Genomics, Transcript Profiling, Proteomics and Metabonomics (GTPM) An Introduction

Document No. 42

Brussels, April 2001

ECETOC Document No. 42

European Centre for Ecotoxicology and Toxicology of Chemicals
4 Avenue E. Van Nieuwenhuyse (Bte 6), 1160 - Brussels, Belgium.

Genomics, Transcript Profiling,
Proteomics and Metabonomics (GTPM) - An Introduction

CONTENTS

ABSTRACT

Genomics, transcript profiling (transcriptomics), proteomics and metabonomics (GTMP) are rapidly developing technologies that enable researchers to study and describe biological events at the level of genetic material (genomics) and its expression in organisms. Expression can be studied at the stage of transfer of genetic information (transcriptomics), at the stage of formation of proteins (proteomics), and by determining the metabolites resulting from the activities of those proteins (metabonomics). Most of the techniques themselves are not new, but can now be applied in such a way that massive amounts of data can be generated, characterising changes in the presence and amounts of potentially thousands of biomolecules simultaneously.

GTPM offers exciting possibilities for research in biology, pharmacology and toxicology and various economically interesting applications in industry are already under development or even actively employed.

There is, on the other hand, the real danger that indiscriminate application of these technologies will lead to the generation of misleading data. Furthermore, the current (relative) lack of reference data could easily lead to mis- or over-interpretation and subsequently to undue concern by regulatory agencies. Therefore, there is an urgent need for the chemical industry as a whole to collaborate with academia and regulators in the development and sharing of such reference data sets. This will help to ensure that the new technologies are wisely applied and that agreement is reached on the appropriate interpretation of the data that they generate.

EXECUTIVE SUMMARY

For decades, the potential for chemicals to affect adversely the health of exposed human beings, has been evaluated using controlled laboratory experiments in animals and *in vitro* systems. These established studies are designed to characterise the adverse effects (behavioural, pathological or biochemical) that result from exposure to the chemical substance. Data from such investigations form the foundation of safety evaluation and risk assessment procedures applied by regulators to ensure the protection of the health of exposed populations.

For the most part, this practice of the science of toxicology has worked well even though there have been many examples where the species of animal tested was subsequently found not to be a good surrogate for humans. Such exceptions have led to the development of more and more sophisticated and specialised techniques to study the nature of the adverse changes and the mechanisms by which chemicals exert their effects. Detailed analysis of the mode of action, through studies of, for example, comparative metabolism, kinetics, pathology, cellular biology, using both *in vivo* and *in vitro* systems, improves the reliability of the predictions of the likely consequences of human exposure to the chemicals under study. However, application of these specialised investigations is laborious and costly, often investigating the potential role of one molecule or pathway at a time, and thus may only be undertaken for relatively few 'industrial' chemicals.

The successes of large-scale genome sequencing programmes such as the Human Genome Project have stimulated the development of new technologies that facilitate the simultaneous measurement of thousands of biological variables in test material. These technologies, termed genomics, transcript profiling (transcriptomics), proteomics and metabonomics (collectively GTPM or 'omics') and their enlightened application are the subjects of this paper and are outlined in more technical detail in Appendix 1. In addition to their applications to descriptive and comparative biology and pharmacology, they allow the effects of xenobiotics on various processes within cells (or tissues) to be characterised, and are used in pure and applied research both as mechanistic tools and for facilitating the early identification and characterisation of possible hazard. In short, these technologies can permit the visualisation, with an unprecedented level of detail, of change in cells and tissues at the molecular level. It should be borne in mind, however, that they facilitate hypothesis building around the potential role of individual genes, proteins or metabolites in biological processes, rather than establish a causative role for these changes in exerting an effect (either harmful or beneficial). For the present there is still a great deal to learn but in time, these tools will bring exciting capabilities in the characterisation of biological mechanisms, and, when used appropriately in conjunction with the more traditional approaches, are destined to revolutionise exploratory biology and, thereby, toxicology.

These emerging technologies that profile changes in such biochemical processes share a number of common features, most pre-eminently being that their usage is, at present, relatively specialised. For the most part they are also expensive to set up and implement and therefore, with some exceptions, are only used routinely in large pharmaceutical

and agrochemical companies, rather than in the academic world. However, a niche market has already arisen for commercial enterprises as a consequence of the enormous interest generated in the scientific community over the sensational volumes of data generated by these approaches. As of now, technology can be supplied to monitor gene expression changes in a limited format and scale that does not require investment in large amounts of equipment or infrastructure. With little financial outlay, even small academic laboratories can thus create and publish preliminary findings obtained with these technologies, possibly without validating that the observed changes are in any way directly involved in the processes under study.

There are many benefits to be gained by the enlightened application of these new technologies that can deliver unprecedented opportunities to characterise the mechanisms by which chemicals interact with cellular processes that are the essence of life. There is potential social as well as financial gain, stemming from advantages such as the ability to make more informed choices in the development of new compounds, the ability to prioritise justifiably needs for conventional (costly) testing and the opportunities for refinement in the usage of animals. For example, the discovery, characterisation and application of new predictive markers of toxicity could have highly significant implications for saving in costs (and time) of product development. Using these results to conduct additional mechanistic studies at an early stage of development could avoid the need to commission uninformative long-term toxicity studies.

Conceivably, once these techniques have helped to reveal a specific mode of action, they could be employed to elucidate the dose-time relationships that constitute the threshold for toxicity. This would facilitate dose selection for conventional studies or, at a later stage, abolish the need for some forms of additional toxicity testing. This again may help to reduce the time taken to get a new product to the market and could potentially help in the refinement and reduction of animal usage.

