Curr. Issues Mol. Biol. (2002) 4: 45-56.

An Overview of Toxicogenomics

Hisham K. Hamadeh, Rupesh P. Amin, Richard S. Paules, Cynthia A. Afshari*

National Institute of Environmental Health Sciences,
Intramural Microarray Center, PO Box 12233 MD2-04,
111 Alexander Drive Bldg. 101,
Research Triangle Park, NC 27709, USA

Abstract

Toxicogenomics is a rapidly developing discipline that promises to aid scientists in understanding the molecular and cellular effects of chemicals in biological systems. This field encompasses global assessment of biological effects using technologies such as DNA microarrays or high throughput NMR and protein expression analysis. This review provides an overview of advancing multiple approaches (genomic, proteomic, metabonomic) that may extend our understanding of toxicology and highlights the importance of coupling such approaches with classical toxicity studies.

Background and Definition of Toxicogenomics

The field of toxicology is defined as the study of stressors and their adverse effects. One sub-discipline deals with hazard identification, mechanistic toxicology, and risk assessment. Increased understanding of the mechanism of action of chemicals being assayed will improve the efficiency of these tasks. However, the derivation of mechanistic knowledge traditionally evolves from studying a few genes at a time in order to implicate their function in mediation of toxicant effects. Undoubtedly, this process has to be accelerated to monitor and discern the effects of the thousands of new compounds developed by the chemical and pharmaceutical industries. There is a need for a screening method that can offer some insight into the potential adverse outcome(s) of new drugs allowing the intelligent advancement of compounds into late stages of safety evaluation.

The rapid development and evolution of genomic- (DeRisi, *et al.,* 1996; Duggan, *et al.,* 1999), proteomic- (Lueking, *et al.,* 1999; Page, *et al.,* 1999; Rubin and Merchant, 2000; Steiner and Anderson, 2000; Weinberger, *et al.,* 2000; Huang, 2001), and metabonomic- (Foxall, *et al.,* 1993; Corcoran, *et al.,* 1997; De Beer, *et al.,* 1998) based technologies has accelerated the application of gene expression for understanding chemical and other environmental stressors' effects on biological systems. These technological advances have led to the development of the field of "toxicogenomics", which proposes to apply global mRNA, protein and metabolite analysis related-technologies to study the effects of hazards on organisms (Afshari, *et al.,* 1999; Farr, 1999; Henry, 1999; Nuwaysir, *et al.,* 1999; Rockett and Dix, 1999; Hamadeh and Afshari, 2000; Pennie, *et al.,* 2000; Rockett and Dix, 2000; Hooker, 2001; Iannaccone, 2001; Olden, 2001; Smith, 2001; Tennant, 2001; Hamadeh *et al.,* 2001; Hamadeh *et al.,* 2002d). These collective approaches will allow the development of a knowledge base of compound effects that will aid in improving the efficiency of safety and risk assessment of drugs and chemicals by facilitating better understanding of the mechanisms by which chemical- or stressor-induced injury occurs.

*For correspondence. Email afshari@niehs.nih.gov; Tel. (919) 541-1310; Fax. (919) 541-0146.

Technologies in Toxicogenomics

Gene Expression Profiling

Gene expression changes associated with signal pathway activation can provide compound-specific information on the pharmacological or toxicological effects of a chemical. A standard method used to study changes in gene expression is the Northern blot (Sambrook *et al.,*1989). An advantage of this traditional molecular technique is that it definitively shows the expression level of all transcripts (including splice variants) for a particular gene. This method, however, is labor intensive and is practical for examining expression changes for a limited number of genes. Alternate technologies, including DNA microarrays, can measure the expression of tens of thousands of genes in an equivalent amount of time (DeRisi, *et al.,* 1996; Duggan, *et al.,* 1999; Hamadeh and Afshari, 2000; Hamadeh *et al.,* 2001). DNA microarrays provide a revolutionary platform to compare genome-wide gene expression patterns in dose and time contexts. There are two basic types of microarrays used in gene expression analyses: oligonucleotide-based arrays (Lockhart, *et al.,* 1996) and cDNA arrays (Schena, *et al.,* 1995). Both yield comparable results, though the methodology differs. Oligonucleotide arrays are made using specific chemical synthesis steps by a series of photolithographic masks, light, or other methods to generate the specific sequence order in the synthesis of the oligonucleotide. The result of these processes is the generation of high-density arrays of short oligonucleotide (~ 20-80 bases) probes that are synthesized in predefined positions. cDNA microarrays differ in that DNA sequences (0.5-2 kb in length) that correspond to unique expressed gene sequences, are usually spotted onto the surface of treated glass slides using high speed robotic printers that allow the user to configure the placement of cDNAs on a glass substrate or chip. Spotted cDNAs can represent either sequenced genes of known function, or collections of partially sequenced cDNA derived from expressed sequence tags (ESTs) corresponding to messenger RNAs of genes of known or unknown function.

Any biological sample from which high quality RNA

© 2002 Caister Academic Press

can be isolated may be used for microarray analysis to determine differential gene expression levels. For toxicology studies, there are a number of comparisons that might be considered. For example, one can compare tissue extracted from toxicant treated organism versus that of vehicle exposed animals. In addition, other scenarios may include the analysis of healthy versus diseased tissue or susceptible versus resistant tissue. For spotted cDNA on glass platforms, differential gene expression measurements are achieved by a competitive, simultaneous hybridization using a two-color fluorescence labeling approach (Schena, *et al.,* 1995; DeRisi, *et al.,* 1996). Multi-color based labels are currently being optimized for adequate utility. Briefly, isolated RNA is converted to fluorescently labeled "targets" by a reverse transcriptase reaction using a modified nucleotide, typically dUTP or dCTP conjugated with a chromophore. The two RNAs being compared are labeled with different fluorescent tags, traditionally either Cy3 or Cy5, so that each RNA has a different energy emission wavelength or color when excited by dual lasers. The fluorescently labeled targets are mixed and hybridized on a microarray chip. The array is scanned at two wavelengths using independent laser excitation of the two fluors, for example, at 632 and 532 nm wavelengths for the red (Cy5) and green (Cy3) labels. The intensity of fluorescence, emitted at each wavelength, bound to each spot (gene) on the array corresponds to the level of expression of the gene in one biological sample relative to the other. The ratio of the intensities of the toxicant-exposed versus control samples are calculated and induction/repression of genes is inferred. Optimal microarray measurements can detect differences as small as 1.2 fold increase or decrease in gene expression.

Although the theoretical applications seem endless, DNA microarrays have certain limitations. These measurements are only semiquantitative due to a number of factors, including cross hybridization and sequence-specific binding anomalies. Another limitation is the number of samples that can be processed efficiently at a time. Processing and scanning samples may take several days and generate large amounts of information that can take considerable time to analyze. Automation is being applied to microarray technology, and new equipment such as the automated hybridization stations and auto-loaded scanners will allow higher throughput analysis. To overcome these limitations, one can combine microarrays with quantitative polymerase chain reaction (QPCR) or Taqman and other technologies in development (Kreuzer, *et al.,* 1999; Tokunaga, *et al.,* 2000) to monitor the expression of hundreds of genes in a high throughput fashion. This will provide more quantitative output that may be crucial for certain hazard identification processes. In the QPCR (Walker, 2001) assay one set of primers is used to amplify both the target gene cDNA and another neutral DNA fragment, engineered to contain the desired gene template primers, which competes with the target cDNA fragment for the same primers and acts as an internal standard. Serial dilutions of the neutral DNA fragment are added to PCR amplification reactions containing constant amounts of experimental cDNA samples. The neutral DNA fragment utilizes the same primer as the target cDNA but yields a PCR product of different size. QPCR can offer more quantitative measurements than microarrays do because measurements may be made in "real time" during the time of the amplification and within a linear dynamic range. The PCR reactions may be set up in 96 or 384-well plates to provide a high throughput capability.

Expression Profiling of Toxicant Response

The validity and utility of analysis of gene expression profiles for hazard identification depends on whether different profiles correspond to different classes of chemicals (Waring, *et al.,* 2001; Waring, *et al.,* 2001; Hamadeh *et al.,* 2002c) and whether defined profiles maybe used to predict the identity/properties of unknown or blinded samples derived from chemically treated biological models (Hamadeh *et al.,* 2002b). Gene expression profiling may aid in prioritization of compounds to be screened in a high throughput fashion and selection of chemicals for advanced stages of toxicity testing in commercial settings.

In one effort to validate the toxicogenomic strategy, Waring and coworkers (Waring, *et al.,* 2001; Waring, *et al.,* 2001) conducted studies to address whether compounds with similar toxic mechanisms produced similar transcriptional alterations. This hypothesis was tested by generating gene expression profiles for 15 known hepatotoxicants *in vitro* (rat hepatocytes) and *in vivo* (livers of male Sprague-Dawley rats) using microarray technology. The results from the *in vitro* studies showed that compounds with similar toxic mechanisms resulted in similar but distinguishable gene expression profiles (Waring, *et al.,* 2001). The authors took advantage of the variety of hepatocellular injuries (necrosis, DNA damage, cirrhosis, hypertrophy, hepatic carcinoma) that were caused by the chemicals and compared pathology endpoints to the clustering output of the compounds' gene expression profiles. Their analyses showed a strong correlation between the histopathology, clinical chemistry, and gene expression profiles induced by the various agents (Waring, *et al.,* 2001). This suggests that DNA microarrays may be a highly sensitive technique for classification of potential chemical effects.

In another study, gene expression alterations in Sprague Dawley rat livers were measured for known and unknown compound treatments. This exercise revealed that it is possible to use previously derived gene expression profiles to characterize unknown compounds. In this study, correct, positive predictions regarding the nature of 12 out of 13 of the blinded samples were made (Hamadeh *et al.,* 2002c). Multiple statistical and computational approaches such as hierarchical clustering (Eisen, *et al.,* 1998), principal component analysis (Johnson and Wichern, 1998), and set pair-wise correlation (Johnson and Wichern, 1998; Neter, 1996) were used to distinguish gene expression profiles derived from rat livers treated with different class chemicals and different durations of exposure (Hamadeh *et al.,* 2002b). Other computational methods such as linear discriminant analysis (Solberg, 1978), single gene ANOVA (Neter, 1996), and genetic algorithm/K-nearest neighbor (Li *et al.,* In Press) were useful in revealing single or groups of highly discriminatory/informative genes whose

expression pattern could distinguish gene expression patterns corresponding to different chemical treatments. Blinded samples that exhibited high similarity (r>0.8) to known samples, as determined by set pair-wise correlation, were considered to tentatively share similar properties/ identities.

Mechanistic Inference from Toxicant Profiling

An extension of the use of toxicogenomics approaches is the better understanding of the mechanisms of toxicity. Bulera and coworkers (Bulera, *et al.,* 2001) identified several groups of genes reflective of mechanisms of toxicity and related to a hepatotoxic outcome following treatment. An example of the advantage of using a toxicogenomics approach to understand mechanisms of chemical toxicity was the observation that microcystin-LR and phenobarbital, both of which are liver tumor promoters, induced a parallel set of genes (Bulera, *et al.,* 2001). Based on this information the authors speculated that liver tumor promotion by both compounds may occur by similar mechanisms. Such observations derived through the application of microarrays to toxicology will broaden our understanding of mechanisms and our ability to identify compounds with similar mechanisms of toxicity. The authors also confirmed toxicity in the animals using conventional methods such as histopathology, modulations in liver enzymes and bilirubin levels and related these effects to gene expression changes; however, it would have been advantageous to utilize gene expression data to map relevant pathways depicting mechanism(s) associated with the hepatotoxicity of each compound (Hamadeh *et al.,* 2001). Collectively, in the future, researchers may attempt to build "transcriptome" or "effector maps" that will help to visualize pathway activation (Tennant, 2001).

Finally, Huang and coworkers (Huang, *et al.,* 2001) utilized cDNA microarrays to investigate gene expression patterns of cisplatin-induced nephrotoxicity. In these studies, rats were treated daily for 1 to 7 days with cisplatin at a dose that resulted in necrosis of the renal proximal tubular epithelial cells but no hepatotoxicity at day 7. Gene expression patterns for transplatin, an inactive isomer, was examined and revealed little gene expression change in the kidney, consistent with the lack of nephrotoxicity of the compound. Cisplatin-induced gene expression alterations were reflective of the histopathological changes in the kidney i.e. gene related to cellular remodeling, apoptosis, and alteration of calcium homeostasis, among others which the authors describe in a putative pathway of cisplatin nephrotoxicity.

Protein Expression

Gene expression alone is not adequate to serve the understanding of toxicant action and the disease outcomes they induce. Abnormalities in protein production or function are expected in response to toxicant exposure and the onset of disease states. To understand the complete mechanism of toxicant action, it is necessary to identify the protein alterations associated with that exposure and to understand how these changes affect protein/cellular function. Unlike classical genomic approaches that discover genes related to toxicant induced disease, proteomics can aid to characterize the disease process directly by capturing proteins that participate in the disease. The lack of a direct functional correlation between gene transcripts and their corresponding proteins necessitates the use of proteomics as a tool in toxicology.

Proteomics is the systematic analysis of expressed proteins in tissues, by isolation, separation, identification and functional characterization of proteins in a cell, tissue, or organism (Lueking, *et al.,* 1999; Page, *et al.,* 1999; Anderson, *et al.,* 2000; Rubin and Merchant, 2000). Proteomics, under the umbrella of toxicogenomics, involves the comprehensive functional annotation and validation of proteins in response to toxicant exposure. Understanding the functional characteristics of proteins and their activity requires a determination of cellular localization and quantitation, tissue distribution, post-translational modification state, domain modules and their effect on protein interactions, protein complexes, ligand binding sites and structural representation.

Currently, the most commonly used technologies for proteomics research are 2-dimensional (2-D) gel electrophoresis for protein separation followed by mass spectrometry analysis of proteins of interest (Rasmussen, *et al.,* 1994; Shaw, *et al.,* 1999; Carroll, *et al.,* 2000; Fountoulakis, *et al.,* 2000; Kaji, *et al.,* 2000; Watarai, *et al.,* 2000). Analytical protein characterization with multidimensional liquid chromatography/mass spectrometry improves the throughput and reliability of peptide identity. Matrix-Assisted Laser Desorption Mass Spectrometry (MALDI-MS) (Stults, 1995; Liang, *et al.,* 1996) has become a widely used method for determination of biomolecules including peptides. Other technologies such as Surface-Enhanced Laser Desorption/Ionization (SELDI) (Kuwata, *et al.,* 1998; Li, *et al.,* 2000; Merchant and Weinberger, 2000; Rubin and Merchant, 2000) and antibody arrays (Borrebaeck, *et al.,* 2001; Haab, *et al.,* 2001; Paweletz, *et al.,* 2001; Sreekumar, *et al.,* 2001) are also proving to be useful.

Cutler and coworkers conducted a study aimed at the investigation of biochemical changes and identification of biomarkers associated with acute renal injury following a single dose of puromycin aminonucleoside to Sprague Dawley rats using a combination of 2-D PAGE, reverse-phase HPLC, mass spectrometry, amino acid analysis and ^{1}H-NMR spectroscopy of urine as well as routine plasma clinical chemistry and tissue histopathology (Cutler *et al.,* 1999). The 2-D PAGE of urine showed patterns of protein change which were in accord with the limited profiles for glomerular toxicity derived by use of other techniques and allowed a more detailed understanding of the nature and progression of the proteinuria associated with glomerular toxicity. Interestingly, the 2-D PAGE approach taken by the investigators, coupled with computational analysis of the accompanying data gleaned on the collected samples, lead to the detection of proteinuria at a considerably earlier time point than has typically been reported following puromycin aminonucleoside exposure, thus potentially defining relatively early biomarkers which are superior to the traditional gross urinary protein determination procedure

(Cutler *et al.,* 2001).

