ELSEVIER

Biochemical Pharmacology

Biochemical Pharmacology 62 (2001) 1311–1336

Commentary

Gene expression microarray analysis in cancer biology, pharmacology, and drug development: progress and potential

Paul A. Clarke[a], Robert te Poele[a,b], Richard Wooster[b,c], Paul Workman[a,*]

[a]*Cancer Research Campaign Centre for Cancer Therapeutics, E Block, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK*
[b]*Section of Molecular Carcinogenesis, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK*
[c]*Sanger Centre, Welcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK*

Abstract

With the imminent completion of the Human Genome Project, biomedical research is being revolutionised by the ability to carry out investigations on a genome wide scale. This is particularly important in cancer, a disease that is caused by accumulating abnormalities in the sequence and expression of a number of critical genes. Gene expression microarray technology is gaining increasingly widespread use as a means to determine the expression of potentially all human genes at the level of messenger RNA. In this commentary, we review developments in gene expression microarray technology and illustrate the progress and potential of the methodology in cancer biology, pharmacology, and drug development. Important applications include: (a) development of a more global understanding of the gene expression abnormalities that contribute to malignant progression; (b) discovery of new diagnostic and prognostic indicators and biomarkers of therapeutic response; (c) identification and validation of new molecular targets for drug development; (d) provision of an improved understanding of the molecular mode of action during lead identification and optimisation, including structure–activity relationships for on-target versus off-target effects; (e) prediction of potential side-effects during preclinical development and toxicology studies; (f) confirmation of a molecular mode of action during hypothesis-testing clinical trials; (g) identification of genes involved in conferring drug sensitivity and resistance; and (h) prediction of patients most likely to benefit from the drug and use in general pharmacogenomic studies. As a result of further technological improvements and decreasing costs, the use of microarrays will become an essential and potentially routine tool for cancer and biomedical research.

Keywords: Microarray; Drug development; Pharmacology; Cancer

1. Introduction and rationale

1.1. Advent of post-genomic biology

In February of this year, we witnessed arguably one of the most monumental achievements in biology—the simultaneous publication of around 93% of the sequence of the human genome by a public sector consortium and a private company [1,2][1]. The sequence has revealed many surprises, not least that the 10 billion base pairs likely encode around 30,000–40,000 genes—rather fewer than some of the earlier predictions of 100,000 or more [3] and barely twice as many as the fruit fly. On the other hand, the structure of human genes is more complex, incorporating multiple vertebrate-specific functional domains into sophisticated protein products, with further diversity being provided by alternative splicing. One of the aims behind the strategy of genomic sequencing is to provide an inventory of all the genes and regulatory sequences required to build an organism. The sequencing continues apace in both the public and private sectors, and complete sequences of mouse, rat, and other genomes will follow shortly (see http://www.sanger.ac.uk/or http://www.ncbi.nlm.nih.gov/Genomes/index.html). However, the dep-

* Corresponding author. Tel.: +44-20-8722-4301; fax: +44-20-8642-1140.

E-mail address: paulw@icr.ac.uk (P. Workman).

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; Cy, Cyanine; DLCL, diffuse large cell lymphoma; dNTPs, deoxyribonucleotides; ESTs, expressed sequence tags; mRNA, messenger RNA; NHL, non-Hodgkin's lymphoma; ORF, open reading frame; RT–PCR, reverse transcription–polymerase chain reaction; and 17AAG, 17-allylamino,17-demethoxygeldanamycin.

[1]Due to the large number of contributors to Ref.[2] (more than 250 authors), the reader is referred to the original article for a complete listing of the authors' names.

0006-2952/01/$ – see front matter
PII: S0006-2952(01)00785-7

osition of billions of bases of sequence into the databases cannot be considered the final goal as we are unable to define biological function—or disease pathology—purely from a genomic DNA sequence. The repositories of genes and their regulatory sequences represent the starting point of the new challenge of post-sequence functional genomics, which is to understand how these components interact and function [4].

Importantly, in terms of practical application, annotation of the human genome is certain to produce benefits not only for understanding basic biology, but also for identifying the molecular basis of disease and for accelerating the rate of drug discovery and development. Alongside high throughput sequencing to determine normal and abnormal gene structure, the use of microarray technology to measure gene expression patterns on a global scale is now poised to revolutionise the discovery and use of new medical treatments. The application of gene expression microarrays in cancer biology, pharmacology, and drug development is the subject of this review.

1.2. Gene expression profiling in cancer biology and treatment

The use of gene expression microarrays is particularly important in cancer. This is because the accumulation and combinatorial effects of abnormalities that drive the initiation and malignant progression of cancer result from the altered sequence or expression level of cancer-causing genes. These genetic abnormalities, which may be inherited or acquired, lead to the 'big six' hallmark traits of cancer, namely: (a) self-sufficiency in proliferative growth signals; (b) insensitivity to growth inhibitory signals; (c) evasion of apoptosis; (d) acquisition of limitless replicative potential; (e) induction of angiogenesis; and (f) induction of invasion and metastasis [5,6].

As will be discussed in subsequent sections, the use of microarrays can be extremely valuable both in understanding the basic biology and in the treatment of cancer. Important applications include: (a) development of a more global understanding of the gene expression changes that contribute to malignant progression; (b) discovery of diagnostic and prognostic indicators and biomarkers of response; (c) identification and validation of new molecular targets; (d) provision of an improved understanding of the molecular mode of action during lead identification and optimisation; (e) prediction of potential side-effects during preclinical development and toxicology studies; (f) confirmation of the molecular mode of action during hypothesis-testing early clinical trials; (g) identification of genes involved in conferring drug sensitivity and resistance; and (h) prediction of patients most likely to benefit from the drug and use in general pharmacogenomic studies (Fig. 1).

The use of gene expression microarrays in the above-mentioned ways can be carried out regardless of the nature of the molecular target. It should be noted, however, that as a result of our improved understanding of the genetics and molecular biology of cancer, there is an increasing move away from relatively non-selective cytotoxic drugs towards the new generation of molecular therapeutic agents that target the key molecular abnormalities that drive malignant progression and which, as a result, have an impact on one or more of the six hallmark traits of cancer listed above [7,8]. Proof of principle is now emerging that these selective agents show biological and therapeutic activity by the desired mechanism, not only in preclinical models but also in the cancer patient; leading examples are the monoclonal antibody Herceptin (trastuzumab) in erbB2-positive breast cancer, the bcr-abl tyrosine kinase inhibitor Glivec (STI571) in Philadelphia chromosome-positive chronic myelogenous leukemia and acute lymphocytic leukemia, and the epidermal growth factor receptor tyrosine kinase inhibitor Iressa (ZD1839) in non-small cell lung cancer [7,8]. Microarray technology will be especially useful for the discovery, development, and clinical use of such genome-based molecular therapeutics.

1.3. Why measure global gene expression?

The gene expression profile of a cell determines its phenotype, function, and response to the environment. The complement of genes expressed by a cell is very dynamic and will respond rapidly to external stimuli. Therefore, measurement of gene expression can potentially provide clues about regulatory mechanisms, biochemical pathways, and broader cellular function. In addition, the determination of genes expressed in disease tissue compared with the normal counterpart will further the understanding of disease pathology and identify potential points for therapeutic intervention. The underlying genetic basis of cancer makes this especially valid as tumor behavior is likely to be dictated in a combinatorial fashion by the mutation and abnormal expression of hundreds of genes. Thus, analysis of global gene expression patterns has the potential to predict biological behavior and clinical consequences, an expectation that will revolutionise cancer diagnosis and treatment.

In traditional biological and pharmacological experiments, it is usual to measure the expression of only a single gene or, at most, a handful of genes. A simple example of the potential benefit of measuring gene expression in cancer pharmacology is illustrated by the study of Wosikowski *et al*. [9], who measured the expression of *epidermal growth factor receptor, transforming growth factor-*α, and c-*erb-B2* in the 60 human tumor cell line panel of the US National Cancer Institute. A database of 49,000 compounds was then searched for agents for which cytotoxicity correlated with high level expression of the receptor and ligand mRNAs, and a number of correlations were noted. These were indicative of compounds that were potentially acting as inhibitors of the above mentioned tyrosine kinase receptors and related pathways. Though successful, there is a considerable limitation to the approach. Given that there are 30,000–

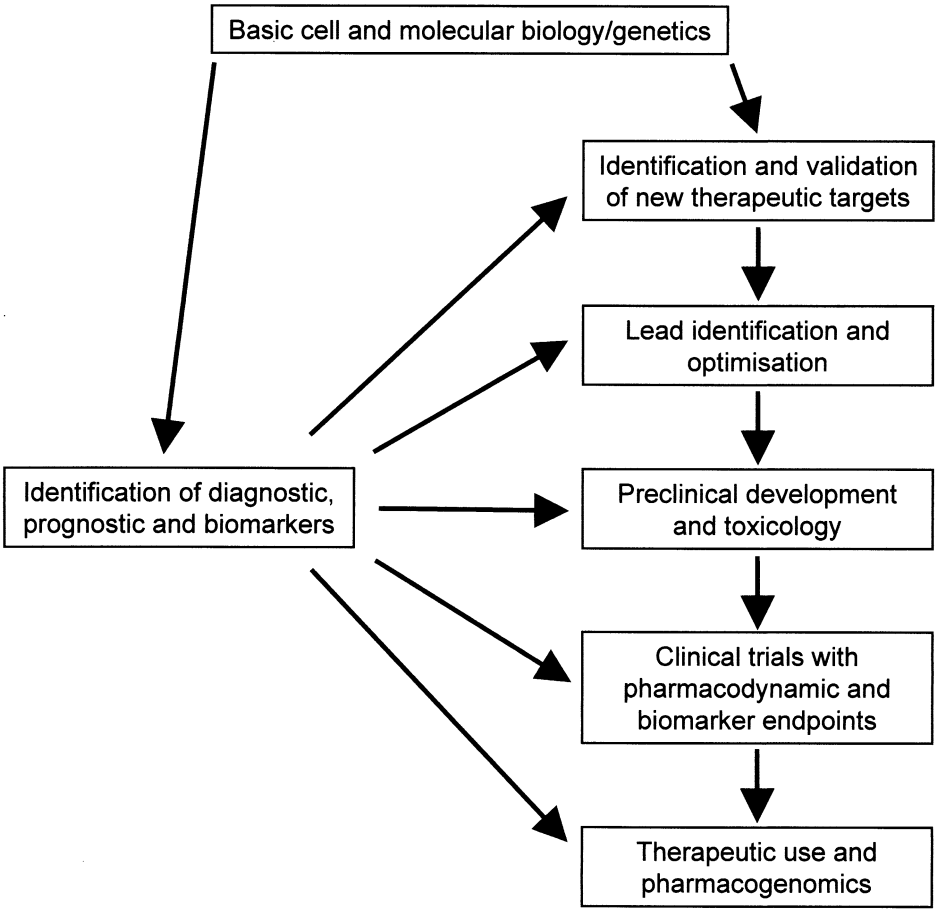

Fig. 1. Various phases in the discovery and development of therapeutic agents and of diagnostic, prognostic, and other biomarkers. Microarray expression profiling can be used to help advance all stages of this process.

40,000 human genes, screening the entire inventory of genes with techniques that measure only a few genes at one time is impractical. A more efficient and complete strategy would be to measure simultaneously all the genes, or a large defined subset, encoded by the human genome. In addition to obtaining a global view of gene expression, this strategy could also identify genes previously unassociated with particular cellular processes and would also remove the intellectual bias that is inevitably generated when examining a single gene or small numbers of genes. A subsequent study attempted to address this challenge by correlating the expression across the 60 cell line panel of 140 genes, as measured at the protein level by two-dimensional gels, with the cytotoxicity of approximately 4000 different compounds [10]. A number of drug sensitivity/expression correlations were noted, involving genes encoding products such as MDR-1, heat shock proteins, and p53, and these suggested a number of testable hypotheses. More importantly, this study demonstrated the utility that an expression profiling approach could potentially have in the development of cancer therapeutics. Ideally then, we would like to be able to profile total gene expression patterns in large numbers of experimental samples.

1.4. Why measure mRNA?

Gene expression can be assessed by measuring the quantity of the final product, i.e. the protein, or its intermediate, the mRNA template. At face value it would appear to be more rational to measure either the expression or activity of the final product, as this is the unit of function within the cell. Consistent with this, the few studies that have directly compared the relationship between mRNA and protein levels have found a poor correlation between the expression of selected proteins and their mRNAs. For example, the levels of the protein products of genes with similar abundances of mRNAs could vary up to 20-fold in *Saccharomyces cerevisiae;* likewise there was a 30-fold variance in the levels of mRNAs encoding proteins that were expressed with comparable abundance [11,12]. These observations, coupled with the fact that the final functional product of gene expression is measured rather than the intermediate mRNA species, imply that protein-based rather than transcript-based methods would be preferable. Analysis of protein also allows assessment of the influence of post-translational modifications such as phosphorylation, glycosylation, and proteolytic processing. However, the efficiency with which

one can currently obtain global and quantitative information of gene expression by measuring mRNA far outstrips the available proteomic technologies, which are generally lower throughput, more technologically challenging, and less readily compatible with routine material from clinical studies. Current RNA-based approaches to assess gene expression have been demonstrated to be both robust and reproducible. In addition, numerous studies have identified remodeled gene expression, assessed by measuring mRNA, during all the key cellular processes. Transcription factors that regulate mRNA production are frequently the targets of pathways that transduce environmental signals. Furthermore, there is also an increasing number of reports of altered expression of individual genes following drug exposure [13].

Changes in the molecular phenotype of the cell should be accurately reflected by its transcriptional profile, and the evaluation of gene expression by measuring mRNA should provide a molecular signature of the state of the activity of the cell and by extension the activity of proteins that regulate that state. However, although analysis of global mRNA expression (the transcriptome) undoubtedly generates much valuable information, the ideal scenario would be one where this is measured alongside global protein expression (the proteome) within the same experiment.

Gene expression can be profiled using a number of different technologies that have been described as 'open' or 'closed' systems [14]. Open systems are generally better suited to identifying uncharacterised genes within biological systems, and include technologies such as differential display [15], subtractive hybridisation approaches, serial analysis of gene expression [16], or massively parallel signature sequencing [17]. Closed systems such as gene arrays rely on the availability of gridded clones, cDNAs, or ESTs and are well suited to high throughput screening of multiple samples.

2. Microarray technology

2.1. Basic principles

Gene expression microarray technology has been founded on a number of important developments that have emerged over the last 3 decades. These include blotting methodologies, key recombinant DNA technologies such as the PCR, the availability of gridded libraries, and the accumulation of tens of thousands of sequenced cDNA and EST clones from the various genome sequencing initiatives.

Historically, mRNA abundance and hence gene expression have been determined by immobilising bulk RNA, electrophoretically separated or spotted, on nitrocellulose or charged nylon membranes, followed by incubation with radiolabeled gene-specific probes and detection of hybridised probes by autoradiography [18]. The array approach reverses this procedure and is essentially a miniaturised, massively parallel version of the conventional dot blot. In an array experiment, unlabeled nucleic acid probes are deposited onto a solid support so that the exact location of each spotted sample on the support is known and can be easily identified at the end of the experiment. The arrays are then probed with labeled single-strand cDNA representations of the cellular RNA pool. In the case of arrays printed on glass slides, RNA from test and reference samples can be compared simultaneously on a single array by using fluorescent labels that emit light at different excitation wavelengths [18]. This gives information on relative RNA abundance based on a direct comparison between the test and reference samples. Arrays printed on nylon membranes require radiolabeled test and reference samples and, in this case, are hybridised to separate arrays that are processed in parallel. Having identified differentially regulated genes, it is essential that the differences be validated by alternative methodologies that measure RNA, such as northern blotting, RNase protection or RT–PCR, or alternatively by protein analysis.

Much as any other new and exciting technology, gene expression microarrays have been subjected to extensive literature review over the last 2 years; many of the reviews deal with the technological aspects of their fabrication and application [19–22]. In general, the conduct of an array experiment can be divided into the following processes: (a) fabrication of the array; (b) RNA preparation and labeling; (c) application of the labeled RNA to the array, measurement of hybridisation, and image aquisition; and finally (d) data analysis. The establishment of an array facility requires considerable troubleshooting and validation. The conditions and procedures for conducting an array experiment vary between laboratories; however, most websites provide detailed protocols (see Table 1). Figure 2 shows a schematic of the array process, and the important aspects of the sequential procedures involved are dealt with in the following sections.

2.2. The array

There are two main types of array: (a) those where oligonucleotides are synthesised on the array *in situ* using photolithographic or other techniques [23,24], and (b) arrays where nucleic acids (PCR products, plasmids, or oligonucleotides) are robotically deposited onto the solid support. Initially, the number of elements that could be arrayed was limited; however, major technological advances over the last 3–4 years now allow on the order of 100,000 oligonucleotides or 10,000 PCR products to be arrayed per $cm^{2.}$ Arrays of spotted nucleic acids generally have low nanogram (usually 5–15 ng) quantities of nucleic acid per spot and are spaced at 100- to 300-μm distances, while arrays of oligonucleotides synthesised *in situ* have approximately 10^7 oligonucleotides/spot [4].

Another key aspect of array fabrication is selection of the sequences to be arrayed. In the case of organisms such as *S. cerevisiae* or *Caenorhabditis elegans*, where all the genes encoded by their genomes have been identified, it is possible to generate an array encompassing every ORF. The human

Table 1
Useful genome and microarray websites

Website	Location or other detail
http://www.sanger.ac.uk/	Sanger Centre
http://genome.ucsc.edu/	Human genome annotation
http://www.ensembl.org/	Human genome annotation
http://www.ncbi.nlm.nih.gov/UniGene/	UniGene gene clustering
http://www.tigr.org/tdb/hgi/index	The Institure for Genomic Research
http://www.ebi.ac.uk/	European Bioinformatic Institute
http://www.discover.nci.nih.gov/textmining/filters	Medminer
http://www.ncbi.nlm.nih.gov/Omim/	Online Mendelian Inheritance in Man
http://www.genome.ad.jp/kegg/	KEGG: Kyoto Encyclopedia of Genes and Genomes
http://www.hmorf.com	Gene expression and annotation
http://www.icr.ac.uk/array/array	Institute of Cancer Research, UK
http://www.nhgri.nih.gov/DIR/Microarray/main	Cancer Genetics Branch, National Human Genome Research Institute
http://cmgm.stanford.edu/pbrown/	Brown lab
http://genome-www4.stanford.edu/MicroArray/SMD/	Stanford microarray database
http://cmgm.stanford.edu/cgibin/cgiwrap/taebshin/dcforum/dcboard.cgi	Microarray discussion forum
http://rana.lbl.gov/	Eisen lab
http://www.ebi.ac.uk/microarray/	European Bioinformatic Institute - microarray site
http://www.ncbi.nlm.nih.gov/geo/	Gene expression omnibus
http://www.discover.nci.nih.gov/	Genomics and bioinformatics group, National Cancer Institute, USA
http://www.microarrays.org/	General information
http://www.umich.edu./~caparray/	University of Michigan-Comprehensive Cancer Center
http://w95vcl.neuro.chop.edu/vcheung/	The Genomics group at Children's Hospital, Philadelphia
http://sequence.aecom.yu.edu/bioinf/funcgenomic	Functional Genomics at the Department of Molecular Genetics, Albert Einstein College of Medicine, New York
http://www.mged.org	Microarray gene expression database group
http://www.affymetrix.com/	Affymetrix
http://www.resgen.com/	Research Genetics
http://www.ambion.com/	Ambion
http://www.clontech.com/	Clontech
http://www.rii.com/	Rosetta Inpharmatics Inc.
http://www.axon.com/	Axon Instruments Inc.
http://www.apbiotech.com/	Amersham Pharmacia
http://www.chem.agilent.com/	Agilent Technologies

See Ref. 95 for additional information.

genome project has not yet reached this stage, although the goal may not be far away. Therefore, construction of a human gene-specific array requires a choice to be made from a range of gene sequences (probes) selected from public databases, such as GenBank, dbEST, UniGene, or proprietary databases [25–27]. The UniGene database (http://www.ncbi.nlm.nih.gov/UniGene/) is an excellent starting point for identifying and choosing the DNA sequences to be arrayed. This database is an experimental system for automatically partitioning clone sequences into non-redundant sets of gene-oriented clusters that represent unique genes. It is a very useful gene discovery resource, as in addition to sequences of well-characterised genes, hundreds of thousands of novel ESTs corresponding to genes of unknown function are included. cDNAs from local sources can also be included or, for gene discovery projects, cDNA libraries of uncharacterised clones can be used. A good alternative starting point for assembling a list of clones for arraying is the 15K set (http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/) that comprises approximately 15,000 UniGene clusters.

Several factors govern the choice of clones to be arrayed. The presence within the clone of repeat sequences, such as *Alu* or *LINE* repeats, can influence the hybridisation signal and should be avoided. Another factor to be considered is that some ESTs exhibit weak UniGene clustering and, as the procedures for automated sequence clustering are still under development, the results may change from time to time as improvements are made. In some instances, clones will be reclustered from one gene cluster to another and in essence change identity. This effect will become less of an issue as all the genes encoded by the human genome are definitively identified (estimated completion 2004). In the interim, it is necessary to update putative gene identities on a regular basis. Alternatively, mapping clone sequences directly onto the human genome sequence using the Ensembl package at European Bioinformatic Institute (http://ensembl.ebi.ac.uk/) or searching the Institute for Genomic Research human gene index (http://www.tigr.org/tdb/hgi/), which assembles ESTs into tentative human consensus sequences, can sometimes identify conflicts or clustering artifacts. There has also been some evidence of discrepancy between the actual and designated sequences of some clones and also cases of clones being mixed. Mixed clones can be detected and

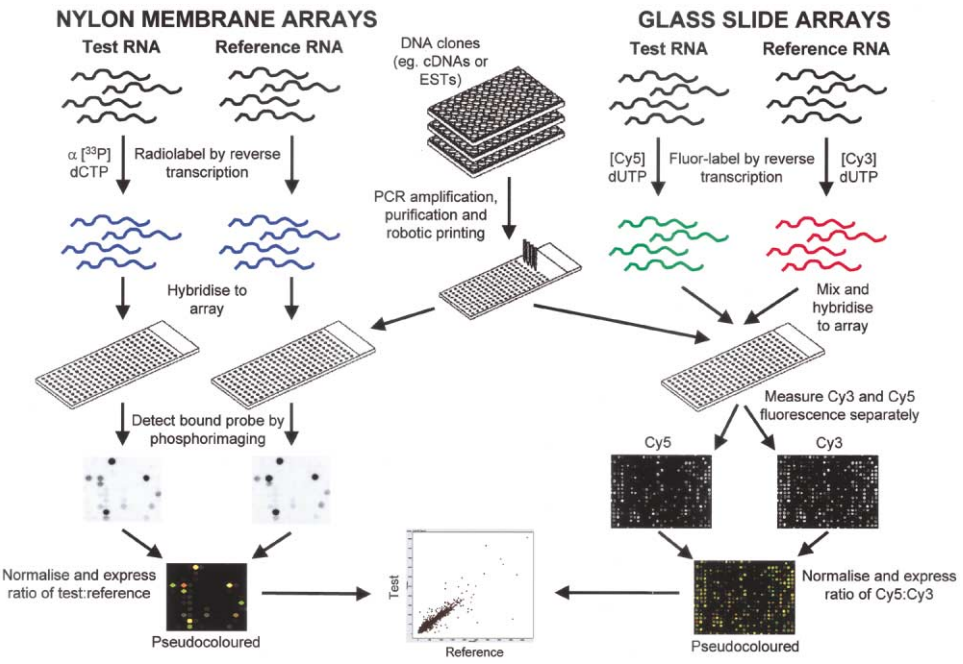

Fig. 2. Schematic of the various steps in a microarray experiment. Plasmid clones are propagated in bacteria, and the cloned inserts are amplified by PCR and then purified. The purified PCR products are then robotically printed onto glass or nylon solid supports. Modifications of this approach include the use of oligonucleotides instead of PCR products or the *in situ* synthesis of oligonucleotides directly onto the glass support using photolithographic or other techniques. Separate nylon based arrays are hybridised with ^{33}P-radiolabeled cDNA prepared from the test and reference sample, whereas glass slide arrays are hybridised simultaneously with Cy5 and Cy3 fluorescently labeled test and reference samples, respectively. Following stringency washes, hybridisation to nylon arrays is detected by phosphorimaging. Hybridisation to glass slides is detected by excitation of the two fluors at the relevant wavelength and the fluorescent emission collected with a charge-coupled device. The test and reference images are overlaid using specialist software and can be displayed in a number of ways, including as a scatter plot of the ratio of test:reference gene expression. Figure modified from Ref. [18].

avoided by including a quality control step using gel electrophoresis to exclude PCR reactions that yield multiple products. Most clones available for arraying are now sequence validated, but some discrepancies have been reported with sequence-verified mouse clones [28]. These potential problems reinforce the need to validate important microarray observations using complementary methods.

Another important decision is the choice of support on which to array the probes. Nucleic acids can be arrayed on a solid support such as nitrocellulose, charged nylon, or glass. The kinetics of duplex formation can be complicated by diffusion of solvent or solutes into and out of pores and by local effects within a pore. In this respect glass has some practical advantages, as it is non-porous, and the rigidity of glass makes gridding and reading easier. In the future these factors may allow miniaturisation and automation by incorporation into flow cells [29]. As mentioned earlier, glass slides allow test and control samples to be compared directly on the same array, whereas nylon membranes require parallel processing of samples on separate arrays. On the other hand, arrays on glass slides are restricted to a single use, whereas nylon arrays are generally probed with radiolabeled cDNAs and, with care, can be used a number of times, a consideration that may be important in smaller institutions or laboratories where access to core facilities is restricted or when resources are limited.

Oligonucleotides synthesised *in situ* use synthetic linkers attached to the surface of the glass arrays, while nucleic acids can be gridded onto charged nylon membranes or glass slides coated with poly-L-lysine or aminosilane. The gridding can be performed in batches using custom-built robotic arrayers (for example see http://cmgm.stanford.edu/pbrown/mguide/index.html; http://sequence.aecom.yu.edu/bioinf/funcgenomic.html; http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/aboutreader.html). Alternatively, gridding can be carried out using commercially available robotic systems. The clones, PCR products, or oligonucleotides can be printed in various solvents ranging from high salt solutions such as 3× saline–sodium citrate buffer to solvents such as 50% DMSO. After gridding, the array is briefly UV-irradiated to cross-link DNA thymines to the amine group on the support and then denatured, although this is unnecessary for DNA gridded in DMSO.

2.3. Sample preparation and labeling

The arrays are probed with labeled representations of the cellular mRNA pool that are prepared from the biological samples of interest, such as tissue culture cells or tissue biopsies. In tissue culture experiments, it is clear that the condition of the cells is absolutely critical as subtle growth differences can greatly influence the expression profile. Hughes *et al.* [30] identified a number of transcripts that exhibited inherent fluctuation in isogenic untreated yeast

cultures that appeared to have no phenotypic differences. These fluctuations represent biological noise. In parallel experiments with yeast mutant strains that lacked a growth defect, these fluctuations accounted for virtually all of the > 2-fold changes in gene expression that were detected. Thus, it is critical that culture conditions should be standardised as far as possible. It is also essential that all the appropriate controls are included, for example when comparing the effects of drug treatment regimens. These include vehicle controls and, ideally, the use of inactive drug analogues. Appropriate control vectors and transfection controls should be included during gene transfer experiments. In addition, sample handling must also be standardised to avoid artifacts resulting from prolonged manipulation. The effects on gene expression of infective agents, such as mycoplasma that can frequently contaminate cell cultures, are unknown, and appropriate screening should be performed regularly. Hughes *et al.* [31] noted that aneuploidy in some yeast strains resulted in spurious gene expression correlations due to gene dosage effects from an altered gene copy number. This factor could be important in the continuous passaging of tumor cells that commonly have unstable genomes. The influence of tumor cell aneuploidy on gene expression profile is unknown, but this creates the potential for variation when the same cell line is compared between different experiments and laboratories. This is especially relevant to the aim of establishing worldwide databases for comparison of gene expression profiles generated in different institutions. However, it is possible to use the same microarray for comparative genomic hybridisation, using labeled genomic DNA rather than cDNA, and this may provide the opportunity for an internal standard or control for gene copy number [32].

Expression profiling of clinical biopsy material also has a number of potential pitfalls. Biopsy material is often limited in availability, is generally not homogeneous, and, depending on the biopsy type, may contain different cell populations. In the case of tumor biopsies, there will be underlying gene expression patterns from normal tissue, including components such as surrounding normal tissue stroma, immune elements, and vasculature [33]. However, this can be taken into account by comparison of gene expression patterns with those in relevant cell lines that have transcriptome profiles characteristic of muscle, vasculature, or stroma [33,34]. An alternative is to dissect out the tumor cells, for example by using laser capture microdissection, although this can be a time-consuming process [35–37]. An additional factor to be considered is the potential for tissue handling effects. A number of genes, including c-*fos* and *junB*, have been reported to be induced by prolonged handling of the sample [33]. Thus, the time between the biopsy being taken and snap-freezing in liquid nitrogen or disruption in lysis buffer may also be critical.

The labeled representations of cellular mRNA are generated using reverse transcriptase, an enzyme that generates a single-strand DNA copy of each RNA. Generally, [α^{33}P] dNTPs are used for radiolabeling protocols, and fluorescent Cy dNTPs, namely Cy3-dNTPs and Cy5-dNTPs, are used together in the case of fluorescent methodologies. Other less common labeling strategies include the use of psoralen-biotin derivatives or ligation to an RNA molecule carrying biotin that can subsequently be detected with labeled streptavidin. Both total cellular RNA and poly(A)$^+$ mRNA can be used for array experiments. Reverse transcriptase requires an oligonucleotide primer to initiate the labeling reaction. There are a number of options: (a) oligo(dT) primers that bind the poly(A)$^+$ tail of mRNA; (b) random hexamer oligonucleotides that will bind anywhere on the RNA; or (c) a series of gene-specific or semi-degenerate primers that are specific to the arrayed set of clones [38].

The quality of the RNA preparation is critical for efficient labeling and optimal signal. This is especially true for glass slide arrays probed with fluorescently labeled cDNA, as these can be very susceptible to background fluorescence induced by non-specific binding to carbohydrate, lipid, or protein contaminants that can co-purify with RNA. In addition, the methodology of RNA extraction can influence the proportion of different RNA species isolated, a factor that may affect labeling efficiency when comparing different methodologies [39]. Another limitation of the current labeling techniques is that they require relatively large amounts of RNA. Fluorescent labeling requires 20–100 μg of total RNA or 1–2 μg of poly(A)$^+$ mRNA, while radioactive labeling is more sensitive but still generally requires 2–10 μg of total RNA or 0.1–0.5 μg of poly(A)$^+$ mRNA. These requirements potentially make microarrays incompatible with the very limited RNA yields obtained from certain types of biopsy or microdissected samples. Sgroi *et al.* [35] have found that using two radionucleotides and labeling both the first- and second-strand cDNA synthesis steps were sufficient to detect gene expression in 100,000 microdissected breast tumor cells. An alternative is to develop an efficient and reproducible method for RNA amplification. One approach uses an RT–PCR based method that can be performed on single cells; this method can be efficient and reproducible, but suffers from the observation that cDNA abundance does not correlate with original mRNA levels [4]. An alternative is to use a linear amplification method based on cDNA synthesis by reverse transcription from primers that are tagged with a bacteriophage RNA polymerase promoter sequence [37]. The cDNA is amplified by *in vitro* transcription from the incorporated bacteriophage promoter sequence, using the appropriate bacteriophage RNA polymerase, and the linearly amplified RNA is then labeled and used for microarray analysis. This approach has been reported to be reproducible and allows the use of as few as 500-1000 cells or a little as 1–50 ng of total RNA; however, below 100 pg, the complexity of the sample decreases as low copy number RNAs are lost [21].

2.4. Hybridisation and image acquisition

Microarrays exploit sequence complementarily, a remarkable feature of nucleic acids that allows them to as-

semble into duplex in a process that is reversible but occurs with absolute fidelity [4]. In effect, each labeled cDNA searches out and pairs with its complementary sequence on the array. The rules of recognition and the major factors that govern them, i.e. base composition, temperature, concentration of monovalent and divalent ions, and sequence complexity, are reasonably well understood. The signal obtained following hybridisation to the arrays gives both a measure of the number of molecules bound and also their identity. The labeled cDNA is generally purified to remove unincorporated nucleotides, salt, detergents, PCR primers, proteins, and RNA template prior to hybridisation to the array. Prehybridisation, hybridisation, and stringency wash conditions are similar to those frequently used for Southern or northern blotting. Blocking reagents are used to prevent non-specific interactions. Those commonly used include Denhardt's solution, SDS, sonicated salmon sperm DNA, tRNA, COT_1 DNA, and poly(A) oligonucleotides. The potential for background fluorescence when using glass slide arrays necessitates the extra precautions of filtering all solutions and taking care that hybridisation buffer components, such as SDS or urea, do not precipitate out while setting up the hybridisation or washing the arrays. Hybridisation of radiolabeled cDNAs can be detected by autoradiography and densitometry or more preferably by phosphorimaging using commercially available phosphorimagers. The hybridisation of fluorescently labeled cDNAs is detected using laser scanners that excite the fluorescent dyes and collect their emission at the relevant wavelengths.

2.5. Image and data analysis

A number of steps are necessary to obtain gene expression data following acquisition of an array image. The first is to correctly identify the spots on the array. Most commercial readers or arrayers provide software for this step. In addition, several public sites provide unsupported software also developed for this purpose (e.g. http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/). Spot identification requires overlaying and aligning of a grid specifying spot location onto the array image, a process that generally necessitates considerable operator input. Having located the spots, the background signal has to be calculated and subtracted from the hybridisation signal. This is done using algorithms that predict the expected position, size, and shape of the spot and then calculate local background in the vicinity of each spot. The solid format of glass has some particular advantages over nylon membranes in this respect, as nylon membranes can exhibit some degree of creasing or deformation that can confound grid alignment. The nature of radioactive decay also makes identification of the border of the spot and calculation of background difficult; this is because within the spot there is smooth graduation from high signal to background signal, as compared with the sharp defined spot resulting from a fluorescent signal. An additional problem with radiolabeled probes is the effect of 'blossoming', where a high hybridisation signal encroaches on a neighboring spot and influences the signal detected from that spot. This effect can be reduced by using ^{33}P-labeled probes rather than ^{32}P. However, one advantage of nylon membranes is that they are less prone to background effects caused by contaminants within the probe, precipitation of components of the hybridisation buffer, and also dust or irregularities on the slide surface that can effect fluorescent signals. These points also emphasise the fact that for all array types there is a need for operator interaction with the data, in effect to curate the data by flagging up bad spots. This curation process is currently one of the rate-limiting steps in the acquisition of array data.

Comparison of data from multiple arrays or multiple samples on a single array requires the data to be normalised. This can be achieved using two different approaches. One strategy is to normalise to a set of genes that do not vary under the experimental conditions of choice (e.g. the 90 control genes described by DeRisi *et al.* [40]). Alternatively, one can make the assumption that, under the conditions studied, the expression of the majority of genes will remain unchanged, allowing normalisation by comparison with global gene expression. This approach works well for closely related samples, but obviously yields poorer normalisation as the samples diverge and the differences in gene expression increase. Another option is to 'spike' each RNA sample with an equal amount of an internal standard RNA. For arrays using both fluorescent and radiolabeled probes there are commercially and publicly available software packages that can overlay array images, calculate background, normalise the data, and produce an output of absolute expression or ratio of expression comparing test sample to the reference sample.

Having obtained absolute values or ratios of gene expression, it is necessary to establish a database that allows the management and comparison of the information obtained from multiple experiments. The aim is to take the raw hybridisation data and simplify this information down to a table of gene/clone identity, expression value, or ratio of test:reference, and then to integrate this information with databases that contain genomic data, functional information, or literature references [41]. Database design is important as large quantities of information have to be managed before (data pertinent to clone location, identity, and array fabrication conditions) and after (data on experimenter, experimental aims and conditions, sample description, labeling conditions, raw hybridisation images, intensities, ratios, and background) carrying out the experiment [42]. It is becoming increasingly clear that biological function generally results from complex interactions between many components; for example, transcriptional regulation is more sophisticated than the traditional view of gene expression being a simple on–off event. In addition, biological systems generally incorporate features such as feedback, feed-forward, error checking, and redundancy [43,44]. Therefore, the acquisition of sufficiently large datasets is essential to address complex biological systems or genome-wide function.

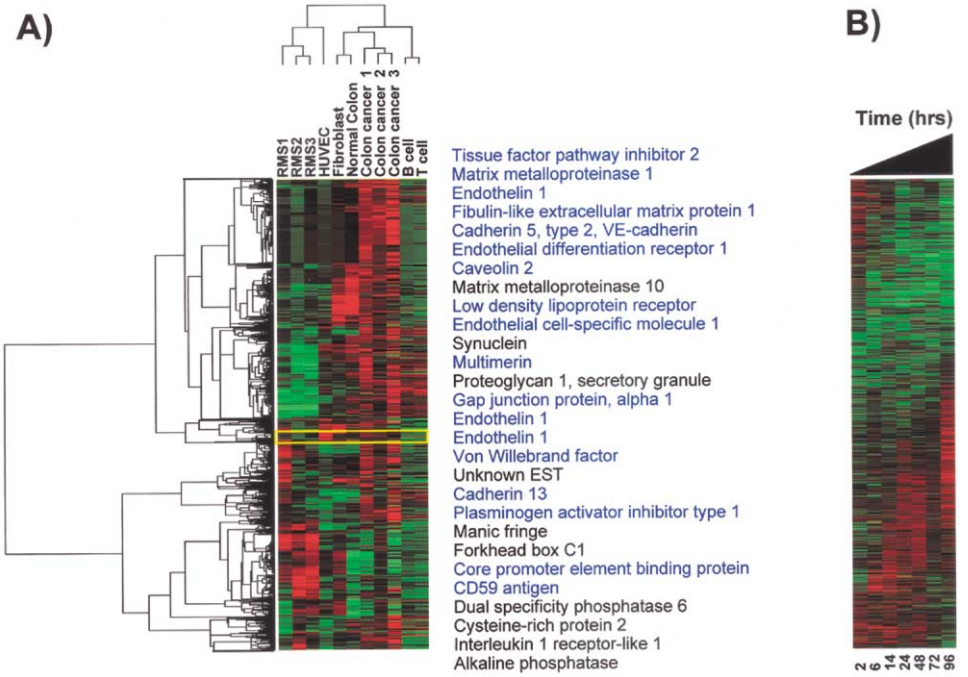

Fig. 3. Examples of hierarchical clustering of microarray data. (A) mRNA isolated from human endothelial cells (HUVEC), normal fibroblasts, B cells, T cells, a normal bowel biopsy, three rhabdomyosarcoma lines, and three colon cell lines was profiled against a nylon membrane array carrying 4132 clones (Research Genetics). Images were analysed using proprietary software to generate a gene list with raw hybridisation data and other parameters. Background signal was removed, and the signals were normalised using all the data points that were universally expressed in all cell types. For each cell type the expression data were expressed as a ratio relative to the median. The ratios were converted to a logarithmic value, the data were then arranged using a self-organising map, and analysis carried out by hierarchical clustering [45]. The figure shows a typical output: a green signal represents a gene whose expression is below the median, a red signal indicates a gene whose expression is above the median, and a black signal indicates no difference from the median. The intensity of the colour is proportional to the degree of difference from the median. The yellow box highlights a cluster of genes whose expression is increased in endothelial cells, the text indicates the identity of genes within that cluster, and the blue text identifies genes whose products have a specific role in endothelial cells or whose expression is restricted to or increased in endothelial cells. (B) HCT116 colon adenocarcinoma cells were treated with 100 μM LY294002 (a phosphatidylinositide-3-kinase inhibitor). mRNA was prepared from cells isolated at the time points indicated, labeled with Cy5-dUTP and hybridised in parallel with Cy3-dUTP-labeled mRNA isolated from control HCT116 cells to a glass slide array carrying 5808 clones (http://www.icr.ac.uk/array/array.html). After the stringency washes, the slides were scanned, and a list of ratios of test:reference was obtained. Data were analysed using the clustering algorithm of Eisen *et al.* [45]. The output demonstrates the time-dependence of gene expression regulation following exposure to LY294002.

Once a suitable dataset has been established, the process of mining the data for meaningful information begins [41–43]. The easiest form of analysis is simply to list all of the genes that differ in expression level, for example between the test sample and the control. Although useful, this approach does not uncover the large amounts of complex information contained within the data. More sophisticated analysis involves the identification of non-random groups of genes associated with a particular biological situation, so that these can be examined by further approaches. Multidimensional scaling in two or three dimensions of Euclidean distances allows the assessment of the approximate degree of similarity in gene expression between samples. Another common approach is to look at multiple experiments and to arrange or cluster the expression data into small homogeneous groups. This can be done manually with small datasets where identification of the extremes between two samples or over a short time course is relatively straightforward. However, this approach fails to extract all the potential information in genome-scale experiments with hundreds or thousands of samples. Therefore, mathematical approaches that essentially organise the data by grouping genes with similar expression patterns have been developed.

One strategy is to use hierarchical clustering, an approach commonly used in sequence/phylogeny experiments, coupled with outputs that facilitate visual examination of the data (Fig. 3) [45]. Several studies, especially in yeast where greater than 50% of the ORFs have an ascribed function, have noted co-regulation of the expression of genes encoding proteins from a common biochemical pathway, that are functionally related, or that form multi-protein complexes. This has led to the basic proposal that genes with similar expression profiles and behavior are likely to be co-regulated and functionally related [4]. Thus, it is possible to cluster the data from a given set of conditions or cell types and use a ‘guilt by association’ strategy to identify functional clusters [44]. In this way, the function of novel or known genes can be predicted, and tentative function can then be tested rigorously by conventional biochemical approaches. In addition, co-regulation may also allow the identification of common regulatory elements within the promoter sequences of genes that cluster together. Although

this strategy lacks some logical rigor, the approach can be made more systematic by incorporating statistical analysis of the probability of a particular cluster forming. The application of this type of clustering analysis has major implications for cancer biology and drug development. For example, this approach may identify groups of genes associated with drug sensitivity/resistance or genes that influence outcome or disease progression/metastasis. However, there are some limitations to this approach, as hierarchical clustering requires well-separated data to give unambiguous results. For such well-separated data, with clear-cut differences in gene expression, most clustering algorithms will generally give the same result; however, less well-separated data will give variable results, and this makes interpretation difficult [43]. Another mathematical approach commonly used for data mining is the application of self-organising maps, and this approach has also been applied successfully to gene expression data [46].

An additional limitation in the analysis of microarray data is the ability to measure small changes in gene expression. Most investigators typically apply an arbitrary global threshold of a 2- to 4-fold change for differences in expression that might be considered biologically interesting or significant. As described earlier, Hughes *et al.* [30] noted that in a large number of control experiments, many genes fluctuated, possibly due to biological noise within the system. They calculated a scaling factor that accounted for such variation; this could then be used to correct experimental data and allow the identification of changes below the commonly used 2-fold threshold. This study suggests that conditions that are apparently identical to the experimenter may not necessarily be identical from the cell's perspective, and thus scaling factors for biological noise or natural variation may become a necessity for all array experiments.

3. Applications of microarrays in cancer biology, pharmacology, and drug development

If the transcriptional profile provides a molecular phenotype of the specific state of the cell, it is reasonable to propose that the transcriptional response could be used to measure a change in cellular state. Such changes could result from a normal cellular process, a disease, or a response to genetic, chemical, or therapeutic perturbation. The ability to obtain quantitative information from a transcriptional profile would thus be an exceptionally powerful means to explore basic biology, diagnose disease, facilitate drug development, tailor therapeutics to specific pathologies or genetic profile, and also to generate databases relevant to biological or therapeutic processes and pathways [44]. As mentioned in the introduction, gene expression profiling using microarrays can provide major benefits across the whole of basic cancer research, as well as in the diagnosis and treatment of the disease [14,47,48]. In the following sections, we review the progress and potential of gene expression microarray technology with respect to the basic biology and therapy of cancer.

3.1. Basic biology of cancer

The mechanisms that cells employ to regulate normal processes, such as proliferative signal transduction and cell division, are frequently subverted during tumorigenesis. Changes in gene expression have been associated with many key cellular processes including those required for tumor growth and survival [6]. Therefore, gene expression profiling may identify genes required for these cellular processes. In the context of cancer and drug development, the application of microarray analysis may identify genes whose products are involved in tumorigenesis and malignant progression and hence help to identify potential drug targets. In addition, it is clear that the genetic basis of cancer is multi-factorial and that some oncogenes operate in a non-dominant, cooperative manner that is only apparent in certain combinations and cell backgrounds. The contribution of these genes may remain silent when examined individually and will only be revealed when gene expression is analysed on a global scale [48].

Cell cycle-dependent mRNA fluctuation has been described in numerous settings and was initially studied with gene expression profiling by Cho *et al.* [49] in *S. cerevisiae.* Temperature-sensitive *cdc28* and *cdc15* mutants arrest in the G1 and G2 phases of the cell cycle when incubated at the restrictive temperature. Gene expression was profiled at regular intervals over two cell cycles following release of these mutants from the cell cycle block by incubation at the permissive temperature. Of the 6218 genes encoded by the *S. cerevisiae* genome, 1348 genes fluctuated by more than 2-fold in both mutants during the time course; of these 416 genes exhibited a periodic fluctuation consistent with cell cycle phase-specific expression. The expression patterns were tight enough to distinguish between early and late events within the different phases, e.g. between events in early and late G1. Consistent with the literature, cell cycle-dependent changes were observed for the cyclin family of genes, which ultimately regulate transcription factors, and also for genes involved in DNA replication and packaging. Genes encoding constituents of a given protein complex were generally co-regulated. More than 60% of the genes detected had been implicated previously in the cell cycle, suggesting that measuring RNA can indeed be an indicator of biological function. However, on a cautionary note, it was observed that of the cell cycle genes known to be involved in budding, DNA replication, and other cell cycle roles, 75% did not exhibit fluctuation at the RNA level. This was possibly because some of these gene products were modulated at the post-transcriptional level. The data from this study were subjected subsequently to a more sophisticated analysis [50]. The most variable 3000 ORFs were arranged using a K means algorithm into 30 gene clusters that were particularly enriched for genes with similar function. These included clusters of genes, exhibiting periodic fluctuation, that encoded cell cycle phase-specific proteins

involved in cell cycle regulation, budding, cell polarity, and centrosome organisation.

In addition, the distribution of regulatory sequences upstream of the gene promoter was analysed in both of the yeast studies described above [50,51]. A number of gene clusters had selective upstream motifs; for example, 55 of the 101 genes that clustered for polarity and budding had an early cell cycle box sequence motif, which occurred with an incidence of less than 4% in the other gene clusters. Additional analysis also revealed the potential presence of two novel regulatory sequence motifs. Half of the gene clusters were enriched for functional categories or for a specific upstream motif. Clusters with functional relationships were statistically tighter than those that contained genes that appeared to be biologically unrelated, and those with related regulatory motifs were also more tightly clustered. One important observation was the identification of previously uncharacterised ORFs that exhibited periodic fluctuation, and these were assigned to clusters containing known genes with related function. This work suggested that the function of unknown genes could be predicted by examining their clustering pattern, an observation that will be discussed later in this review.

Transcription factors are frequently targets of many signaling pathways, and a number of studies have profiled gene expression following responses mediated by external signals. The expression of 8600 distinct human genes has been profiled by microarray analysis at regular intervals following the stimulation of normal human diploid fibroblasts by serum [51]. The response to serum was rapid and widespread. In all, the expression of 517 genes was altered by more than 2.2-fold, with genes such as c-*fos, junB,* and *MAPK phosphatase* being induced within 15 min of exposure to serum. The timing of gene expression was coincident with progression through the cell cycle, and distinct clusters of genes involved in cell division were identified. A less expected finding was the observation that genes involved in wound healing were also induced, although in hindsight this may not be that surprising given the role of fibroblasts and serum in that process. The results suggested that fibroblasts, which are not normally in contact with serum, are programmed to respond abruptly to serum and to orchestrate the healing response by promoting chemotaxis and differentiation of the various cell types involved in the immune response and angiogenesis.

In another study of signal transduction, the induction of immediate early genes by receptor tyrosine kinase-activated signaling pathways was examined using receptors which lacked key binding sites for the intracellular adapter molecules that are required for the regulation of different signaling cascades [52]. This study used NIH3T3 cells transfected with a gene encoding a fusion protein, consisting of the wild-type or mutated cytoplasmic portion of β-platelet-derived growth factor receptor fused to the extracellular portion of the macrophage colony-stimulating receptor. High-density oligonucleotide arrays were used to measure the expression of 5938 genes following exposure of NIH3T3 cells expressing the fused receptor to macrophage colony-stimulating factor. Sixty-six genes were induced by more than 3-fold following an exposure for a few hours to macrophage colony-stimulating factor, of which 50% had been identified previously in the literature as immediate early genes induced by receptor tyrosine kinase activation. The mutant form of the β-platelet-derived growth factor receptor that lacks binding sites for phospholipase C-γ1, phosphoinositide-3-kinase, shp2, and rasgap could still induce these genes, but to a slightly lower level than the wild-type receptor. Removal of an additional site that was bound by grb2 resulted in a greatly reduced induction, but did not eliminate completely the induction of immediate early genes. These observations suggested that the distinct pathways emanating from the β-platelet-derived growth factor receptor exhibited a certain degree of functional redundancy for the induction of immediate early genes and also indicated that diverse pathways could exert overlapping effects on the induction of these genes. However, there was some evidence for specificity, as restoring a rasgap binding site on the β-platelet-derived growth factor receptor resulted in the induction of genes normally regulated by interferon-γ, although the physiological significance of this was unclear. An additional set of experiments compared gene expression following the stimulation of NIH3T3 cells with either fibroblast growth factor or platelet-derived growth factor. Again the same pattern of immediate early gene induction was detected. These observations suggested considerable overlap and cross-talk in the regulation of gene expression by different signaling pathways. However, subtle differences may potentially have been missed as high, non-physiological concentrations of macrophage colony-stimulating factor were employed, as well as a relatively high 3-fold cut-off for the gene expression changes.

Other studies have used microarray analysis to examine the downstream effectors of signaling pathways. Roberts *et al.* [53] profiled gene expression in *S. cerevisiae* during the pheromone response of wild-type and mutants of the various mitogen-activated protein kinases involved in a number of different signal transduction pathways. They found subsets of co-expressed genes that reflected the activity, cross-talk, and overlap of the different signaling pathways regulated by specific mitogen-activated protein kinases, particularly two distinct mitogen-activated protein kinase mutants that revealed overlap between filamentous growth and mating responses. Guo *et al.* [54] used c-*myc* null and wild-type rat fibroblasts to identify c-myc responsive genes. Myc was found to regulate genes involved in protein synthesis and metabolism, suggesting that its role may be to prepare the synthetic apparatus for the demands of proliferation. The effects of one of the immediate early genes, *egr1,* a regulator of transcription that is frequently overexpressed in prostate tumors, were examined using an adenoviral-mediated expression of *egr1* in a prostate cancer cell line [55]. The expression of a number of genes encoding proteins involved in neuroendocrine differentiation, as well

as several growth factors (platelet-derived growth factor-A, insulin-like growth factor-II, and transforming growth factor β1), was increased following *egr1* expression. These observations suggested an early role for *egr1* in prostate malignancies, although further validation is required.

The studies described thus far in this section illustrate the potential power of gene expression profiling by microarrays and exemplify the concept of co-regulation of banks of functionally related genes. As mentioned, genes whose products have similar cellular functions, those that encode products which form large complexes and those which operate on the same biochemical pathway frequently exhibit co-regulation of expression. Observations of this type of relationship can also be used to generate testable hypotheses when previously uncharacterised genes cluster with genes of a specific function or biochemical pathway. In addition, microarray expression profiling has also revealed the potential interaction between distinct signaling pathways, has identified potential regulatory sequence motifs upstream of gene promoters, and has provided a broader view of genome structure and function than would have been possible with conventional methodology.

Among the first microarray studies directly relevant to cancer biology were those of Schena *et al.* [56] and DeRisi *et al.* [40]. Schena *et al.* examined heat shock and phorbol ester treatment of a leukaemic T cell line using a 1065 element array and found a number of changes that would be expected following those stimuli. DeRisi *et al.* used a 1161 element array to examine the effects of reintroducing chromosome 6 into a melanoma cell line that lacked this chromosome. They detected significant changes in the expression of 78 genes, of which 16 were re-analysed and corroborated by northern blotting. One of the first attempts to classify cancer cell lines using microarrays was reported by Khan *et al.* [57]. They chose to examine alveolar rhabdomyosarcoma as those tumors are relatively uniform genetically; 70–80% have a translocation that creates a novel oncogenic transcription factor by fusing the *pax3* or *pax7* gene with *fkhr*. Also important was the fact that these tumors are aggressive soft tissue tumors that are frequently difficult to classify. A set of cell lines established from alveolar rhabdomyosarcoma tumors were profiled with an array of 1238 genes and compared to a reference sample from a diploid foetal myofibroblast cell line. The data were analysed using multidimensional scaling that clustered the alveolar rhabdomyosarcoma cell lines together when compared with Ewings sarcoma, melanoma, prostate, and breast cancer cell lines. The genes associated with the alveolar rhabdomyosarcoma cluster showed a consistent pattern of expression of 37 genes, including those known to be deregulated in alveolar rhabdomyosarcoma, e.g. *pax3-fkhr* and *cdk4*. These observations were followed up with a subsequent study transfecting *pax3* and *pax3-fkhr* into NIH3T3 cells [58]. *Pax3* induced the expression of only a single gene and repressed the expression of three additional genes; in contrast, the oncogenic fusion protein induced the expression of a number of genes, including a high proportion of muscle-specific genes, some of which were confirmed by northern blotting in several alveolar rhabdomyosarcoma cell lines.

A smaller study of prostate cancer used nylon membrane arrays to compare hormone-sensitive and hormone-insensitive cell lines grown as xenograft tumors in mice [59]. Several genes were induced in the hormone-insensitive tumors and two, *igfbpP2* and *hsp27*, were followed-up at the protein level by immunohistochemistry, using an array of tissue biopsies from 238 prostate tumors and 26 benign prostate tissues. Igfbp2 was expressed in all recurrent hormone refractory tumors, some primary tumors, and none of the benign tumors. Hsp27 had a similar pattern, but was less widely expressed. However, there was no statistical association between igfbp2 or hsp27 expression and tumor stage, nor did expression influence response to treatment. Another study used a similar strategy to identify a potential marker for anaplastic large-cell lymphoma [60]. Thirty-one hematopoietic cell lines were examined with a 588-element nylon membrane array. The expression of one gene, *clusterin*, was found to be restricted to the anaplastic-large cell lymphoma cell lines. This was confirmed by western blotting and immunohistochemical analysis of 198 primary lymphoma biopsies representing most major lymphoma subtypes. With two exceptions, none of the non-anaplastic large cell lymphomas expressed clusterin; in contrast, all 36 of the anaplastic large cell lymphomas expressed clusterin. The authors concluded that although the function of clusterin in the disease is unknown, it may have potential as a diagnostic marker for anaplastic large cell lymphoma. Another study of two clonally related T cell lymphoma cell lines derived from the same patient at different stages of tumor progression detected changes in genes involved in signal transduction, transcription, adhesion, proliferation, and cell death [61]. Of particular interest was the observation that expression of the gene encoding bleomycin hydrolase was increased with progression, despite the fact that the patient had not been exposed to bleomycin; this suggested that resistance genes are not always induced by exposure to chemotherapy, but can also be induced by tumor progression.

In an attempt to understand the mechanisms of metastasis, Clark *et al.* [62] profiled melanoma gene expression using a mouse xenograft model of tumor metastasis. Human and mouse cell lines were profiled using an array of 7070 human and 6347 mouse genes. Three genes, *fibronectin*, *rhoC*, and *thymosin β4*, were expressed at higher levels in the metastases of the two cell lines. Fibronectin is an extracellular glycoprotein that is a ligand for the integrin family, rhoC is a GTPase that regulates cytoskeletal organisation in response to extracellular factors, and thymosin β4 is an actin-sequestering protein that regulates actin polymerisation. These genes have all been associated or correlated previously with tumor metastasis. Other genes that were up-regulated to a lesser extent included those whose products are associated with cytoskeletal organisation, such as α-catenin, α-actinin, and α-centractin, together with

genes encoding extracellular matrix components such as collagen $\alpha 2$ and $\alpha 1$, matrix Gla protein, fibromodulin, and biglycan. Having identified potential candidates involved in metastasis, the authors investigated the role of rhoC expression using a gene transfer strategy. Expression of rhoC enhanced metastases, while expression of a dominant negative rhoC suppressed metastases, implying that rhoC was important for tumor invasion.

The observations of the cancer microarray studies outlined above demonstrate that it may be possible to classify tumor cell lines by type using gene expression profiles. They also show that it is possible to associate particular genes with cancer types or biological processes such as metastasis. Again, the results generated provide testable hypotheses for further investigation. The studies discussed above were generally restricted to 500-8000 genes, the majority of which were characterised previously. The use of larger arrays, incorporating uncharacterised cDNAs or ESTs and coupled with the 'guilt by association' premise, may allow the identification of novel genes associated with particular cancer types, those genes critical for tumorigenesis and metastasis, and also those genes involved in the development of resistance or failure to respond to treatment. In addition, they may yield novel targets for drug discovery and development.

3.2. Drug discovery and development

Following the identification of a specific molecular target and the development of a series of small molecule inhibitors, it is necessary to confirm that such inhibitors do indeed act by the desired mechanism of action on their intended target. In addition, the early identification of potential on-target and off-target effects, and also of pharmacodynamic markers of these effects, is highly beneficial for the subsequent preclinical and clinical studies required for the development of anticancer agents [7,8,47]. The use of microarrays can also generate a database that could allow the mechanism of action of a given agent to be predicted from the changes in gene expression that they induce. Such databases also allow compounds likely to act by a given mechanism to be identified for further study. The application of array technology in this way has been reported recently in a number of studies, and these are discussed in detail below.

One study in particular is worth reviewing at some length, as it illustrates many of the important issues concerning the use of microarrays in pharmacological mechanism of action work. As described earlier, the growth inhibitory activity of approximately 70,000 compounds has been assessed using the US National Cancer Institute's panel of 60 human tumor cell lines, and the effects of some of these agents have been compared to the expression of 140 proteins [10]. This strategy has been extended recently to incorporate gene expression profiling by microarray [34, 63]. Ross *et al.* [34] profiled gene expression of the tumor cell line panel using an array consisting of 9703 elements that corresponded to 3700 previously characterised genes, 1900 identified by sequence similarity to genes encoded by other organisms, and also 2400 uncharacterised ESTs. The 1167 genes that showed the greatest variation and the 6831 genes that demonstrated a good, consistent hybridisation signal were subjected to hierarchical cluster analysis. This arranged the cell lines by their presumptive tissue of origin, the only exceptions being some non-small cell lung and breast cancer lines that were distributed in multiple branches. Characteristic clusters of genes distinguished each cell line type. For example, melanomas were defined by a 90-gene cluster that included genes whose products were specific to melanoma biology, such as *tyrosinase*, *dopachrome tautomerase*, and the *mart1* tumor surface antigen, although rather paradoxically two breast cell lines shared expression of these melanoma genes. Nevertheless, it was generally possible to find a molecular signature of the tissue from which the cell lines originated, implying that the physiological or experimental adaptation for growth in culture was not sufficient to overwrite the tissue-specific gene expression program. Cluster analysis identified a number of functionally related genes, for example those that were markers of proliferation, or were involved in progression through the cell cycle, regulation of protein synthesis, drug metabolism, or interferon response.

Having constructed this database of constitutive gene expression patterns, Scherf *et al.* [63] then analysed the data for a relationship to compound sensitivity. The gene expression data were related to the cytotoxicity following a 48-hr exposure to 70,000 compounds from the National Cancer Institute collection. The array data were filtered down to 1376 genes that clustered the cell lines by tissue of origin. Initially, the drug sensitivity data were clustered for a subset of 118 agents with relatively well-established mechanisms of action. The resulting clusters corresponded to the likely modes of action and were divided into DNA and DNA/RNA antimetabolites, tubulin inhibitors, DNA-damaging agents, topoisomerase I inhibitors, and topoisomerase II inhibitors. 5-Fluorouracil, an inhibitor of RNA and DNA syntheses, clustered with the RNA synthesis inhibitors, suggesting that in these cytotoxicity assays the main mechanism of this agent was inhibition of RNA synthesis.

The clustering pattern was altered when the 1376 genes were analysed in relation to the activity of the 118 drugs. The antimetabolite and alkylating agent clusters were changed such that the antimetabolites appeared in five distinct classes, while the alkylating agents were divided in two classes. The antifols, purine analogues, and pyrimidine analogues were separated into separate branches, whereas the alkylating agents separated into N-7 reactive nitrogen mustards and a group containing ethylamines, nitrosoureas, and alkyl alkane sulfonates. The antitubulin cluster was unchanged, whereas the topoisomerase inhibitors were rearranged in a manner that revealed mechanistic differences among the subclasses of compound. The topoisomerase I

inhibitors clustered by whether or not they required activation, and the topoisomerase II inhibitors separated into an anthracycline node and another node that contained mitoxantrone and the bioreductive agents. This suggested that the ability to produce double-strand DNA breaks might be an important feature of the latter. Etoposide clustered with the alkylating agents, implying that drug metabolism as opposed to direct mechanism of action also featured in this method of drug activity–gene expression correlation.

The same study also used a clustered image map to visualise the data and summarise the relationship between drug activity and gene expression. Two examples, 5-fluorouracil and L-asparaginase, were cited. Eighteen cell lines, including all seven colon cell lines, showed high sensitivity to 5-fluorouracil; of these, fourteen expressed low levels of mRNA encoding dihydropyrimidine dehydrogenase, an enzyme that catabolises 5-fluorouracil. Some cells, including acute lymphoblastic leukaemias, lack asparagine synthase and require exogenous L-asparagine. L-Asparaginase depletes extracellular L-asparagine and has been used in the treatment of acute lymphoblastic leukaemia. Comparison of L-asparaginase sensitivity and asparagine synthase gene expression revealed a moderate negative correlation for the entire cell line panel; however, the leukaemic subpanel gave a very high negative correlation coefficient. The data described in the above study are available in an open public database (http://dtp.nci.nih.gov/) that can be searched for a number of different parameters. For example, searching the database for a relationship between epidermal growth factor receptor expression and increased cytotoxicity returns the top five hits identified by the study of Wosikowski *et al.* cited earlier ([9]; Clarke PA, unpublished observations).

In some ways, the study of Scherf *et al.* [63] was analogous to gene expression profiling studies in clinical tumors that look for predictors of classification or outcome. However, there are several potential limitations associated with this study: (a) the cell lines have been selected for growth in culture and should really be considered surrogates for *in situ* tumors; (b) the database is generated from a single endpoint of growth delay at 48 hr, which is a measure of short-term growth inhibition and/or cytotoxicity; and (c) the relationship between drug activity and expression is correlative and not necessarily causal, as changes in gene expression following drug exposure were not measured. Most of the drugs, but not all, clustered by presumed mechanism of action. The exceptions may be due to experimental variability or the loss of information as a result of compressing 60 dimensions of drug activity across the cell lines into a single dimension. In addition, only 20–30% of the possible genes encoded by the human genome were examined; it is possible that clustering may be further improved by the inclusion of a greater number of genes or potentially more cell lines. Other effects, such as incorrect or incomplete assignment of drug mechanism, influences of transporters/efflux pumps, metabolizing enzymes and other sensitivity/resistance genes, and secondary off-target effects of the agent, will also influence clustering. It remains to be seen whether this database could be used to predict off-target as well as on-target effects.

The benefit of establishing databases, such as those described above, is that they allow subsequent comparison with data for novel compounds with unknown molecular mechanisms. This facility is already available for drug cytotoxicity and specific molecular target expression data using the COMPARE algorithm (http://dtp.nci.nih.gov/). The addition of global gene expression data should further enhance the power of this informatics tool. However, the next major goal will be the establishment of a database of expression profile changes that occur in response to treatment with anticancer agents or following drug target modulation by genetic means (see following section).

As described earlier, the entire gene complement of *S. cerevisiae* has been established, and its genome has been subjected to extensive mutational analysis. Therefore, *S. cerevisiae* may be an ideal starting point as a model in which to investigate the value of gene expression profiling in response to drug treatment. A number of studies have begun to address this strategy. Gray *et al.* [64] incorporated a yeast array study into the development of cyclin-dependent kinase inhibitors. Compounds based on a trisubstituted purine demonstrated some selectivity towards cdk2/cyclin complexes and inhibited the activity of two yeast cyclin-dependent kinases, cdc28p and pho85p. The gene expression response of yeast cells following exposure to an existing cyclin-dependent kinase inhibitor, flavopiridol, compound 52 (a trisubstituted purine cyclin-dependent kinase inhibitor), and 52Me, an inactive but closely related compound, was measured using high-density oligonucleotide arrays. Of the 6200 genes monitored, 132, 194, and 2 genes were altered by more than 2-fold following treatment with flavopiridol, compound 52, and compound 52Me, respectively. Of the genes that changed, 63 were altered by exposure to both compound 52 and flavopiridol; of these only 9 were down-regulated, but they included 5 genes associated with cell cycle progression. Changes consistent with inhibition of pho85p were also detected. In addition, increased expression of metabolic genes and genes whose products were associated with transport, cell wall synthesis, and stress response were apparent. The effects of genetically manipulating cyclin-dependent kinase activity were also tested by profiling the expression of a *cdc28p* temperature-sensitive mutant. At the restrictive temperature, 100 genes were altered by more than 2-fold. A number of these genes were also altered by the active small molecule inhibitors; however, another set of genes were altered only in the *cdc28p* mutant and were not affected by the cyclin-dependent kinase inhibitors. The inclusion of an inactive analogue in these experiments was important, as few changes were detected, suggesting that chemotype effects due to the chemical backbone were minimal, and strongly supporting the view that the gene expression changes detected with the active analogue 52 resulted directly from cyclin-dependent

kinase inhibition. Although there were some similarities in the gene expression profile between small molecule inhibition and genetic inhibition, the differences between the two strategies suggest that the small molecules may have additional off-target effects or that there may be an intrinsic methodological difference between chemical and genetic manipulation of a target. Dose–response effects may also be involved.

A comparison between manipulating a target by either genetic or small molecule approaches was also carried out by Marton *et al.* [65], who used various deletion mutants of *S. cerevisiae* to compare the effects of specific mutations with those of known inhibitors of the same pathway. The calcineurin-signaling pathway, which can be inhibited by FK506 or cyclosporin A, and inhibition of his3 by 3-aminotriazole, were used as model pathways. The gene expression changes induced by the inhibitors were examined in wild-type yeast and compared with the effects seen in mutant isogenic strains lacking the calcineurin catalytic subunits or *his3*. In both cases, there was a correlation between drug inhibition of the target and inactivation of the target by mutation. A more sophisticated approach using a mechanism 'decoder' strategy could also be applied. Initially, the expression profile of drug-treated wild-type cells is compared with the expression profile of a panel of mutant strains. The mutant strains with the most similar expression profile are then selected and treated with the drug. For a perfect drug with absolute specificity, treatment of a mutant strain that lacks the drug target should not alter the gene expression profile. However, in reality no drug shows perfect target specificity, so any changes in expression profile detected in the drug-treated mutant will give clues to off-target effects. Treatment with FK506 gave a similar profile to the calcineurin mutants; however, subsequent treatment of these mutant strains with FK506 altered gene expression in a manner consistent with an off-target effect. This profile corresponded to an effect dependent on the gcn4 transcriptional activator. The expression of these genes was unchanged following FK506 treatment of *gcn4*-null cells and confirmed the requirement for gcn4. An additional subset of genes encoding drug efflux pumps, similar to the multidrug resistance family of proteins, were still induced in the *gcn4*-null cells, suggesting a secondary off-target effect.

Hughes *et al.* [30] extended this approach using a 'compendium' of expression profiles to demonstrate that the gene expression profile of a mutation successfully serves as a molecular phenotype that corresponds, in turn, to a phenotype defined by conventional biochemical assays. They achieved this by creating a reference database of expression profiles from 300 full genome expression profile experiments using *S. cerevisiae* mutated in 276 ORFs and maintained in a single growth condition. Although not every mutation gave an identifiable phenotype in this single growth condition, the expression of at least one gene, other than the deleted gene, changed by more than 2-fold. As described earlier, a large number of parallel negative controls of untreated isogenic yeast cultures were also profiled to ascertain whether particular transcripts had an inherent fluctuation that exceeded that for other genes. In each one of these negative control experiments, at least one gene changed by more than 2-fold. Hierarchical clustering of this dataset identified genes that were regulated by nutritional status or stress. These genes displayed small magnitude, coordinated differences between seemingly identical control cultures, and were thought to represent biological noise within the culture system. The data generated were then used to generate an error model that could provide biological correction. In some mutants with no apparent growth defect, these fluctuations accounted for virtually all of the more than 2-fold changes; however, of those mutants that lacked a growth defect, a third had at least five genes whose expression was altered significantly when the gene error model correction factor was applied. Of those mutants that affected growth, 90% showed significant changes in the expression of at least five genes. Thus, the single growth condition evoked a response in about half of the mutants studied.

Several classes of co-regulated genes were identified, e.g. those whose products were involved in ergosterol biosynthesis, mitochondrial respiration, protein kinase C/calcineurin signaling, amino acid biosynthesis, DNA damage/S phase arrest, and mating [30]. A general observation was that specific mutations altered the expression of genes whose products were involved in the same cellular process that was affected by the mutation; also, different mutants that affected the same pathway frequently exhibited similar gene expression profiles. Additionally, cells treated with a relevant small molecule inhibitor showed a gene expression profile similar to that resulting from mutation of the target; for example, inhibition of HMG-CoA reductase by lovastatin gave a similar profile to an *hmg2* mutant. This important study also demonstrated that the function of some unknown ORFs could be predicted by cluster analysis, and several predictions were confirmed subsequently by biochemical analysis. One example was the clustering pattern of an ORF, which suggested a role in sterol biosynthesis; this was confirmed biochemically and by complementation with its human homologue. Another example was revealed by exposure to dycyclone, an anesthetic, the gene expression profile of which resembled those of mutations that affected the ergosterol pathway. Biochemical analysis confirmed that this pathway was inhibited by dycyclone. One mutant, *erg2*, was hypersensitive to the drug, while overexpression of *erg2* resulted in decreased drug sensitivity. In contrast to *erg2*, other mutants of this pathway were unaffected by dycyclone. Biochemical analysis demonstrated that both *erg2* mutants and dycyclone-treated cells accumulate the same intermediates of this biosynthetic pathway. The human homologue of erg2 is the sigma receptor, a neurosteroid-interacting protein that regulates K^+ conductance and binds several neuroactive drugs, such as haloperidol. Erg2 was also inhibited by haloperidol, an observation

consistent with the sigma receptor and erg2 being related gene products.

There were also many genes whose expression varied by less than 2-fold. In expression profiling experiments, changes of this magnitude are largely ignored as they are considered unreliable. However, it has been widely recognised that changes of this degree may still be important, and there is a requirement for protocols to identify noise and bias and to assess whether these small but potentially critical changes are biologically significant. Hughes *et al.* [30] used their error model to apply a mask to the data, and this facilitated the reliable identification of genes whose expression varied by 1.5-fold. This strategy successfully assigned eight previously uncharacterised ORFs to four pathways, a result that would have been unlikely had conventional cut-off parameters been employed.

Although by no means fully comprehensive, the yeast study has been discussed in detail because it has established the principle of the method. A crude estimate anticipates that a total of 300–700 distinct, full genome transcription patterns would be obtained from the complete set of gene mutations in yeast, when profiling under a single condition [30]. However, one major limitation is the use of a single growth condition, as only half the mutations give a response. Therefore, it may be necessary to look at the remaining mutations under more restrictive conditions when establishing a panel of conditions and mutations.

Application of this approach to mammalian cells will present additional challenges. One major challenge will be the production of 'targetless' cells that are either permanent knockouts or alternatively conditional knockouts that can be switched on and off. Possible strategies could include the use of antisense gene or oligonucleotide technology, the expression of dominant-negative inhibitors, or alternatively the generation of gene knockouts by homologous recombination. In addition, gene expression pattern is dependent upon the tissue of origin [34], with the additional complication that there can be considerable variation between cell lines from the same originating tissue. Screening the growth inhibitory activity of anticancer agents against the National Cancer Institute panel of 60 cell lines has already demonstrated that cell lines derived from the same class of cancer or tissue type can exhibit a wide range of drug sensitivities [63]. Similarly, there are also cell-line dependent differences in the changes in global gene expression profile in response to a given drug. For example, microarray studies in our laboratory have shown that treatment of four different human colon cancer cell lines with an hsp90 molecular chaperone inhibitor results in differing gene expression profiles in response to treatment [66]. Thus, care must be taken when choosing a cell line(s) for profiling experiments.

Despite the potential complications of examining drug action in mammalian cells, a number of simple studies have investigated the utility of global gene expression profiling following exposure of cancer cell lines to individual agents. The response to DNA damage has been assessed following exposure of human myeloid cells to methanosulfonate, γ irradiation, or UV irradiation [67]. Genes expected to be induced by these treatments, such as those regulated by p53, were detected; in addition, the expression of several other genes previously unassociated with DNA damage has also been shown to be responsive to genotoxic stress. Also of significance was the observation that gene expression changes in response to the DNA-damaging treatments varied widely between cell types, indicating that, as discussed above, cellular context can play an important role in therapeutic response. Other studies have also examined the effects on gene expression profile of agents that either directly or indirectly result in DNA damage. Kudoh *et al.* [68] compared doxorubicin-sensitive and -resistant forms of the MCF-7 breast cancer cell lines. Many genes were altered following exposure of the sensitive line to doxorubicin; in contrast, the resistant line exhibited fewer changes, although the genes that did change were also altered in the sensitive line. These included the induction of epoxide hydrolase, a drug-metabolising enzyme frequently overexpressed in breast and other tumors. Analysis of two fibrosarcoma cell lines that were resistant to a diverse range of DNA-interacting agents demonstrated altered expression of genes involved in DNA repair and replication, signal transduction, cell cycle control, and transcription [69].

Studies of agents other than DNA-interactive drugs are now appearing in the literature. One study investigated the effects of a non-specific protein kinase inhibitor, staurosporine, on interleukin-3-dependent murine pro-B cells, and compared the gene expression pattern to that following the induction of apoptosis subsequent to interleukin-3 deprivation [70]. A number of genes were altered following death induced by both stimuli. Interestingly, the apoptotic stimuli influenced a number of genes previously unlinked to cell death pathways. For example, staurosporine treatment caused the induction of genes involved in inflammation.

In our own laboratory, we have used microarrays to examine alterations in gene expression pattern following exposure to the novel agent 17AAG [66]. This is the first hsp90 molecular chaperone inhibitor to enter clinical trial and shows considerable potential as an anticancer agent because of its ability to reduce the cellular levels of several important oncogenic hsp90 client proteins [66,71,72]. The expression profile of a number of human colon adenocarcinoma cell lines was obtained following exposure to 17AAG. The response of each cell line to this agent varied widely at the gene expression level and also at the protein level, again indicating that cellular context has an important role in response to anticancer agents. A number of interesting changes were detected. These included induction of hsp90, the molecular target of the drug, in cell lines that had reduced sensitivity to 17AAG, contrasting with low hsp90 expression in cell lines that are particularly sensitive to 17AAG (Fig. 4). Other gene expression changes, such as those in cytoskeletal and signaling genes, also showed considerable variation between different cell lines. Of particular

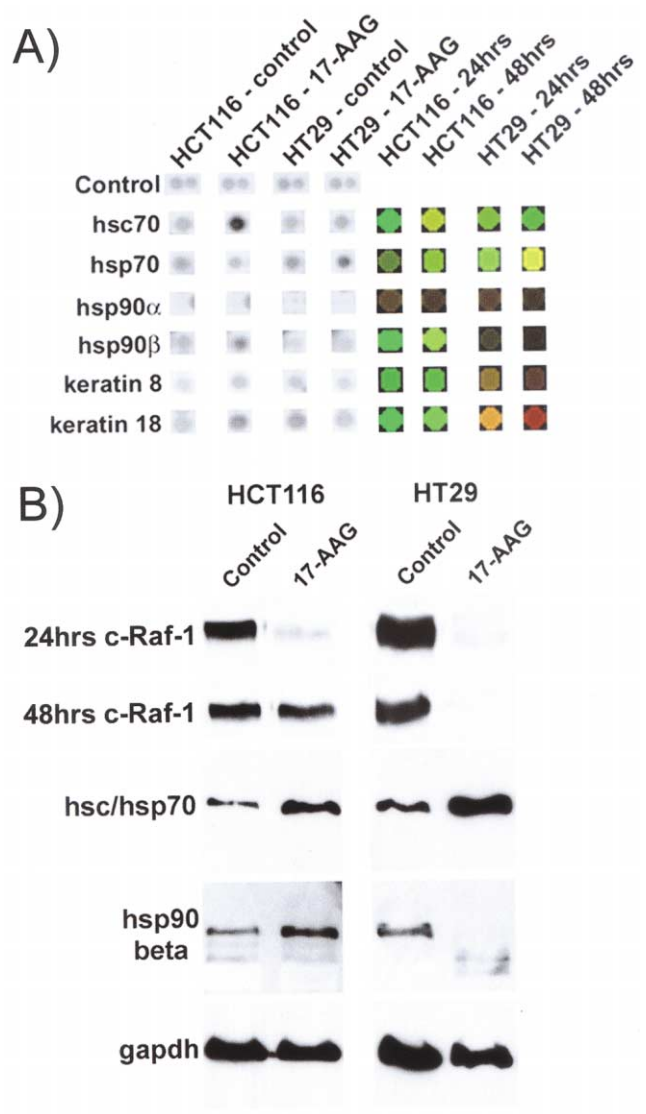

Fig. 4. Demonstration of the use of microarray analysis to investigate the effects of the hsp90 inhibitor 17AAG in human colorectal cancer cells. This figure demonstrates the biological effects of 17AAG in two human colon cancer cells and shows the differential regulation of hsc/hsp70 and hsp90β at the RNA and protein level. The top panel (A) shows raw and processed array data from HT29 and HCT116 human colon adenocarcinoma cells treated with 0.5 and 1 μM ($1 \times IC_{50}$ at 72 hr) 17AAG, respectively. The four columns on the left-hand side show the phosphorimage data following a 24-hr treatment with 17AAG and the four pseudo-coloured columns on the right-hand side show data from the 24 hr and also the 48 hr treatment displayed as a ratio of control:17AAG treatment for each time point. Green = increased by 17AAG treatment, yellow = unchanged, red = decreased by 17AAG treatment, brown/black = undetectable. (B) Western blotting of c-raf-1, hsc/hsp70, hsp90β, and GAPDH expression following 17AAG treatment. In one cell line, HCT116, c-raf-1 is depleted at 24 hr, but has recovered at 48 hr. In the other cell line, HT29, c-raf-1 is depleted at both 24 and 48 hr. Hsc/hsp70 is induced in both cell lines, while hsp90β is induced in HCT116 cells that recover, but is depleted in the more sensitive HT29 cell line. Modified from Ref. 66 (Oncogene 2000;19;4125–33), with permission from the Nature Publishing Group.

note was the consistent induction by 17AAG of the expression of members of the *hsp70* molecular chaperone gene family (*hsp70* and *hsc70*) in all of the cell lines examined. The induction of *hsp70* family genes was confirmed at the protein level and was also observed in human peripheral blood lymphocytes treated with 17AAG *ex vivo*. Using a combination of microarray and western blotting analysis, we defined a molecular signature of hsp90 inhibition, consisting of hsp70 induction and client protein depletion. This signature is being used to provide a pharmacodynamic readout of hsp90 inhibition in peripheral blood lymphocytes and tumor biopsies in our ongoing Phase I clinical trial of 17AAG. Gene expression changes will also be determined by microarray analysis.

Microarrays may also have a role in studies of structure–activity relationships. We have used a microarray strategy to compare inactive and active derivatives of 17AAG and radicicol, a structurally dissimilar hsp90 inhibitor. As mentioned above, expression profiling by microarray identified molecular signatures associated with hsp90 inhibition. These changes were seen with both 17AAG and radicicol and thus were independent of chemotype (Fig. 5); however, additional changes in gene expression specific to the particular chemotype could also be detected (Maloney A, Clarke PA and Workman P, unpublished observations). Such information is extremely useful during preclinical drug development, as it provides feedback to the medicinal chemists as to which chemical features are likely to confer molecular and, hence, therapeutic selectivity.

Although currently there is little information in the literature, another important application of this approach, particularly for pharmaceutical companies, is prediction of the toxicity of drug candidates [73]. The utility of this approach will require a comprehensive database of reference compounds with known pharmacological and toxicological properties. Once this is established, a new compound would be compared to the database to predict compound-related or mechanism-based toxicity.

The investigations outlined thus far in this section demonstrate the potential of the gene expression profiling strategy using microarrays to enhance drug discovery and development, and the study of experimental therapeutics. This approach can: (a) identify the molecular targets of drugs with unknown mechanisms of action; (b) identify potential pharmacodynamic or toxicological markers; (c) distinguish between active and inactive analogues in structure–activity relationship studies; and (d) provide clues about genes involved in drug sensitivity and resistance. As with the application of most new technologies, there are a number of caveats that will need to be addressed. The minimum 2-fold cut-off for gene expression changes widely quoted in the literature is a restriction that may result in more subtle changes below this threshold, although perhaps equally important, being overlooked. This point may be especially relevant in view of frequent anecdotal reports which suggest that test:reference ratios for expression levels derived from arrays are generally lower than those obtained from more conventional approaches, such as northern blotting or RT–PCR. Another complication is the ability to distinguish

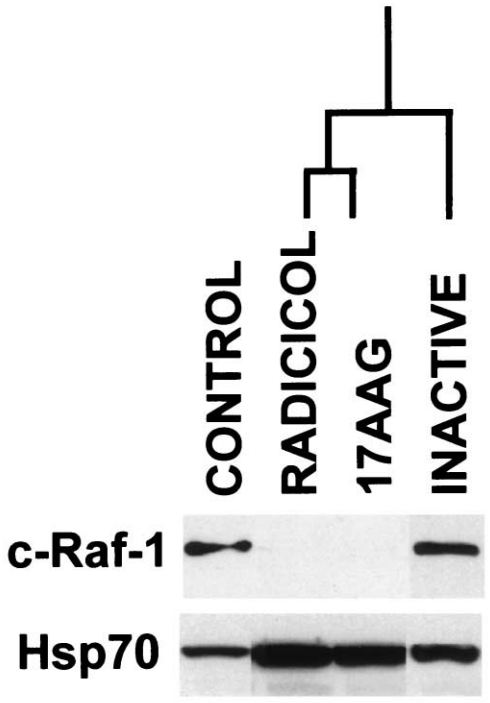

Fig. 5. Validation at the protein level of hsp70 as a gene and pharmacodynamic marker responsive to inhibition of hsp90 by the drug 17AAG. A2780 ovarian adenocarcinoma cells were treated with equimolar doses of 17AAG and an inactive analogue of 17AAG, and also with an equitoxic dose of radicicol, a structurally dissimilar hsp90 inhibitor. The western blots show that the two active compounds depleted c-raf-1 and induced hsp70, while the inactive 17AAG derivative did not deplete c-raf-1 and did not induce hsp70 significantly. The dendrogram at the top of the figure is derived from hierarchical clustering of microarray-derived gene expression data from a parallel experiment. This demonstrates that the chemically dissimilar but active hsp90 inhibitors (17AAG and radicdicol) have more similar gene expression profiles than the structurally similar but inactive 17AAG analogue.

between primary effects on the drug target and those that arise as a downstream consequence of the biological effects that follow inhibition of the target activity. For example, a cyclin-dependent kinase inhibitor could directly influence transcription factor activity and gene expression, while the downstream biological effects of cyclin-dependent kinase inhibition may include cell cycle arrest, an event that will also change the expression of cell cycle-regulated genes, or apoptosis, which will also lead to gene expression changes. The establishment of comprehensive databases, including biological outcomes, should eventually allow us to distinguish between the gene expression changes that are either immediately dependent upon, or alternatively downstream of, target inhibition. Another issue in the use of microarrays during drug development relates to the observation that gene expression varies widely between cell types, and in most cases these differences are dominant over changes induced by drug treatment. Therefore, choice of cell type is an issue, and the use of a standard cell line panel may be essential. In many experiments, it is clear that the expression of hundreds or thousands of genes may change, so dissecting out the important changes may become increasingly difficult. As mentioned earlier, experimental design remains extremely important; attention must be paid to including comprehensive negative controls and also to conducting sufficient repeats. In drug treatment experiments, appropriately detailed analysis of time and dose dependence must be carried out. In our experience and that of others [64], the inclusion of inactive drug analogues as negative controls is increasingly important.

3.3. Cancer classification and molecular pharmacology

The current challenge of cancer treatment is to target specific therapies to pathogenetically distinct tumor types [47,73–75]. It has long been apparent that therapeutics which work against one type of cancer are frequently inactive in other types. For example, drugs that are active in breast cancer may be inactive in lung or colon cancers. Moreover, patients with cancers of an apparently similar type, for example colon, vary widely in response to treatment with the same drug. Individualisation of treatment according to molecular pathology is likely to become even more important with the introduction of new molecular therapeutics that target specific genetic and biochemical abnormalities responsible for malignant progression [7,8, 47,73–75]. Individualisation of cancer therapy is dependent, in part, upon advances in the molecular classification of tumors, and this is discussed below.

Improvements in cancer classification have already proved central to developments in cancer treatment and include technological advances, such as enzyme-based histochemical assays, immunophenotyping, and cytogenetics. Although these approaches have resulted in improved subdivision and classification of cancer, it is highly likely that many cancer subclasses are yet to be defined by molecular markers. Gene expression profiling of clinical tumor samples by microarrays has the potential to contribute in a major way to improved approaches to classification. In addition, with the shift of drug development away from general cytotoxics towards agents that act on specific molecular targets, there is even more of a need for methodologies that provide a molecular profile of the tumor that has prognostic significance, i.e. that predicts treatment outcome. Furthermore, profiling gene expression changes in tumors and surrogate tissues following drug treatment has great potential in the molecular pharmacology of anticancer agents. In addition to providing prognostic information, such gene expression changes could identify, as mentioned earlier, pharmacodynamic markers for use in confirming the molecular mode of action of drugs during clinical trials.

One major challenge for expression profiling of solid tumors is the heterogeneity of biopsies, due to the presence of normal cells of different types. Leukaemia and lymphoma biopsies are less affected by this problem. Also, these cancers are extensively classified and are relatively

easy to sample for tumor biopsy. Two key studies, discussed in detail below, have taken advantage of this and have tested a number of hypotheses concerning gene expression profiling and the molecular classification of cancer.

Although the distinction between the acute leukaemias (ALL and AML) is well established, there is no single test to distinguish between ALL and AML. Classification still requires interpretation of morphology, histochemistry, immunophenotyping, and cytogenetics. Distinguishing between these two leukaemias is critical as their treatment regimens are different. Moreover, despite improved therapy and better response rates, patients that relapse have a poor outlook, and mortality from treatment is still a factor. Golub *et al.* [76] set out to determine whether gene expression profiling by microarray could be used to classify leukaemia samples. The gene expression profile of 6817 genes was measured in bone marrow samples from 27 ALL and 11 AML patients using a high-density oligonucleotide array. The data were initially analysed by neighborhood analysis, an approach that defines the idealised expression pattern corresponding to a gene that is uniformly high in one condition and low in another. The dataset was then tested to find whether there was a high density of other genes nearby, as compared to equivalent random patterns. For the samples analysed, approximately 1100 genes were better correlated with the AML–ALL class distinction than would be expected to happen by chance alone. The known samples were then analysed to create a class predictor, capable of assigning a new sample to one of the two classes. A procedure was developed with a set number of informative genes that could potentially distinguish between classes, so that the expression of these genes in a new sample of unknown pathology could be used to predict its classification. In effect, each informative gene casts a weighted vote in favor of one or another classification, and the sum of these votes is used to set a threshold. The informative genes included: cell surface markers for which known antibodies are used; genes critical for S-phase progression; chromatin remodeling genes; cell adhesion genes; and some known oncogenes. Of the 38 samples examined, the 50 genes that correlated most closely with the AML–ALL distinction could successfully predict the classification of 36/38 samples. In the next stage, 34 unknown samples from peripheral blood and bone marrow were examined from several different reference laboratories. The gene predictor successfully identified 29 of 34 samples with high prediction scores. Interestingly, one laboratory used a different sample preparation protocol that resulted in lower predictive strengths, implying that standardisation of protocols will be essential for studies of this type.

Following this initial success, the approach was then used to examine response and outcome following treatment with an anthracycline plus cytosine arabinoside regimen. Fifteen AML patients with long-term follow-up were examined. However, neighborhood analysis could not predict response, and no strong multigene signatures that correlated with outcome were identified. Explanations for the failure to predict outcome could be that potentially informative transient changes in gene expression profile were missed, as only pre-treatment samples and not post-treatment were examined, or that the relapsing clone had grown out from a single or a small group of cells, the expression profile of which was masked by the bulk of the leukaemic cells.

Another, perhaps more important, question was also asked: Could gene expression profiling alone be used to classify and distinguish between ALL and AML without prior knowledge of the pathology? Two steps were necessary: the first was to determine algorithms to cluster the data by gene expression and the second was to determine whether the classes identified by clustering are real rather than a result of random aggregation or factors such as differences in sample isolation, storage, or preparation. Self-organising maps were used to establish the optimal set of centres around which to cluster the data. The total dataset was then clustered to the nearest centre. Employing a two-centre model, 24/25 ALL and 10/13 AML cases were identified correctly. Using a 20-gene predictor based on these data, 34/38 samples were assigned correctly, with one error and three unassignable samples. With a four-centre model the samples were subdivided into AML, T cell AML, and two different B cell ALL classes; the significance of the two B cell ALL classes was unknown, but these could represent different transformation mechanisms.

Finally, this study also presented an interesting anecdotal case of a patient with a leukaemic presentation that was diagnosed as AML with atypical morphology. The gene expression profile of a marrow sample from this patient was analysed using the class predictor. This gave a low prediction score for both AML and ALL, as neither lymphoid- nor myeloid-specific genes were expressed. In fact, the gene expression pattern detected was more typical of mesenchynal tissue, such as muscle. Cytogenetics subsequently identified a t2:13(q35;q14) translocation characteristic of alveolar rhabdomyosarcoma, a muscle tumor.

In the second study, Alizadeh *et al.* [77] asked whether expression profiling could be used to generate a molecular portrait of distinct types of B cell malignancy and whether distinct B cell malignancies not recognised by current classification systems could be identified. Despite the clinical, morphological, and molecular parameters used to classify lymphoma, patients with similar diagnoses can experience a very different response to treatment. Lymphoma classification has steadily evolved; however, it was significant that recent schemes such as the Revised European-American Lymphoma Classification scheme unified various morphological subtypes into single groups despite the suspicion that they ‘include more than one disease entity’ [77,78]. Each of the currently recognised categories of B cell malignancy can be traced to a particular stage of B cell differentiation, although the extent to which this relationship is maintained in the malignant cell is unclear. DLCL cells have hypermutated immunoglobulin genes, implying that they arise from germinal centre B cells or a later stage of differentiation.

These lymphomas account for approximately 40% of all NHLs. They have a variable clinical course and, although most respond to therapy, fewer than 50% achieve a durable remission. A 'lymphochip' with 17,856 clones was derived from normal lymphoid and lymphoma cDNA libraries, genes regulated by mitogens or cytokines, and a curated set of genes of predicted importance to lymphocyte or cancer biology. Ninety-six normal and malignant biopsy samples were compared to an RNA reference sample derived from a pool of nine lymphoma cell lines. Initial hierarchical clustering separated the malignancies by class, with a few exceptions. Several significant observations were made: (a) comparison of unpurified and purified follicular lymphoma cells from three patients did not change the resulting clustering pattern, suggesting that there was no benefit from purifying the lymphoma cells; (b) two samples from a chronic lymphocytic lymphoma patient taken 18 months apart gave similar results, suggesting a degree of stability in the gene expression profile; and (c) the latter two samples were more similar than samples from other chronic lymphocytic lymphoma patients, demonstrating that each patient can have a unique expression profile, despite being in the same disease category.

Following hierarchical clustering, three branches of the clustering dendrogram captured 43/46 of the DLCLs that were defined by proliferation, lymph node, T cell cluster, and germinal centre B cell gene clusters. Reclustering the data using the genes that defined the germinal centre B cell signature gave two distinct branches of DLCL: one branch had a germinal centre pattern, and the other had a profile similar to an activated B cell pattern. This pattern was not obviously related to the histology of these tumors, as only two of the tumors were classified by the Kiel formula to be immunoblastic-like (activated B cell-like), and there was no evidence for normal germinal centres within the tumor biopsies. There was considerable heterogeneity in the DLCLs, despite the same diagnosis, and no individual gene was diagnostic of the germinal centre or activated B cell type; rather, a shared profile was important. The overall rate of response following treatment with an anthracycline-based multi-agent chemotherapy regimen did not correlate with any particular cluster. However, the group with the activated B cell signature had a significantly poorer overall survival; 16/21 of these died, as compared with the germinal centre type where 6/19 died. Importantly, the classification was independent of the international prognostic indicators for this type of lymphoma. Generally, DLCLs are treated with multi-agent chemotherapy as first line treatment, followed by bone marrow transplantation if a complete remission is not maintained. The observations of Alizadeh *et al.* [77] suggested the possibility that DLCLs might actually consist of two separate entities and that patients who have a profile indicating a poor prognosis might conceivably benefit from moving directly to high dose therapy with bone marrow transplantation, rather than undergoing the potentially debilitating first-line regimen. However, this will require confirmation by large-scale prospective studies.

The two studies discussed above are important in that they demonstrate that haematological malignancies, which can be readily subclassified by cell type and differentiation stage, can be further subdivided using the molecular approach of gene expression profiling. However, it is noteworthy that response to chemotherapy could not be predicted in either study. In the lymphoma study, it was also clear that there was a residual heterogeneity in the two DLCL subgroups that were identified. Despite an overall favorable prognosis, five of the patients with the germinal centre type DLCL died within 2 years of diagnosis, while three of the patients in the unfavorable group survived for 5 years after treatment. The numbers are small, but this could possibly imply a hidden signature in these subgroups that might be revealed by analysis of larger numbers of tumors. Alternatively, the limitation of gene expression profiling may have been reached.

Compared to leukaemias and lymphomas, gene expression profiling of solid tumors is even more of a challenge, as the normal tissue content will influence the expression pattern. Solid cancers are relatively less well classified, other than by the obvious distinction of tissue of origin. In addition, biopsy material is harder to obtain and is generally only available following surgical resection, or when the tumor is at an accessible site, e.g. skin nodules. In addition, the availability of repeat biopsies is very limited, making drug treatment studies relatively difficult.

Two breast cancer biopsies were analysed in the study of Ross *et al.* [34], and their expression profile was compared to the cancer cell lines. A gene cluster defining epithelial cells was apparent from the tumors and epithelial cancer cell lines. Several other clusters were also identified. One was a cluster of stromal genes, whose expression was shared with stromal-like cell lines—this was probably due to the stromal component of the tumor. Another cluster shared genes with a multiple myeloma and was due to the presence of B cells infiltrating the tumor [34].

Perou *et al.* [33] examined tumors from 42 individuals with breast cancer. A tumor sample was taken by open surgical biopsy, the patients were treated with doxorubicin for an average of 16 weeks, and the remaining tumor was subsequently resected. In addition, primary tumor and lymph node metastases from two patients were directly compared. The expression profile was assessed with an 8102 gene array and compared to a reference RNA sample from a pool of eleven different cell lines. The analysis also included seventeen different cell culture lines to aid identification of different cell types within the tumor. The data were filtered down to 1753 genes whose expression varied 4-fold from the median in at least three samples. A great variation in expression was detected; however, following hierarchical clustering 15/20 of the pre- and post-treatment samples clustered together and, likewise, the primary and lymph node tumors also clustered together. This implied that the expression of the majority of tumor genes remains unchanged following either chemotherapeutic treatment or metastatic spread. In three instances, the post-treatment

samples clustered with normal breast tissue, an observation consistent with the fact that these tumors had responded to treatment and there was no tumor in the sample.

The largest gene cluster was a proliferation cluster that correlated with common immunohistochemical markers of proliferation and increased mitotic index. Other clusters included a large group of interferon-regulated genes, a c-*erb-B2*-related cluster of genes located to the region of chromosome 17 which is frequently amplified, and a cluster containing genes, including c-*fos* and *junB*, that were induced by prolonged handling of the sample. In addition, there were clusters of genes consistent with different cell types that may also be present in the tumor: these included endothelial cells, stromal cells, adipose-enriched/normal breast cells, B cells, T cells, and macrophages. The analysis also identified four tumor subgroups with expression profiles characteristic of estrogen receptor positive/luminal epithelial cells, basal-like epithelial cells, erb-B2 positive cells, and normal breast epithelial cells. These again implied the presence of diseases within disease.

An additional breast cancer study has compared sporadic breast tumors to hereditary *BRCA-1*- and *BRCA-2*-related tumors [79]. Multidimensional scaling successfully separated the three types of tumor; of 3226 genes that were analysed, 51 were found to best differentiate between these different tumor types. Examination of the data suggested that *BRCA-1*-related tumors differed significantly from *BRCA-2*-related and sporadic tumors, exhibiting transcriptional activation of pathways involved in apoptosis and DNA repair. The *BRCA-2*-related and sporadic tumors exhibited similar profiles. Interestingly, one patient appeared to have been misclassified into the *BRCA-1* group, but had no discernible mutation. Subsequent analysis of the promoter region revealed aberrant methylation of the *BRCA-1* promoter regions. Hypermethylation of the *BRCA-1* promoter is known to silence *BRCA-1* expression, and the observation was corroborated when *BRCA-1* expression was found to be low in this patient. These observations demonstrated that expression profiling could be used to characterise *BRCA-1* and *BRCA-2* driven tumors and also demonstrated that they are molecularly distinct.

A similar approach has been used in ovarian cancer, comparing normal and malignant ovarian tissue samples. Welsh *et al.* [80] used high-density oligonucleotide arrays to profile and classify 27 serous ovarian adenocarcinomas. The tumors could be split into several groups. One group clustered with normal ovarian tissue, overexpressed high levels of the ribosomal genes, and contained tumors that were generally well-differentiated histologically. Another group clustered with the ovarian cell lines, were poorly differentiated, and underexpressed genes involved in metabolism. A third group exhibited expression patterns consistent with the presence of stroma and activated B cells. Exclusion of the last group of biopsies resulted in a list of genes enriched for potential markers of epithelial ovarian malignancy. A separate study compared the gene expression profiles of serous and mucinous ovarian cancer biopsies [81]. Increased expression of a number of genes common to both types of tumor was detected; in addition, expression patterns specific to serous and mucinous tumors were also detected. Both of these ovarian studies identified potential tumor-specific markers that will require extensive further validation.

To try and avoid the problem of contamination by normal cell types and yet still obtain sufficient material for analysis, Bittner *et al.* [82] profiled expression in melanoma tumor biopsies that were passaged in culture for 2–52 weeks (median 8 weeks passage) prior to analysis. Five samples were analysed straight from biopsy, and one sample was analysed directly from biopsy with a portion also cultured for 3 passages prior to profiling. Expression was compared to a standard cell line reference using an array of 8150 elements, corresponding to 6971 unique genes. The data were analysed by hierarchical clustering using a matrix of Pearson correlation coefficients, and results were plotted using multidimensional scaling. A cluster containing 19 of the melanoma samples was identified, although no clinical or tumor cell characteristics were specifically associated with this cluster. Different samples taken from the same patient were the most closely related, even, in one case, when sampled a year apart. The patient population had uniformly poor prognosis, and outcome data were available for some patients in this study; ten were in the identified melanoma cluster and of these four died, whereas all four patients that were outside this cluster died. However, this difference was not significant. The extent to which melanomas can be subclassified using gene expression microarrays remains to be elucidated, but the fact that this study could group the tumors by molecular markers suggests that such an approach does have promise.

A number of studies have addressed the problem of heterogeneity in solid tumors using the alternative approach of laser capture microdissection. Sgroi *et al.* [35] isolated approximately 100,000 normal, invasive, and metastatic cells using laser capture microdissection of a single breast tumor. Rather than using amplification protocols that may be prone to representative biases, total RNA was extracted and radiolabeled by reverse transcription primed using oligo (dT) and then further radiolabeling by second strand cDNA synthesis. This approach detected a number of tumor-specific genes, and the expression of several of these was subsequently confirmed by RT–PCR. While this study was restricted to a single tumor sample, it demonstrates that microdissection approaches can be feasible and potentially useful. Another study of several squamous cell carcinomas of the head and neck used a similar approach to identify genes showing altered expression that are involved in cell cycle regulation, signal transduction, angiogenesis, and cell death regulation [36].

It is clear from these studies in various tumor types that gene expression profiling of clinical biopsy samples by microarrays can be used to classify tumors and to indicate the presence of previously unidentified molecular subtypes. This approach may also provide information on the under-

lying biology and also lead to the identification of potential novel drug targets, for example, a significantly overexpressed gene. The problem of sample heterogeneity in solid tumors may be circumvented by comparing expression profiles to a panel of cell lines representative of the different cell types expected to be present in the tumor, or alternatively by microdissection of the sample to isolate tumor and normal tissue components. In the case of the lymphoma study, the eventual outcome (survival) could be predicted by expression profiling; however, it is clear that in at least three of these studies, the gene expression profile prior to treatment could not predict response to treatment. As discussed earlier, correlation of pre-treatment basal gene expression profile with the outcome of drug treatments may have limited use, as potentially important but transient responses to treatment will be missed. For example, we have seen transient responses *in vitro*, at both the protein and mRNA level, to hsp90 inhibition by 17AAG, with reversion to basal levels following drug removal [66]. As yet there have been no rigorous investigations that profile gene expression in serial samples, before, during, and after treatment, either in *in vivo* xenograft tumor models or in clinical studies. There are two reasons why these types of expression profiling studies would be important: (a) prediction of outcome with established agents, and (b) demonstration or confirmation of molecular mechanisms of action during the clinical development of new agents. For conventional cytotoxics that are given periodically at acute doses with a period of recovery, serial biopsying of patients is likely to be difficult if not impossible, and indeed may not be particularly informative. However, for the clinical development of new molecular therapeutic agents, such as the signal transduction inhibitors that may be given chronically over extended periods, it may be more feasible to obtain serial biopsies that could give clues both to expected mechanism of action and to prediction of response. As an alternative to serial biopsying, an approach could be used similar to that employed by Ellis *et al.* [83] who measured an apoptotic response following a single pre-operative dose of chemotherapy. In this scenario, expression profiling of biopsies before and after a single dose of chemotherapy could be used predictively. Bittner *et al.* [82] noted that the expression profiles of tumor tissue and of cells cultured from the tumor were similar. This suggests the possibility of using *ex vivo* drug treatments as a predictor. Another possibility would be to use circulating tumor cells for microarray analysis, although an amplification step would be required. Most studies have focused on tumor tissue; however, the profiling of a surrogate tissue such as peripheral blood lymphocytes or buccal mucosa has yet to be explored in clinical studies.

4. Conclusions and future directions

4.1. Technological advances in the hardware

The microarray field is evolving rapidly, and new technologies, methods, and applications are emerging very quickly. Advances in a number of areas might be expected. It has been noted recently that only 62% of the 1000 sequence-verified mouse cDNAs obtained commercially were uncontaminated and actually contained inserts with significant similarity to the published data for the relevant clone [28,84]. In view of the doubt over the identity of the clones arrayed, and given the time and expense associated with the multiple steps required to isolate clones, to perform PCR, and to purify the products, it could be concluded that oligonucleotide arrays may have a brighter future.

Oligonucleotide-based microarrays offer additional advantages of specificity, the ability to incorporate multiple oligonucleotides encompassing a single ORF, and also the ability to include mismatch control oligonucleotides. However, considerable effort will be required to identify the optimal oligonucleotide for each ORF within the human genome. A recent study has described the use of ink-jet printing technology coupled with improved oligonucleotide chemistry that allows *in situ* synthesis of 60mer oligonucleotides at a sufficiently high density of gridding to permit a representation of all human genes on a pair of slides [85].

At the moment, the list of human genes is still incomplete, and a substantial amount of the annotation continues to rely on hypothetical prediction using known features of gene structure or similarity with other organisms. Gene expression profiling could aid the process of gene identification and annotation. An example is the study of Penn *et al.* [86] who used three gene-predicting algorithms to identify ORFs. Primers were designed to amplify ORFs of interest, and from 350 megabases of genomic DNA, 9498 amplicons were identified and spotted onto glass slides. Thirty-one percent were novel and 29% exhibited some similarity to other database sequences; of these, approximately 50% were expressed. The use of oligonucleotide arrays coupled with data from the human genome sequence offers the prospect of arrays either encompassing every gene encoded by the human genome or alternatively the development of exon-specific arrays. Shoemaker *et al.* [87] applied this approach using an exon array to scan the human genome; 50 arrays were employed, containing a total of 1,090,408 oligonucleotides and corresponding to 442,785 predicted exons, of which one-fifth had been defined previously by the Ensemble package. In a single profiling experiment, 58% of the Ensemble defined exons and 34% of the remaining exons were positive. The application of ink-jet printing technology, as described earlier, could substantially accelerate the process of identifying exons in the human genome and will have an important role in annotating gene structure. The ultimate objective of developing a gene chip containing all human genes should be achieved in the next 4–5 years.

The microarray process can also be improved in other ways. For example, a recent study has described the use of a single large glass slide or wafer onto which 49 identical arrays had been gridded [80]. The samples were hybridised, washed, and scanned using a custom-built apparatus that

allowed simultaneous processing of all 49 samples in under 1 week; this also had the advantage that all hybridisations were subject to identical conditions. Currently, the detection limit of arrays using an amplification step is of the order of five mRNA copies/cell. The development of better labeling and detection systems will also be required to improve sensitivity; in addition this will also reduce the amount of input RNA required for array experiments and would ideally obviate the need for an amplification step prior to labeling. Examples of new technologies in this area include the development of infrared excitable phosphor particles that exhibit a phenomenon known as up-conversion when illuminated, a process that has been reported to increase sensitivity over conventionally used fluors [88]. Alternative approaches include electronic DNA detection using capture probes attached via conducting 'molecular wire'. Hybridisation of cDNA labeled with ferrocene, which can transfer electrons, is detected by applying a voltage to this system [89]. Microtransponders consisting of a 0.0125 nm^3 silicon photocell/radio transmitter attached to a nucleic acid probe are also being developed; following hybridisation the photocell is activated by laser excitation of a fluorescently labeled cDNA and the resulting radio signal detected [89]. These approaches could be coupled with miniaturisation and microfluidics (which reduce sample and reagent consumption) and also the use of flow cells. Such integration of the many different features of array technology could provide a single, simple-to-use piece of equipment [reviewed in Ref. 89].

4.2. Developments in software and bioinformatics

Until recently, efforts have concentrated on developing array hardware. More recently, attention has begun to focus on what to do with the data and how to analyse and present it. Although considerable progress has been made, further major advances in data analysis and visualisation are still required. An example is the use of systems such as neural networks, which can be trained with established data and then applied to examine unknown systems. There are currently a limited number of publicly available tools for data storage, processing, retrieval, and integration of microarray data in the context of existing knowledge. Efforts are now underway to establish public databases of gene expression profile results (e.g. see the gene expression omnibus at http://www.ncbi.nlm.nih.gov/geo/or ArrayExpress at http://www.ebi.ac.uk/arrayexpress/). Although there have been attempts to standardise data and experimental format (see http://www.mged.org/), there is still no consensus on how to merge data from different array types (such as oligonucleotide versus cDNA, glass versus nylon, radioactive versus fluorescently labeled) or how to include data from alternative gene expression profiling approaches, such as differential display and serial analysis of gene expression. Other factors that also have to be considered and improved include the choice of reference RNAs and the standardisation of experimental design. In addition, not only will the sheer abundance of data generated by expression profiling require better methods for data handling, but this will also require us to rethink how we interpret the data and generate hypotheses. Interpreting the biological significance of a gene cluster still presents a formidable challenge; however, a recent study has demonstrated methodologies for associating microarray data with the literature so that expression patterns can be more rapidly understood [90].

Early gene expression profiling studies have come from an observational or at least a less hypothesis-driven approach. The inferences that are drawn from expression profiling data are not the endpoint, as they require validation and further evaluation using biochemical approaches. Although it is conceivable that gene expression microarrays could be developed in such a way as to become routine analytical methods in their own right, at present they are mainly used as a screening tool with which to generate hypotheses. The need for validation of a high volume of findings will no doubt refocus efforts on the development or improvement of higher throughput cell-based assays that can assess multiple biochemical parameters. An interesting recent example was reported by Ideker *et al.* [91] who combined microarrays with quantitative proteomics and databases of known physical interactions to build a biochemical pathway, test the hypothesis, and refine the model generated.

4.3. Practical impact

It is clear that gene expression microarrays are already having a major impact on cancer biology, pharmacology, and drug development. As reviewed here, there are many examples of this in the literature, and there will be an extensive body of unpublished information in the pharmaceutical and biotechnology industries that is not available for proprietary reasons. The major limiting factor in the further application of microarrays is probably cost and access to the technology. Costs are likely to decrease and access will grow as a result of the expanding provision of core facilities and the increasing friendliness of the technology. Microarrays will continue to make a major contribution to the progressively complete molecular description and understanding of the biology of normal and cancer cells. Our comprehension of the changes in gene structure and expression levels during cancer progression will become increasingly thorough over the next few years, so that understanding the proteome, rather than the transcriptome, will become critical. An intermediate approach could be to analyse mRNAs being actively translated by isolating mRNA associated with the ribosomal apparatus [reviewed in Ref. 92].

Gene expression arrays have a key role to play in all phases of drug discovery and development (Fig. 1). This includes the identification and validation of new targets, the profiling of on-target and off-target effects during the op-

timisation of new therapeutic agents, understanding molecular mechanisms of action and structure–activity relationships and the prediction of side-effects, and the discovery of diagnostic, prognostic, and pharmacodynamic biomarkers. Microarrays will be used increasingly during early clinical development to confirm that the desired mode of action is occurring and to profile the molecular factors responsible for drug sensitivity and toxicity. In addition, microarrays will have an important role to play in the molecular elucidation of drug resistance [93,94]. As the role of pharmacogenomics becomes more extensive, it is not yet clear to what extent microarrays will be employed in the clinic, for example, in the diagnosis, selection, and monitoring of patients during cancer therapy. Will microarrays be used routinely or will they predominantly be employed to supply diagnostic, prognostic, and biomarker endpoints for more conventional analyses such as immunohistochemistry and ELISA assays? Most likely some combination of these will be used.

Over the next 5–10 years we will have an increasing number of new molecular therapeutics targeted to the major abnormalities that are responsible for cancer progression. Current examples include Herceptin, Glivec (STI571), and Iressa (ZD1839) [7,8,47,73–75]. These new agents are likely to find optimal activity in particular subgroups of patients. Identifying these subgroups will be a challenge, and microarrays will play an important role in this process. There is even the potential to move towards the vision of individualised, genome-based, combinatorial cancer therapy, targeted to the genetics of particular patients [47,73–75]. Microarray analysis will play a leading role alongside other new technologies in testing the ability to achieve this vision.

Acknowledgments

Work in the authors' laboratories is funded by the UK Cancer Research Campaign (CRC), and P.W. is a CRC Life Fellow. We thank Alison Maloney for unpublished data and our colleagues in the Signal Transduction and Molecular Pharmacology Team and Molecular Pathology Team for valuable discussion. We apologise to many colleagues in the field whose work we have been unable to cite due to space restrictions.

References

[1] The International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001;409: 860–921.
[2] Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science 2001;291:1304–51.
[3] Liang F, Holt I, Pertea G, Karamycheva S, Quackenbush J. Gene index analysis of the human genome estimates approximately 120,000 genes. Nat Genet 2000;25:239–40.
[4] Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. Nature 2000;405:827–36.
[5] Brent R. Genomic biology. Cell 2000;100:169–83.
[6] Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100: 57–70.
[7] Workman P. Towards genomic cancer pharmacology: innovative drugs for the new millenium. Curr Opin Oncol Endocr Metab Invest Drugs 2000;2:21–5.
[8] Garrett MD, Workman P. Discovering novel chemotherapeutic drugs for the third millenium. Eur J Cancer 1999;35:2010–30.
[9] Wosikowski K, Schuurhuis D, Johnson K, Paull KD, Myers TG, Weinstein JN, Bates SE. Identification of epidermal growth factor receptor and c-erbB2 pathway inhibitors by correlation with gene expression patterns. J Natl Cancer Inst 1997;89:1505–15.
[10] Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ, Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, Buolamwini JK, van Osdol WW, Monks AP, Scudiero DA, Sausville EA, Zaharevitz DW, Bunow B, Viswanadhan VN, Johnson GS, Wittes RE, Paull KD. An information-intensive approach to the molecular pharmacology of cancer. Science 1997;275:343–9.
[11] Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances. Electrophoresis 1997;18:533–7.
[12] Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19:1720–30.
[13] Bailey DS, Bondar A, Furness LM. Pharmacogenomics—it's not just pharmacogenetics. Curr Opin Biotechnol 1999;9:595–601.
[14] Stratowa C, Wilgenbus KK. Gene expression profiling in drug discovery. Curr Opin Mol Ther 1999;1:671–9.
[15] Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 1992;257: 967–71.
[16] Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995;270:484–7.
[17] 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. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 2000;18:630–4.
[18] Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. Nat Genet 1999;21(Suppl):10–4.
[19] Eisen MB, Brown PO. DNA arrays for analysis of gene expression. Methods Enzymol 1999;303:179–205.
[20] Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J. A concise guide to cDNA microarray analysis. Biotechniques 2000;29:548–62.
[21] Pilarsky CH, Schmitt AO, Dahl E, Rosenthal A. Microarrays: chances and challenges. Curr Opin Mol Ther 1999;1:727–36.
[22] Cohen B (Ed.). The Chipping Forecast. Nat Genet 1999;21(Suppl): 1–60.
[23] Fodor S, Read JL, Pirrung MC, Styer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science 1991;251:767–73.
[24] Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. Nat Genet 1999;21(Suppl):20–32.
[25] Benson DA, Boguski MS, Lipman DJ, Ostell J. Gen Bank. Nucleic Acids Res 1997;25:1–6.
[26] Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tomé P, Aggarwal A, Bajorek E, Bentolila S, Birren BB, Butler A, Castle AB, Chiannilkulchai N, Chu A, Clee C, Cowles S, Day PJR, Dibling T, Drouot N, Dunham I, Duprat S, East C, Edwards C, Fan J-B, Fang N, Fizames C, Garrett C, Green L, Hadley D, Harris M, Harrison P, Brady S, Hicks A, Holloway E, Hui L, Hussain S, Louis-Dit-Sully C, Ma J, MacGilvery A, Mader C, Maratukulam A, Matise TC, McKusick KB, Morissette J, Mungall A, Muselet D, Nusbaum HC, Page DC, Peck A, Perkins S, Piercy M, Qin F, Quackenbush J, Ranby S, Reif T, Rozen S, Sanders C, She X, Silva J, Slonim DK, Soderlund C, Sun W-L, Tabar P, Thangarajah T,

Vega-Czarny N, Vollrath D, Voyticky S, Wilmer T, Wu X, Adams MD, Auffray C, Walter NAR, Brandon R, Dehejia A, Goodfellow PN, Houlgatte R, Hudson JR Jr, Ide SE, Iorio KR, Lee WY, Seki N, Nagase T, Ishikawa K, Nomura N, Phillips C, Polymeropoulos MH, Sandusky M, Schmitt K, Berry R, Swanson K, Torres R, Venter JC, Sikela JM, Beckmann JS, Weissenbach J, Myers RM, Cox DR, James MR, Bentley D, Deloukas P, Lander ES, Hudson TJ. A gene map of the human genome. Science 1996;274:540–6.

[27] Boguski MS, Lowe TM, Tolstoshev CM. DbEST—database for expressed sequence tags. Nat Genet 1993;4:332–3.

[28] Halgren RG, Fielden MR, Fong CJ, Zacharewski TR. Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. Nucleic Acids Res 2001;29:582–8.

[29] Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. Nat Genet 1999;21(Suppl):5–9.

[30] Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, D'Armour C, Bennett HA, Coffey E, Dai H, He YE, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH. Functional discovery via a compendium of expression profiles. Cell 2000;102:109–26.

[31] Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH, Marton MJ. Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet 2000;25:333–7.

[32] Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 1999;23:41–66.

[33] Perou CM, Serlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lenning PE, Berresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature 2000;406:747–52.

[34] Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JCF, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D, Brown PO. Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 2000;24:227–35.

[35] Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JR, Elkahloun AG. *In vivo* gene expression profile analysis of human breast cancer progression. Cancer Res 1999;59:5656–61.

[36] Leethanukul C, Patel V, Gillespie J, Pallente M, Ensley JF, Kootongkaew S, Liotta LA, Emmert-Buck M, Gutkind JS. Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. Oncogene 2000;19:3220–4.

[37] Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG. Gene expression profiles of laser-captured adjacent neuronal subtypes. Nat Med 1999;5:117–22.

[38] Talaat AM, Hunter P, Johnston SA. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. Nat Biotechnol 2000;18:679–82.

[39] Carlsson B, Jernas M, Lindell K, Carlsson LMS. Total RNA and array-based expression monitoring. Nat Biotechnol 2000;18:579.

[40] DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 1996;14:457–60.

[41] Bassett DE, Eisen MB, Boguski MS. Gene expression informatics—it's all in your mine. Nat Genet 1999;21(Suppl):51–5.

[42] Ermolaeva O, Rastogi M, Pruitt KD, Schuler GD, Bittner ML, Chen Y, Simon R, Meltzer P, Trent J, Boguski MS. Data management and analysis for gene arrays. Nat Genet 1998;20:19–23.

[43] Bittner M, Meltzer P, Trent J. Data analysis and integration: of steps and arrows. Nat Genet 1999;22:213–5.

[44] Young RA. Biomedical discovery with DNA arrays. Cell 2000;102:9–15.

[45] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–8.

[46] Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovosky E, Lander ES, Golub TR. Interpreting the patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 1999;96:2907–12.

[47] Workman P, Clarke PA. Innovative cancer drug targets: genomics, transcriptomics and clinomics. Expert Opin Pharmacother 2001;2:911–5.

[48] Szallasi Z. Gene expression patterns and cancer. Nat Biotechnol 1998;16:1292–3.

[49] Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ, Davis RW. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 1998;2:65–73.

[50] Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. Nat Genet 1999;22:281–5.

[51] Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO. The transcriptional program in the response of human fibroblasts to serum. Science 1999;283:83–7.

[52] Fambrough D, McClure K, Kazlauskas A, Lander ES. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. Cell 1999;97:727–41.

[53] Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, He YD, Dai H, Walker WL, Hughes TR, Tyers M, Boone C, Friend SH. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science 2000;287:873–80.

[54] Guo QM, Malek RL, Kim S, Chiao C, He M, Ruffy M, Sanka K, Lee NH, Dang CV, Liu ET. Identification of c-Myc responsive genes using rat cDNA microarray. Cancer Res 2000;60:5922–8.

[55] Svaren J, Ehrig T, Abdulkadir SA, Ehrengruber MU, Watson MA, Milbrandt J. EGR1 target genes in prostate carcinoma cells identified by microarray analysis. J Biol Chem 2000;275:38524–31.

[56] Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 1996;93:10614–9.

[57] Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, Smith PD, Jiang Y, Gooden GC, Trent JM, Meltzer PS. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res 1998;58:5009–13.

[58] Khan J, Bittner ML, Saal LH, Teichmann U, Azorsa DO, Gooden GC, Pavan WJ, Trent JM, Meltzer PS. cDNA microarrays detect activation of a myogenic transcription program by the *PAX3-FKHR* fusion oncogene. Proc Natl Acad Sci USA 2000;96:13264–9.

[59] Bubendorf L, Kolmer K, Kononen J, Koivisto P, Mousses S, Chen Y, Mahlamaki E, Schraml P, Moch H, Willi N, Elkahloun AG, Pretlow TG, Gasser TC, Mihatsch MJ, Sauter G, Kallioniemi K. Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays. J Natl Cancer Inst 1999;91:1758–64.

[60] Wellmann A, Thieblemont C, Pittaluga S, Sakai A, Jaffe ES, Siebert P, Raffeld M. Detection of differentially expressed genes in lymphomas using cDNA arrays: identification of clusterin as a new diagnostic marker for anaplastic large-cell lymphomas. Blood 2000;96:398–404.

[61] Li S, Ross DT, Kadin M, Brown PO, Wasik MA. Comparative genome-scale analysis of gene expression profiles in T cell lymphoma cells during malignant progression using complementary DNA microarray. Am J Pathol 2001;158:1231–7.

[62] Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. Nature 2000; 406:532–5.
[63] Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, Scudiero DA, Eisen MB, Sausville EA, Pommier Y, Botstein D, Brown PO, Weinstein JN. A gene expression database for the molecular pharmacology of cancer. Nat Genet 2000;24:236–44.
[64] Gray NS, Wodicka L, Thunnissen A-MWH, Norman TC, Kwon S, Espinoza FH, Morgan DO, Barnes G, LeClerc S, Meijer L, Kim S-H, Lockhart D, Shultz PG. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. Science 1998;281:533–8.
[65] Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard J, Slade D, Hongyue D, Bassett DE, Hartwell LH, Brown PO, Friend SH. Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat Med 1998;4:1293–301.
[66] Clarke PA, Hostein I, Banerji U, Di Stefano F, Maloney A, Walton M, Judson I, Workman P. Gene expression profiling of human colon cancer cells following inhibition of signal transduction by 17-allylamino-17-demethoxygeldanamycin, an inhibitor of the hsp90 molecular chaperone. Oncogene 2000;19:4125–33.
[67] Amundson SA, Bittner M, Chen Y, Trent J, Meltzer P, Fornace AJ. