Toxicogenomics: a new revolution in drug safety

Arthur L. Castle, Michael P. Carver and Donna L. Mendrick

New drugs are screened for adverse reactions using a laborious, costly process and still some promising therapeutics are withdrawn from the marketplace because of unforeseen human toxicity. Novel higher throughput methods in toxicology need to be developed. These new approaches should provide more insight into potential human toxicity than current methods. Toxicogenomics, the examination of changes in gene expression following exposure to a toxicant, offers the potential to identify a human toxicant earlier in drug development and to detect human-specific toxicants that cause no adverse reaction in rats.

Arthur L. Castle and Donna L. Mendrick*
Gene Logic, Gaithersburg, MD, USA
and Michael P. Carver
Wyeth, Andover, MA, USA
*tel: +1 301 987 1741
fax: +1 301 987 1701
e-mail: dmendrick@genelogic.com

▼ The biggest improvement needed in the drug development process is in the field of toxicology, which is the point where most developmental bottlenecks occur. One promising area of advancement is the new field of toxicogenomics. Detection of changes at the molecular level provides insight into a toxicant's mechanism-of-action and its potential to cause human toxicity. Toxicogenomics has grown quickly with the number of articles published approximately tripling in 2001 (estimated) over that of 2000. However, the field is still in its infancy, as shown by the majority of review articles describing the promise of this technology. As toxicogenomics data grows, a developing challenge is the analysis of large datasets and the building of predictive toxicogenomic databases. Here, this challenge is addressed together with how toxicogenomics can predict toxicity faster than classical measurements, such as serum chemistry perturbations.

Gene Logic (Gaithersburg, MD, USA) is building a reference database called the ToxExpress™ Module. Examples of its use to predict toxicity of drugs and chemicals before standard signs of toxicity in rats are observable are presented. Particularly promising is an example of how this database correctly predicted a drug that is toxic in humans, yet failed to elicit obvious adverse reactions in rats. This suggests that toxicogenomics could predict a compound's potential human toxicity better than existing methods.

Analysis methods

Although specific examples suggest that microarray technology holds the promise of being more predictive then current toxicological methods, the accuracy and reproducibility of gene expression information for various toxins will ultimately determine the efficacy of these predictive methods. At Gene Logic, an accurate gene expression profile for a toxin comes from a time-course study conducted with a minimum of three biological replicates. These individual studies also need to be examined in a meta-analysis to understand general molecular mechanisms of toxicology.

Studies within our reference database profile the expression changes of thousands of genes. Multiple time points, doses and vehicles are used in these studies. Marker genes that can classify the types of toxicity and the potential mechanisms-of-action need to be identified within studies and among related studies. Not all genes are expected to change expression after toxicant exposure and all gene expression is subject to normal biological variability. We can consider the genes whose expression changes because of toxic exposure as signals and the biological expression variability as noise. By building predictive models or grouping only genes that are affected by toxicity we can increase our signal-to-noise ratio and our understanding of the mechanism of toxicity. The key to good target selection is deciphering which genes are primarily affected by toxicity signals rather then noise. Care is taken to discern a pharmacological from a toxicological effect by using multiple

 1359-6446/02/$ – see front matter PII: S1359-6446(02)02327-9

doses of each compound and drugs that have a similar mechanism-of-action but that differ in their toxicological profile.

Marker genes

Selection of marker genes often entails ranking with respect to differences between experimental groups. Therefore, the definition of difference becomes important. Ideally, this should be based on statistical inference where genes are ranked by the probability that their expression values are different in a reproducible sense and not the result of random noise [1]. However, the ability to infer significance depends on the sample size and population assumptions. Not all genes are expected to have the same dynamic range or variability; therefore, they do not belong to the same population. A large database enables direct determination of biological variability and dynamic range for every gene. For this reason, the company generally collects data from at least a thousand biological samples per organ to directly determine dynamic range and variability. Information on each gene can thus stand alone without the assumption that all genes behave in a similar way and therefore belong to the same population.

Both parametric and non-parametric methods are available to help rank genes based on inference or discrimination ability [1–6]. When the gene expression data is made normally distributed, the common student t-test, Z-test or analysis of variance (ANOVA) are used to rank genes [1,6]. Simply accepting all genes with p values <0.05 as targets will cause another problem; large numbers of false-positives will occur as a result of multiple tests, each with a 5% chance of being a false-positive. Randomization is often used to adjust p values for multiple tests because it is most likely to give a good estimate of false-positive rate [1,5,6].

Although the ideal scenario is where one has a large database, often few or no replicates are run. In those cases, an arbitrary fold change cut-off is often used [7–9], where the significance can only be interpreted in reference to normal gene expression variation; however, some attempts have been made to improve gene expression variability estimates. Variability between different genes has been measured and reapplied to each individual gene [4,9,10]. A probability for a fold change or relative difference is then calculated based on this pooled variability estimate. Genes that are most variable between selected conditions or successive rounds of clustering have also been used to select target genes. However, all these methods assume that all genes belong to a population with similar biological variability and dynamic range [11,12].

