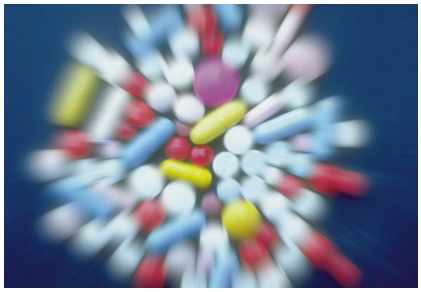
DNA microarrays in drug discovery and development

Christine Debouck[1] & Peter N. Goodfellow[2]

SmithKline Beecham Pharmaceuticals. [1]709 Swedeland Road, King of Prussia, Pennsylvania 19406, USA. [2]New Frontiers Science Park, Third Avenue, Harlow, Essex, CM19 5AW, UK.

DNA microarrays can be used to measure the expression patterns of thousands of genes in parallel, generating clues to gene function that can help to identify appropriate targets for therapeutic intervention. They can also be used to monitor changes in gene expression in response to drug treatments. Here, we discuss the different ways in which microarray analysis is likely to affect drug discovery.

Most pharmaceuticals sold today have their origin in a biochemistry laboratory. Drug discovery has traditionally started with a biochemical pathway implicated in a pathophysiological process. An appropriate enzymatic activity, preferably the rate-limiting step in the pathway, was characterized and purified, most often from animal tissues. The purified enzyme was then screened against collections of structurally diverse small molecules. Occasionally, sufficient information was known about the mechanism of action and structure of the enzyme to allow particular classes of small molecules to be targeted. Finally, medicinal chemists worked at optimizing the lead compounds conferring desirable properties such as bioavailability and removing undesirable properties such as poor specificity for the target enzyme. A similar process was applied to the identification of receptors and their use as drug targets. This biochemistry-based approach has served the pharmaceutical industry well and has provided many effective drugs for a variety of diseases.



The advent of molecular biology and gene cloning techniques transformed drug discovery in the following ways.

- Gene cloning and *in vitro* expression technology provides human targets when access to human tissue samples is limited or impossible. Use of human targets as opposed to animal proteins is critical, because a single amino acid difference can render a compound ineffective against a target of interest[1].
- Cloning methodology can be used to produce targets that are difficult or dangerous to isolate from natural sources. The protease from the human immunodeficiency virus, HIV, is present in only minute amounts in virions; a biochemical approach would have required massive cultivation of this lethal pathogen[2].
- Cross-hybridization with cloned sequences is a rapid method for the identification of related targets. The testing of compound selectivity against related targets can help to minimize unwanted side effects of drugs.
- Site-directed mutagenesis can be used to test hypotheses about drug target interaction and guide medicinal chemistry efforts.

Although the molecular biology revolution changed the processes of drug discovery, two major barriers remained in place. First, the number of potential targets was limited by the number of cloned genes. Second, linking targets with therapeutic utility, a process known as target validation, required a detailed understanding of pathophysiological processes. The first barrier has been removed, initially by the application of high throughput EST gene sequencing[3] and more recently, by the combined efforts of participants in the Human Genome Project. At first approximation, the majority of human genes are available as targets and all genes will be available when the human genome has been completely sequenced. The second barrier is yet to be surmounted. An ideal technique would equate human genes with disease on a whole genome scale. Potential approaches include genome scans in linkage and association studies as well as gene expression profiling methods such as DNA microarrays and proteomics. DNA microarrays can be used for both genotyping and measuring mRNA levels. Here, we consider the potential of DNA microarrays to rapidly generate information necessary for the identification and validation of novel therapeutic targets.

Gene expression patterns in normal cells and tissues

The expression pattern of a gene provides indirect information about function. A gene expressed only in the kidney is unlikely to be directly involved in the pathology of schizophrenia. Several effective drugs have been developed against protein targets that are widely expressed in the body (for example, the angiotensin receptor[4–5]). However, highly selective tissue expression of a drug target is attractive, as the potential for unwanted side effects may be more restricted. Expression data have their own limitations; mRNA levels may not reflect protein levels and expression of a protein may not always have a physiological consequence.

