microarrays. *Hepatology* **33**, 1239–1258 (2001).
27. Bartosiewicz, M. J., Jenkins, D., Penn, S., Emery, J. & Buckpitt, A. Unique gene expression patterns in liver and kidney associated with exposure to chemical toxicants. *J. Pharmacol. Exp. Ther.* **297**, 895–905 (2001).
28. Cutler, P. *et al.* An integrated proteomic approach to studying glomerular nephropathy. *Electrophoresis* **20**, 3647–3658 (1999).
29. Coen, M. *et al.* Integrated application of transcriptomics and metabolomics yields new insight into the toxicity due to paracetamol in the mouse. *J. Pharm. Biomed. Anal.* **35**, 93–105 (2004).
30. Donald, S. *et al.* Hepatobiliary damage and changes in hepatic gene expression caused by the antitumor drug eteinscinidin-743 (ET-743) in the female rat. *Cancer Res.* **62**, 4256–4562 (2002).
31. Fertuck, K. C., Eckel, J. E., Gennings, C. & Zacharewski, T. R. Identification of temporal patterns of gene expression in the uteri of immature, ovariectomized mice following exposure to ethynylestradiol. *Physiol. Genomics* **15**, 127–141 (2003).
This article describes the united expression changes and temporal patterns of physiological responses to an oestrogenic agent, with the aim of producing a mechanism-based understanding of effects on cell cycle, proliferation, tissue remodelling and related processes.
32. Fountoulakis, M. *et al.* Modulation of gene and protein expression by carbon tetrachloride in the rat liver. *Toxicol. Appl. Pharmacol.* **183**, 71–80 (2002).
33. Hamadeh, H. K. *et al.* Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.* **67**, 219–231 (2002).
34. Huang, Q. *et al.* Assessment of cisplatin-induced nephrotoxicity by microarray technology. *Toxicol. Sci.* **63**, 196–207 (2001).
35. Kramer, J. A. *et al.* Transcription profiling distinguishes dose-dependent effects in the livers of rats treated with clofibrate. *Toxicol. Pathol.* **31**, 417–431 (2003).
36. Peterson, R. L. *et al.* Mechanistic toxicogenomic analysis of WAY-144122 administration in Sprague-Dawley rats. *Toxicol. Appl. Pharmacol.* **196**, 80–94 (2004).
37. Ruepp, S. U., Tonge, R. P., Shaw, J., Wallis, N. & Pognan, F. Genomics and proteomics analysis of acetaminophen toxicity in mouse liver. *Toxicol. Sci.* **65**, 135–150 (2002).
38. Waring, J. F. *et al.* Identifying toxic mechanisms using DNA microarrays: evidence that an experimental inhibitor of cell adhesion molecule expression signals through the aryl hydrocarbon nuclear receptor. *Toxicology* **181–182**, 537–550 (2002).
This paper reports for the first time the methodology of toxicant profiling; a panel of responses to 15 reference hepatotoxicants was used to identify the probable mechanism of toxicity of a novel agent. The prediction was confirmed through other laboratory tests.
39. Wagenaar, G. T. M. *et al.* Gene expression profile and histopathology of experimental bronchopulmonary dysplasia induced by prolonged oxidative stress. *Free Radic. Biol. Med.* **36**, 782–801 (2004).
40. Lu, T. *et al.* Application of cDNA microarray to the study of arsenic-induced liver diseases in the population of Guizhou, China. *Toxicol. Sci.* **59**, 185–192 (2001).
41. Hamadeh, H. K. *et al.* Integration of clinical and gene expression endpoints to explore furan-mediated hepatotoxicity. *Mutat. Res.* **549**, 169–183 (2004).
42. Iida, M. *et al.* Changes in global gene and protein expression during early mouse liver carcinogenesis induced by non-genotoxic model carcinogens oxazepam and Wyeth-14,643. *Carcinogenesis* **24**, 757–770 (2003).
43. Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bornmann, W. & Ahr, H. J. Characteristic expression profiles induced by genotoxic carcinogens in rat liver. *Toxicol. Sci.* **77**, 19–34 (2004).
44. Twigger, S. *et al.* Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res.* **30**, 125–128 (2002).
45. Higgins, M. A. *et al.* Gene expression analysis of the acute phase response using a canine microarray. *Toxicol. Sci.* **74**, 470–484 (2003).
The authors describe the development of the first canine array. It was used to characterize the acute phase response to inflammation that is seen in parallel with other phenotypic markers in the beagle, an important model organism in late stage, pharmaceutical toxicity testing.
46. Ulrich, R. G., Rockett, J. C., Gibson, G. G. & Pettit, S. D. Overview of an interlaboratory collaboration on evaluating the effects of model hepatotoxicants on hepatic gene expression. *Environ. Health Perspect.* **112**, 423–427 (2004).
47. Pennie, W., Pettit, S. D. & Lord, P. G. Toxicogenomics in risk assessment: an overview of an HESI collaborative research program. *Environ. Health Perspect.* **112**, 417–419 (2004).
This provides an overview of the toxicogenomics consortium supported by the Health and Environmental Sciences Institute: the ILSI Committee on the Application of Toxicogenomics to Risk Assessment.
48. Lindon, J. C. *et al.* Contemporary issues in toxicology: the role of metabolomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharmacol.* **187**, 137–146 (2003).
This paper led COMET to apply metabolomics studies to toxicogenomics
49. Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C. & Afshari, C. A. Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinog.* **24**, 153–159 (1999).
50. Heller, R. A. *et al.* Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl Acad. Sci. USA* **94**, 2150–2155 (1997).
51. Loeb, W. F. & Quimby, F. W. (eds) *The Clinical Chemistry of Laboratory Animals* (Taylor and Francis, 1999).
52. Travlos, G. S. *et al.* Frequency and relationships of clinical chemistry and liver and kidney histopathology findings in 13-week toxicity studies in rats. *Toxicology* **107**, 17–29 (1996).
53. Paules, R. Phenotypic anchoring: linking cause and effect. *Environ. Health Perspect.* **111**, A338–A339 (2003).
The author describes the use of the term 'phenotypic anchoring', first mentioned by Tennant in reference 7 to describe the process of using a pathology finding or toxicological endpoint to provide context for expression changes.
54. Searfoss, G. H. *et al.* Adipsin, a biomarker of gastrointestinal toxicity mediated by a functional γ -secretase inhibitor. *J. Biol. Chem.* **278**, 46107–46116 (2003).
For the first time, the authors apply the process of identifying a molecular marker of toxicity to support pre-clinical studies, and define the mechanism of toxicity associated with a novel class of toxicant.
55. Newton, R. K., Aardema, M. & Aubrecht, J. The utility of DNA microarrays for characterizing genotoxicity. *Environ. Health Perspect.* **112**, 420–422 (2004).
56. Petricoin, E. F. 3rd *et al.* Serum proteomic patterns for detection of prostate cancer. *J. Natl Cancer Inst.* **94**, 1576–1578 (2002).
57. Thompson, K. L. *et al.* Identification of platform-independent gene expression markers of cisplatin nephrotoxicity. *Environ. Health Perspect.* **112**, 488–494 (2004).
58. Kramer, J. A. *et al.* Overview on the application of transcription profiling using selected nephrotoxicants for toxicology assessment. *Environ. Health Perspect.* **112**, 460–464 (2004).
59. Baker, V. A. *et al.* Clofibrate-induced gene expression changes in rat liver: a cross-laboratory analysis using membrane cDNA arrays. *Environ. Health Perspect.* **112**, 428–438 (2004).
60. Chu, T. M., Deng, S., Wolfinger, R., Paules, R. S. & Hamadeh, H. K. Cross-site comparison of gene expression data reveals high similarity. *Environ. Health Perspect.* **112**, 449–455 (2004).
61. Waring, J. F. *et al.* Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyriene. *Environ. Health Perspect.* **112**, 439–448 (2004).
62. Smith, L. L. Key challenges for toxicologists in the 21st century. *Trends Pharmacol. Sci.* **22**, 281–285 (2001).
63. Hogstrand, C., Balesaria, S. & Glover, C. N. Application of genomics and proteomics for study of the integrated response to zinc exposure in a non-model fish species, the rainbow trout. *Comp. Biochem. Physiol. B* **133**, 523–535 (2002).
64. Juan, H.-F. *et al.* Biomic study of human myeloid leukemia cells differentiation to macrophages using DNA array, proteomic, and bioinformatic analytical methods. *Electrophoresis* **23**, 2490–2504 (2002).
