

Expert Opinion

1. Introduction
2. Biomarkers defined
3. Challenges in qualifying molecular fingerprints as biomarkers
4. Application of molecular profiling in disease and toxicity biomarker development
5. Modelling transcriptional response towards predictive metabolism and toxicity
6. Conclusions
7. Expert opinion

For reprint orders, please
contact:
reprints@ashley-pub.com

Ashley Publications
www.ashley-pub.com



Molecular profiling approaches for identifying novel biomarkers

Wendy J Bailey[†] & Roger Ulrich

[†]*Merck & Co., Inc., 770 Sumneytown Pike, West Point, PA 19486, USA*

An unprecedented interest in biomarker development has arisen from the increasing use of genomic information and high-throughput technologies in the field of drug development. Monitoring global cellular responses to perturbation due to disease, drug treatment or toxicity is achieved using molecular profiling methods such as DNA microarrays, proteomics and metabonomics. Unique fingerprints composed of molecular changes are captured and subjected to interpretation with the goal of class discovery, comparison or prediction. Each fingerprint reflects a cumulative response of complex molecular interactions, and if these interactions can be significantly correlated to an end point, the molecular fingerprint may be qualified as a predictive biomarker. Furthermore, in cases where the predictive power of any single response or set of responses is statistically significant, a molecular fingerprint can provide novel information related to the underlying disease biology or mechanism of toxicity. There is an acute need for effective biomarkers in every phase of drug development, from discovery, to preclinical studies, through to clinical trials. The context in which these molecular biomarkers are used will depend upon the nature of the biological problem being addressed. This review will summarise experimental and computational efforts in the field of molecular profiling and discuss the significant challenges in interpreting molecular profiling data and qualifying novel transcriptional biomarkers.

Keywords: biomarker, microarray, molecular profiling, toxicogenomics, transcriptional fingerprint

Expert Opin. Drug Saf. (2004) 3(2):137-151

1. Introduction

There is a widespread expectation that with the completion of sequencing the human, mouse and rat genomes, biomedical research will advance at an accelerated rate. This expectation is heightened by the advent of high-throughput technologies covering everything from genome annotation to screening chemical compounds and the molecular profiling of DNA, mRNA, proteins and metabolites. Molecular profiling is defined as platform technologies capable of recording the entire cellular response to perturbation [1,2]. How close are we to the promise of a novel paradigm shift in this era of systems biology, and will the pay-off be more rapid development of effective therapies, custom-made for what ails each of us? There is no question that the scientific community can generate data at an unprecedented rate, but the challenge is to transform that information into meaningful, interpretable knowledge.

The bottleneck in terms of drug discovery and development has moved downstream beyond hypothesis generation, target identification and lead discovery, to choosing which hypotheses, targets and compounds warrant follow-up and critical resource allocation. A central component in this process is biomarker discovery and development. It is one thing to accelerate the pace at which therapeutic candidates advance through the pipeline, but it is another to see more of them make it through clinical trials and reach the market. Biomarkers have long played a crucial role during drug development through clinical trials in determining therapeutic efficacy,

drug toxicity and disease progression. However, the process of biomarker identification and validation has historically been an arduous one. If high-throughput technologies are going to succeed in delivering safer, more efficacious drugs, they must also deliver superior biomarkers ensuring that success.

This review begins by defining the term 'biomarker' and the contextual application of biomarkers in the drug discovery and development pipeline. Section 3 lays out the challenges in interpreting the data from transcriptional expression arrays, with a focus on evaluating biomarkers for toxicity and determining the appropriate context for the use of those biomarkers. Section 4 details specific examples of how expression profiling has been used for disease and compound classification and how application of this technology differs depending on the biological question being posited. Section 5 addresses the confounding factors in data interpretation and discusses various computational approaches to data analysis.

2. Biomarkers defined

Because the use of the term 'biomarker' is broadening and has been used to describe a range of measurable end points (e.g., drug, device, surgery, vaccine, biological agent and behavioural modification), the National Institutes of Health (NIH) formed a committee to provide working definitions for specific terms and a conceptual model of how biomarkers could be utilised [3]. According to the NIH Initiative on Biomarkers and Surrogate End Points, a biological marker (biomarker) is defined as:

- A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention [3].

A biomarker is distinguished from a clinical end point, which is defined as:

- A characteristic or variable that reflects how a patient feels, functions or survives [3].

Whilst a surrogate end point is defined as:

- A biomarker that is intended as a substitute for a clinical end point. A surrogate end point is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiological, therapeutic, pathophysiological or other scientific evidence [3].

Apparent in these definitions is a distinction between biomarker uses in a non-clinical versus a clinical setting. The degree to which a biomarker candidate is qualified will vary depending on the setting for its intended use. The evidence needed to qualify an analyte as a surrogate end point requires that the biomarker measurement correlates with a clinical end point and that the correlation is statistically reproducible. In the words of the FDA, a drug submitted for approval should have '...an effect on a surrogate end point that is

reasonably likely, based on epidemiological, therapeutic, pathophysiological or other evidence, to predict clinical benefit or on the basis of an effect on a clinical end point other than survival or irreversible morbidity' [4].

Furthermore, the NIH Biomarkers Definitions Working Group discourages the use of the term 'validation' to describe performance characteristics of a measurement or assay technique because it implies that a surrogate end point used for one therapeutic intervention can be generalised to include other interventions that may or may not relate to a specific end point. Therefore, the recommendation is to use the terms 'qualification' or 'evaluation' to describe the process of determining surrogate end point status as a substitute for a clinical end point.

Biomarkers can be classified based on the degree to which they are related to a clinical end point. First, a therapeutic intervention can have a direct effect on a putative biomarker, which provides either a reliable response making it a good candidate for a surrogate end point or an unreliable response due to large variations in response or because it cannot be assayed accurately. Second, the effect can be indirect but can provide a reliable response because the biomarker is associated with a secondary physiological, pathological or pharmacological process that is impacted by the intervention. Such a biomarker may be conditionally associated with an end point; for example, creatine kinase (CK) elevation can be an indicator of muscle myopathy unless an individual has participated in vigorous exercise. Such biomarkers become more useful as the relevant exceptions are defined. Third, the biomarker response is unlinked to a therapeutic intervention or disease process and is not predictive of clinical outcome [5].

The traditional paradigm of biomarker development must change given the growing number of drug candidates entering preclinical screening. Not only have genomic technologies allowed refinement of compound discovery in well-understood areas such as cardiovascular disease, they have also allowed headway to be made in less-understood therapeutic areas such as obesity and neurodegeneration [6]. The need to evaluate novel biomarkers in parallel with novel drug targets is exemplified by disorders that have a long presymptomatic period, such as Alzheimer's and Parkinson's disease [7-9]. This parallel approach should reduce the attrition of drugs in development by better predicting which compounds will fail due to suboptimal pharmacokinetic or pharmacodynamic properties, toxicity or lack of efficacy. Biomarker development is a multi-stage process in which validated targets or drugs can be used to identify biomarkers in the development phase. Next, in the preclinical phase, biomarkers can be used to monitor pharmacological and toxicological effects in *in vitro* or *in vivo* model systems. Established end points in animal studies can serve to qualify novel biomarkers in concert with the evaluation of novel compounds. Finally, during clinical trials and beyond, biomarkers are used to monitor the efficacy and safety of a compound in patients and, if approved as a surrogate end point, can serve as a diagnostic indicator.

A single biomarker need not function in all phases of development but rather, depending on what parameters are being measured or monitored, a sequential set of biomarkers could facilitate drug development more effectively than a single assay. Furthermore, the complexity of many disease processes and mechanism of action for therapeutic intervention makes it likely that a single biomarker will not account for all pathological or treatment effects. Therefore, it is equally unlikely that a single criterion can be used to link a surrogate end point to a disease. In such cases, the use of multiple biomarkers to provide a read-out of efficacious or adverse events could serve as a more reliable surrogate end point. This is where the promise of high-throughput and genomic technologies can advance the process of biomarker discovery. Several recent review articles explore issues related to biomarker assay development and provide a number of excellent examples of this process [5,7,10]. Redefining what is considered to be a biomarker and how it can be used throughout the drug development process is important now in light of the range of new approaches for predicting or monitoring disease progression or adverse drug effects. Molecular profiling strategies that monitor mRNA, protein and/or metabolic response to perturbation are all potential sources of novel biomarkers. The field of imaging biology is emerging as another potentially rich source for identifying novel biomarkers [7]. In fact, the current increased rate of novel target identification can be attributed to an increasing understanding of mechanisms of disease and toxicity as well as effects of therapeutic intervention brought about by the 'omics' technologies.