The appropriate application of these techniques is, however, more demanding of careful experimental design than ever, as the potential to generate incomplete and misleading data is great. A challenge for all scientists, inside and outside industry, engaged in research employing these techniques, is the avoidance of over-interpretation of results that are generated in the absence of good experimental design and hypothesis testing. This could result in misinterpretations and erroneous extrapolation to other species, including man. In this regard it is important for there to be developed a common understanding amongst scientists, whether they are from academia, industry or regulatory organisations, regarding the interpretation and application of data generated by these approaches.

The attainment of common ground through collaboration involving the generation, sharing and publication of suitable, high-quality, pilot data should be the prime goal for scientists and institutions engaged in researching the new technology and its appropriate application towards improving the knowledge of the interaction of chemicals with living things. It is in the interests of all stakeholders, including the industry, to seek

opportunities for investing resources in those initiatives that promote the development of a sound scientific database. Such a database would enable the technologies to be evaluated from an informed perspective, assisting those engaged in assessing chemical safety, including regulatory agencies, to place these data into context with other, more conventional data. It will also help to counterbalance the impact of inadequately interpreted data published by researchers from laboratories with insufficient resources to do justice to these technologies.

1. INTRODUCTION

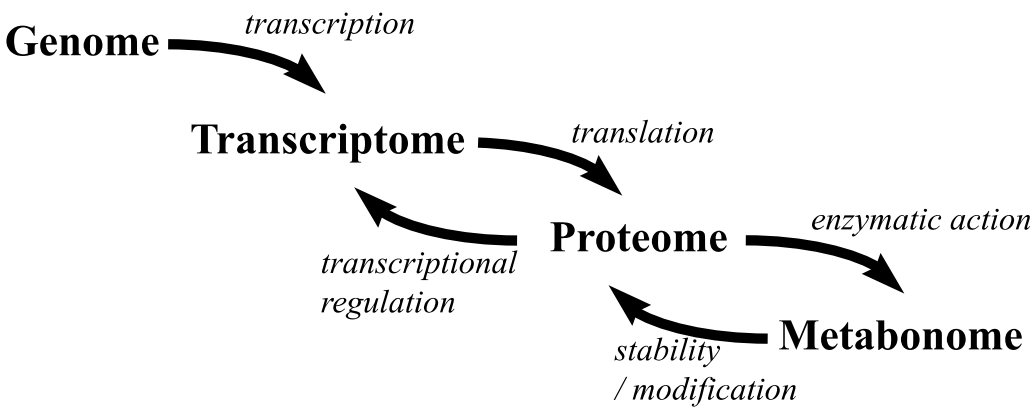
The last decade has seen major developments in large-scale genome sequencing and in the development of technical platforms to support this enterprise. Several genomes, such as that of yeast (DeRisi *et al*, 1997), have now been completely resolved at the nucleotide level, and it is anticipated that high resolution sequencing of the entire human genome could be completed early this decade. The availability of the sequence information for many thousands of genes has facilitated the development of a number of technologies which profile changes in genes (and gene products) at a scale that will greatly facilitate the discovery of how various biological processes are regulated. These techniques are generally termed genomics, transcript profiling (transcriptomics), proteomics and metabonomics (GTPM) and are discussed briefly in this paper. An overview of these technologies is given in Appendix 1. It should be noted that many of the examples cited in this discussion refer to transcript profiling, which is currently the most widely used of these technologies and thus the best represented in the literature.

Xenobiotic exposure has the potential to cause alterations at different organisational levels of a cell or tissue:

- Genome: the chromosomal DNA information.
- Transcriptome: the messenger RNA from actively transcribed genes.
- Proteome: the entire protein complement of a biological sample.
- Metabonome: the constituent metabolites in a biological sample.

Potentially, xenobiotics can have a primary mode of action that affects any of these compartments although it is unlikely that a phenotypic change, with the possible exception of necrosis, can take place without measurable alterations of all compartments downstream of the genome. Some of processes by which changes in one compartment can influence other compartments are outlined simplistically in Figure 1.

Figure 1: Example Processes by which Information flows through Different Categories of Molecules in a Cell



Potentially, these technologies permit simultaneous and quantitative measurement of thousands of biomarkers in biological samples (Schena *et al*, 1995). Many benefits can be envisaged by their appropriate use in investigative biology (Khan *et al*, 1999), with much emphasis placed on their potential to revolutionise drug discovery (Marton *et al*, 1998).

Their careful application could also assist research associated with safety assessment by:

- enhancing ability to extrapolate accurately between experimental animals and humans in the context of risk assessment;
- enabling a more detailed appreciation of molecular mechanisms of toxicity;
- facilitating more-rapid screens for substance toxicity;
- allowing compound selection decisions to be based on safety, rather than dominated by efficacy;
- providing new research leads.

Naturally, these technologies also have the capacity to generate misleading data if not employed appropriately. The careful interpretation of results from such data-rich technologies represents a major challenge to investigators employing them. In particular, over-interpretation of such results, without the discipline of good experimental design and hypothesis testing, could result in raising unfounded concern over the potential of a chemical to cause adverse health effects in humans or other species. As these technologies become more widely accessible, the importance of establishing and communicating their strengths, weaknesses and appropriate usage becomes more urgent.

The most effective way to maximise the development and effective application of the new technologies is to encourage the exchange of data and collaborative initiatives between researchers and institutes engaged in pioneering the techniques.