A serious limitation of proteomic analysis using 2-D gel electrophoresis is the sensitivity of detection. Analysis of low abundance proteins by 2-D electrophoresis is challenging due to the presence of high abundant proteins such as albumin, immunoglobulin heavy and light chains, transferrin, and haptoglobin in the sera or actin, tubulin, and other structural proteins when analyzing tissue. Selective removal of these proteins from protein samples via column-based immunoaffinity procedures allows for more sample to be loaded on gels thereby facilitating visualization of low abundant proteins that would otherwise be obscured by more abundant ones (Kennedy, 2001).

Metabolite Analysis by NMR

Genomic and proteomic methods do not offer the information needed to gain understanding of the resulting output function in a living system. Neither approach addresses the dynamic metabolic status of the whole animal. The metabonomic approach is based on the premise that toxicant-induced pathological or physiological alterations result in changes in relative concentrations of endogenous biochemicals. Metabolites in body fluids such as urine, blood, or cerebrospinal fluid (CSF), are in dynamic equilibrium with those inside cells and tissues, thus toxicant-induced cellular abnormalities in tissues should be reflected in altered biofluid compositions. An advantage of measuring changes in body fluids is that these samples are much more readily available from human subjects. High resolution NMR spectroscopy (^{1}H NMR) has been used in a high-throughput fashion to simultaneously detect many cellular biochemicals in urine, bile, blood plasma, milk, saliva, sweat, gastric juice, seminal, amniotic, synovial and cerebrospinal fluids (Holmes, *et al.,* 1995; Robertson, *et al.,* 2000; Bundy, *et al.,* 2001; Griffin, *et al.,* 2001; Nicholls, *et al.,* 2001; Waters, *et al.,* 2001). In addition, intact tissue and cellular suspensions have also been successfully analyzed for metabolite content using magic-angle-spinning ^{1}H NMR spectroscopy (Garrod, *et al.,* 1999).

Metabolic Profiling of Toxicant Response

Robertson and coworkers evaluated the feasibility of a toxicogenomic strategy by generating NMR spectra of urine samples from male Wistar rats treated with different hepatotoxicants (carbon tetrachloride, α-naphthylisothiocyanate) or nephrotoxicants (2-bromoethylamine, 4-aminophenol) (Robertson, *et al.,* 2000). Principal component analysis (PCA) of the urine spectra was in agreement with clinical chemistry data observed in blood samples taken from the chemically exposed animals at various time points of chemical exposure. Furthermore, PCA analysis suggested low dose effects with two of the chemicals, which were not evident by clinical chemistry or microscopic analyses. This conclusion was demonstrated with the 150 mg/kg 2-bromoethanolamine treated animals where only 5 of 8 of the animals had creatinine or BUN levels, at day 1, that were outside the normal range, while all animals exhibited diuresis and principal component analysis was clearly indicative of a consistent effect in all 8 animals.

In another seminal study, ^{1}H NMR spectroscopy was used to characterize the time-dependency of urinary metabolite perturbation in response to toxicant exposure. Male Han Wistar or Sprague Dawley rats were treated with either control vehicle or one of 13 model toxicants or drugs that predominantly target liver or kidney. The resultant ^{1}H NMR spectra were analyzed using a probabilistic neural network approach (Holmes, *et al.,* 2001). A set of 583 of the 1310 samples were designated as a training set for the neural network, with the remaining 727 independent cases employed as a test set for validation. Using these techniques, the 13 classes of toxicity, together with the variations associated with strain, were highly distinguishable (>90%). An important aspect of this study is the sensitivity of the methodology towards strain differences that will be useful in investigating the genetic variation of metabolic responses across multiple animal models and may also prove useful in identifying susceptible subpopulations.

Localization of Gene Expression

In order to help understand the role of genes or proteins in toxic processes, specific cellular localization of these targets is needed. Pathological alterations such as necrosis and vasculitis are often localized to specific regions of an organ or tissue. It is not known whether subtle gene or protein expression alterations associated with these events are detectable when the whole organ is used for preparation of samples for further analyses. Laser capture microdissection (LCM) (Emmert-Buck, *et al.,* 1996; Bonner, *et al.,* 1997; Fend, *et al.,* 2000; Murakami, *et al.,* 2000) is one method used to precisely select affected tissue thereby enhancing the probability of observing gene or protein expression changes associated with pathologically altered regions. For example, profiling specific pathological lesions that are considered to be precursors to cancer may help in understanding how chronic chemical exposure leads to tumor development.

However, for some tissues or laboratories, LCM may not be technically feasible to discern gene expression in cellular subtypes. A technical challenge may be that the affected area or region is too small for enough RNA or protein to be extracted for later analysis, or the extra manipulation compromises the quality of harvested samples. Therefore, when deriving samples from gross organ or tissue samples for expression analysis, one often has no measure of specific gene or protein expression alterations attributable to the pathological change that was diluted in the assayed organ or tissue. When an organ, or part thereof, is harvested from a chemically exposed animal, the response to the insult is almost always diluted to a certain extent because not every area or cell is responsive to treatment. Similarly, tumor samples or other diseased tissues may contain other significant cell types including stroma, lymphocytes, or endothelial cells. Dilution effects are also involved when a heterogeneous expression response occurs. For example, even in a homogeneous cell population, each individual cell may have a very different quantitative response for each gene expression

change. In order to address this problem, we evaluated the sensitivity of cDNA microarrays in detecting diluted gene expression alterations thus simulating relatively minor changes in the context of total organ or tissue. We found statistically significant differences in the expression of numerous genes between two cell lines (HaCaT and MCF-7) that continued to be detected even after a 20-fold dilution of original changes (Hamadeh *et al.,* 2002a) showing that microarray analyses, when conducted in a manner to optimize sensitivity and reduce noise, may be used to determine gene expression changes occurring in only a small percentage of cells sampled.

Finally, once important biomarkers are hypothesized from genomics and proteomics technologies, candidate target genes or proteins can then be monitored using more high-throughput, cost-effective immunohistochemical analyses in the form of tissue microarrays. Tissue microarrays are microscope slides where thousands of minute tissue samples from normal and diseased organisms can be tiled in an array fashion. The tissue microarrays can then be probed with the same fluorescent antibody to monitor the expression, or lack of, certain candidate markers for exposure or disease onset.

Database Requirements

Profiles corresponding to gene, protein, or metabolite measurements should be housed in a relational database that will facilitate the query of data depending on different criteria. Technical requirements of the database are beyond the scope of this discussion. From a biological perspective, the ideal database will not only house the aforementioned data, but will also hold additional toxicology information describing various parameters of the stressor-subjected biological systems. The parameters might include body and organ weights, mortality, histopathological results, and clinical chemistry and urinalysis measurements in animal studies or cell viability, cell cycle analyses, cell density, culture conditions and cell morphology reports in the case of *in vitro* studies. Chemical purity, solubility, stability, and volatility, also are important to archive. These additional data are of importance when conducting pattern recognition-oriented toxicogenomic studies because they facilitate the understanding of similarities between genomic, proteomic, or metabonomic profiles. These others will aid in the interpretation of different profiles as suggested by pattern recognition tools such as clustering algorithms or principal components analysis (Hamadeh *et al.,* 2002b; 2002c).

Toxicogenomic Components: Comparative/Predictive and Functional

Comparative/Predictive Toxicogenomics

There are two main applications for a toxicogenomic approach, comparative/predictive and functional. Comparative genomic, proteomic, or metabonomic studies measure the number and types of genes, protein, and metabolites respectively that are present in normal and toxicant-exposed cells, tissues, or biofluids. This approach is useful in defining the composition of the assayed samples in terms of genetic, proteomic or metabolic variables. Thus a biological sample derived from toxicant, or sham treated animals can be regarded as an n-dimensional vector in gene expression space with genes as variables along each dimension. The same analogy can be applied for protein expression or NMR analysis data thereby providing n-dimensional fingerprints or profiles of the biological sample under investigation. Thus, this aspect of toxicogenomics deals with automated pattern recognition analysis aimed at studying trends in data sets rather than probing the individual genes for mechanistic information. The need for pattern recognition tools is mandated by the volume and complexity of data generated by genomic, proteomic and metabonomic tools, and human intervention, in required repetitive computation, is kept to a minimum. Automatic toxicity classification methods are very desirable and prediction models are well suited for this task.