The field of biomarker discovery and development is complex and broad. This review will not attempt to present a comprehensive assessment of all aspects relating to this topic, but rather will focus on the second phase of biomarker development, preclinical compound evaluation, primarily looking for indicators of toxicity (defined in the broad sense as including adverse events that may not kill cells) using *in vivo* or *in vitro* models. The use of gene and/or protein expression technologies to monitor responses to adverse xenobiotic exposure is termed toxicogenomics. A qualified toxicity biomarker should predict an adverse event earlier than could otherwise be done without the biomarker. In order for the biomarker to be used in clinical trials, monitoring must be non-invasive (e.g., blood, urine and imaging). A proposed paradigm to accelerated compound evaluation for liabilities before they become drug development candidates is to build up a reference database of expression profiles from known toxicants. The success of such a database relies on the principle that shared mechanisms of toxicity induced by structurally related or unrelated compounds have similar transcriptional responses (e.g., guilt by association). Signatures from novel compounds in development would be queried against this toxicity compendium and 'go/no go' decisions will be based upon similarities between transcriptional signatures [11]. This is somewhat analogous to forensic techniques using images of human fingerprints or DNA microsatellite patterns searched

against national repositories. The difference is that there would likely not be an exact match unless the compound or a close analogue already exists in the database, but rather there would be a match to a set of predictive transcriptional changes (i.e., a molecular fingerprint) that serves as a biomarker indicative of a known toxicity [12-14]. The database could also reveal clues to unknown toxic mechanisms.

3. Challenges in qualifying molecular fingerprints as biomarkers

Molecular profiling yields a global readout from tissue or cultured cells in response to perturbation. For compound evaluation, this cellular response includes transcriptional or protein changes due to pharmacology, drug metabolism, adaptive homeostasis or toxicity. These changes can number several thousand; the magnitude of each can be large or small. Since platforms that monitor transcriptional changes are more widely used in toxicity screening than those for proteins, the following discussion will focus specifically on the application of gene expression microarray technologies [15,16]. In interpreting such data, the initial challenge is to parse the transcriptional changes that are a genuine response to the drug (i.e., the biological signal) from the background noise (e.g., variability inherent to the platform or between animals). Next, one must assign responding genes, identified as a relevant biological signal above, to specific toxicity classes as distinguished by different compounds. An additional challenge is to predict toxicity before any histological damage has occurred and before elevations are seen in standard clinical chemistry markers. For an appreciation of how transcriptional profiling can impact the assessment of drug safety, it is first necessary to understand the goals of the preclinical process and the protocols in current use for detecting toxicities.

Sistare [17] outlines the objectives of non-clinical studies as follows:

- Identify the dose-limiting and associated toxicities and discover whether or not the toxicities seen are reversible.
- Identify the safety pharmacology risks of the drug (and its metabolites) to the function of the CNS, cardiovascular system, and other systems.
- Define the toxicokinetics, that is, dose/exposure/time relationship.
- Define the safe starting dose for human trials.
- Define the appropriately safe patient-monitoring strategy for clinical trials.

The traditional biomarkers used in non-clinical studies include histopathology and the measurement of specific enzyme activity levels in blood (clinical chemistry). The accepted 'gold standard' for identifying toxicity in safety assessment studies is the histopathology evaluation of tissue after compound treatment, because often, tissue damage occurs prior to any change detected in routinely measured clinical chemistry parameters. These traditional biomarkers

have been in use for > 50 years with little modification [18]. Furthermore, histopathology is time- and labour-intensive and the accuracy of the evaluation can be subjective [19,20]. In order for molecular biomarkers to supplant histopathology as a preferred method for evaluating the safety of a compound in animals with application to the clinic, it must be demonstrated that they can be reliable surrogates of toxicity as well as accurately and reproducibly measured. There is no advantage over histopathology if the novel markers are not early predictors of toxicity and/or are not non-invasive. Factors that contribute to the difficulty in identifying a single molecular biomarker for toxicity that is both specific and sensitive include:

- Heterogeneity of animal response to drug treatment for any number of reasons, including genetic background, ADME (absorption, distribution, metabolism and excretion), physiology or gender.
- Complexity of the mechanism of action.
- Target organ specificity.
- Species specificity.
- Variability in the magnitude and duration of the response.

The problem of specificity can be addressed by the identification of multiple transcripts that correlate to a toxicological end point and represent independent cellular pathways. Any one transcript may contribute very little to the overall phenotype but multiple transcripts responding in concert to a cellular insult provide a unique molecular fingerprint. Specificity can be increased by a deeper understanding of the underlying biology and qualifying the biomarkers in a defined context. For example, advances in automated microscopic digital enhancement of tissue samples are allowing, for the first time, the ability to conduct unsupervised quantitative image analysis [21]. Kriete *et al.* combined this technology with gene expression profiling in a toxicity study using carbon tetrachloride (CCl₄)-treated rats [22]. The authors were able to identify transcripts correlated with cytological tissue metrics of pathology and a number of these genes are known to be associated with the mechanism of CCl₄ toxicity [23,24].

Evaluating a molecular fingerprint as a preclinical biomarker necessitates determining the 'correct' context for its use. A single profiling experiment is a snapshot in time; that is why collecting high quality data requires a detailed survey of transcriptional changes occurring over multiple doses, times, tissues and model organisms. Careful experimental design followed by appropriate data analysis may allow accurate temporal, spatial and mechanistic reconstruction of the toxicological response. The defined context may be restricted to a single species in a 'proof-of-principle' study targeted at screening closely related compounds, or it may be applicable to any compound resulting in a specific end point in, for example, kidney toxicity, which functions in a range of pre-clinical model organisms. The impediments for moving a safety biomarker into clinical trials as a surrogate end point are considerable [25]. In fact, much more so than for a marker

of efficacy due to the danger of missing a potentially life-threatening toxicity or an adverse effect. Conversely, a false positive response for a safety biomarker could unnecessarily result in withdrawing the candidate from further development. Thus, the application of a toxicogenomic approach in drug discovery before preclinical evaluation could result in more drugs advancing into clinical trials. Using molecular fingerprints identified from a reference database of expression profiles for known toxicants could serve as non-clinical biomarkers in this context.

4. Application of molecular profiling in disease and toxicity biomarker development

A simple Medline literature search illustrates the growing importance of genomic sciences to the field of toxicology. The number of research and review articles on the topic has grown from the tens to the hundreds in the last 5 years (Figure 1). In fact, it was not until 1999 that the term 'toxicogenomics' appeared in the literature [26]. In the late 1990s, the emphasis was on the use of pharmacogenetic approaches to customise therapeutic intervention based on single nucleotide polymorphisms (SNPs). Inherent variation in ADME genes, for example, can determine drug exposure levels [27,28]. Whilst characterising these SNPs in model organisms and humans has yielded clear successes, the interest in genomic technologies has expanded to include transcriptional profiling as a means of understanding mechanisms of disease and toxicity. The degree of success in identifying novel biomarkers is dependent upon the biological question of interest. For instance, two very different goals are: to stratify cancer patient populations based upon specific molecular markers from tumour tissue to predict prognostic outcome; or to classify compounds based on specific mechanisms of action for toxicity. The relative success of molecular profiling for biomarker discovery and qualification for practical use will depend upon the extent to which perturbation of the underlying biological pathways is captured in a transcriptional profile. The following section contrasts two such applications of molecular profiling in biomarker discovery, one for the prediction of long-term survival in cancer patient populations and the second for the prediction of a toxicological response to known toxicants or drugs. The purpose of this comparison is to illustrate that the relative success of generating predictive biomarkers is dependent upon the aetiology of the biological process under evaluation.