2. APPLICATIONS TO MECHANISM-BASED TOXICOLOGY RESEARCH AND BIOMARKER IDENTIFICATION

An obvious application of these technologies is the investigation of the regulation events that underpin the response of a cell to toxic insult. Where possible, an appropriate model (such as an *in vitro* cultured cell line) can be designed to recapitulate the toxic endpoint. An example of such an endpoint, non-genotoxic carcinogenesis, is usually assayed by using a long-term cancer bioassay in rodents (Chabra *et al*, 1990), which is expensive in terms of both time and animal usage. GTPM applications may assist the identification of surrogate markers for the development of this phenotype in cultured cells. The exposure of rodent hepatocytes to the non-genotoxic carcinogen, phenobarbitone, has been studied by using microarray and gel-based expression technologies, with the result that in excess of 300 genes were found whose expression is modulated by this reagent (Rhodi *et al*, 1999). Many other toxic endpoints could be profiled with this approach, and combinatorial approaches such as using cell lines developed from transgenic or knockout animals (Ryffel, 1997) could provide powerful insights into the role of specific genes.

Treatment of test systems with known reference toxicants (e.g. with similar toxic endpoint, mechanism, chemical structure or target organ) could permit the identification of diagnostic gene expression, or proteome and/or metabolite patterns. Such pattern recognition could facilitate the discovery and subsequent validation of biomarkers useful for application in higher throughput technologies to characterise or detect specific toxicity endpoints.

The throughput requirements of such an approach, allied with the limited availability of test substance, e.g. in the early phases of product development, may necessitate employing *in vitro* culture systems. Although the use of *in vitro* approaches can simplify the technical application of these approaches, it is of paramount importance that the model system recapitulates the biology being studied as accurately as possible. For this reason many researchers choose to employ *in vivo* systems due to absence of appropriate cell line models. In all cases, the validity of the test system is of greater importance than the technology being applied. Thus, while such *in vitro* systems have greater practical advantages, there are major considerations to be made:

- Where the results are likely to influence the derivation of No Observed Adverse Effect Levels (NOAELs) the relevance to *in vivo* situation must be established by correlating the observed changes with 'classical' adverse effects.
- Specific metabolism may be required to produce the active species, dependent on strain, sex or route of administration.
- Compound-induced transcript changes may not reflect accurately the response of the corresponding organ *in vivo*, potentially due to the microenvironment and to cellular interactions that might be required to lead to the observed mode of toxicity.
- Availability of appropriate cell lines may be limited (although where mechanistic information is not sought, generic cell lines may still be useful).

In vivo systems are probably also needed to establish base-line system noise levels, the effects of background events, such as food composition, as well as the effects of physiological patterns of activity e.g. circadian rhythm.

Whatever the final test system will be, characterising and using predictive markers of toxicity as flags could deliver highly significant savings in the costs of product development. Using these results to prompt and guide additional mechanistic studies at an early stage of development could lessen the chances of generating uninformative long-term toxicity studies and the need for them to be repeated. In addition, these suitably-characterised markers have the potential to point to threshold values for toxicity, thereby facilitating dose selection for long-term studies and might help shorten the time of the final product to market.

The development of the reference data sets to allow the 'pattern recognition' approach to toxicology is likely to require the application of complex computer algorithms and statistical approaches. For example, statistical clustering techniques have been applied to microarray data to analyse the temporal patterns of gene expression, which characterise serum-responsiveness and wound repair (Iyer *et al*, 1999), and to distinguish cancerous tissue from normal tissues and cell lines (Alon *et al*, 1999). A number of resources exist in both the academic and commercial sectors to assist in managing reference compound gene expression data sets to facilitate 'pattern recognition' (Bassett *et al*, 1999). Scientists at The National Human Genome Research Institute have developed one example, the software platform ArrayDB. The system facilitates the storage, retrieval and analysis of microarray data along with information linking some 15,000 genes to public domain sequence and pathway databases (Ermolaeva *et al*, 1998). In addition there are collaborative efforts involving pharmaceutical, agrochemical and chemical industries, such as the International Life Science Institute (ILSI) Health and Environmental Sciences Institute (HESI) subcommittee (application of genomics and proteomics to mechanism based risk assessment) to generate publicly accessible datasets of this type using reference toxicants (see www.ilsi.org). The construction of such baseline datasets will undoubtedly assist in determining the normal variability in levels of individual genes or gene products, which will in turn assist in distinguishing background noise from adaptive or mechanistically important change.

A question commonly posed is to what extent we can be certain that the changes observed in the level or state of any individual biomolecule may contribute to a phenotypic change. Without detailed mechanistic follow-up work, the changes measured can only be considered correlative at best. In the context of mechanism-based research it is probably best to regard results obtained using GTPM technologies as the springboard to more detailed and focused investigations (using other experimental approaches) that would confirm or refute the significance of the observed changes. Furthermore, it is certainly not the case that induced changes in gene expression that are associated with no altered protein production are of no interest or without relevance.

However if changes are proven to represent reliable, sensitive and selective biomarkers of a toxic process or the actions of a particular class of toxicant then mechanistic support for their involvement may no longer be necessary. No single test method will ever be absolute or comprehensive for all toxic endpoints. It is a tool that provides data that require interpretation and is not an end in itself. It is for this reason that the toxicologist should remain the final arbiter, employing experience and expertise to interpret accurately the results of all the studies.