The data profiles reflect the pharmacological or toxicological effects, such as disease outcome, of the drug or toxicant being utilized. The underlying goal is that a sample from an animal exposed to an unknown chemical, or displaying a certain pathological endpoint, can then be compared to a database of profiles corresponding to exposure conditions with well-characterized chemicals, or to well defined pathological effects, in order to glean/predict some properties regarding the studied sample. These predictions, as we view them, fall into 2 major categories, namely, classification of samples based on the class of compound to which animals were exposed to, or classification of samples based on the histopathology and clinical chemistry that the treated animals displayed. Such data will allow insight into the gene, protein, or metabolite perturbations associated with pharmacologic effects of the agent or toxic endpoints that ensue. If array data can be "phenotypically anchored" to conventional indices of toxicity (histopathology, clinical chemistry etc.) it will be possible to search for evidence of injury prior to it's clinical or pathological manifestation. This approach could lead to the discovery of potential early biomarkers of toxic injury.

"Supervised" predictive models (Zhou and Bennett, 1997; Jonic, *et al.,* 1999; Tafeit and Reibnegger, 1999) have been used for many years in the financial sectors for evaluating future economic prospects of companies, and in geological institutes for predicting adverse weather outcomes using past or historical knowledge. They have also been utilized to make predictions, using clinical and radiographic information, regarding the diagnosis of active pulmonary tuberculosis at the time of presentation at a health-care facility that can be superior to physicians' opinion (El-Solh, *et al.,* 1999). Predictive modeling will undoubtedly revolutionize the field of toxicology by recognizing patterns and trends in high-density data, and forecasting gene-, protein-, or metabolite-environment interactions relying on historical data from well studied compounds and their corresponding profiles.

During the development of a predictive model, a number of issues must be considered. These include the representativeness of the variables to the entity being modeled and the quality of databases consulted. The National Center for Toxicogenomics (NCT), at the National

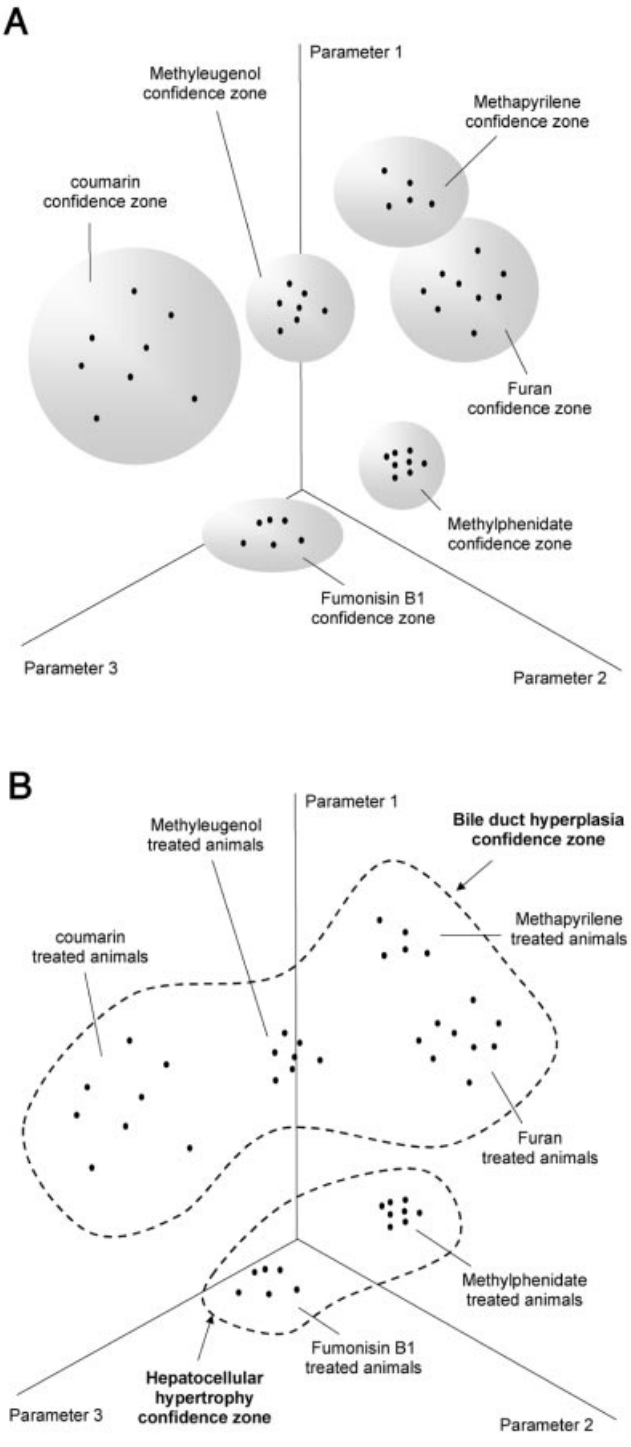


Figure 1. Hypothetical 3-dimensional representation of samples, derived from biological systems subjected to various exposure conditions, based on the expression levels of gene, protein or metabolite levels. Computational algorithms can form prediction zones (A) circumscribing sets of samples derived from the same exposure conditions (*in vivo*, *in vitro*) or (B) zones that encompass samples based on user defined endpoints associated with these samples.

Institute of Environmental Health Sciences, is building a database to store many variables (ex. dose, time, biological system) and observations (ex. histopathology, body weight, cell cycle data) that accompany the process of compound evaluation studies (*in vivo* or *in vitro*) (Tennant, 2001). Recording these parameters will greatly enhance the process of parameter selection in subsequent efforts such as predictive modeling or mechanism of action interpretation. Predictive modeling can be fragmented into a multistage process. The primary stage predictive modeling includes hypothesis development, organization and data collection. Secondary stage modeling includes initial model development and testing. Tertiary stage modeling includes continued application of the model, ongoing refinement, and validation. Ideally, tertiary stage modeling is a perpetual process whereby lessons learned from previous model applications are incorporated into new and future applications maintaining or increasing the predictive robustness of the model.

First Stage of Comparative/Predictive Model: Data Collection

The development of the primary stage of a predictive model involves activities such as data collection strategies based on proposed hypotheses. Data can be generated from *in vivo* or *in vitro* experiments, depending on the suitability of the biological system for studying effects of the targeted compound. In the case of *in vivo* studies, hypotheses must be generated regarding the compounds and endpoint effects so that other measures, such as pathology, serum markers, and carcinogenicity potential, are made and can contribute to the ensuing model development. Data on animal weight fluctuations, serum markers, pathological alterations, and mortality rates corresponding to a chemical exposure study should be documented and be the primary source of such information for the constructed predictive model. Pertinent data and analytically useful variables gathered from other sources (ex. National Toxicology Program) can be evaluated and incorporated into the model. These data are important in developing a theoretical framework in which to interpret the results of the predictive model as well as to provide a guide for the data to be collected.

Second Stage of Comparative/Predictive Model: Model Development

The next step in the predictive model construction involves a deductive phase that incorporates collected data into the second stage of the model. The degree of correlation between gene-, protein- or metabolite-related profiles of different compounds or different toxicological/pathological outcomes and the accompanying variables can be measured and ranked. Computational and statistical approaches would be applied to the data set to glean relationships and dissimilarities among the variables studied.