4.1 Molecular profiling and cancer

Within the field of cancer research, understanding the aetiology of a tumour is critical for elucidating disease progression and determining the prognosis of a patient. Transcriptional profiling has proven to be a powerful tool in disease staging and classification that surpasses traditional histopathological evaluation [29]. In addition to the long-term goal of understanding the genetic mechanisms underlying cancer, an

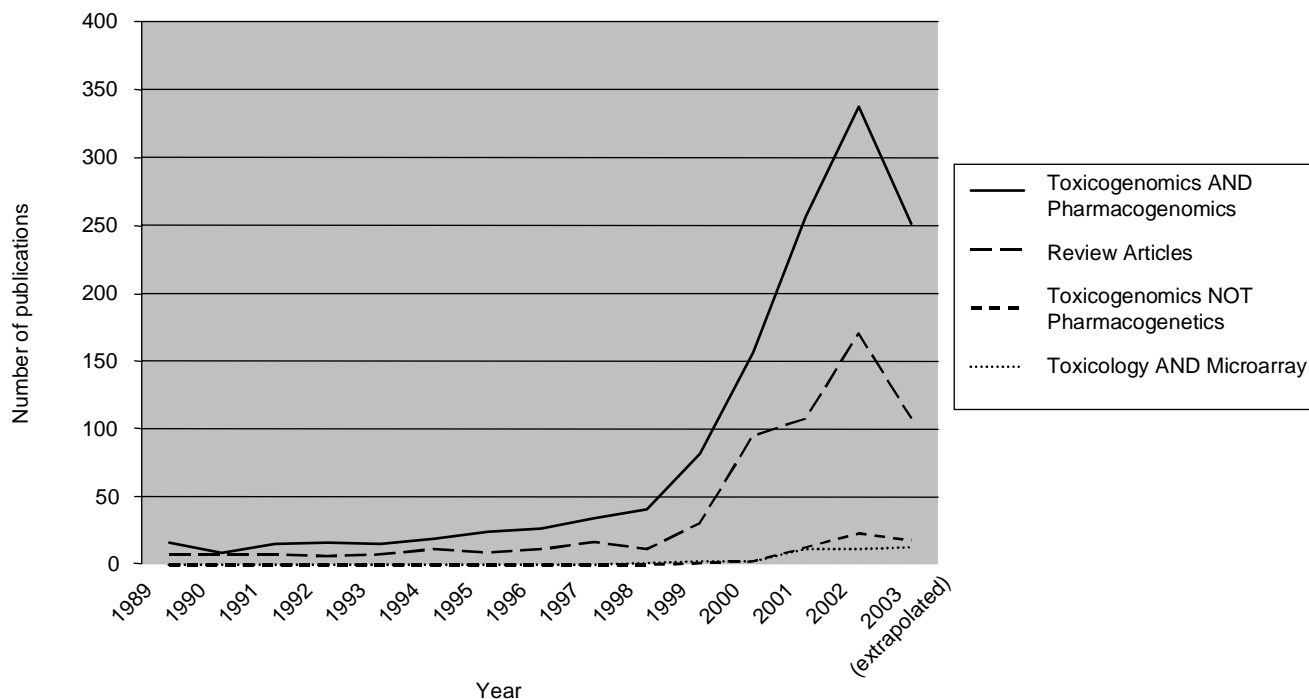


Figure 1. Medline search of toxicogenomic literature. Search conducted on 23 September 2003. Review articles are those associated with Toxicogenomics AND Pharmacogenetics.

immediate goal is to identify novel biomarkers that measure response to a drug treatment in a clinical setting [30]. In so doing, decisions can be made regarding the most appropriate course of treatment.

The success of using expression data to predict long-term patient survival was demonstrated in a non-Hodgkin's lymphoma study [31]. Gene expression patterns were diagnostic for disease originating either from the germinal centres, a B cell-rich region of the lymph node, or from *in vitro* activation of peripheral B cells. Patients with profiles showing a germinal centre origin of cancer had a much higher 10-year survival rate (6 deaths out of 19 patients) than those with activated B cell profiles (16 deaths out of 21 patients). A second example comes from a recent breast cancer study using a 70-gene molecular fingerprint termed 'prognosis-classifier genes' to classify 295 women with either Stage I or II breast carcinomas [32,33]. This signature gene set was more successful at predicting the likelihood of tumour recurrence within 5 years than any clinical measurement. This study illustrates that improved classification of patient populations using gene expression profiling may ultimately enable more appropriate treatments, such as avoidance of unnecessary chemotherapy.

Buckhaults *et al.* [34] demonstrated that the expression patterns from as few as five genes can be used to discriminate between four types of adenocarcinoma (ovarian, breast, colon and pancreatic). In this study, 11 SAGE (serial analysis of

gene expression) libraries representing 21,000 unique transcripts were subjected to supervised and unsupervised statistical analyses to identify a panel of five diagnostic genes. Gene expression-based probability maps correctly segregated the expression profiles correlating to the tumour of origin in 80.6% (50 out of 62) of cases. Further evaluation of these biomarkers must be completed before they are used for diagnosing metastatic carcinomas of unknown origin. Radmacher *et al.* [35] point out that the level of accuracy of a predictive rule based on significantly regulated transcripts for clinical use will likely need to be more stringent than that needed to discriminate between two types of cancer. However, these studies illustrate the potentially effective use of molecular profiling towards practical applications in a clinical setting for cancer diagnosis and prognosis. It remains to be seen as to whether or not transcriptional responses will lead to new cancer targets and an improved granularity in defining the genetic aetiology of the disease.

4.2 Molecular profiling and compound classification

All of the gene expression microarray studies conducted to classify compounds are based on the premise that molecular pathways are affected in a similar way by groups of compounds with shared modes of toxicity (e.g., cholestasis) and that the manifestation of toxicity *in vivo* or *in vitro* is reproducibly measurable by transcriptional profiling. If these assumptions are correct, the field of toxicogenomics will

revolutionise drug development and will have a profound effect on human health. A survey of the scientific literature will provide an indication of how far the field has come towards the application of this approach for screening compounds in early development.

By using hierarchical or nearest neighbour clustering, Thomas *et al.* [36] failed to obtain the expected correlations between gene expression profiles and known toxicological classes (peroxisome proliferators, aryl hydrocarbon receptor [AhR] agonists, non-coplanar polychlorinated biphenyls, inflammatory agents and hypoxia-inducing agents) based on transcriptional response. Screening 12 compounds in mice with a single acute dose and profiling liver RNA using a 1200-gene array, they were, however, able to identify 12 predictive transcripts using Bayesian statistics and a 'leave one out' cross-validation procedure. Using this subset of transcripts, the compounds were correctly classified into the five toxicity categories. Of the 12 diagnostic transcripts, at most 4 distinguished one category from the next, and 6 out of the 12 were cytochrome P450 (CYP450) drug-metabolising enzymes. Induction of these genes after a single dose often reflects a pharmacological response rather than a toxicity response [37]. Furthermore, inclusion of any transcripts beyond the diagnostic set to the analysis split up the treatment groupings. This result indicates a lack of robustness in those groupings and the lack of strength in the data overall to correctly partition these classifiers. Furthermore, it illustrated the importance of reducing the list of responding genes to an informative set or 'fingerprint' in order to accurately classify toxicities. For this purpose, end point data are needed in early profiling studies to guide selection of informative genes via supervised clustering. Multiple time points postdose were collected for three of the compounds. It is of interest to note that the transcriptional level of CYP1A1 and CYP1A2 induction remained constant for nine profiles measured over a 2-week period after a single acute dose of dioxin. The authors further note that the AhR agonist classification does not change over a range of doses of dioxin from as low as 0.05 µg/kg to as high as 100 µg/kg. Determining the biological relevance of these observations seems critical since induction of these two genes alone was responsible for grouping the two AhR agonists. It was not possible to determine the statistical significance of the classifications due to the lack of replicate treatments.

Hamadeh *et al.* [38] tested the hypothesis that structurally unrelated compounds from the same chemical class will produce similar but distinct expression profiles that can be discriminated from profiles produced by compounds from a different chemical class. Three peroxisome proliferators, peroxisome proliferator-activated receptor (PPAR) agonists (Clofibrate, Wyeth, gemfibrozil), were compared to phenobarbital, a CYP450 enzyme-inducer. In this rat study, a single dose of each compound was used, targeting the maximum tolerated dose (MTD). Liver RNA from three animals per dose was profiled for two time points (24 h and 14 days) on an approximately 1700-gene rat microarray. Two-dimensional

hierarchical clustering and principal component analysis performed on the 24-h time point grouped replicate samples together followed by compound class, with the three PPAR agonists grouping together before joining phenobarbital samples. Most gene expression changes occurring 24 h after a single acute dose of a compound primarily reflect pharmacokinetic responses, and if a toxicity signature is present, parsing it from the on-target response may be difficult in the absence of other toxicity end points. The same analysis was not run on the 2-week samples when independent signs of toxicity may have occurred (e.g., elevated alanine aminotransferase [ALT] or aspartate aminotransferase [AST]). The fact that the 24-h samples clustered based on an early liver response to a single dose is reassuring that use of the technology and experimental design are valid, but falls short of providing much insight into the ability to perform toxicological classification of expression profiles.

In an accompanying paper, Hamadeh *et al.* [39] compare blinded samples against a subset of 22 discriminating genes derived from the aforementioned three PPAR agonist and phenobarbital data set by linear discriminant analysis and genetic algorithm/K-nearest neighbours. The blinded samples were from rat liver collected at 24 h, 3 days and 2 weeks. Three compounds, phenytoin, hexobarbital (both in the phenobarbital family) and diethylhexylphthalate (DEHP, a peroxisome proliferator) were classified based on their expression profiles compared to the previously characterised expression profiles using the 22 discriminating gene set. These samples were correctly delineated into either peroxisome proliferators or CYP450 enzyme-inducers by pair-wise correlation analysis. The phenytoin samples could be distinguished as 'phenobarbital-like' or as 'not peroxisome proliferator' based on high- or low-dose, respectively. The one exception, a high-dose animal designated as 'not peroxisome proliferator', was from a 24-h sample. The 3-day DEHP (high-dose only) samples showed the highest correlation to the 24-h Clofibrate and 2-week Wyeth signatures, thus designated as 'Clofibrate/Wyeth-like', whilst the 2-week samples correlated best with the Wyeth signature, therefore labelled 'Wyeth-like'. The 24-h transcriptional response for the PPAR compounds in the training set using the 22 discriminating genes showed a very high induction relative to controls, whilst the induction levels for the 2-week samples were relatively low except for gemfibrozil. The fold induction for the barbiturates was not consistent for either the 24-h or 2-week time point. An illustration of these types of response is shown in Figure 2. This temporal heterogeneity in the magnitude of the response may have contributed to a lack of within-compound class resolution. The lack of within-class resolution was revealed in the two-dimensional hierarchical clustering where the early and late time points were not discriminated among the blinded and known fibrates. Within the phenobarbital node, compounds failed to cluster based on time, dose or structure. It remains to be seen whether or not the power of a negative correlation such as the 'not peroxisome proliferator'

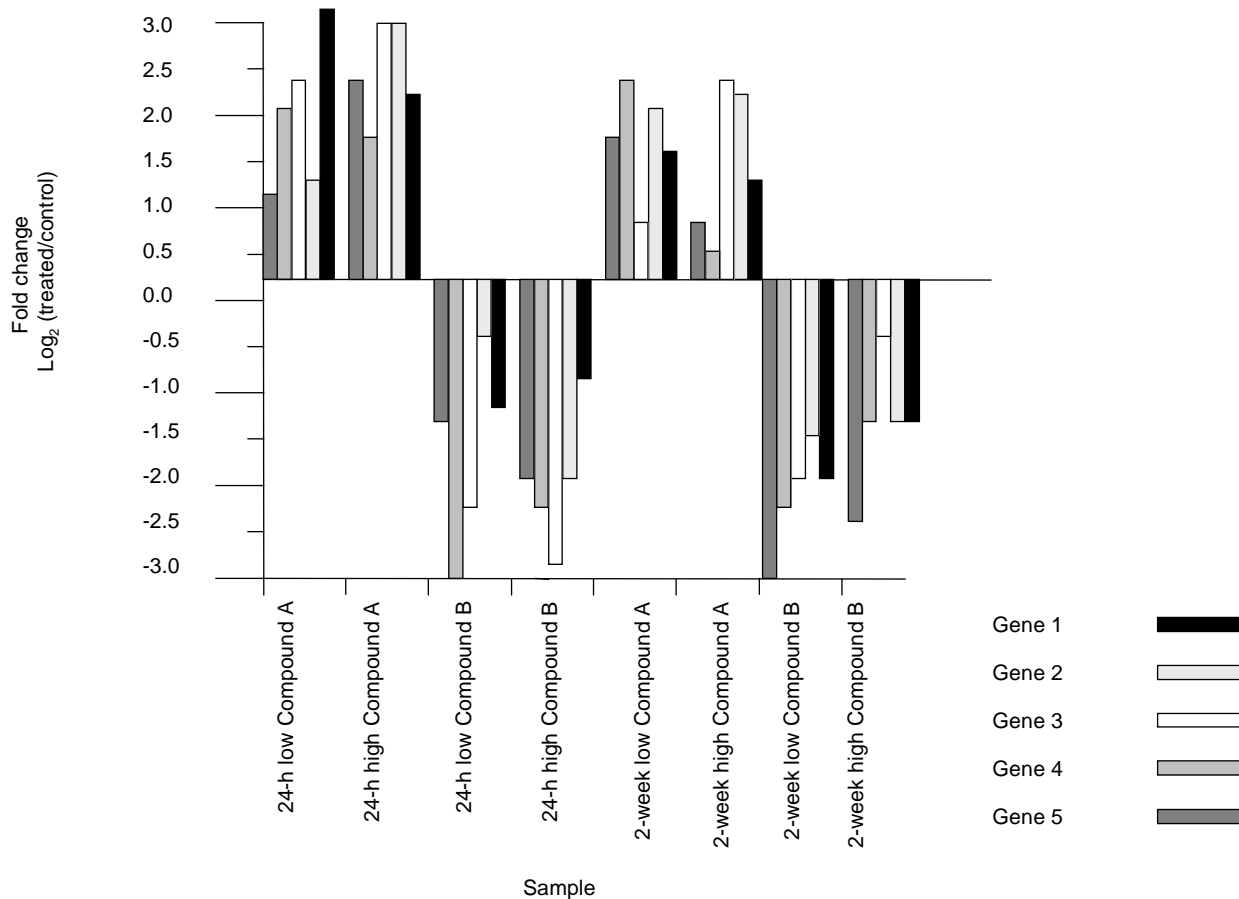


Figure 2. Example of temporal and compound-specific transcriptional response. A set of genes can effectively discriminate between compounds with different modes of action (e.g., CYP450 enzyme inducer versus PPAR agonist). However, this example illustrates how variation in transcriptional response for different doses or times within a compound class can prevent accurate within-class predictions.

CYP450: Cytochrome P450; PPAR: Peroxisome proliferator-activated receptor.

designation is sufficient for screening closely related compounds against a reference database. This study validates the use of a comparative analysis for expression profiles from disparate chemical classes but does not go far enough in providing a predictable designation within a chemical class when time or dose are variables. Again, it is also unclear as to which gene changes are associated with the pharmacological response versus the toxicological response.

Waring *et al.* [40,41] conducted complementary studies using 15 known hepatotoxins in an attempt to classify compounds based upon common mechanisms of action in cultured rat hepatocytes (*in vitro*) and live animals (*in vivo*). In the *in vitro* model, cells were cultured for 24 h with a single exposure level for each compound. In the *in vivo* study, three rats were treated with a single dose per compound for 3 days. The day 3 time point was selected in order to observe early gene changes in response to toxicity before the initiation of secondary responses such as inflammation or fibrosis that could confound biological interpretation. Specific doses were selected based on the expectation of a histopathological outcome by

day 7 but not before. The three liver samples per treatment group were pooled prior to expression profiling, although some individual samples were analysed for comparison. Pooling samples provides an average expression result, eliminating animal-to-animal variability. Unsupervised two-dimensional agglomerative hierarchical clustering was performed on significantly regulated genes showing a +/- 2-fold change in at least one *in vitro* experiment or three *in vivo* experiments. The resulting clusters from each study shared a single clade (sub-group) in common, Aroclor grouping with 3-methylcolanthrene. The *in vivo* study used divisive and K-means clustering as well as agglomerative clustering analysis and the different clustering methods resulted in different compound groupings. An illustration of that effect is shown in Figure 3. Furthermore, inclusion of individual animal samples from two of the compounds with the pooled samples changed the groupings of all but two subclasses representing four compounds (Aroclor/3-methylcolanthrene [3MC], and allyl alcohol/CCl₄). This is an indication of the lack of robustness among the shared gene sets to maintain specific branching

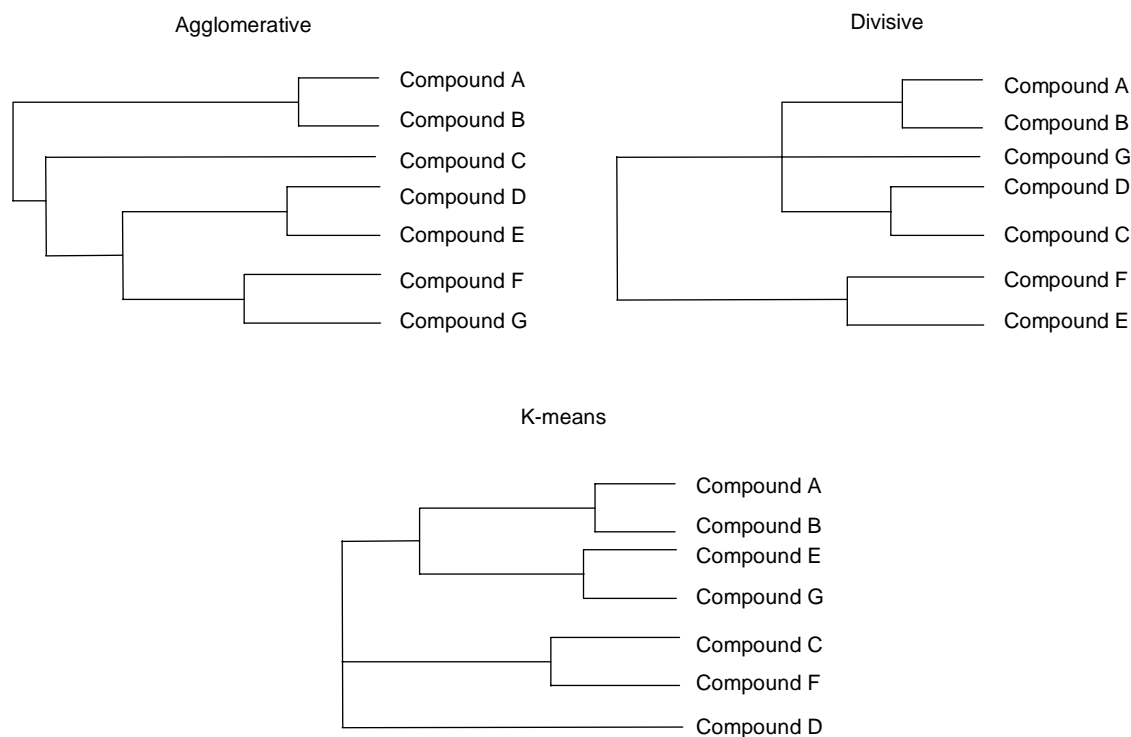


Figure 3. Example of clustering-specific topology changes. Different clustering algorithms can result in different topologies based on the same data. If subclades are not maintained under different clustering methods, the underlying data lack sufficient strength of grouping and little weight can be placed on any single topology.

relationships within the tree. An evaluation of the common gene set for the Aroclor/3MC grouping reveals that many are drug-metabolising enzymes, consistent with what is known about the interaction of these compounds with the aryl hydrocarbon nuclear receptor. Induction of CYP450 genes is an important part of the detoxification response and xenobiotic metabolism mirrors endogenous ligand-specific interactions with nuclear receptors by activating the same drug-metabolising enzymes. For this reason, it is unlikely that regulation of CYP450 genes in response to a xenobiotic or a drug could function as a predictive biomarker for a specific mechanism of toxicity. Comparison of the *in vivo* and *in vitro* results revealed an overall lack of agreement in the compound classifications, with the exception of the Aroclor/3MC subclade. From this work, there is no clear transcriptional biomarker set that could be used for both *in vitro* and *in vivo* studies, with the possible exception of genes commonly regulated between Aroclor/3MC in both models.

Another study aimed at assessing the effect of the nephrotoxin cisplatin on the kidney also used both an *in vitro* and *in vivo* model [42]. Cisplatin toxicity is specific to the kidney because of its active uptake by probenecid-inhibitable organic ion transporters in the proximal tubular epithelial cells [43,44]. The experimental design included transplatin, an inactive

isomer, that is taken up in the same way as cisplatin but is not toxic at equivalent doses. In the *in vivo* study, two dose groups were used with two to five rats per group for each compound. Kidneys and livers were collected after 24 h or 7 days of dosing. Histological evidence of necrosis and apoptosis was found only in the kidneys from the day 7 animals treated with cisplatin. Using a 250-gene microarray, 22 transcriptional changes were observed to be statistically significant at the ≥ 2 -fold change level in the 7-day kidney, compared to 3 from the 24-h samples. Many of these changes agreed with known mechanisms of toxicity from the literature. Among these, regulation of multi-drug-resistant genes and tissue remodelling transcripts may indicate a downstream adaptive response to the toxic insult and tissue regeneration. Few gene changes occurred in the liver with either compound. For the *in vitro* experiments, normal rat renal epithelial and hepatocyte cells were cultured with increasing concentrations of cisplatin over a 24-h period. Statistical analysis of the cultured samples was not performed as only two replicates were used per compound, but the observed gene changes were distinct from those found in the *in vivo* study. Furthermore, a dose response was observed in the liver cells but not in kidney cultures, and the number of gene expression changes > 2 -fold equalled 33 in the liver cells versus 9 in the kidney cells. Cell shrinkage occurred in the liver

cells, whilst the kidney cultures had no apparent morphological change. These data suggest that *in vitro*, cisplatin has a more profound effect on liver than on kidney cells. The authors caution that *in vitro* models may not substitute for or be predictive of toxicity responses *in vivo*.

The strategy of building a comparative database of expression profiles from known toxicants is being attempted at the commercial level. This effort is described in a study from Burczynski *et al.* [45] in work done with Phase I Molecular Toxicology, Inc. A transcriptional expression survey was conducted on 100 compounds in human HepG2 hepatoma cells. Initially, no replicates were run, with each compound being profiled a single time. Comparisons of expression changes over approximately 250 genes were performed using the Pearson's correlation coefficient. A pair-wise distance matrix was produced using single similarity scores representing an entire profile for each compound. Eight pre-designated categories delineated the compounds based on toxic response. No significant correlations were found among compounds known to have similar toxicities or pharmacology. When replicate samples were run using the nephrotoxin cisplatin, only a small percentage of the expression changes were reproducible, however, many of these genes are known to be associated with cisplatin toxicity, including those that respond to DNA-damaging agents. A second set of genes reproducibly associated with NSAIDs was combined with the cisplatin-specific set and used to repeat the Pearson's correlation across all compounds. Further refinement of the method involved running the correlation analysis using genes:

- Common to the DNA-damaging agents.
- Common to the NSAIDs.
- Showing a difference in response between the two.

Twenty genes were identified that selectively discriminate the DNA-damaging agents from NSAIDs. This study demonstrates the importance of replicate samples for the identification of invariant genes associated with a toxicity class. In addition, selecting a subset of relevant genes in a supervised fashion also decreases the likelihood of clustering on background noise.

Each of these studies demonstrates the difficulty of generating a transcriptional profile that is truly predictive rather than diagnostic of an expected outcome. Noise inherent in the system coming from animal-to-animal variability, primary versus secondary transcriptional response and adaptive responses unrelated to toxicity, contribute to the lack of robustness in grouping compounds based on a shared mechanism of toxicity. Furthermore, few toxicants have a specific molecular target and unique mechanism of action that could provide a clean response detectable above background. Identifying predictive biomarkers of toxicity using transcriptional profiling will require supervised approaches with focused analyses on subsets of genes that are well-correlated to a specific toxicity or end point classifier.

4.3 Discerning biological mechanisms

The previous sections highlighted two different applications of transcriptional profiling. The first, predicting the long-term survival of cancer patients based on gene expression fingerprints, and the second, classification of transcriptional responses to predict mechanisms of toxicity using either cultured cells or animal studies. The degree of complexity in interpreting profiling data is dependent upon the biological question one is trying to address. Inherently, the question is whether or not the observed mRNA regulation is a direct cause of a perturbation (e.g., disease, toxicity) or a secondary response to it. It has long been understood that genetic alterations are the underlying causative agents in cancer [46-49]. Therefore, the link to gene regulation is a direct one, either through mutations, chromosomal breakage or by epigenetic factors that alter gene expression, such as the methylation state of regulatory regions. For example, a direct cause-and-effect can be observed between the silencing of a tumour suppressor gene and the absence of its corresponding mRNA in the transcriptome. Moreover, a cascade of downstream events ensues from the loss of this gene, resulting in alterations of normal biological pathways. The sequence of these events appears to be somewhat predictable and accurately reflected in expression profiles from tumour biopsies.

In contrast, toxicants act at multiple molecular and cellular levels and the degree of damage is dependent upon a number of factors including dose, duration of treatment and ADME, and for drugs, pharmacokinetics and pharmacodynamics. This makes prediction of a potential toxicological outcome based on transcriptional readout alone particularly challenging, especially when toxicity is driven by the therapeutic target. In fact, many of the observed transcriptional changes will be either induction of ADME pathways in the short-term or secondary responses to the initial toxicity reflecting generalised cell death or inflammation at later time points (Figure 4). Although an array of compounds are well understood in terms of their toxicities, such as target tissue and mode of action (e.g., mitochondrial poisons), often little is understood about the underlying molecular mechanisms. However, these molecular responses are what are being monitored in microarray experiments.