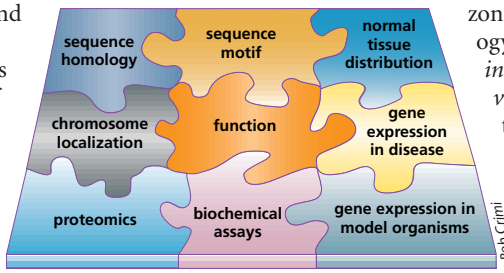
Knowledge of highly selective gene expression, as well as sequence homology to a known gene family, can provide a convenient shortcut for implicating a target in a given pathway or disease. For example, cathepsin K, a novel cysteine protease originally identified in an EST library, is expressed selectively in osteoclasts[6]. (Osteoclasts are responsible for bone resorption; an imbalance between bone formation and bone resorption is the underlying pathophysiology of osteoporosis.) Inhibition of cathepsin K might be predicted to interfere with the function of osteoclasts and prevent osteoporosis. To test this hypothesis, it

was necessary to refine the expression data at both the mRNA and the protein level. *In situ* hybridization was used to demonstrate high levels of expression in osteoclasts and not other bone cells and immunohistochemistry was used to show that cathepsin K protein is found in the 'pits' where proteolytic degradation of bone occurs[7]. The tissue distribution of cathepsin K was first indicated by a database scan of EST sequences derived from an osteoclast library, although it could as easily have been identified by microarray[7]. It also illustrates some of the limitations. The expression pattern is only as specific as the source of mRNA, and tissues may be composed of many different cell types which can be difficult to separate (see page 38; ref. 8). Ultimately, it may be necessary to measure more sophisticated indicators of function such as protein location and rates of turnover.

Differential gene expression patterns in disease

The most attractive application of microarrays is in the study of differential gene expression in disease. Other techniques depend on either DNA sequencing (of EST libraries or SAGE elements) or on PCR-based differential display methods—these techniques are laborious and often insensitive.

Diversion from normal physiology is frequently accompanied by a panoply of histological and biochemical changes, including changes in gene expression patterns. The up- or downregulation of gene activity can either be the cause of the pathophysiology or the result of the disease. Although targeting disease-causing gene products is desirable to achieve disease modification, interfering with genes that are expressed as a consequence of disease can lead to alleviation of symptoms. The opportunity to compare the expression of thousands of genes between 'disease' and 'normal' tissues and cells will allow the identification of multiple potential targets.



Examples of monitoring differential gene expression by microarrays are beginning to appear in the literature. Rheumatoid tissue was recently analysed using a microarray of about 100 genes known to have a role in inflammation[9]. Among others, genes encoding interleukin-6 and several matrix metalloproteinases, including matrix metallo-elastase (HME), were markedly upregulated. The latter result was unexpected, as the distribution of HME was previously thought to be limited to alveolar macrophages and placental cells. The advantage of profiling the expression of large numbers of genes is illustrated by another study in which a DNA microarray with about 1,000 cDNAs of unknown sequence affixed to its surface queried expression in phorbol ester-treated cells and untreated cells[10]. After DNA sequencing of the cDNAs corresponding to induced signals, the two most highly induced corresponded to genes known to be upregulated by phorbol esters. One moderately induced signal did not match any known gene. This novel gene is now implicated in the processes of tumour promotion.

Gene expression in model systems

Disease models in inbred animals are attractive, as they are more amenable to well-controlled experimental design. In all cases, relevance to human disease needs to be carefully gauged as gene sequence and gene function are not always faithfully conserved between species. Nonetheless, detailed profiling of gene expression in model systems will yield additional insight into cellular, animal and human physiology critical to the discovery and validation of therapeutic targets.

Most analyses of differential gene expression in model organisms have been conducted using lower throughput approaches, such as differential display PCR. Although these techniques have uncovered physiologically relevant gene regulations[11,12], microarray analysis allows a much higher throughput. Applications will include comparison of gene expression patterns in the normal mouse to those in mice overexpressing selected genes (transgenics) or defective for selected genes (point mutations, knockouts and so on) as well as other animal models of human diseases. This type of analysis has already been described using the yeast *Saccharomyces cerevisiae*[13] and will soon be deployed for *Caenorhabditis elegans*.

Gene expression patterns in pathogens

A new bacterial genome seems to be sequenced in its entirety every month. The small size of these genomes will allow the easy construction of a single microarray where every gene from a given microbe is represented. For the microbiologist, confined for years to studying bacteria one gene at a time in a test tube under artificial growth conditions, the horizons appear unlimited. Microarray technology will identify genes that are turned on *in vitro* but not at the site of infection *in vivo*, and vice versa—and those genes that are only turned on during infection *in vivo*. Such genes will encode virulence determinants that are regulated by environmental signals such as the transition from ambient temperature to body temperature[14]. In fact, traditional genetic techniques to identify virulence genes are time consuming and will be quickly supplanted by microarray methods. A similar approach will be used to study viral gene expression during the time course of acute infection or during latency. Microarrays can also be used to study the response of the host to challenge with the pathogen.