Neural networks, which have been used in models predicting health status of HIV/AIDS patients (Giacomini,

et al., 1997; Kwak and Lee, 1997; Ioannidis, *et al.,* 1998), can be trained with a set of available profiles from previously studied compounds or pathophysiological states. This allows the automation of all the actions aimed at searching the interrelationships and producing predictions regarding unknown or new profiles. Every compound or effect is characterized by various parameters describing its gene expression pattern. Thus, a pattern may be represented by a vector in space whose components could represent various parameters that drive the decision of classification. Dimensionality of this space is the number of vector components or parameters involved and is based on the analysis of multiple parameters that can correlate similar expression profiles. As a simplified example, if we consider each compound or adverse endpoint we are modeling to have only three attributes, these three parameters can represent vector coordinates in a 3-dimensional space. Figure 1A shows how the treated animals, or cells, could be spatially disposed, so that one can easily notice where they are grouped, i.e. have similar parameters, for which reason they most probably belong to the same group. Now we proceed to defining which objects are situated in particular nodes of the map. A multitude of available algorithms satisfactorily cluster objects in 3- or n-dimensional space based on computational approaches (ex. PCA). We can then construct similarity zones around various preset chemical (Figure 1A) or adverse endpoint (Figure 1B) nodes. Such similarity zones would allow the classification, with a defined level of confidence, of the identity of unknown samples which neighbor samples in the training data set. Thus possessing the map and information about the analyzed compounds, we can reliably judge the compounds with which we are less familiar.

The initial predictive model can be tested using the data collected in the primary stage. Based upon the outcome of this exercise, variables such as toxicant induced lesion severity or organ weight fluctuations can be introduced or removed from the process, or the weighting of the variables can be adjusted until the model is able to predict the highest percentage of chemicals possible. This highlights the need for the consulted database to contain enough parameters such as histopathological observations or clinical chemistry data that accompany an experimental design to facilitate this dynamic model optimization process. Developed models should ideally allow the distinction of gene expression profiles associated with chemical exposure or pathological outcome depending on the querying preferences of the user and the question being asked. Once this has been achieved, tertiary stage modeling may begin.

Third Stage of Comparative/Predictive Model: Utility

The use of genomic resources such as DNA microarrays in safety evaluation will facilitate an emerging type of experimentation termed "*in silico*" testing. For example, if compound A was found to bear similarity to compound B, and B had some aspects that were close to compound C, then a relationship could be defined between compounds A and C based on the their common link to B. *In silico* experimentation can define this relationship through rigorous computation and mining of high-density gene expression data. Developments in computer modeling and expert systems for the prediction of biological activity and toxicity will revolutionize the process of drug discovery and development, by reducing the need to use animals for the pre-screening of almost limitless numbers of potential drug candidates.

It is not foreseeable that in the near future predictive models will take the place of actual testing. However, in the context of toxicogenomics, and with the increasing number of chemicals to be tested, better prioritization can be used to select the compounds for animal testing. The most promising efficacious compounds with the least probability of an adverse outcome would be selected for further development.

Functional Toxicogenomics

Functional toxicogenomics is the study of genes' and proteins' biological activities in the context of compound effects on an organism. Gene and protein expression profiles are analyzed for information that might provide insight into specific mechanistic pathways. Mechanistic inference is complex when the sequence of events following toxicant exposure is viewed in both dose and time space. Gene and protein expression patterns can indeed be highly dependent on the toxicant concentrations furnished at the assessed tissue and the time of exposure to the agent. Expression patterns are only a snapshot in time and dose space. Thus, a comprehensive understanding of potential mechanisms of action of a compound requires establishing patterns at various combinations of time and dose. This will minimize the misinterpretation of transient responses and allow the discernment of delayed alterations that could be related to adaptation events or may be representative of potential biomarkers of pathophysiological endpoints.

Studies that target temporal expression of specific genes and protein in response to toxicant exposure will lead to a better understanding of the sequence of events in complex regulatory networks. Algorithms, such as self-organizing maps (Kohonen, 1999), can categorize genes or proteins based on their expression pattern across a continuum of time points. These analyses might suggest relationships in the expression of some genes or proteins depending on the concerted modulation of these variables.

An area of study which is of great interest to toxicologists is the mechanistic understanding of toxicant induced pathological endpoints. The premise that perturbations in gene, protein, or metabolite levels are reflective of adverse phenotypic effects of toxicants offers an opportunity to phenotypically anchor these perturbations. This is quite challenging due to the fact that phenotypic effects often vary in the time-dose space of the studied agent and may have regional variations in the tissue. Furthermore, very few compounds exist that result in only one phenotypic alteration at a given coordinate in dose and time. Thus, objective assignment of measured variables to multiple phenotypic events is not possible under these circumstances. However, by studying multiple structurally and pharmacologically unrelated agents that

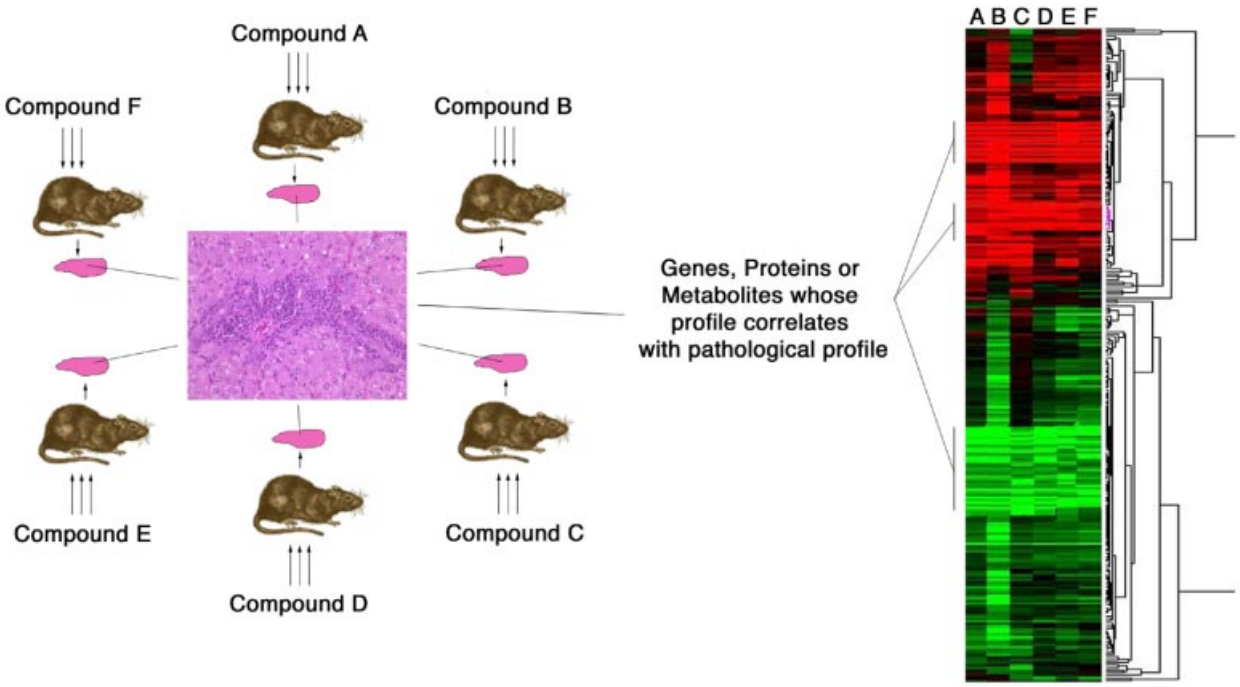


Figure 2. Schematic diagram outlining a strategy for the phenotypic anchorage of gene, protein, or metabolite alteration data. A hypothetical histopathological analysis of livers derived from rats treated with either of compounds A, B, C, D, E, or F reveals an overlap among the effects manifested in one common pathological endpoint. Commonality across animals revealed by cluster analysis of gene, protein, or metabolite levels would indicate a potential association between the altered parameters and the shared histopathological endpoint.

share pathological endpoints of interest, one could tease out gene, protein, or metabolite modulations that are in common between the studied compounds (Figure 2). Laser capture microdissection may also be used to capture regional variations such as zonal patterns of hepatotoxicity. This concept will allow the objective assignment of measurable variables to phenotypic observations that will supplement traditional pathology.

It is noteworthy to mention that, stand-alone, gene and protein expression, or metabolite fluctuation analyses are not expected to produce decisive inferences on the role of genes or proteins in certain pathways or regulatory networks. However, these tools constitute powerful means to generate viable and testable hypotheses that can direct future endeavors on proving or disproving the involvement of genes, proteins, and metabolites in cellular processes. Ultimately, hypothesized mechanistic inferences have to be validated by the use of traditional molecular biology techniques that include the use of specific enzyme inhibitors, and the examination of the effects of over-expression or deletion of specific genes or proteins on the studied toxic endpoint or mechanism of compound action.

Future of Predictive Toxicology

From the rapid screening perspective, it is neither cost effective, nor is it practical to survey the abundance of all genes, proteins, or metabolites in a sample of interest. It would be prudent to conduct cheaper, more high-throughput measurements on variables that are of most interest in the toxicological evaluation process. Thus, this reductionist strategy mandates the selection of subsets of genes, proteins or metabolites that will yield useful information in regards to classification purposes such as hazard identification or risk assessment. The challenge is finding out what these minimal variables are and what data we need to achieve this knowledge. Election of these subsets by surveying existing toxicology literature is inefficient because the role of most genes or proteins in toxicological responses is poorly defined. Moreover, there exists a multitude of undiscovered or unknown genes (ESTs) that might ultimately be key players in toxicological processes.

We propose the use of genes, proteins, or metabolites, that are found to be most discriminative between stressor induced-specific profiles, for efficient screening purposes. Discriminative potential of genes, proteins, or metabolites is inferred when comparing differences in the levels of these parameters across toxicant exposure scenarios. In the case of samples derived from animals treated with one of few chemicals, the levels of one gene, protein, or metabolite might be sufficient to distinguish samples based on the few classes of compounds used for the exposures. However, multiple parameters are needed to separate samples derived from exposures to a larger variety of chemical classes. Finding these discriminatory parameters requires the use of computational and mining algorithms that extract this knowledge from a database of chemical effects.

Linear discriminant analysis (LDA) (Johnson and Wichern, 1998) and single gene ANOVA (Neter, 1996) can be used to test single parameters (ex. genes) for their ability to separate profiles corresponding to samples derived from different exposure conditions (ex. chemical identity, biological endpoint). Higher order analyses such as genetic algorithm/K-nearest neighbor (GA/KNN) (Li *et al.,* In Press) are able to find a user defined number of parameters that would, as a set, highlight the most difference between biological samples based on the levels of genes, proteins, or metabolites. Once the profile of a parameter, or a set of parameters, is found to distinguish between samples in a data set, it can be used to interrogate the identity of unknown samples for screening purposes in a high-throughput fashion. It is important to keep in mind that since these discriminatory parameters are derived from historical data, it is possible that their status might not hold once significant volumes of new data is inputted in the database that computations are run. It is prudent to view discriminatory parameters (genes, protein, metabolites) as dynamic entities that can be updated periodically depending on the availability of new toxicant related profiles used.

Summary

Toxicogenomic tools will inevitably improve the way data is extracted from classical toxicology studies. Ultimately, through the use of computational tools encompassed within the comparative branch of toxicogenomics, environmental hazard identification may be performed in a high-throughput and efficient fashion. These achievements will be facilitated through the development of gene, protein, or metabolite markers whose levels can be monitored in samples derived from exposed populations. Compound profiling will also improve our understanding of toxicant induced adverse endpoints in biological systems (pathological lesions, cell cycle alterations) by providing information about the underlying molecular pathways that are involved in response to compound exposure. This knowledge will lead to a more informed and precise classification of compounds for their safety evaluation.

Acknowledgements

The authors would like to thank Drs B. Alex Merrick and Gary Boorman for critical review of the manuscript and Ms. Jennifer Collins for editorial assistance.

References

Afshari, C.A., Nuwaysir, E.F., and Barrett, J.C. 1999. Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. Cancer Res. 59: 4759-4760.

Anderson, N.L., Matheson, A.D., and Steiner, S. 2000. Proteomics: applications in basic and applied biology. Curr. Opin. Biotechnol. 11: 408-412.

Bonner, R.F., Emmert-Buck, M., Cole, K., Pohida, T., Chuaqui, R., Goldstein, S., and Liotta, L.A. 1997. Laser capture microdissection: molecular analysis of tissue. Science. 278: 1481-1483.

Borrebaeck, C.A., Ekstrom, S., Hager, A.C., Nilsson, J., Laurell, T., and Marko-Varga, G. 2001. Protein chips based on recombinant antibody fragments: a highly sensitive approach as detected by mass spectrometry. Biotechniques. 30: 1126-1130, 1132.

Bulera, S.J., Eddy, S.M., Ferguson, E., Jatkoe, T.A., Reindel, J.F., Bleavins, M.R., and De La Iglesia, F.A. 2001. RNA expression in the early characterization of hepatotoxicants in wister rats by high-density DNA microarrays. Hepatology. 33: 1239-1258.

Bundy, J.G., Osborn, D., Weeks, J.M., Lindon, J.C., and Nicholson, J.K. 2001. An NMR-based metabonomic approach to the investigation of coelomic fluid biochemistry in earthworms under toxic stress. FEBS Lett. 500: 31-35.

Carroll, K., Ray, K., Helm, B., and Carey, E. 2000. Two-dimensional electrophoresis reveals differential protein expression in high- and low-secreting variants of the rat basophilic leukaemia cell line. Electrophoresis. 21: 2476-2486.

Corcoran, O., Spraul, M., Hofmann, M., Ismail, I.M., Lindon, J.C., and Nicholson, J.K. 1997. 750 MHz HPLC-NMR spectroscopic identification of rat microsomal metabolites of phenoxypyridines. J. Pharm. Biomed. Anal. 16: 481-489.

De Beer, R., Van den Boogaart, A., Cady, E., Graveron-Demilly, D., Knijn, A., Langenberger, K.W., Lindon, J.C., Ohlhoff, A., Serrai, H., and Wylezinska-Arridge, M. 1998. Absolute metabolite quantification by in vivo NMR spectroscopy: V. Multicentre quantitative data analysis trial on the overlapping background problem. Magn. Reson. Imaging. 16: 1127-1137.

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., and Trent, J.M. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer [see comments]. Nat. Genet. 14: 457-460.

Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P., and Trent, J.M. 1999. Expression profiling using cDNA microarrays. Nat. Genet. 21: 10-14.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA. 95: 14863-14868.

El-Solh, A.A., Hsiao, C.B., Goodnough, S., Serghani, J., and Grant, B.J. 1999. Predicting active pulmonary tuberculosis using an artificial neural network. Chest. 116: 968-973.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., and Liotta, L.A. 1996. Laser capture microdissection. Science. 274: 998-1001.

Farr, S.R.n.D. 1999. Concise review: gene expression applied to toxicology. Toxicol. Sci. 50: 1-9.

Fend, F., Kremer, M., and Quintanilla-Martinez, L. 2000. Laser capture microdissection: methodical aspects and applications with emphasis on immuno-laser capture microdissection. Pathobiolo. 68: 209-214.

Fountoulakis, M., Berndt, P., Boelsterli, U.A., Crameri, F., Winter, M., Albertini, S., and Suter, L. 2000. 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.

Foxall, P.J., Mellotte, G.J., Bending, M.R., Lindon, J.C., and Nicholson, J.K. 1993. NMR spectroscopy as a novel approach to the monitoring of renal transplant function. Kidney Int. 43: 234-245.

Garrod, S., Humpfer, E., Spraul, M., Connor, S.C., Polley, S., Connelly, J., Lindon, J.C., Nicholson, J.K., and Holmes, E. 1999. High-resolution magic angle spinning 1H NMR spectroscopic studies on intact rat renal cortex and medulla. Magn. Reson. Med. 41: 1108-1118.

Giacomini, M., Ruggiero, C., Maillard, M., Lillo, F.B., and Varnier, O.E. 1997. Application of neural networks to the follow-up of AIDS patients. Stud. Health Technol. Inform. 43: 386-390.