Currently, it is not known how often transcriptional changes resulting from compound treatment *in vivo* are direct effects or downstream effects of toxicity. Furthermore, it is not clear how confounding these secondary and tertiary effects are on generating an accurate molecular fingerprint of a specific toxicity. For example, induction of cytochrome P450s mediated by the xenobiotic receptor CAR (constitutive active receptor) converts acetaminophen into a reactive quinone form that covalently binds to cellular macromolecules and causes oxidative stress via production of free radicals [50,51]. The inactivation of the reactive quinone intermediate is dependent on the availability of reduced glutathione to which it is conjugated by glutathione S-transferase. In fact, oxidative stress is a major contributory factor in precipitating the

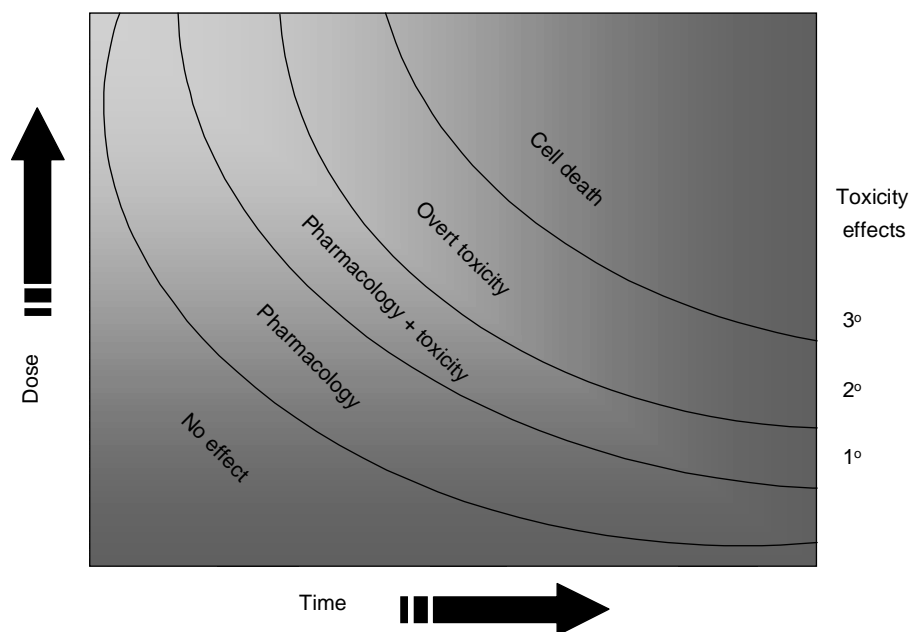


Figure 4. The continuum of effects elicited by a compound relative to dose and time. The ability to parse a direct or 1° effect of a compound in terms of toxicity from 2° or even 3° effects is complicated by the overlap of the cellular response.

depletion of glutathione in response to drug treatment or even in the process of disease [52]. Although the activation of CYP450 genes produces a toxic intermediate, the mechanism by which reactive quinone damages tissue is not understood. Thus, the CYP450 induction in response to acetaminophen is not the direct cause of the toxicity but does play a role in generating tissue-damaging agents. Therefore, induction/repression of drug-metabolising enzymes alone will likely not serve as specific toxicological descriptors with broad application.

The identification of adipsin as a non-invasive biomarker of gastrointestinal (GI) toxicity associated with off-target effects of an amyloid precursor protein γ -secretase inhibitor was recently reported [53]. If the upregulation of this protein in the GI tract is qualified as an indicator of Notch-1 signaling disruption in follow-up studies, it could represent the first example of a novel biomarker discovered and developed for detecting compound-specific toxicity using transcriptional expression technologies.

In order to categorise compounds based on transcriptional response, accurate toxicity classifiers need to be generated. Those based on how a toxicity is manifested (e.g., necrosis, oxidative stress) or on the physiological action of a drug (e.g., peroxisome induction, anti-inflammatory) likely lack the specificity needed to be effective descriptors. Molecular profiles can reveal common mechanistic details that are both

relevant and irrelevant to understanding the toxicity of a compound and keying in on the relevant similarities is the challenge. Moreover, these molecular fingerprints must outperform traditional biochemical assays or chemical biomarkers in order to be accepted as a new standard for measuring toxicity. Monitoring induction of CYP450 enzymes has been a critical component in the drug safety assessment process for decades. The hope is that the entire transcriptional response of a cell to a drug will provide unique insights into mechanisms of toxicity beyond what is already understood in terms of ADME enzymes. An essential way forward in toxicogenomics will be the development and application of predictive mathematical models. Rigorously defined classifiers applied to thoroughly evaluated samples in terms of their inclusion in a training set, will result in better predictive rules. The final section discusses advances in bioinformatics and computational approaches to microarray data analysis and the identification of transcriptional fingerprints as biomarkers.

5. Modelling transcriptional response towards predictive metabolism and toxicity

In order for transcriptional fingerprints to have value as biomarkers for toxicity, the transcriptional responses must fulfil the criteria of being objectively measured and evaluated

as an indicator of toxicity. Since compound-specific toxicity profiles are typically buried in a background of gene changes that are not directly related to the cause of toxicity, it is important to apply appropriate data analysis methods to identify and confirm the specific indicator fingerprints. Microarray data can be analysed in essentially two ways, using either unsupervised or supervised methods. An unsupervised approach makes no *a priori* assumptions nor requires expert knowledge, whilst a supervised approach places predetermined restrictions on the analysis with knowledge of between-class differences (e.g., PPAR agonists versus PPAR antagonists). Supervised methods are generally employed for class comparisons or class predictions. An unsupervised method is more commonly used for class discovery when it is not clear whether or not within-class variation reflects distinct biological entities (e.g., disparate aetiology of disease within a subpopulation of breast tumours). The goal of either method is to define those genes that are significantly differentially expressed and then look for correlations among those changes that most accurately characterise class membership. An overview of the various methods routinely applied to microarray data analysis is presented by Butte [54].

In analysing microarray data, a common error is the use of unsupervised cluster analysis and simple fold change statistics for problems of class comparison or prediction. Whilst this can yield useful information, it is not necessarily the most valuable approach for discriminating transcriptional biomarkers. There are numerous clustering algorithms and parameters to choose from in running an unsupervised analysis. The approach that will accurately represent the data structure depends upon many variables including experimental design. Simon *et al.* [55] point out that significance cutoffs for genes are determined by including either all genes with relatively high intensity or all genes that have meaningful variation across samples. Only a small subset of these genes are likely to be relevant for class discrimination, yet the distances used in the clustering take into account all gene changes resulting in poor class resolution. That is, average fold change indices fail to discriminate between random and significant variation among samples within the same group. A 2-fold change may be significant in one context but not in another. Strictly relying on fold change indices or visual inspection of a cluster display is not statistically rigorous. Simon *et al.* stress the importance of cross-validation in order to obtain an unbiased estimate of the true error rate of the predictor. Through simulation analysis, the authors demonstrate how leaving out cross-validation for any step in the class prediction process (selection of informative genes, weighting informative genes and creation of a prediction rule) can result in gross miscalculation of the probability of correctly classifying a randomly selected future sample. Furthermore, when the candidate predictors greatly outnumber the cases, even complex methods do not perform well [56]. Thus, the problem with clustering experiments and/or genes based on all significantly regulated transcripts is that many of those gene changes,

although statistically significant, may not be relevant to an observed end point.

Ding [57] developed an unsupervised feature selection algorithm that is based on gene similarities alone. By ordering expression data and experimental samples based on an optimal similarity criterion, adjacent genes/samples are most similar and these non-discriminant genes (i.e., genes common to both clusters) are identified and discarded from the clustering analysis. This approach can be applied in a pairwise fashion to select genes that best discriminate between subclasses of experiments. The authors compared a subset of 50 genes identified using their method with a set of 50 genes selected using a supervised t-test based on information known about colon cancer cells versus normal cells. An overlap of 40 genes between these sets indicates that the unsupervised method captured comparable knowledge about the class structure as it did about human expertise. This unsupervised method could be most valuable when applied to data sets where a clear outcome is not evident and expert knowledge is limited or absent.

A second method based on Rissanen's MDL (minimum description length) principle simultaneously performs gene clusters and subsets a selection of gene clusters for sample classification [58]. This algorithm generates 'active' clusters that are class predictive and 'inactive' clusters that possess cross-sample variation but do not necessarily function as good class predictors. Applying this method to a leukaemia profiling data set and to the NCI60 data, the simultaneous method generated active gene clusters that were 'cleaner' than a method that clusters and selects subsets of genes separately. The simultaneous method exhibited a more homogeneous regulation pattern within the genes from the active clusters. The centroids (the centre of the multidimensional means of individual transcripts) of the clusters and genes closest to the centroids were also shown to be well-correlated when comparing genes from the full data set to those in cross-validation sets. The authors demonstrate that model selection instability is decreased by selecting gene clusters rather than individual genes.

A different unsupervised clustering method simultaneously clusters genes and conditions revealing 'checkerboard' patterns of expression data [59]. Five different cancer microarray data sets were analysed using this procedure and results were compared to several benchmarks to assess the effectiveness of identifying pairs of piece-wise constant eigenvectors (a special set of vectors associated with a linear system of equations) of genes and conditions. That is, applying a gene expression matrix to a step-like condition classification vector will result in a step-like gene classification vector. Then it is a matter of determining whether or not the structure of the expression matrix reveals clusters of co-regulated genes and experimental conditions associated with the effects of those genes. Overall, the method performed well, effectively portioning tumour types and patient populations and sometimes discriminating subpatient populations within a given disease class. The

authors acknowledge that removal of irrelevant genes prior to clustering may improve the discriminatory capabilities of their method.