GTPM technologies are not a promise for the future; they are tools available to us now, which if used correctly and within the guiding principles of good experimental biology, will bring huge dividends. Concern has been voiced already that a potential problem is the misinterpretation, or over-interpretation, of such high-volume data analyses particularly in the context of determining product safety. It must be recognised that the interaction of xenobiotics with biological systems will in many instances result in some changes in gene expression, even under circumstances where such interactions are not associated with adverse effects. The challenge again is to ensure that sound judgement and the appropriate toxicological skills and experience are brought to bear on the data generated so that toxicologically-relevant changes in biomolecules are distinguished from those that are of no concern. It is therefore important that emerging data from these technologies is followed through with traditional approaches and analysed fully to establish if the measured changes are background noise, adaptive, beneficial or potentially harmful. Only when in possession of the full picture can individual changes be scored as relevant to the mechanism under study. In this regard it is important to recognise that changes in biomarkers cannot be viewed as either Potentially Referable Findings, or as changes which compromise previously characterised No Observed Adverse Effect Levels (NOAELs) of exposure under circumstances where there are no physiological or pathological indicators of adverse effect. These are points on which it is critically important to foster the development of consensus and common understanding across the stakeholder constituencies including industry, academia and regulators.

3. CONCLUSIONS

The GTPM technologies have the exciting potential to support toxicology and safety assessment by generating large quantities of data, potentially useful in mechanistic research. The technologies may also be applied in a 'predictive' context to identify biomarkers diagnostic for particular toxicant classes.

Experimental design, including full dose and time course is clearly critical, and lack of rigour here may generate misleading and potentially damaging data. The technologies do not distinguish causative events from adaptive response (or even system noise) and therefore the potential involvement of any gene, protein or metabolite in a given toxic response requires dedicated follow-up work. In this regard the best application of these technologies may be in concert with more traditional approaches (e.g. biochemical toxicology, clinical chemistry, pathology). Simply put, it is not possible to do other than hypothesise as to whether any given change is harmful, neutral or potentially beneficial to the organism by employing these technologies in isolation. In addition these technologies are capable of characterising changes at low levels of exposure (or at early time points), under circumstances where more 'traditional' indicators of toxicity are not observed.

At the present time, it is not possible to recognise how many and which of these changes represent the onset of a genuine adverse effect and which are healthy adaptive or repair responses. Without further substantiation, such preliminary indications could form the basis for over-conservative regulatory action, for example by application of the Precautionary Principle; a situation demonstrated in the past with the advent and application of the Ames test as a new approach to toxicological evaluation of chemicals. It will therefore be essential to correlate these molecular changes with classical adverse effects and define those exposure levels that also evoke the changes (pathological, biochemical, developmental or functional) hitherto accepted as the basis for establishing NOAELs.

Development and applications of the technology will be facilitated by collaborations on technical issues, such as platform comparisons, data analysis/bioinformatics, broad species coverage. In this regard, it is encouraging to see the emergence of academia and industry consortia, such as the ILSI HESI sponsored Genomics Subcommittee, with commitment to begin building and populating the databases needed for pattern recognition. The magnitude of this problem, including timescales needed before the realisation of databases that are truly predictive of a broad range of toxicities are developed, remains to be seen. Finally, it is important for both scientists and regulators to appreciate what the technologies can and cannot do, and to work together to define their appropriate use in regulatory processes and discussions. The reaching of common ground would undoubtedly be facilitated by the sharing and publication of appropriate pilot data testing the utility of these technologies in helping understand the mechanism

by which chemical species exert biological effects. It would be advantageous for the chemical industries to commission such studies for the following reasons:

- Enable informed discussion on appropriate applications of these technologies.
- Consider impact of the technologies from an informed perspective.
- Increase understanding among companies as to the pros and cons of the individual approaches.
- Communicate the chemical industries' commitment to the appropriate application of these technologies to the regulatory agencies.
- Place 'good science' examples of the utility of these approaches in the public domain to promote the establishment of common understanding and a responsible code of practice in the generation, use and interpretation of GTPM analyses.

APPENDIX 1: TECHNOLOGY OVERVIEW

Genomics

Characterisation (at the level of DNA sequencing or chromosome mapping) of the DNA sequence of an organism provides the primary information used for a diverse range of technologies and applications. Completed DNA sequencing of a number of species has now been achieved, with both the human and murine high-resolution maps due imminently. The technologies to accomplish such large-scale enterprises are evolving rapidly, as demonstrated by the recent announcement that Celera Genomics (www.celera.com) has characterised 95% of the murine genome (approx 9.3 billion base pairs) in approximately seven months of effort.

The investigation of variable, or polymorphic, regions of genes in an attempt to characterise their potential role in idiosyncratic responsiveness is a distinct discipline, and is often referred to as 'pharmacogenomics' or 'pharmacogenetics'. These technologies allow the mapping of single nucleotide polymorphisms (SNPs) for specific genome regions. SNP maps are being generated in both the academic (http://snp.cshl.org) and commercial (www.celera.com) domains with approx 3 million individual SNPs now characterised. Clearly, genetic polymorphism is a critical consideration in exploring inter-individual differences in response to toxic insults and in identifying susceptible sub-populations. For this reason, attempts to map common polymorphisms that may be associated with sensitivity or resistance to chemical insults are of particular importance and represent the driving force for the evolution of 'toxicogenetics' (Guengerich, 1998; Wang *et al*, 1998). These technologies are maturing at a rapid rate, as evidenced by the recent announcement by Affymetrix inc (www.affymetrix.com) of the formation of a spin-off company, Perlegen Sciences, which will capture the entire genomes of 50 human individuals on microarray-style chips, for use in linking different SNPs to disease or adverse reaction to xenobiotic exposure.