Griffin, J.L., Walker, L.A., Shore, R.F., and Nicholson, J.K. 2001. Metabolic profiling of chronic cadmium exposure in the rat. Chem. Res. Toxicol. 14: 1428-1434.

Haab, B.B., Dunham, M.J., and Brown, P.O. 2001. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol. 2.

Hamadeh, H., and Afshari, C.A. 2000. Gene chips and functional genomics. American Scientist. 88: 508-515.

Hamadeh, H.K., Bushel, P., Paules, R.S., and Afshari, C.A. 2001. Discovery in toxicology: Mediation by gene expression array technology. J. Biochem. Mol. Toxicol. 15: 231-242.

Hamadeh, H.K., Bushel, P., Tucker, C.J., Martin, K., Paules, R., and Afshari, C.A. 2002a. Detection of diluted gene expression alterations using cDNA microarrays. Biotechniques. In Press.

Hamadeh, H.K., Bushel, P., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J.C., Blanchard, K., Paules, R.S., and Afshari, C.A. 2002b. Prediction of compound signature using high density gene expression profiling. Toxicol. Sci. In press.

Hamadeh, H.K., Bushel, P., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J.C., Blanchard, K., Paules, R.S., and Afshari, C.A. 2002c. Gene expression analysis reveals chemical-specific profiles. Toxicol. Sci. In press.

Hamadeh, H.K., Nuwaysir, E.F., Paules, R.S., Barrett, J.C., and Afshari, C.A. 2002d. cDNA microarray technology: Merging toxicology and genomics. Comments Toxicol. In Press.

Henry, C.M. 1999. DNA microarrays in toxicology. Anal. Chem. 71: 462A-464A.

Holmes, E., Caddick, S., Lindon, J.C., Wilson, I.D., Kryvawych, S., and Nicholson, J.K. 1995. 1H and 2H NMR spectroscopic studies on the metabolism and biochemical effects of 2-bromoethanamine in the rat. Biochem. Pharmacol. 49: 1349-1359.

Holmes, E., Nicholson, J.K., and Tranter, G. 2001. Metabonomic characterization of genetic variations in toxicological and metabolic responses using probabilistic neural networks. Chem. Res. Toxicol. 14: 182-191.

Hooker, T.M. 2001. The Brave New World of Toxicogenomics. Environmental Compliance Litigation. 16: 5.

Huang, Q., Dunn, R.T., 2nd, Jayadev, S., DiSorbo, O., Pack, F.D., Farr, S.B., Stoll, R.E., and Blanchard, K.T. 2001. Assessment of cisplatin-induced nephrotoxicity by microarray technology. Toxicol. Sci. 63: 196-207.

Huang, R. 2001. Detection of multiple proteins in an antibody-based protein microarray system. J. Immunol. Meth. 255: 1-13.

Iannaccone, P.M. 2001. Toxicogenomics: "The call of the wild chip". Environ. Health Perspectives. 109: A8-A11.

Ioannidis, J.P., McQueen, P.G., Goedert, J.J., and Kaslow, R.A. 1998. Use of neural networks to model complex immunogenetic associations of disease: human leukocyte antigen impact on the progression of human immunodeficiency virus infection. Am. J. Epidemiol. 147: 464-471.

Johnson, R.A. and Wichern, D.W. 1998. Applied Multivariate Statistical Analysis. 4th ed. Prentice-Hall, Upper Saddle River, NJ.

Jonic, S., Jankovic, T., Gajic, V., and Popovic, D. 1999. Three machine learning techniques for automatic determination of rules to control locomotion. IEEE Trans. Biomed. Eng, 46: 300-310.

Kaji, H., Tsuji, T., Mawuenyega, K.G., Wakamiya, A., Taoka, M., and Isobe, T. 2000. Profiling of Caenorhabditis elegans proteins using two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis. 21: 1755-1765.

Kohonen, T. 1999. Comparison of SOM point densities based on different criteria. Neural Comput. 11: 2081-2095.

Kreuzer, K.A., Lass, U., Bohn, A., Landt, O., and Schmidt, C.A. 1999. LightCycler technology for the quantitation of bcr/abl fusion transcripts. Cancer Res. 59: 3171-3174.

Kuwata, H., Yip, T.T., Yip, C.L., Tomita, M., and Hutchens, T.W. 1998. Direct detection and quantitative determination of bovine lactoferricin and lactoferrin fragments in human gastric contents by affinity mass spectrometry. Adv. Exp. Med. Biol. 443: 23-32.

Kwak, N.K., and Lee, C. 1997. A neural network application to classification of health status of HIV/AIDS patients. J. Med. Syst. 21: 87-97.

Li, L., Darden, T.A., Weinberg, C.R., and Pedersen, L.G. Gene assessment and sample classification for gene expression data using a genetic algorithm/k-nearest neighbor method. Combinatorial Chemistry and High Throughput Screening, In press.

Li, X., Mohan, S., Gu, W., Miyakoshi, N., and Baylink, D.J. 2000. Differential protein profile in the ear-punched tissue of regeneration and non-regeneration strains of mice: a novel approach to explore the candidate genes for soft-tissue regeneration. Biochim. Biophys. Acta. 1524: 102-109.

Liang, X., Zheng, K., Qian, M.G., and Lubman, D.M. 1996. Determination of bacterial protein profiles by matrix-assisted laser desorption/ionization mass spectrometry with high-performance liquid chromatography. Rapid Commun. Mass Spectrom. 10: 1219-1226.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E.L. 1996. Expression monitoring by hybridization to high-density oligonucleotide