With a thousand or more observations for every gene in our database, direct examination of gene expression distributions are possible, as are the selection of targets for determining both the presence of toxicity and its probable mechanism.

Classification of samples

Once target genes have been selected, information from multiple genes is used to classify a drug as toxic and to determine its potential toxicity types. Many methods have been used to classify samples and genes in microarray experiments. Sample classification starts with known samples called a training set. Combinations of known sample genes that change reproducibly are identified, then used to determine the treatment of unknown samples. Classification success depends on the reliability that changes in known samples accurately reflect what will be seen in the population being tested. An additional challenge is proving that these combinations of gene expression changes are relatively unique to the classification being attempted. Classification as normal or pathological must consider all possible normal states and abnormal pathologies. Therefore, a large number of toxin-treated and control samples must be used to approximate the uniqueness of the markers to measuring the toxic events.

Many methods of classification have been applied to microarray data and there are several public data sets that can be used to compare methods [13–16]. Each method classifies samples by different criteria and no one method should be expected to work best on all types of data. The goals of the experiment and its design should dictate classification methodology. An overview is presented in Table 1.

Several methods exist to classify samples based on the expression of several genes. These methods include logistic regression, linear discriminant analysis (LDA), neural networks, support vector machine and co-clustering approaches [2,3,17–24]. In these methods, a training set of data is generated for use in rule or formula building; these rules or formulas can then be applied to classify samples into one or more groups.

Logistic regression and LDA weight the contribution of each gene expression value to predict a sample based on the distribution of that gene in the training set [2,3,25]. Therefore, these methods require enough samples to adequately estimate a distribution. Tens to hundreds of samples might be needed depending on the variability of the gene expression data. Both logistic regression and LDA use statistical inference to weight the contributions of each gene expression value in sample prediction. Logistic regression estimates a linear log likelihood function based on the distribution of values between groups, whereas LDA uses probability functions based on normal distributions and can have non-parametric forms [2,3].

Table 1. Multivariate methods of sample classification

Method	How it works
Linear discriminant analysis	Weights assigned to each gene based on distribution of known samples in the training set. Requires enough samples to adequately estimate distribution. Uses probability functions based on normal distributions and requires the assumption of normality.
Logistic regression	Weights assigned to each gene based on distribution in training set and requires enough samples to adequately estimate distribution. Estimate a linear log likelihood function for distribution values. Robust against departures from normal distributions.
Neural networks	Use functional units called nodes that calculate inputs, process relationships between input nodes and calculate output. Weights assigned are learned from a training set. Requires many samples to accurately estimate weights. No assumptions on underlying distributions are needed. Can learn from new information.
Support vector machines	Draws an optimal complex boundary through multidimensional space that separates sample groups in the training set. Depends on each sample being good representatives of groups. Highly sensitive to measurement errors or outliers.
Clustering approaches	Uses a guilt by association method. Samples are classified as to which they most closely resemble. Can be a useful method when the number of samples is small. Assumes the same measure of closeness can be applied equally to all genes. The measure of closeness (metric) can greatly influence results.

Neural networks do not use statistical inference to weight genes. Instead they have functional units called nodes that calculate inputs, process relationships or rules between input nodes or calculate outputs. The weights used for these processes are learned from a training set of data [25,26]. Neural networks need many samples for accurate node weights estimation and network functionality but have the advantage of being able to learn complex patterns and adapt to new information.

Clustering approaches use a 'guilt by association' method to classify samples and have been used to classify types of toxicity [17,18,20–22,27–29]. New samples are classified by which samples they most resemble. The methods for calculating similarity vary, but the same measure is applied to each gene. Difficulties can occur because gene expression is not weighted by statistical inference or through a training regime; therefore, many insignificant relationships can drive the classification [5,25,30]. Despite these disadvantages, clustering methods are often used for sample classification, especially when too few samples are present for more robust statistical inference or rule-based methods.

Classification of genes

Genes can be classified according to how close their expression profiles are related across many samples. Clustering is a method that is widely used to analyze gene expression. However, a general problem with clustering is that the same measure of similarity is used with every gene [25]. The differences in dynamic range and variability between genes are often not known and assumed to be similar so no accurate level of confidence can be assigned to distinguish an expression pattern from biological noise. Most clustering methods also assume that a gene can belong to only one cluster; a condition that is inconstant with the intricate interconnections of pathways known to occur in biology. In addition, hierarchical clustering assumes that a parent–child type relationship exists in the regulation of all genes. Although some of these assumptions might not be valid for gene expression, clustering can identify new relationships between genes that can then be validated by experimental manipulation.

Some of the first papers on gene classification used hierarchical clustering to classify genes [31–33]. This method has been used to create a tree diagram where genes that most resemble each other in expression patterns are grouped together; these groups are then sub-grouped by how closely they resemble other groups. These trees can be built divisively by progressively splitting from one all-inclusive group or by pairing samples and building upon these pairs in an agglomerative approach [17,30,33,34]. The trees generated by hierarchical clustering create a parent–child structure where individual samples or groups are treated as subgroups of larger groups. Genes in a hierarchical cluster are compared to each other using a formula called the metric method (which determines how different any groups of genes are from each other) and a linkage method (which determines how comparisons are made) [30,33,35]. Each group of genes has a representative group average pattern,

which could be unlike any single constituent gene. A key assumption is that this pattern of overall similarity is representative of its constituents [25,30]. If the underlying biology being examined is expected to have a hierarchical structure then this method is appropriate [30]. However, where hierarchical structure is not supported by the phenomena being studied, other methods of clustering could be more effective.

K-means clustering groups of genes together by sorting each gene expression pattern into a predetermined number of clusters [34,36,37]. Because genes are not forced into a classification hierarchy, this method has an advantage over hierarchical clustering if only a few patterns of expression are expected to dominate the analysis and the relationships between these clusters does not fit into a parent–child pattern. However, K-means clustering estimates no between-cluster relationships. Also, each gene must belong to one cluster regardless of the overall fit to that cluster and each gene can belong only to one cluster. Other methods (such as self-organizing maps and multidimensional scaling) can map genes or samples relative to each other in 2D or 3D space but rely on the user supervision to then define clusters [2,21,34,38–40].

Clustering will always classify genes and it is important to compare the results from multiple methods [17,21,30, 34,35,41]. Results can be cross-validated or the stability of clusters tested by adding random noise or doing randomization experiments [5,16,22,42]. If the results are unstable, and highly sensitive to changes in the above parameters, then experimental noise could be driving the creation of the cluster. As with other methods, replication of experiments will improve the likelihood of generating stable clusters [5].

Genes or groups of genes can also be classified by how they interact with each other. Genetic networks model the interactions between genes enabling a gene to be in more than one cluster, which is most consistent with biological observations [43–50]. The simplest form is a Boolean network where all genes are related to each other by an 'on or off' pattern [49,51]. More complex networks include modeling of a saturating S shape response among genes, linear equation of each gene's function based on weighted effects of all other genes, and self-evolving algorithms [32,44,45,50–52]. Although these methods could help reverse engineer complex mechanisms, they require large amounts of accurate data [51].

Genomic platforms

Gene expression can be monitored one gene at a time or by the thousands. Quantitative real-time PCR (Q-RT-PCR) enables the quantitation of one gene at a time and is robust; useful traits when the genes of interest are known. However, to identify genes of interest, a broad survey approach is more useful. Microarrays (closed systems) and open platforms, such as READS™ (Gene Logic) [53–55] differential display technology, enable the simultaneous evaluation of thousands of genes.

Microarray platforms offer greater speed and more quantitative information than the differential display technology but this platform provides advantages: (1) detecting novel as well as known genes and (2) enabling the profiling of any species, even plants. Microarray platforms use either cDNA clones, or oligonucleotides, which can discriminate small sequence variations among highly homologous genes (e.g. the cytochrome P450 family), DNA mutations and polymorphic variations between individuals [56,57]. Technical aspects of these platforms have been the subject of a recent review and will not be discussed here [58].

The combination of READS™ and ToxExpress™ (Gene Logic) and GeneChip® (Affymetrix, Santa Clara, CA, USA) enables the detection of both novel and known genes that respond to toxicants, as well as validation of interesting genes using Q-RT-PCR.

Predictive toxicology

Two recent papers have demonstrated the ability to cluster hepatotoxins using the gene expression profiles from animals treated with these agents [29,59]. Bulera and co-workers. identified a blinded sample treated with one of the toxicants using clustering analysis [59]. Waring *et al.* reported the ability to show strong correlation between classical toxicity endpoints (e.g. clinical chemistry values) and gene expression changes [29]. Both of these papers illustrate the promise of toxicogenomics.

As will be discussed later, the enormous costs of developing a large, predictive database for determining potential compound toxicity have relegated this type of project to the private sector. We are building ToxExpress™ with expression data from marketed pharmaceuticals, classical chemical toxicants, and a few proprietary drugs, which are donated to the program by customers. These compounds span pharmacological and structural classes and some are rat or human specific, which results in a database that is useful for predicting toxicity during development of therapeutics.

The levels of mRNA expression in target tissues are examined using multiple platforms as described previously. The focus here will be limited to data generated using the GeneChip® microarray platform, whose current rat genome array examines the expression of >26,000 genes and expressed sequence tags (ESTs) simultaneously. As

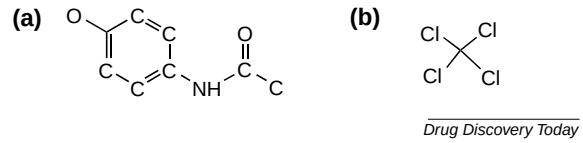

Figure 1. Chemical structure of acetaminophen **(a)** and carbon tetrachloride **(b)**. Images reproduced from The Merck Index.