The term 'functional genomics' refers to a host of technologies that enables the biological role of genes to be investigated. These include the well-established technologies of transgenic or knockout animal models. More recently the use of heterologous expression systems have allowed a higher throughput approach to assessing gene function. In this system a family of unique viral vectors that can carry genes of interest are expressed in a host organism to monitor the phenotypic effects of gene expression.

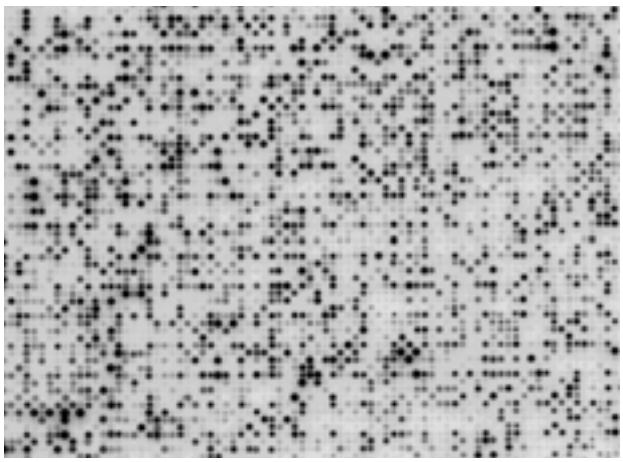
Comparable approaches for knocking out the genes in simple organisms also exist, and while these are primarily commercial enterprises using proprietary technology, initiatives in the academic domain continue to be developed. For example, the *C. elegans* Gene Knockout Project (http://elegans.bcgsc.bc.ca) is a worldwide consortium whose ultimate goal is to produce null alleles of all known genes in the *C. elegans* genome as a model system. Such approaches are taken one step further with comparative genomics, an approach that leverages functional and structural information from one biological system across all other biological systems, pioneered by commercial organisations such as Exelixis (www.exelixis.com). This is likely to be a significant growth area in molecular toxicology research given its potential to help in extrapolating mechanistic data from laboratory species to potential adverse effect in man.

Transcriptomics (Transcript Profiling)

DNA chips, or microarrays, permit the quantitative comparison of the expression levels of thousands of individual genes in different biological samples, thus facilitating comparisons, for instance, between normal tissue and diseased tissue, or between control cell lines and toxicant-treated cell lines. Practical usage of cDNA microarrays in toxicology research has been termed toxicogenomics.

The physical construction of cDNA arrays involves the immobilisation of DNA sequences, representative of the coding sequence of genes of interest, on a solid medium such as a glass slide or nylon membrane filter (Bowtell, 1999). Although there are a number of different technical platforms, the basic principles are comparable, involving the quantitation of complementary hybridisation of these immobilised single cDNA sequences with labelled messenger RNA (mRNA) prepared from the sample being tested. Software analysis then allows the determination of the extent of hybridisation of the labelled probes to the corresponding arrayed cDNA spots. A comparison of control samples with treated samples allows quantitative measurement of treatment-associated changes in gene expression (Brown and Botstein, 1999). Microarray platforms have been developed by commercial vendors, pharmaceutical or agrochemical companies and academic institutions (Nuwaysir *et al*, 1999) including custom cDNA microarrays designed specifically to measure genes of potential toxicological relevance. The application of toxicogenomics to mechanistic and predictive toxicology could assist in the identification of more-efficient markers for adverse health effects. These markers could highlight potential 'show stopping' toxicity earlier in a new compound's development than is possible at present, potentially enhancing the predictive power of *in vitro* model systems (Davila *et al*, 1998).

Figure 2: Illustration of a Microarray used in Transcript Profiling



The picture illustrates a result obtained by the DNA microarray technology employed in transcription profiling (Transcriptome).
Hybridisation with labelled cDNA preparations to pieces of complementary gene sequences (fixed on the support) from treated and untreated tissue will show different intensities of the spots when gene activities have changed due to (toxic) treatment. The relative intensities on the same array show the difference of messenger RNA levels within a given sample. RNAs of high copy number (e.g. for structural proteins like albumin) yield highly intensive spots.

The microarrays used to address these issues are either broad coverage 'discovery' style arrays, which include an unbiased selection of gene sequences (which may include probes for genes of as yet uncharacterised function), or 'hypothesis-driven' where the arrays are designed to focus on genes of relevance to a particular discipline such as toxicology (Pennie *et al*, 2000).