65. Amin, R. P. *et al.* Genomic interrogation of mechanism(s) underlying cellular responses to toxicants. *Toxicology* **181–182**, 555–563 (2002).
66. Kita, Y. *et al.* Implications of circadian gene expression in kidney, liver and the effects of fasting on pharmacogenomic studies. *Pharmacogenomics* **12**, 55–65 (2002).
67. Kaput, J. & Rodriguez, R. L. Nutritional genomics: the next frontier in the postgenomic era. *Physiol. Genomics* **16**, 166–177 (2004).
68. Kaput, J. Diet-disease gene interactions. *Nutrition* **20**, 26–31 (2004).
69. Klein, T. E. & Altman, R. B. PharmGKB: the pharmacogenetics nad pharmacogenomics knowledge base. *Pharmacogenomics J.* **4**, 1 (2004).

70. Wittliff, J. L. & Erlander, M. G. Laser capture microdissection and its applications in genomics and proteomics. *Methods Enzymol.* **356**, 12–25 (2002).
71. Jain, K. K. Application of laser capture microdissection to proteomics. *Methods Enzymol.* **356**, 157–167 (2002).
72. Emmert-Buck, M. R. *et al.* Laser capture microdissection. *Science* **274**, 998–1001 (1996).
73. Bonner, R. F. *et al.* Laser capture microdissection: molecular analysis of tissue. *Science* **278**, 1481–1483 (1997).
74. Karsten, S. L., Van Deerlin, V. M., Sabatti, C., Gill, L. H. & Geschwind, D. H. An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. *Nucleic Acids Res.* **30**, E4 (2002).
75. Hamadeh, H. K. *et al.* Detection of diluted gene expression alterations using cDNA microarrays. *Biotechniques* **32**, 322, 324, 326–329 (2002).
76. Huang, R. P. Detection of multiple proteins in an antibody-based protein microarray system. *J. Immunol. Methods* **255**, 1–13 (2001).
77. Merchant, M. & Weinberger, S. R. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* **21**, 1164–1177 (2000).
78. Liotta, L. & Petricoin, E. Molecular profiling of human cancer. *Nature Rev. Genet.* **1**, 48–56 (2000).
79. Quackenbush, J. Computational analysis of microarray data. *Nature Rev. Genet.* **2**, 418–427 (2001).
80. Farland, W. H. The U. S. Environmental Protection Agency's Risk Assessment Guidelines: current status and future directions. *Toxicol. Ind. Health* **8**, 205–212 (1992).
81. Farland, W. H. Cancer risk assessment: evolution of the process. *Prev. Med.* **25**, 24–25 (1996).
82. Larsen, J. C., Farland, W. & Winters, D. Current risk assessment approaches in different countries. *Food Addit. Contam.* **17**, 359–369 (2000).
83. Begley, T. J., Rosenbach, A. S., Ideker, T. & Samson, L. D. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol. Cancer Res.* **1**, 103–112 (2002).
84. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cel* **102**, 109–126 (2000).
85. Lewis, F., Maughan, N. J., Smith, V., Hillan, K. & Quirke, P. Unlocking the archive — gene expression in paraffin-embedded tissue. *J. Pathol.* **195**, 66–71 (2001).
86. Wetmore, B. A. & Merrick, B. A. Toxicoproteomics: proteomics applied to toxicology and pathology. *Toxicol. Pathol.* **32**, 619–642 (2004).
87. Stuart, G. W. & Berry, M. W. A comprehensive whole genome bacterial phylogeny using correlated peptide motifs defined in a high dimensional vector space. *J. Bioinform. Comput. Biol.* **1**, 475–493 (2003).
88. Chaussabel, D. & Sher, A. Mining microarray expression data by literature profiling. *Genome Biol.* **3**, research0055.1–research0055.16 (2002).
89. Sluka, J. P. In *Methods of Microarray Data Analysis II* (eds Lin, S. M. & Johnson, K. F.) (Kluwer Academic Publishers, Boston, 2002).
90. Patterson, S. D. & Aebersold, R. H. Proteomics: the first decade and beyond. *Nature Genet.* **33**, S311–S323 (2003).
91. Nicholson, J. K., Connelly, J., Lindon, J. C. & Holmes, E. Metabonomics: a platform for studying drug toxicity and gene function. *Nature Rev. Drug Discov.* **1**, 153–161 (2002).
92. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
93. Mattingly, C. J., Colby, G. T., Rosenstein, M. C., Forrest, J. N. Jr & Boyer, J. L. Promoting comparative molecular studies in environmental health research: an overview of the comparative toxicogenomics database (CTD). *Pharmacogenomics J.* **4**, 5–8 (2004).
- The authors proposed and developed a public database to house gene sequences and toxicology information in support of comparative toxicogenomics across species, and to inform human risk assessment.**
94. Brazma, A. *et al.* ArrayExpress — a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31**, 68–71 (2003).
95. Ball, C. A. *et al.* Standards for microarray data. *Science* **298**, 539 (2002).
96. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. *Nature Genet.* **29**, 365–371 (2001).
97. Ikeo, K., Ishi-i, J., Tamura, T., Gojobori, T. & Tateno, Y. CIBEX: center for information biology gene expression database. *C. R. Biol.* **326**, 1079–1082 (2003).
98. Orchard, S. *et al.* Common interchange standards for proteomics data: public availability of tools and schema. *Proteomics* **4**, 490–491 (2004).
99. Bartosiewicz, M., Trounstein, M., Barker, D., Johnston, R. & Buckpitt, A. Development of a toxicological gene array and quantitative assessment of this technology. *Arch. Biochem. Biophys.* **376**, 66–73 (2000).
100. Jiang, X. S. *et al.* A high-throughput approach for subcellular proteome: identification of rat liver proteins using subcellular fractionation coupled with two-dimensional liquid chromatography tandem mass spectrometry and bioinformatic analysis. *Mol. Cell. Proteomics* **3**, 441–455 (2004).
101. Vitorino, R. *et al.* Identification of human whole saliva protein components using proteomics. *Proteomics* **4**, 1109–1115 (2004).
102. Talamo, F. *et al.* Proteins from bovine tissues and biological fluids: defining a reference electrophoresis map for liver, kidney, muscle, plasma and red blood cells. *Proteomics* **3**, 440–460 (2003).
103. Reddy, P. H. *et al.* Gene expression profiles of transcripts in amyloid precursor protein transgenic mice: up-regulation of mitochondrial metabolism and apoptotic genes is an early cellular change in Alzheimer's disease. *Hum. Mol. Genet.* (2004).
104. Hejine, W. H., Stierum, R. H., Slijper, M., van Bladeren, P. J. & van Ommen, B. Toxicogenomics of bromobenzene hepatotoxicity: a combined transcriptomics and proteomics approach. *Biochem. Pharmacol.* **65**, 857–875 (2003).
105. Weiss, A., Delproposto, J. & Giroux, C. N. High-throughput phenotypic profiling of gene-environment interactions by quantitative growth curve analysis in *Saccharomyces cerevisiae*. *Anal. Biochem.* **327**, 23–34 (2004).
106. Fountoulakis, M. *et al.* Two-dimensional database of mouse liver proteins: changes in hepatic protein levels following treatment with acetaminophen or its nontoxic regioisomer 3-acetamidophenol. *Electrophoresis* **21**, 2148–2161 (2000).
107. Lee, W. M. Drug-induced hepatotoxicity. *N. Engl. J. Med.* **349**, 474–485 (2003).
108. Xie, T. *et al.* Changes in gene expression linked to methamphetamine-induced dopaminergic neurotoxicity. *J. Neurosci.* **22**, 274–283 (2002).
109. Dam, K., Seidler, F. J. & Slotkin, T. A. Transcriptional biomarkers distinguish between vulnerable periods for developmental neurotoxicity of chlorpyrifos: implications for toxicogenomics. *Brain Res. Bull.* **59**, 261–265 (2003).
110. Hamadeh, H. K., Trouba, K. J., Amin, R. P., Afshari, C. A. & Germolec, D. Coordination of altered DNA repair and damage pathways in arsenite-exposed keratinocytes. *Toxicol. Sci.* **69**, 306–316 (2002).
111. Hu, D., Cao, K., Peterson-Wakeman, R. & Wang, R. Altered profile of gene expression in rat hearts induced by chronic nicotine consumption. *Biochem. Biophys. Res. Commun.* **297**, 729–736 (2002).
112. Waring, J. F., Ciurlionis, R., Jolly, R. A., Heindel, M. & Ulrich, R. G. Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol. Lett.* **120**, 359–368 (2001).