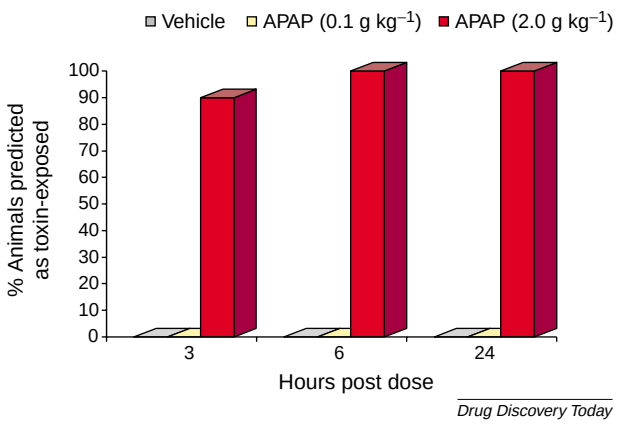

Figure 2. Animals were given one oral treatment with vehicle or acetaminophen (APAP) and sacrificed 3, 6 or 24 h later. The modeling software correctly predicted that none of the animals treated with vehicle (white bars) or a low dose of APAP (yellow bars) were exposed to a toxicant. By contrast, the animals that received a toxic dose of APAP were seen as toxin-exposed, shown by the red bars.

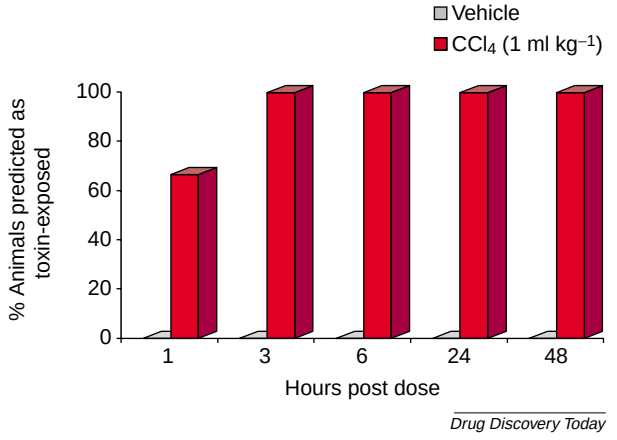

Figure 3. Animals were given one oral treatment with vehicle or a toxic dose of carbon tetrachloride (CCl_4) and sacrificed 1, 3, 6, 24 or 48 h later. The modeling software correctly predicted that none of the animals treated with vehicle (white bars) were exposed to a toxicant. By contrast, all of the animals that received a toxic dose of carbon tetrachloride (red bars) were seen as toxin-exposed when sacrificed 3–48 h post-exposure. Even at 1 h post-exposure, the modeling software correctly predicted the majority of carbon tetrachloride treated rats.

discussed above, it is crucial to employ statistical methods for microarray data analysis; thus, 3–10 biological replicates are built into every experiment. Resultant gene expression data are processed using both proprietary-modeling algorithms, as well as commonly used clustering and statistical methods, such as linear discriminant analysis [60]. There are two basic aims of this program: (1) to provide databases that will drive early-stage compound toxicity ranking using *in vitro* or *in vivo* testing; and (2) to enable users to leverage this database during pre-IND (Investigational New Drug) submission toxicity studies when multiple *in vivo* rat tissues are exposed to the drug under study.

The depth of our database in terms of the magnitude of genes surveyed and the number of replicates (3–10), doses (2) and time points (3–5) enables the user to obtain detailed information on their compound. The modeling software ranks the potential toxicity, provides information on the type of pathology likely to occur and whether this might be a rat- or human-specific event, and matches the unknown compound to the closest agent in the reference database (enabling further understanding of its mechanism-of-action [60]). Using models built with gene expression data, toxicity can be predicted at early phases before classical toxicological changes occur (e.g. alterations in serum parameters or pathological evidence of toxicity). These models are robust enough to predict toxicity across multiple structural classes, such as acetaminophen (acetyl-*p*-aminophenol; APAP) and carbon tetrachloride (CCl_4) (Fig. 1) [61]. The modeling software accurately predicts that rats treated with high doses of APAP and CCl_4 have been exposed to hepatotoxicants as shown in Figs 2 and 3. Furthermore, data obtained in a blinded fashion have been correctly classified as either exposed to a toxicant or non-toxicant (data not shown).