arrays. Nat. Biotechnol. 14: 1675-1680.
Lueking, A., Horn, M., Eickhoff, H., Bussow, K., Lehrach, H., and Walter, G. 1999. Protein microarrays for gene expression and antibody screening. Anal Biochem, 270: 103-111.
Merchant, M., and Weinberger, S.R. 2000. Recent advancements in surface-enhanced laser desorption/ ionization- time of flight-mass spectrometry. Electrophoresis. 21: 1164-1177.
Murakami, H., Liotta, L., and Star, R.A. 2000. IF-LCM: laser capture microdissection of immunofluorescently defined cells for mRNA analysis rapid communication. Kidney Int. 58: 1346-1353.
Neter, J., M.H. Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. 1996. Applied Linear Statistical Models. 4th ed. Irwin Press, Chicago.
Nicholls, A.W., Holmes, E., Lindon, J.C., Shockcor, J.P., Farrant, R.D., Haselden, J.N., Damment, S.J., Waterfield, C.J., and Nicholson, J.K. 2001. Metabonomic investigations into hydrazine toxicity in the rat. Chem. Res. Toxicol. 14: 975-987.
Nuwaysir, E.F., Bittner, M., Trent, J., Barrett, J.C., and Afshari, C.A. 1999. Microarrays and toxicology: the advent of toxicogenomics. Mol. Carcinog. 24: 153-159.
Olden, K.G. 2001. Genomics: Implications for toxicology. Mutation Res. 473: 3-10.
Page, M.J., Amess, B., Rohlff, C., Stubberfield, C., and Parekh, R. 1999. Proteomics: a major new technology for the drug discovery process. Drug Discov. Today. 4: 55-62.
Paweletz, C.P., Charboneau, L., Bichsel, V.E., Simone, N.L., Chen, T., Gillespie, J.W., Emmert-Buck, M.R., Roth, M.J., Petricoin, I.E., and Liotta, L.A. 2001. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene. 20: 1981-1989.
Pennie, W.D., Tugwood, J.D., Oliver, G.J., and Kimber, I. 2000. The principles and practice of toxigenomics: applications and opportunities. Toxicol. Sci. 54: 277-283.
Rasmussen, H.H., Mortz, E., Mann, M., Roepstorff, P., and Celis, J.E. 1994. Identification of transformation sensitive proteins recorded in human two-dimensional gel protein databases by mass spectrometric peptide mapping alone and in combination with microsequencing. Electrophoresis. 15: 406-416.
Robertson, D.G., Reily, M.D., Sigler, R.E., Wells, D.F., Paterson, D.A., and Braden, T.K. 2000. Metabonomics: evaluation of nuclear magnetic resonance (NMR) and pattern recognition technology for rapid in vivo screening of liver and kidney toxicants. Toxicol. Sci. 57: 326-337.
Rockett, J.C., and Dix, D.J. 1999. Application of DNA arrays to toxicology. Environmental Health Perspectives. 107: 681-685.
Rockett, J.C., and Dix, D.J. 2000. DNA arrays: technology, options and toxicological applications. Xenobiotica. 30: 155-177.
Rubin, R.B., and Merchant, M. 2000. A rapid protein profiling system that speeds study of cancer and other diseases. Am. Clin. Lab. 19: 28-29.
Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning, A Laboratory Manual, Second Edition. Cold Spring Harbor Laboratory Press.
Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 270: 467-470.
Shaw, A.C., Rossel Larsen, M., Roepstorff, P., Holm, A., Christiansen, G., and Birkelund, S. 1999. Mapping and identification of HeLa cell proteins separated by immobilized pH-gradient two-dimensional gel electrophoresis and construction of a two-dimensional polyacrylamide gel electrophoresis database. Electrophoresis. 20: 977-983.
Smith, L.L. 2001. Key challenges for toxicologists in the 21st century. Trends Pharmacol. Sciences. 22: 281-285.
Solberg, H.E. 1978. Discriminant analysis. CRC Crit. Rev. Clin. Lab. Sci. 9: 209-242.
Sreekumar, A., Nyati, M.K., Varambally, S., Barrette, T.R., Ghosh, D., Lawrence, T.S., and Chinnaiyan, A.M. 2001. Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. Cancer Res. 61: 7585-7593.
Steiner, S., and Anderson, N.L. 2000. Pharmaceutical proteomics. Ann. N.Y. Acad. Sci. 919: 48-51.
Stults, J.T. 1995. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI- MS). Curr. Opin. Struct. Biol. 5: 691-698.
Tafeit, E., and Reibnegger, G. 1999. Artificial neural networks in laboratory medicine and medical outcome prediction. Clin. Chem. Lab. Med. 37: 845-853.
Tennant. R.W. The National Center for Toxicogenomics: Using new technologies to inform mechanistic toxicology. Environ. Health Perspectives. In Press.
Tokunaga, E., Maehara, Y., Oki, E., Koga, T., Kakeji, Y., and Sugimachi, K. 2000. Application of quantitative RT-PCR using “TaqMan” technology to evaluate the expression of CK 18 mRNA in various cell lines. J. Exp. Clin. Cancer Res. 19: 375-381.
Walker, N.J. 2001. Real-time and quantitative PCR: applications to mechanism-based toxicology. J. Biochem. Mol. Toxicol. 15: 121-127.
Waring, J.F., Ciurlionis, R., Jolly, R.A., Heindel, M., and Ulrich, R.G. 2001. Microarray analysis of hepatotoxins in vitro reveals a correlation between gene expression profiles and mechanisms of toxicity. Toxicol Lett. 120: 359-368.
Waring, J.F., Jolly, R.A., Ciurlionis, R., Lum, P.Y., Praestgaard, J.T., Morfitt, D.C., Buratto, B., Roberts, C., Schadt, E., and Ulrich, R.G. 2001. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. Toxicol. Appl. Pharmacol. 175: 28-42.
Watarai, H., Inagaki, Y., Kubota, N., Fuju, K., Nagafune, J., Yamaguchi, Y., and Kadoya, T. 2000. Proteomic approach to the identification of cell membrane proteins. Electrophoresis. 21: 460-464.
Waters, N.J., Holmes, E., Williams, A., Waterfield, C.J., Farrant, R.D., and Nicholson, J.K. 2001. NMR and pattern recognition studies on the time-related metabolic effects of alpha-naphthylisothiocyanate on liver, urine, and plasma in the rat: an integrative metabonomic approach. Chem. Res. Toxicol. 14: 1401-1412.

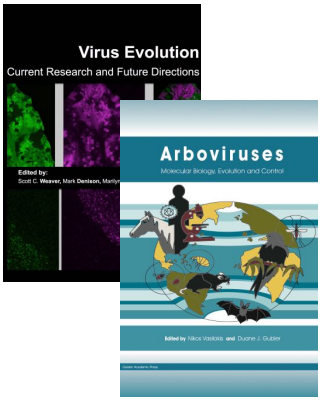
Weinberger, S.R., Morris, T.S., and Pawlak, M. 2000. Recent trends in protein biochip technology. Pharmacogenomics. 1: 395-416.

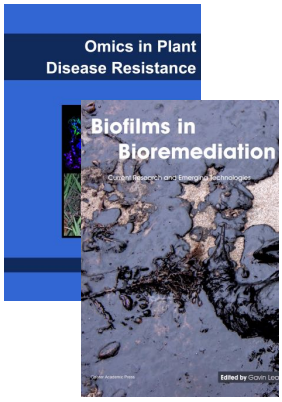
Zhou, J., and Bennett, S. 1997. A supervised learning network based on adaptive resonance theory. Int. J. Neural. Syst. 8: 239-246.

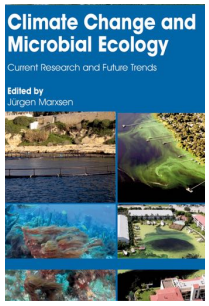
Further Reading

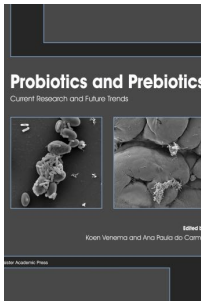
Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at **caister.com**

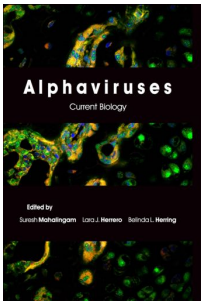
- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof
- ***Aspergillus* and *Penicillium* in the Post-genomic Era**
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2
- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorit, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins
- **Omics in Plant Disease Resistance**
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr
- **Acidophiles: Life in Extremely Acidic Environments**
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles
- **Climate Change and Microbial Ecology: Current Research and Future Trends**
  Edited by: J Marxsen (2016)
  www.caister.com/climate
- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
  Edited by: G Lear (2016)
  www.caister.com/biorem
- **Microalgae: Current Research and Applications**
  Edited by: MN Tsaloglou (2016)
  www.caister.com/microalgae
- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma
- **Virus Evolution: Current Research and Future Directions**
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol
- **Arboviruses: Molecular Biology, Evolution and Control**
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo
- ***Shigella*: Molecular and Cellular Biology**
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella
- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
  Edited by: AM Romaní, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms
- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha
- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile



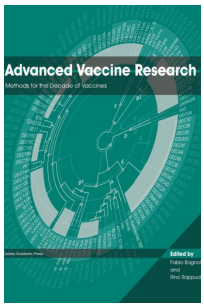


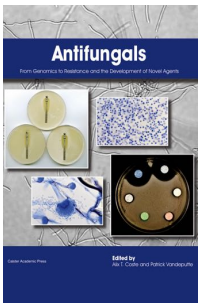












- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow
- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics
- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015
- ***Corynebacterium glutamicum*: From Systems Biology to Biotechnological Applications**
  Edited by: A Burkovski (2015)
  www.caister.com/cory2
- **Advanced Vaccine Research Methods for the Decade of Vaccines**
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines
- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals
- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
  Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
  www.caister.com/bacteria-plant
- ***Aeromonas***
  Edited by: J Graf (2015)
  www.caister.com/aeromonas
- **Antibiotics: Current Innovations and Future Trends**
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics
- ***Leishmania*: Current Biology and Control**
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2
- ***Acanthamoeba*: Biology and Pathogenesis (2nd edition)**
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2
- **Microarrays: Current Technology, Innovations and Applications**
  Edited by: Z He (2014)
  www.caister.com/microarrays2
- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order