113. Hamadeh, H. K., Bushel, P., Paules, R. & Afshari, C. A. Discovery in toxicology: mediation by gene expression array technology. *J. Biochem. Mol. Toxicol.* **15**, 231–242 (2001).
114. Witzmann, F. A. *et al.* Analysis of rat testicular protein expression following 91-day exposure to JP-8 jet fuel vapor. *Proteomics* **3**, 1016–1027 (2003).
115. Huang, Q. *et al.* Gene expression profiling reveals multiple toxicity endpoints induced by hepatotoxicants. *Mutat. Res.* **549**, 147–167 (2004).
116. Adachi, T. *et al.* Toxicogenomic difference between diethylstilbestrol and 17 β -estradiol in mouse testicular gene expression by neonatal exposure. *Mol. Reprod. Dev.* **67**, 19–25 (2004).
117. Nadadur, S. S., Schladweiler, M. C. & Kodavanti, U. P. A pulmonary rat gene array for screening altered expression profiles in air pollutant-induced lung injury. *Inhal. Toxicol.* **12**, 1239–1254 (2000).
118. Akhtar, R. A. *et al.* Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr. Biol.* **12**, 540–550 (2002).
119. Hejine, W. H. *et al.* Bromobenzene-induced hepatotoxicity at the transcriptome level. *Toxicol. Sci.* **79**, 411–422 (2004).
120. Bessems, J. G. & Vermeulen, N. P. Paracetamol (acetaminophen)-induced toxicity: molecular and biochemical mechanisms, analogues and protective approaches. *Crit. Rev. Toxicol.* **31**, 55–138 (2001).
121. Reilly, T. P. *et al.* Expression profiling of acetaminophen liver toxicity in mice using microarray technology. *Biochem. Biophys. Res. Commun.* **282**, 321–328 (2001).
122. de Longueville, F. *et al.* Use of a low-density microarray for studying gene expression patterns induced by hepatotoxicants on primary cultures of rat hepatocytes. *Toxicol. Sci.* **75**, 378–392 (2003).
123. Xirasagar, S. *et al.* CEBS object model for systems biology data. *SysBio-OM. Bioinformatics* **20**, 2004–2015 (2004).
124. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
125. Petricoin, E. F. 3rd *et al.* Medical applications of microarray technologies: a regulatory science perspective. *Nature Genet.* **32**, S474–S479 (2002).
126. Hermjakob, H. *et al.* The HUPO PSI's molecular interaction format — a community standard for the representation of protein interaction data. *Nature Biotechnol.* **22**, 177–183 (2004).
127. Jones, A., Hunt, E., Wastling, J., Pizarro, A. & Stoekert, C. J. Jr. An object model and database for functional genomics. *Bioinformatics* (2004).

Acknowledgments
We are indebted to B. Alex Merrick, Richard S. Paules and Raymond W. Tennant for their consultation and assistance with this manuscript, and to Kenneth Olden, Samuel Wilson, Lutz Birnbaumer and the staff of the National Center for Toxicogenomics for their continuing support and involvement with this work.

Competing interests statement
The authors declare no competing financial interests.

 **Online links**

DATABASES
The following terms in this article are linked online to:
OMIM: <http://www.ncbi.nlm.nih.gov/Omim>
Alzheimer disease

FURTHER INFORMATION
Chemical Effects in Biological Systems:
<http://cebs.niehs.nih.gov>
CEBS Development Forum:
<http://niehs.nih.gov/cebs-df>
Clinical Data Interchange Standards Consortium:
<http://www.cdisc.org>
EMBL-European Bioinformatics Institute:
<http://www.ebi.ac.uk/microarray/Projects/tox-nutri>
'From OMICS to systems biology' poster
<http://www.nature.com/nrg/journal/v5/n10/poster/omics>
ILSI Health and Environmental Services Institute:
<http://hesi.ilsil.org>
MGED Reporting Structure for Biological Investigations (RSBI):
<http://www.mged.org/Workgroups/rsbi>
NIEHS National Center for Toxicogenomics:
<http://www.niehs.nih.gov/nct>
Toxicogenomics Research Consortium:
<http://www.niehs.nih.gov/nct/trc.htm>
Access to this interactive links box is free online.