Treatment with high doses of APAP did not elicit clinical chemistry or histological changes at 3 or 6 h post-dose, yet the toxicogenomic evaluation correctly identified those animals treated with this toxic dose. Similarly, animals treated with CCl_4 failed to show clinical chemistry signs of toxicity until 6 h post-dose. Examination of the liver at 1 h showed no alterations and minimal changes at 3 h post-dose, yet toxicogenomic studies identified most of the rats sacrificed at the 1 h time point and all animals studied 3 h post-dose.

Toxicogenomics detection of a human-specific toxicant

It is known that preclinical studies in animals do not always detect drugs that prove to be toxic in humans. To determine if toxicogenomics can detect some of these harmful drugs before they enter the clinic, tacrine (Cognex®;

Sigma, St Louis, MO, USA) was studied. Controls include donepezil (Aricept®; Pfizer, New York, NY, USA) and physostigmine (Antilirium®; Sigma), which have a mechanism-of-action similar to tacrine but do not induce hepatotoxicity. Tacrine and donepezil are used clinically to treat patients with Alzheimer's disease. Physostigmine was used in a clinical trial but has not been approved by the FDA for this indication. These drugs are reversible inhibitors of acetylcholinesterase although selectivity for cholinesterases varies between the drugs (Table 2).

Table 2. Comparisons of acetylcholinesterase inhibitors

	Inhibitory action [62]		Hepatotoxicity	
Drug	**Acetylcholinesterase**	**Butyrylcholinesterase**	**Rats**	**Humans**
Tacrine	Yes	Yes	No	Yes
Donepezil	Yes	No	No	No
Physostigmine	Yes	Yes	No	No

Unfortunately, ~25% of patients treated with tacrine exhibit asymptomatic elevation of serum aminotransferase concentrations greater than threefold within normal limits and biopsies performed on a small number of patients have revealed liver necrosis [62,63]. The adverse events associated with tacrine have severely limited its clinical use [62]. There are no indications of liver toxicity in animals or humans following treatment with donepezil or physostigmine [62,64]. Tacrine is viewed as a human-specific toxicant because it induces elevations of liver transaminase levels in humans but there was no evidence of liver toxicity in mice, rats or dogs during preclinical development [65,66].

In the study shown later, rats were dosed by oral gavage once at time 0 h, then followed for 48 h [67]. There were no serum chemistry changes or alterations in the liver that suggested the rats had been treated with a hepatotoxicant; this is similar to what was seen in preclinical testing [65]. However, the modeling software build upon data in the reference database identified many of the animals treated with high doses of tacrine and sacrificed at 6 h as exposed to a toxicant (Fig. 4a). None of the animals treated with vehicle, low doses of tacrine, donepezil or physostigmine were designated as toxin-exposed.

The predictive ability is based on the identity and magnitude of gene responses, not just on the number of gene expression changes. As an illustration, the gene expression levels from each group were compared with those in control groups. The number of genes at each time point and dose that exceeded a twofold change in expression (with a $p < 0.01$), compared with vehicle-treated animals were graphed (Fig. 4b). As these data illustrate, toxicity prediction is not based solely on the number of genes whose expression is affected by a drug. Rats treated with physostigmine exhibited significant changes in the expression level of many genes but were not predicted as having been exposed to a toxicant. Similarly, rats treated with donepezil and sacrificed 24 h post-dose exhibited more changes in gene expression than did high dose tacrine treated animals at this time point, yet none of the former were predicted as being toxin-exposed while some of the latter were correctly identified. It is not yet clear how often a toxicogenomic analysis in rats will detect a human specific toxicant although there are several such studies being performed.

Public initiatives

Pharmaceutical companies, working through ILSI (International Life Sciences Institute, a nonprofit scientific

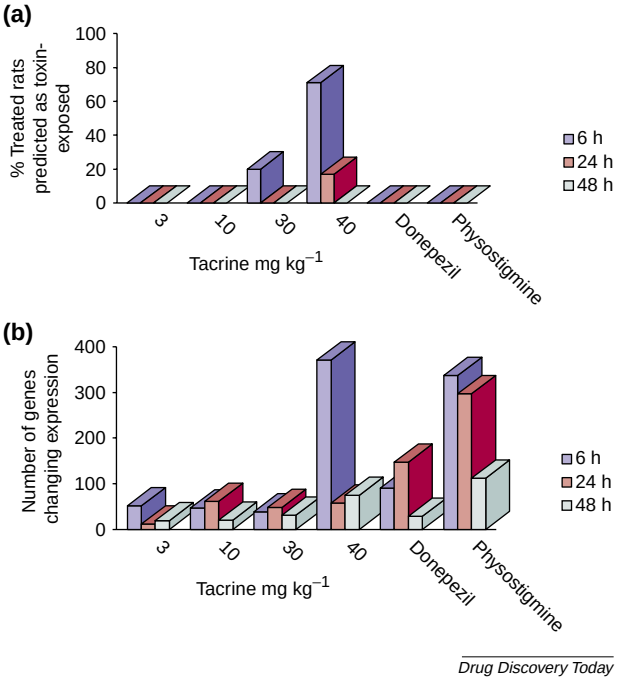

Figure 4. Animals were dosed orally once with vehicle, tacrine, donepezil or physostigmine and then sacrificed 6, 24 or 48 h later. **(a)** The percentage of animals identified, on the basis of ToxExpress™ (Gene Logic, Gaithersburg, MD, USA) modeling software, as exposed to toxicants or non-toxicants is illustrated. **(b)** The number of genes in each drug-treated group whose expression level changed in comparison to vehicle-treated rats is illustrated. Criteria used for this graph were genes exhibiting a twofold or greater change with statistical significant of $p < 0.01$. The illustration is not limited to genes showing predictive toxicity but portrays all genes present on the Affymetrix (Santa Clara, CA, USA) rat A microarray.

research foundation based in Washington, DC, USA; http://www.ilsi.org), have formed a committee to assess the use of genomics and proteomics in mechanism-based risk assessment. The long-term objectives of ILSI are to develop a broad-based multinational collaboration involving academic, governmental and industrial institutions to explore the use of genomic and proteomic technologies for gaining insight into the mechanisms behind the toxic response. Together, they will assist the development of a consensus on potential applications of such data with respect for mechanism-based risk assessment. At this time, there is no mandate to develop a large, predictive database.

Researchers at the National Institute of Environmental Health Sciences (NIEHS; Research Triangle Park, NC, USA) have formed a microarray center using cDNA microarray technologies (http://dir.niehs.nih.gov/microarray/home.htm). This effort is focused on understanding the mechanisms-of-action of environmental agents, studying the effect of low and high dose exposures, using gene expression information as biomarkers to assess human exposure, and developing a public database of gene expression profiles; for this, they are working within the NIEHS intramural and extramural communities, as well as forming industrial partnerships.

Impact on the pharmaceutical industry

The potential benefits of a successful toxicogenomics program have been described briefly. Treatment with animals or cells in culture with new chemical entities (NCE) and examination of the resulting gene expression profiles can influence several areas of drug development. First, it can impact the quality of drug development pipelines by improving the science of toxicology, providing more specific information as to the mechanisms of drug pathologies and providing it earlier in the discovery–development continuum. Second, it can improve the efficiency of the process because toxicogenomics information complements genomic target identification and characterization methods used in discovery and leads to reduced attrition during drug development for unfavorable compounds. Toxicogenomics can be applied at any stage in the drug development process, but appears to have greatest potential use when used in one or more of the following settings:

- The development of molecular screens for candidate selection in early or late discovery phases.
- Better understanding of the mechanisms of target-organ pathologies in animals versus man (so-called 'issue management' in preclinical development).
- Developing short-term biomarkers for subchronic or chronic toxicities in late preclinical development.
- Developing robust and predictive biomarkers for drug-related effects in man, which are likely to occur during Phase II and Phase III clinical trials.

The risk to a pharmaceutical company for misunderstanding incurred with toxicogenomics approaches will depend on multiple factors, such as the technology and the type of studies employed (i.e. *in vivo* and *in vitro*, examining reactions in animals or humans). In general, *in vivo* studies using global gene expression profiling platforms with compounds that have already advanced into clinical trials is considered to have the highest risk of uncovering some unexplained or uninterpretable toxicogenomics data. The least amount of risk to drug development would be seen with *in vitro* studies using only model compounds from the literature and clinical trial failures or less potent analogs from the discovery program of interest. Alternatively, *in vivo* or *in vitro* systems that use a targeted approach in which only a few genes of known function are measured should be of low risk.

By providing a rational basis for interpreting the multitude of gene expression changes that might occur in a single toxicogenomic experiment, a comprehensive database is essential for minimizing these perceived risks to a promising drug. The predictive models described previously can identify patterns of gene expression changes that are relevant to a particular pathology or known toxicant by following the time course of many genes at once. Therefore, it is not necessary to know the identity of individual genes and ESTs or even their functions, to use this approach in drug discovery and development. The key consideration then becomes how accurate the model predictions are, not which individual genes are induced or repressed.

We are proposing to use toxicogenomics to make more mechanistic-based decisions about target-organ toxicities than standard batteries of biochemical toxicity endpoints currently provide. Thus, it seems reasonable to maximize the specificity of the predictions (i.e. reduce the false-positive rate), at the expense of sensitivity (i.e. enable a higher false-negative rate), wherever possible [68,69]. There are at least two strong arguments for building more specific toxicogenomic models to be used in drug discovery and development. First, a high false-positive rate suggests that the model is seriously flawed. A high false-negative rate can often be attributed to a lack of data, which usually means too few gene expression profiles (training sets of compounds, time courses, doses, replicates) are in the database to adequately predict target-organ toxicity for the NCE of interest. Second, a high false-positive rate could eliminate good drug candidates prematurely in the process, particularly if toxicogenomic screens are used early in the discovery

phase. The higher false-negative rates that result from minimizing the false-positives in the predictive models only become problematic when toxicogenomics information is used to make decisions late in preclinical development. Such a strategy could enable toxic compounds to continue into clinical development, where the potential for human suffering and the development costs become quite high, and perhaps even unbearable.

References

1 Kerr, M.K. and Churchill, G.A. (2001) Statistical design and the analysis of gene expression microarray data. *Genet. Res.* 77, 123–128
2 Golub, T.R. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537
3 Raychaudhuri, S. *et al.* (2000) Pattern recognition of genomic features with microarrays: site typing of *Mycobacterium tuberculosis* strains. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 8, 286–295
4 Park, P.J. *et al.* (2001) A nonparametric scoring algorithm for identifying informative genes from microarray data. *Pac. Symp. Biocomput.* 52–63
5 Kerr, M.K. and Churchill, G.A. (2001) Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8961–8965
6 Callow, M.J. *et al.* (2000) Microarray expression profiling identifies genes with altered expression in HDL-deficient mice. *Genome Res.* 10, 2022–2029
7 Butte, A.J. *et al.* (2001) Determining significant fold differences in gene expression analysis. *Pac. Symp. Biocomput.* 6–17
8 Tsien, C.L. *et al.* (2001) On reporting fold differences. *Pac. Symp. Biocomput.* 496–507
9 Manduchi, E. *et al.* (2000) Generation of patterns from gene expression data by assigning confidence to differentially expressed genes. *Bioinformatics* 16, 685–698
10 Claverie, J.M. (1999) Computational methods for the identification of differential and coordinated gene expression. *Hum. Mol. Genet.* 8, 1821–1832
11 Fuhrman, S. *et al.* (2000) The application of shannon entropy in the identification of putative drug targets. *Biosystems* 55, 5–14
12 Hastie, T. *et al.* (2000) 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol.* 2(1), RESEARCH0003.1–RESEARCH0003.21
13 DeRisi, J.L. *et al.* (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686
14 Cho, R.J. *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* 2, 65–73
15 Spellman, P.T. *et al.* (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell.* 9, 3273–3297
16 Alizadeh, A.A. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511
17 Alon, U. *et al.* (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6745–6750
18 Perou, C.M. *et al.* (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9212–9217
19 Furey, T.S. *et al.* (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16, 906–914
20 Alizadeh, A.A. *et al.* (2001) Towards a novel classification of human malignancies based on gene expression patterns. *J. Pathol.* 195, 41–52
21 Bittner, M. *et al.* (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406, 536–540
22 Califano, A. *et al.* (2000) Analysis of gene expression microarrays for phenotype classification. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 8, 75–85
23 Raychaudhuri, S. *et al.* (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac. Symp. Biocomput.* 455–466
24 Brown, M.P. *et al.* (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. U. S. A.* 97, 262–267
25 Raychaudhuri, S. *et al.* (2001) Basic microarray analysis: grouping and feature reduction. *Trends Biotechnol.* 19, 189–193
26 Khan, J. *et al.* (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* 7, 673–679
27 Cheng, Y. and Church, G.M. (2000) Biclustering of expression data. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 8, 93–103
28 Sorlie, T. *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10869–10874
29 Waring, J.F. *et al.* (2001) Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharmacol.* 175, 28–42
30 Planet, P.J. *et al.* (2001) Systematic analysis of DNA microarray data: ordering and interpreting patterns of gene expression. *Genome Res.* 11, 1149–1155
31 Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
32 Michaels, G.S. *et al.* (1998) Cluster analysis and data visualization of large-scale gene expression data. *Pac. Symp. Biocomput.* 42–53
33 Eisen, M.B. and Brown, P.O. (1999) DNA arrays for analysis of gene expression. *Methods Enzymol.* 303, 179–205
34 Sherlock, G. (2000) Analysis of large-scale gene expression data. *Curr. Opin. Immunol.* 12, 201–205
35 Heyer, L.J. *et al.* (1999) Exploring expression data: identification and analysis of coexpressed genes. *Genome Res.* 9, 1106–1115
36 Aronow, B.J. *et al.* (2001) Microarray analysis of trophoblast differentiation: gene expression reprogramming in key gene function categories. *Physiol. Genomics* 6, 105–116
37 Huang, J. *et al.* (2001) Effects of ischemia on gene expression. *J. Surg. Res.* 99, 222–227
38 Dysvik, B. and Jonassen, I. (2001) J-Express: exploring gene expression data using Java. *Bioinformatics* 17, 369–370
39 Tamayo, P. *et al.* (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2907–2912
40 Toronen, P. *et al.* (1999) Analysis of gene expression data using self-organizing maps. *FEBS Lett.* 451, 142–146
41 Altman, R.B. and Raychaudhuri, S. (2001) Whole-genome expression analysis: challenges beyond clustering. *Curr. Opin. Struct. Biol.* 11, 340–347
42 Yeung, K.Y. *et al.* (2001) Validating clustering for gene expression data. *Bioinformatics* 17, 309–318
43 McAdams, H.H. and Shapiro, L. (1995) Circuit simulation of genetic networks. *Science* 269, 650–656
44 Maki, Y. *et al.* (2001) Development of a system for the inference of large scale genetic networks. *Pac. Symp. Biocomput.* 446–458
45 van Someren, E.P. *et al.* (2000) Linear modeling of genetic networks from experimental data. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 8, 355–366
46 Hartemink, A.J. *et al.* (2001) Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. *Pac. Symp. Biocomput.* 422–433
47 D'Haeseleer, P. *et al.* (1999) Linear modeling of mRNA expression levels during CNS development and injury. *Pac. Symp. Biocomput.* 41–52
48 Wessels, L.F. *et al.* (1901) A comparison of genetic network models. *Pac. Symp. Biocomput.* 508–519
49 Akutsu, T. *et al.* (2000) Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function. *J. Comput. Biol.* 7, 331–343