Another approach to class prediction capitalises on the correlation between transcriptional responses to identify subsets of genes using dimension-reduction techniques [60]. These methods, termed sliced average variance estimation (SAVE) and sliced inverse regression (SIR), differ from other algorithms in that they do not require the specification of a model in order to estimate linear combinations of transcripts containing all regression information. Multiple subsets of genes with similar accuracy for class prediction were identified when applied to the problem of discriminating BRCA1 and BRCA2 breast cancer expression profiles.

The few examples highlighted here are, like most of the algorithms, developed for microarray analysis in that they focus on cancer data to validate the methods [61-63]. One computational methods study in breast cancer cell lines stimulated with oestrogen did look at time- and dose-response, taking into account order-restricted inference, and illustrated the effectiveness of the algorithm when treatments are ordered by design [64]. Rarely are these methods compared to other standardly used clustering methods or supervised methods. Furthermore, it remains to be seen how effective these approaches will be when applied to data reflecting compound-induced toxicity.

A concerted effort should be undertaken to co-develop appropriate predictive models of toxicity and coordinate experimental design based on what is learned about existing data structure. Generating a reference database of expression profiles reflecting specific toxicity classes requires within-class and between-class coverage of chemical and physiological space. Generating a predictive model that accurately and reproducibly classifies novel compounds into known toxicity groups based on such a reference database will require appropriately designed experiments as training sets as well as sufficiently large validation sets to provide statistically rigorous confidence intervals. Otherwise, rigorous cross-validation must be performed for each training set at every step, including the selection of informative genes, to avoid overfitting of the model.

6. Conclusions

In the overall process of new drug development, it is clear that the target validation and preclinical safety assessment phases hold the strongest short-term promise for critical impact of genomics; this includes the identification and application of new biomarkers. Enabling drug development by triaging compounds early based on molecular fingerprints of both efficacy and toxicity could result in a higher success rate for drugs making it through to clinical trials. In terms of building a toxicity reference database, it remains to be seen whether one experimental paradigm (e.g., two doses, two time points, three replicates) for all compounds will be sufficient to

capture mechanistic transcriptional changes linked to specific toxic effects, or whether time- and dose-response relationships will have to be established individually for each compound. Furthermore, identifying a toxicity signature from a common metabolic signature among within-class members without fully understanding mechanisms will be a challenge. Toxicogenomic studies conducted to date have set the groundwork for future advances to be made, in both experimental design and improving computational methods for microarray data analysis. Strategies undertaken in this field will serve to validate the use of molecular profiling technologies towards the development of preclinical and clinical biomarkers as well as surrogate end points.

7. Expert opinion

This review has focused on the potential to identify novel biomarkers derived from genomic information as well as emerging high-throughput technologies. Being able to survey global cellular responses to perturbation using molecular profiling allows scientists to observe complex regulatory interactions in ways never before possible. The strategic approaches currently pursued in the biomedical research community hold the promise of identifying, developing and delivering more efficacious, safer drugs to patient populations at an unprecedented rate. Identifying novel biomarkers during critical drug discovery and development stages is an integral activity necessary for driving and tracking this process. Biomarkers can be effectively used at multiple stages of the drug development process and need not necessarily overlap in function. For example, during the discovery and validation phase, a molecular fingerprint could reliably detect a novel compound's effect on a target tissue or specific mechanistic pathway and provide additional information about off-target effects. The use of microarray data for monitoring adverse effects and the use of a toxicogenomics compendium at the interface of the discovery and preclinical phases could further aid in identifying potential drug development liabilities. These early stages do not require that the biomarker be easily accessible as any tissue can be surveyed from model organisms. This is not true for clinical biomarkers. However, insights into mechanisms of efficacy or toxicity gleaned from early stage molecular studies could provide information needed to identify a novel protein in blood or a metabolite from urine to monitor patient populations.

There are a number of challenges yet to overcome in this new era before the genomics promise is realised. Problems of standardisation of platforms and protocols are currently being addressed [65]. Interpretation of observed transcriptional changes is hampered by the lack of functional annotation for many genes and the inadequate understanding of whether the transcriptional changes reflect adaptation due to feedback inhibition/activation or are directly involved in the mechanism of interest. The use of molecular fingerprints for detecting toxicity has unique issues, in that predictive class rules must be accurate and capable of detecting

adverse effects before traditional methods, such as changes in clinical chemistry and histopathology, if they are to become accepted as biomarkers. Predictive models of toxicity must take into account differences between doses, time points, species, gender, animals and classes of compounds. Failure to adequately capture these variables could result in

misclassification of toxicity associated with compounds, which in turn, could translate into a significant loss of time and money. Conversely, if accurate predictive models of toxicity and the biomarkers necessary to track adverse effects can be generated, at least one promise of this new genomic age will be realised.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

- PETRICOIN EF III, HACKETT JL, LESKO LJ *et al.*: Medical applications of microarray technologies: a regulatory science perspective. *Nat. Genet. Suppl.* (2002) **32**:474-479.
- IDEKER T, GALITSKI T, HOOD L: A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* (2001) **2**:343-372.
- BIOMARKERS DEFINITIONS WORKING GROUP: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* (2001) **69**:89-95.
- THE FOOD AND DRUG MODERNIZATION ACT OF 1997: Title 21 Code of Federal Regulations. Part 314, Subpart H, Section 314.500.
- COLBURN WA: Biomarkers in drug discovery and development: from target identification through drug marketing. *J. Clin. Pharmacol.* (2003) **43**(4):329-341.
- SCHADT EE, MONKS SA, FRIEND SH: A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. *Biochem. Soc. Trans* (2003) **31**(2):437-443.
- FRANK R, HARGREAVES R: Clinical biomarkers in drug discovery and development. *Nat. Rev. Drug Discov.* (2003) **2**(7):566-580.
- An excellent review of current advances in the clinical biomarker field.
- MORRISH PK, RAKSHI JS, BAILEY DL, SAWLE GV, BROOKS DJ: Measuring the rate of progression and estimating the preclinical period of Parkinson's disease with [18F]dopa PET. *J. Neurol. Neurosurg. Psychiatry* (1998) **64**(3):314-319.
- DU AT, SCHUFF N, ZHU XP *et al.*: Atrophy rates of entorhinal cortex in AD and normal aging. *Neurology* (2003) **60**(3):481-486.
- WAGNER JA: Overview of biomarkers and surrogate endpoints in drug development. *Dis. Markers* (2002) **18**(2):41-46.
- WARING JF, DAI X, HE Y, LUM P, ROBERTS CJ, ULRICH R: 'If you build it, they will come': design, construction and use of a compendium of expression profiles for mechanism of toxicity prediction. In: *Toxicogenomics: Principles and Applications* Hamadeh HK, Afshari CA (Eds), Wiley Publishers, NY. In Press.
- WARING JF, ULRICH RG: The impact of genomics-based technologies on drug safety evaluation. *Annu. Rev. Pharmacol. Toxicol.* (2000) **40**:335-352.
- CASTLE AL, CARVER MP, MENDRICK DL: Toxicogenomics: a new revolution in drug safety. *Drug Discov. Today* (2002) **7**(13):728-736.
- ULRICH R, FRIEND SH: Toxicogenomics and drug discovery: will new technologies help us produce better drugs? *Nat. Rev. Drug Discov.* (2002) **1**(1):84-88.
- GERHOLD DL, JENSEN RV, GULLANS SR: Better therapeutics through microarrays. *Nat. Genet. Suppl.* (2002) **32**:547-552.
- TUGWOOD JD, HOLLINS LE, COCKERILL MJ: Genomics and the search for novel biomarkers in toxicology. *Biomarkers* (2003) **8**:79-92.
- SISTARE FD: Development and application of interspecies biomarkers in nonclinical safety evaluations. In: *Biomarkers in Clinical Drug Development*. Bloom JC, Dean RA (Eds), Marcel Dekker, Inc., NY (2003):97-117.
- MACGREGOR JT: The future of regulatory toxicology: impact of the biotechnology revolution. *Toxicol. Sci.* (2003) **75**:236-248.
- MORRIS AJ: Information and observer disagreement in histopathology. *Histopathology* (1994) **25**:123-128.
- BOORMAN GA, HASEMAN JK, WATERS MD, HARDISTY JF, SILLS RC: Quality review procedures necessary for rodent pathology databases and toxicogenomic studies: the National Toxicology Program Experience. *Toxicol. Pathol.* (2002) **30**:88-92.
- FURNESS PN: The use of digital images in pathology. *J. Pathol.* (1997) **183**:253-263.
- KRIETE A, ANDERSON MK, LOVE B *et al.*: Combined histomorphometric and gene-expression profiling applied to toxicology. *Genome Biol.* (2003) **4**(5):R32.
- PANDURO A, SHALABY F, WEINER FR, BIEMPICA L, ZERN MA, SHAFRITZ DA: Transcriptional switch from albumin to α -fetoprotein and changes in transcription of other genes during carbon tetrachloride induced liver regeneration. *Biochemistry* (1986) **25**(6):1414-1420.
- CLARKE H, EGAN DA, HEFFERNAN M *et al.*: α -glutathione s-transferase (α -GST) release, an early indicator of carbon tetrachloride hepatotoxicity in the rat. *Hum. Exp. Toxicol.* (1997) **16**(3):154-157.
- ROLAN P, ATKINSON AJ, LESKO LJ: Use of biomarkers from drug discovery through clinical practice: report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development. *Clin. Pharmacol. Ther.* (2003) **73**:284-291.
- NUWAYSIR EF, BITTNER M, TRENT J, BARRETT JC, AFSHARI CA: Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinog.* (1999) **24**(3):153-159.
- OSCARSON M: Pharmacogenetics of drug metabolizing enzymes: importance for personalized medicine. *Clin. Chem. Lab. Med.* (2003) **41**(4):573-580.
- ROGERS JF, NAFZIGER AN, BERTINO JS Jr: Pharmacogenetics affects dosing, efficacy, and toxicity of cytochrome P450-metabolized drugs. *Am. J. Med.* (2002) **113**(9):746-750.

Molecular profiling approaches for identifying novel biomarkers

29. LIOTTA L, PETRICOIN E: Molecular profiling of human cancer. *Nat. Rev. Genet.* (2000) **1**:48-56.
30. DEPRIMO SE, WONG LM, KHATRY DB: Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. *BMC Cancer* (2003) **3**(1):3.
31. ALIZADEH AA, EISEN MB, DAVIS RE *et al.*: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* (2000) **403**:503-511.
32. VAN T VEER LJ, DAI H, VAN DE VIJVER MJ *et al.*: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* (2002) **415**(6871):530-536.
33. VAN DE VIJVER MJ, HE YD, VAN T VEER LJ *et al.*: A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* (2002) **347**:1999-2009.
34. BUCKHAULTS P, ZHANG Z, CHEN YC *et al.*: Identifying tumor origin using a gene expression-based classification map. *Cancer Res.* (2003) **63**:4144-4149.
35. RADMACHER MD, MCSHANE LM, SIMON R: A paradigm for class prediction using gene expression profiles. *J. Comput. Biol.* (2002) **9**(3):505-511.
36. THOMAS RS, RANK DR, PENN SG *et al.*: Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol. Pharmacol.* (2001) **60**(6):1189-1194.
37. DENISON MS, WHITLOCK JP Jr: Xenobiotic-inducible transcription of cytochrome P450 genes. *J. Biol. Chem.* (1995) **270**(31):18175-18178.
38. HAMADEH HK, BUSHEL PR, JAYADEV S *et al.*: Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.* (2002) **67**(2):219-231.
39. HAMADEH HK, BUSHEL PR, JAYADEV S *et al.*: Prediction of compound signature using high density gene expression profiling. *Toxicol. Sci.* (2002) **67**(2):232-240.
40. WARING JF, CIURLIONIS R, JOLLY RA, HEINDEL M, ULRICH RG: Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol. Lett.* (2001) **120**(1-3):359-368.
41. WARING JF, JOLLY RA, CIURLIONIS R *et al.*: Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharmacol.* (2001) **175**(1):28-42.
42. HUANG Q, DUNN RT II, JAYADEV S *et al.*: Assessment of cisplatin-induced nephrotoxicity by microarray technology. *Toxicol. Sci.* (2001) **63**(2):196-207.
43. BAN M, HETTICH D, HUGUET N: Nephrotoxicity mechanism of *cis*-platinum (II) diamine dichloride in mice. *Toxicol. Lett.* (1994) **71**(2):161-168.
44. SAFIRSTEIN R, MILLER P, GUTTENPLAN JB: Uptake and metabolism of cisplatin by rat kidney. *Kidney Int.* (1984) **25**(5):753-758.
45. BURCZYNSKI ME, MCMILLIAN M, CIERVO J *et al.*: Toxicogenomics-based discrimination of toxic mechanism in HepG2 human hepatoma cells. *Toxicol. Sci.* (2000) **58**(2):399-415.
46. ABRAHAM SC, WU TT, KLIMSTRA DS *et al.*: Distinctive molecular genetic alterations in sporadic and familial adenomatous polyposis-associated pancreaticoblastomas: frequent alterations in the APC/ β -catenin pathway and chromosome 11p. *Am. J. Pathol.* (2001) **159**(5):1619-1627.
47. DESMAZE C, SORIA JC, FREULET-MARRIERE MA, MATHIEU N, SABATIER L: Telomere-driven genomic instability in cancer cells. *Cancer Lett.* (2003) **194**(2):173-182.
48. GILES RH, VAN ES JH, CLEVERS H: Caught up in a Wnt storm: Wnt signaling in cancer. *Biochim. Biophys. Acta* (2003) **1653**(1):1-24.
49. BIGNOLD LP: Initiation of genetic instability and tumour formation: a review and hypothesis of a nongenotoxic mechanism. *Cell. Mol. Life Sci.* (2003) **60**(6):1107-1117.
50. ROGERS LK, MOORTHY B, SMITH CV: Acetaminophen binds to mouse hepatic and renal DNA at human therapeutic doses. *Chem. Res. Toxicol.* (1997) **10**(4):470-476.
51. JAESCHKE H, KNIGHT TR, BAJT ML: The role of oxidant stress and reactive nitrogen species in acetaminophen hepatotoxicity. *Toxicol. Lett.* (2003) **144**(3):279-288.
52. MYTILINEOU C, KRAMER BC, YABUT JA: Glutathione depletion and oxidative stress. *Parkinsonism Relat. Disord.* (2002) **8**(6):385-387.
53. SEARFOSS GH, JORDAN WH, CALLIGARO DO *et al.*: Adipsin: a biomarker of gastrointestinal toxicity mediated by a functional γ -secretase inhibitor. *J. Biol. Chem.* (2003) **278**:46107-46116.
54. BUTTE A: The use and analysis of microarray data. *Nat. Rev. Drug Discov.* (2002) **1**(12):951-960.
- **This review provides a clear overview of microarray technology and data analysis options.**
55. SIMON R, RADMACHER MD, DOBBIN K, MCSHANE LM: Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J. Natl. Cancer Inst.* (2003) **95**(1):14-18.
- **This paper provides a nice summary of the complicating factors in interpreting microarray data.**
56. DUDOIT S, FRIDLAND J, SPEED TP: Comparison of discrimination methods for the classification of tumours using gene expression data. *J. Am. Stat. Assoc.* (2002) **97**:77-87.
- **This study details differences among computational methods used to analyse microarray data.**
57. DING CH: Unsupervised feature selection via two-way ordering in gene expression analysis. *Bioinformatics* (2003) **19**(10):1259-1266.
58. JORNSTEN R, YU B: Simultaneous gene clustering and subset selection for sample classification via MDL. *Bioinformatics* (2003) **19**(9):1100-1109.
59. KLUGER Y, BASRI R, CHANG JT, GERSTEIN M: Spectral biclustering of microarray data: coclustering genes and conditions. *Genome Res.* (2003) **13**(4):703-716.
60. BURA E, PFEIFFER RM: Graphical methods for class prediction using dimension reduction techniques on DNA microarray data. *Bioinformatics* (2003) **19**(10):1252-1258.
61. GOLUB TR, SLONIM DK, TAMAYO P *et al.*: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* (1999) **286**(5439):531-537.
62. ALBRECHT A, VINTERBO SA, OHNO-MACHADO L: An Epicurean learning approach to gene-expression data classification. *Artif. Intell. Med.* (2003) **28**(1):75-87.

63. LI W, YANG Y: Zipf's law in importance of genes for cancer classification using microarray data. *J. Theor. Biol.* (2002) **219**(4):539-551.
64. PEDDADA SD, LOBENHOFER EK, LI L, AFSHARI CA, WEINBERG CR, UMBACH DM: Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference. *Bioinformatics* (2003) **19**(7):834-841.
65. BRAZMA A, HINGAMP P, QUACKENBUSH J *et al.*: Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat. Genet.* (2001) **29**(4):365-371.

Affiliation

Wendy J Bailey PhD¹ & Roger Ulrich PhD²

†Author for correspondence

¹Merck & Co., Inc., 770 Sumneytown Pike,
West Point, PA 19486, USA

Tel: +1 215 652 9021; Fax: +1 215 993 1835;

E-mail: wendy_bailey@merck.com

²Rosetta Inpharmatics LLC – Merck Research
Laboratories, Kirkland, WA 98034, USA

Tel: +1 425 636 6405; Fax: +1 425 821 5354;

E-mail: roger_ulrich@merck.com