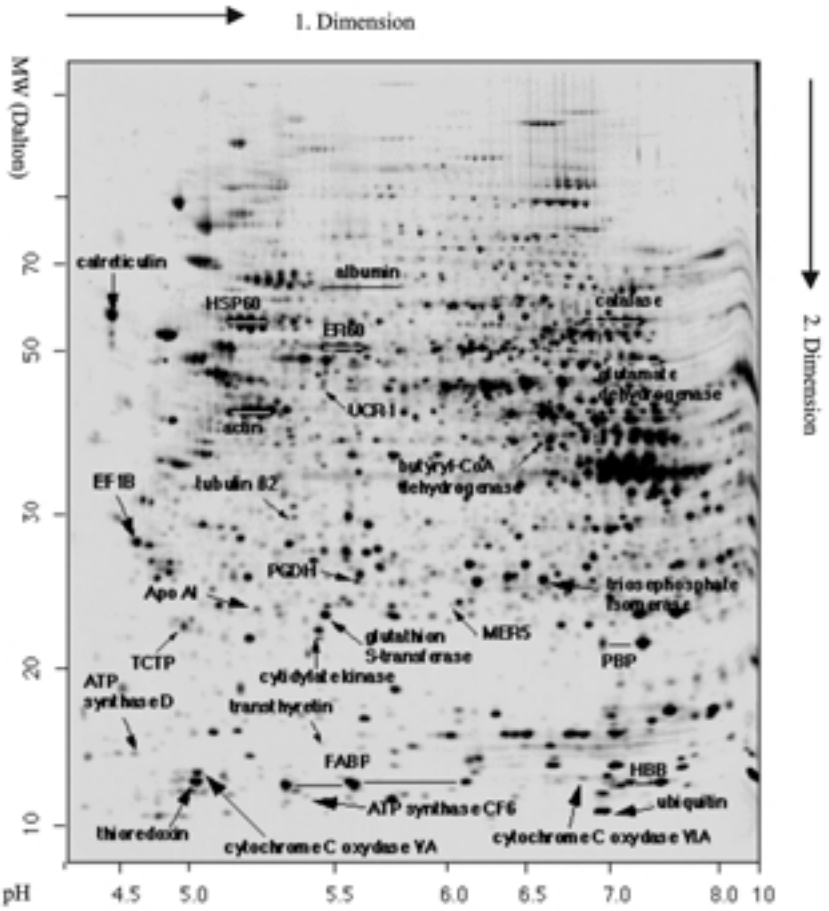
One of the technical caveats associated with microarray usage is incomplete coverage i.e. the microarray technique of gene expression analysis is a closed approach for higher organisms, in that only a defined subset of genes are profiled by a given technical platform. Platforms with complete coverage of simpler genomes, such as yeast or *C. elegans*, have been developed, and mammalian platforms approaching full genome coverage are likely to become available within the next few years. Open approaches to measuring differential gene expression have existed for many years (e.g. differential display polymerase chain reaction) (Liang and Pardee, 1998) but they are generally low throughput, and require considerable follow-up work to confirm the identity of differentially expressed genes. The Serial Analysis of Gene Expression (SAGE) technique has widespread usage in many laboratories (Velculescu *et al*, 1995) and the technology has matured to allow transcript profiling of very small amounts of biological material (Velculescu *et al*, 2000). In addition SAGE expression databases have been established to allow the analysis of gene expression events associated with endpoints such as carcinogenesis (Lal *et al*, 1999; van Kampen *et al*, 2000). Sophisticated high throughput sequencing technologies, such as MPSS (massively parallel signature sequencing) could also be applied to resolve differences in makeup between the transcriptomes of different samples (Brenner *et al*, 2000).

Microarrays only profile gene expression at the mRNA level, so this technology alone cannot confirm whether corresponding changes in the level of functional protein occur. In addition, protein modifications such as phosphorylation, which may be critical for the function of many proteins and pathways, cannot be detected directly by gene profiling techniques, although such protein modifications may result in downstream changes in gene expression, which can be detected.

Proteomics

The complementary technology of proteomics (Anderson *et al*, 1996) can characterise differential protein modification that may lead to changes in the activity of gene products. The traditional approach to proteomics profiling involves two stages. In the first stage the individual proteins in a biological sample are separated and resolved by two-dimensional gel electrophoresis. By comparing the resolved gels from control and treated samples, differentially expressed or modified proteins (often characterised by multiple spots) associated with the treatment can be noted. In the second stage the identity of individual protein spots of interest is established by either mapping their resolved locations based on historic data (Figure 3), or by excising the protein spot from the gel and identifying the protein biochemically (by peptide cleavage and sequencing, or analysis of mass spectra). The technology clearly differs from microarray transcript profiling in that it is an open technology i.e. within certain technical constraints all the proteins in a sample can be profiled. As with transcript profiling, there exist many publicly accessible databases on protein structure and function that can assist proteomics analyses (see for example www.proteome.com; Costanzo *et al*, 2000).

Figure 3: Example of a 2D-gel used in Proteomics



Typically a tissue sample is homogenised and separated by isoelectric focussing in the first and mass in the second dimension. The individual proteins are visualised by staining and can be eluted and characterised by a combination of biochemical processes and mass spectrometry (example taken from http://www.expasy.ch; Published SWISS-2DPAGE maps (Sanchez *et al*, 1995); modified).

Although proteomics has been scaled up in some industrial settings (for example, Oxford Glycosciences (www.ogs.com) processes around 1000 2D-gels per week), the technique generally suffers from low throughput, although microarray-based approaches to protein detection may remedy this (Lueking *et al*, 1999).

Metabonomics

Metabonomics is the approach by which the NMR spectra of biofluids are characterised and used in statistical analysis, typically Principal Component Analysis (PCA), to allow relationships between different compounds or xenobiotics to be estimated. Such associations are made on the basis of reducing an entire NMR spectra result to a single point, plotted in multidimensional space, where the distances between individual points on such a graph are proportional to the statistical similarity between the parent data sets; i.e. spectra of similarly acting compounds will cluster together on a PCA plot. Such approaches have been shown to distinguish toxicities (e.g. hepatotoxicity from nephrotoxicity) by NMR spectra analysis of the urine of treated rats with 98% accuracy (Holmes *et al*, 2000). In addition in these studies this approach is able to characterise subtle differences in metabolism that enable the different strains of animal used to be distinguished on the basis of metabonomic analysis. It is pertinent to note that metabonomics is able to distinguish changes in metabolism at dose or time points before the onset of clinical or pathological indicators of toxicity (Robertson *et al*, 2000).

APPENDIX 2: GLOSSARY OF TERMS

Allele
Alternative form of an gene, e.g. dominant (always expressed if present) or recessive (only expressed if no dominant allele is present)

Biomarker
Observable change (not necessarily pathological) in the function of an organism, related to a specific exposure or event.

cDNA
Complementary DNA enzymatically synthesised as a copy of mRNA.

C. elegans
Caenorhabditis elegans, a nematode or roundworm, the first animal to have its genome completely sequenced and all genes fully characterised.

Coding regions
Those parts of the DNA that contain the information needed to form proteins. Other parts of the DNA may have non-coding functions (e.g. start-stop, pointing or timer functions) or as yet unresolved functions or maybe even 'noise'.

DNA
Deoxyribonucleic Acid - the chemical substance containing the genetic code. See also nucleotide.

Genome
Chromosomal DNA information.

Genomics
Techniques available to identify the DNA sequence of the genome.

Genotype
Full set of genes carried by an individual organism. Note that this is more limited than the genome, since the genome also contains DNA not coding for genes.

Hepatocytes
Liver cells.

Hepatotoxicity
Toxicity to the liver.

Heterologous expression systems
Systems that allow expression of a gene in a different organism.

Hybridisation
Formation of a double strand from two different, more or less complementary single nucleic acid strands.

Idiosyncrasy
Specific (and usually unexplained) reaction of an individual to e.g. a chemical exposure to which most other individuals do not react at all. Example: some people react to their very first aspirin with a potentially fatal shock. General allergic reactions do not fall into this category.

Knockout animals
Genetically engineered animals in which one or more genes, usually present and active in the normal animal, are absent or inactive.

Metabonome
Constituent metabolites in a biological sample.

Metabonomics
Techniques available to identify the presence and concentrations of metabolites in a biological sample.

Murine
Of the mouse.

Nephrotoxicity
Toxicity to the kidney.

NMR
Nuclear Magnetic Resonance, a technique to identify atoms in a sample by measuring the signal given off by the relaxation of e.g. protons previously aligned in a strong magnetic field.

Non-genotoxic carcinogen
A substance that causes cancer, not by primarily damaging the genetic material, but by mechanisms that stimulate cell proliferation, thus increasing the chances for natural mutations to be reproduced, and/or selection of specific cell populations that may derange in a later stage.

Nucleotide
In this case the basic building block of DNA and RNA: a base/sugar/phosphate complex, Three nucleotides form a codon, coding for one amino acid.

Null allele
An inactive form of a gene.

Phenotype
Total of observable features of an organism, as the result of interaction between the genetic material (genotype) and the environment.

Polymerase chain reaction
Technique enabling a rapid multiplication of selected parts of a DNA or RNA strand.

Polymorphism
In this context, the existence of inter-individual differences in DNA sequences coding for one specific gene. The effects of such differences may vary dramatically, ranging from no effect at all to the building of inactive proteins, or not even building the protein.

Proteome
Entire protein complement of a biological sample.

Proteomics
Techniques available to identify the proteins in a biological sample.

mRNA
Messenger Ribonucleic Acid: the substance carrying genetic information from the DNA to the protein production site.

Serum-responsiveness
Cell proliferative reaction to the addition of serum to tissue culture medium after prior deprivation.

Signature sequencing
Sequencing of a short stretch of cDNA close to the end of the complementary mRNA. Sequence stretches of some 20 nucleotides are sufficiently discriminative to identify the transcript of an individual gene in a mammalian tissue.

Single nucleotide polymorphism
(see nucleotide, see polymorphism) Inter-individual variations in the genetic code at the level of one single building block.

Transcription
Formation of mRNA, complementary to a string of DNA.

Transcriptomics
Techniques available to identify the mRNA from actively transcribed genes.

Transcriptome
mRNA from actively transcribed genes.

Transcript profiling
see Transcriptomics

Transgenic animals
Genetically engineered animals carrying genes from a different species.

Xenobiotic(s)
Substance(s) (normally) not present in the reference organism.

BIBLIOGRAPHY

Alon U, Barkai N, Notterman DA, Gish K, Tbarra S, Mack D, Levine AJ. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci* 96:6745-6750.

Anderson NL, Taylor J, Hoffmann JP, Esquer-Blasco R, Swift S, Anderson NG. 1996. Simultaneous measurement of hundreds of liver proteins: application in assessment of liver function. *Toxicol Pathol* 24:72-76.

Bassett DE Jr, Eisen MB, Boguski MS. 1999. Gene expression informatics - it's all in your mind. *Nat Genet* 21:51-55.

Bowtell DD. 1999. Options available - from start to finish - for obtaining expression data by microarray. *Nat Genet* 21:25-32.

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18:630-634.

Brown PO, Botstein D. 1999. Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21:33-37.

Chabra RS, Huff JE, Schwetz BS, Selkirk J. 1990. An overview of prechronic and chronic toxicity/carcinogenicity experimental study designs and criteria used by the National Toxicology Program. *Environ Health Perspect* 86:313-321.

Costanzo MC, Hogan JD, Cusick ME, Davis BP, Fancher AM, Hodges PE, Kondu P, Lengieza C, Lew-Smith JE, Lingner C, Roberg-Perez KJ, Tillberg M, Brooks JE, Garrels JI. 2000. The yeast proteome database (YPD) and Caenorhabditis elegans proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res* 28:73-6.

Davila JC, Rodriguez RJ, Melchert RB, Acosta D Jr . 1998. Predictive value of *in vitro* model systems in toxicology. *Ann Rev Pharmacol Toxicol* 38:63-69.

DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278:680-686.

Ermolaeva O, Rastogi M, Pruitt KD, Schuler GD, Bittner ML, Chen Y, Simon R, Meltzer P, Trent JM, Boguski MS. 1998. Data management and analysis for gene expression arrays. *Nat Genet* 20:19-23.

Guengerich FP. 1998. The environmental genome project: functional analysis of polymorphisms. *Environ Health Perspect* 106:365-368.

Holmes E, Nicholls AW, Lindon JC, Connor SC, Connelly JC, Haselden JN, Damment SJ, Spraul M, Neidig P, Nicholson JK. 2000. Chemometric models for toxicity classification based on NMR spectra of biofluids. *Chem Res Toxicol* 13:471-478.

Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J Jr, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO. 1999. The transcriptional program in the response of human fibroblasts to serum. *Science* 283:83-87.

Khan J, Bittner ML, Chen Y, Meltzer PS, Trent JM. 1999. DNA microarray technology: the anticipated impact on the study of human disease. *Biochim Biophys Act*a 1423:M17-28.

Lal A, Lash AE, Altschul SF, Velculescu V, Zhang L, McLendon RE, Marra MA, Prange C, Morin PJ, Polyak K, Papadopoulos N, Vogelstein B, Kinzler KW, Strausberg RL, Riggins GJ. 1999. A public database for gene expression in human cancers. *Cancer Res* 59:5403-5407.

Liang P, Pardee AB.1998. Differential display. A general protocol. *Mol Biotechnol* 10:261-267

Lueking A, Horn M, Eickhoff H, Bussow K, Lehrach H, Walter G. 1999. Protein microarrays for gene expression and antibody screening. *Anal Biochem* 270:103-111.

Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard J, Slade D, Dai H, Bassett DE Jr, Hartwell LH, Brown PO, Friend SH. 1998. Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat Med* 4:1293-1301.

Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. 1999. Microarrays and toxicology: The advent of Toxicogenomics. *Mol Carcinogenesis* 24:153-159.

Pennie WD, Tugwood JD, Oliver GJ, Kimber I. 2000. The principles and practice of toxigenomics: applications and opportunities *Toxicol Sci* 54:277-283.

Robertson DG, Reily MD, Sigler RE, Wells DF, Paterson DA, Braden TK. 2000. Metabonomics: evaluation of nuclear magnetic resonance (NMR) and pattern recognition technology for rapid *in vivo* screening of liver and kidney toxicants. *Toxicol Sci* 57:326-337.

Rodi CP, Bunch RT, Curtiss SW, Kier LD, Cabonce MA, Davila JC, Mitchell MD, Alden CL, Morris DL. 1999. Revolution through genomics in investigative and discovery toxicology. *Toxicol Pathol* 27:107-110.

Ryffel B. 1997. Impact of knockout mice in toxicology. *Crit Rev Toxicol* 27:135-154.

Sanchez J-C, Appel RD, Golaz O, Pasquali C, Ravier F, Bairoch A, Hochstrasser DF. 1995. Inside SWISS-2DPAGE database. *Electrophoresis* 16:1131-1151.

Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-470.

van Kampen AH, van Schaik BD, Pauws E, Michiels EM, Ruijter JM, Caron HN, Versteeg R, Heisterkamp SH, Leunissen JA, Baas F, van Der Mee M. 2000. USAGE: a web-based approach towards the analysis of SAGE data. *Bioinformatics* 16:899-905.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. *Science* 270:484-487.

Velculescu VE, Vogelstein B, Kinzler KW. 2000. Analysing uncharted transcriptomes with SAGE. *Trends Genet* 16:423-425

Wang DJ, Fan J-B, Siao C-J, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES. 1998. Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077-1082.

MEMBERS OF THE TASK FORCE

B. Pennie*	Syngenta, CTL UK - Macclesfield
H-J. Ahr	Bayer D - Wuppertal
C. Braun	Akzo Nobel NL - Arnhem
R. Bars	Aventis F - Sophia Antipolis
E. Jacob	BASF D - Ludwigshafen
F. Carpanini (Secretary)	ECETOC B - Brussels

* Primary Author

MEMBERS OF THE SCIENTIFIC COMMITTEE

(Peer Review Committee)

N. Carmichael (Co-chairman) Head, Toxicology	Aventis F - Sophia Antipolis
G. Randall (Co-chairman) Director, Environmental Laboratory	AstraZeneca UK - Brixham
C. Braun* Occupational Toxicologist	Akzo Nobel NL - Arnhem
E. Bomhard Head, Industrial Toxicology	Bayer D - Wuppertal
C. d'Hondt Head, Environmental Safety Department	Syngenta CH - Basel
T. Feijtel Manager, Professional and Regulatory Services	Procter & Gamble B - Brussels
J. Jackson Senior Associate, Medical Adviser	Monsanto B - Brussels
R. Millischer Head, Industrial Toxicology Department	Elf Atochem F - Paris
A. Sarrif Director, Toxicology Affairs, Europe	DuPont D - Bad Homburg
J. Solbé Head, SEAC Environment	Unilever UK - Bebington
L. Smith Director, Central Toxicology Laboratory	Syngenta UK - Macclesfield
B. van Ravenzwaay Director, Experimental Toxicology and Ecology	BASF D - Ludwigshafen
H-J. Wiegand Head, Product Safety Department	Degussa-Hüls D - Marl

* Steward