50 Akutsu, T. *et al.* (2000) Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* 16, 727–734

51 Liang, S. *et al.* (1998) Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.* 18–29

52 Koza, J.R. *et al.* (2001) Reverse engineering of metabolic pathways from observed data using genetic programming. *Pac. Symp. Biocomput.* 434–445

53 Prashar, Y. and Weissman, S.M. (1999) READS: a method for display of 3′-end fragments of restriction enzyme-digested cDNAs for analysis of differential gene expression. *Methods Enzymol.* 303, 258–272

54 Lennon, G.G. (2000) High-throughput gene expression analysis for drug discovery. *Drug Discov. Today* 5, 59–66

55 Prashar, Y. and Weissman, S.M. (1996) Analysis of differential gene expression by display of 3′ end restriction fragments of cDNAs. *Proc. Natl. Acad. Sci. U. S. A.* 93, 659–663

56 Afshari, C.A. *et al.* (1999) Application of complementary DNA microarray technology to carcinogen identification, toxicology and drug safety evaluation. *Cancer Res.* 59, 4759–4760

57 Gerhold, D. *et al.* (2001) Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiol. Rev.* 5, 161–170

58 Cunningham, M.J. (2000) Genomics and proteomics. The new millennium of drug discovery and development. *J. Pharmacol. Toxicol. Methods* 44, 291–300

59 Bulera, S.J. *et al.* (2001) RNA expression in the early characterization of hepatotoxicants in Wistar rats by high-density DNA microarrays. *Hepatology* 33, 1239–1258

60 Castle, A.L. *et al.* (2002) Predictive modeling of hepatotoxicants using microarrays and a linear discriminant modeling approach. *Toxicol. Sci.* 66 (1S), 274. Presented at the *41st Annual Meeting of the Society of Toxicology*, 17–21 March 2002, Nashville, TN, USA

61 Mendrick, D.L. *et al.* (2002) Cross compound predictions using gene expression profiles from acetaminophen or carbon tetrachloride, two structurally distinct liver toxicants. *Toxicol. Sci.* 66 (1S), 275. Presented at the *41st Annual Meeting of the Society of Toxicology*, 17–21 March 2002, Nashville, TN, USA

62 Mayeux, R. and Sano, M. (1999) Treatment of Alzheimer's disease. *New Engl. J. Med.* 341, 1670–1679

63 *Physician's Desk Reference* (2002) (56 edn), Medical Economics Company

64 Nordberg, A. and Svensson, A.L. (1998) Cholinesterase inhibitors in the treatment of Alzheimer's Disease. A comparison of tolerability and pharmacology. *Drug Saf.* 19, 465–480

65 Woolf, T.F. *et al.* (1993) Bioactivation and irreversible binding of the cognition activator tacrine using human and rat liver microsomal preparations. Species difference. *Drug Metab. Dispos.* 21, 874–882

66 Fitten, L.J. *et al.* (1987) Long-term oral administration of memory-enhancing doses of tacrine in mice: a study of potential toxicity and side effects. *J. Gerontol.* 42, 681–685

67 Porter, M.W. *et al.* (2002) Identification through gene expression of a toxic response in the rat for a compound that exhibits overt toxicity in humans, but not in rats. *Toxicol. Sci.* 66 (1S), 295–296. Presented at the *41st Annual Meeting of the Society of Toxicology*, 17–21 March 2002, Nashville, TN, USA

68 Cooper, J.A. *et al.* (1979) Describing the validity of carcinogen screening tests. *Br. J. Cancer* 39, 87–89

69 Lave L.B. and Omenn, G.S. (1988) Screening toxic chemicals: how accurate must tests be? *J. Am. Coll. Toxicol.* 7, 565–574