Gene expression in response to drug treatments

Microarrays are potentially powerful tools for investigating the mechanism of drug action. Two recent studies have described the application of high density microarrays to examine the effects of drugs on gene expression in yeast as a model system. In one, the effect of potent kinase inhibitors was analysed on a yeast genome-wide scale by measuring changes in mRNA levels before and after treatment with inhibitors[15]. The second study reported a gene expression pattern (or 'signature') characteristic for the immunosuppressive drug FK506. This same signature was also observed in yeast cells carrying a null mutation in the FK506 target, establishing that genetic and pharmacological ablation of a gene function results in similar changes in gene expression. Treatment of the null mutants with FK506 also revealed additional pathways distinct from the drug's primary target[16]. It is possible that yeast will provide a high-throughput platform for studying the cellular responses to drugs, however, a similar method applied to human cells and tissues would have even more direct utility in identification and validation of novel therapeutics.

In the long term, microarrays will contribute to the analysis of metabolic pathways such as the induction of cytochrome P450 detoxifying enzymes for which new signalling pathways are being identified[17]. They will also be used to delineate and predict adverse events, such as the undesirable upregulation of liver enzymes. Monitoring expression of genes with toxicity potential has elicited the interest of large pharmaceutical companies as well as specialized biotech companies. Traditional pathology and

toxicity studies, however, will remain in place for many years, as the full understanding of the toxicological consequences of over- or under-expression of a given gene are dependent on a much deeper knowledge of the identity and function of the full complement of human genes. Even so, the identification of induced gene products that can be used as surrogate markers to readily follow the effect and dose of a drug in the clinical setting will make drug development easier. This approach will be particularly useful when the drug itself is difficult to track by conventional detection methods or when its clinical outcome takes a long time to manifest itself—for example, the reduction in bone fracture incidence as a consequence of anti-osteoporosis therapy.

Microarrays encompassing at least one element for each expressed gene in a given organism will soon become available for many organisms. The number and variety of applications is unlimited. In the next decade, the microarray may well become as essential as PCR is now, but many limitations currently apply. Common standards will be required before investigators can meaningfully combine data, which will have to be stored in databases that can be effectively mined (see page 51, ref. 18). Microarrays are not the panacea for all gene expression studies; labour-intensive *in situ* hybridization and immunohistochemistry techniques will remain indispensable—although hopefully, they will be joined by high-throughput methods of protein analysis[19].

1. Oksenberg, D. *et al*. A single amino-acid difference confers major pharmacological variation between human and rodent 5-HT1B receptors. *Nature* **360**, 161–163 (1992).
2. Debouck, C. *et al*. Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the gag precursor. *Proc. Natl Acad. Sci. USA* **84**, 8903–8906 (1987).
3. Adams, M.D. *et al*. Sequence identification of 2,375 human brain genes. *Nature* **355**, 632–634 (1992).
4. Bergsma, D.J. *et al*. Cloning and characterization of a human angiotensin II type 1 receptor. *Biochem. Biophys. Res. Comm*. **183**, 989–995 (1992).
5. Smith, R.D. & Timmermans, P.B. Human angiotensin receptor subtypes. *Curr. Opin. Nephrol. Hypertens*. **3**, 112–122 (1994).
6. Drake, F.H. *et al*. Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. *J. Biol. Chem*. **271**, 12511–12516 (1996).
7. Pietu, G. *et al*. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res*. **6**, 492–503 (1996).
8. Cole, K.A., Krizman, D.B. & Emmert-Buck, M.R. The genetics of cancer—a 3D model. *Nature Genet*. **21**, 38–41 (1999).
9. Heller, R.A. *et al*. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl Acad. Sci. USA* **94**, 2150–2155 (1997).
10. Schena, M. *et al*. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl Acad. Sci. USA* **93**, 10614–10619 (1996).
11. Wang, X. *et al*. Discovery of adrenomedullin in rat ischemic cortex and evidence for its role in exacerbating focal brain ischemic damage. *Proc. Natl Acad. Sci. USA* **92**, 11480–11484 (1995).
12. Qu, D. *et al*. A role for melanin-concentrating hormone in the central regulation of feeding behaviour. *Nature* **380**, 243–247 (1996).
13. De Risi, J.L., Iyer V.R. & Brown P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
14. Mekalanos, J.J. Environmental signals controlling expression of virulence determinants in bacteria. *J. Bacteriol*. **174**, 1–7 (1992).
15. Gray, N.S. *et al*. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **281**, 533–538 (1998).
16. Marton, M.J. *et al*. Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med*. **4**, 1293–1301 (1998).
17. Bertilsson, G. *et al*. Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. *Proc. Natl Acad. Sci. USA* **95**, 12208–12213 (1998).
18. Bassett, D.E. Jr, Eisen, M.B. & Boguski, M.S. Gene expression informatics—it's all in your mine. *Nature Genet*. **21**, 51–55 (1999).
19. Persidis, A. Proteomics. *Nature Biotechnol*. **16**, 393–394 (1998).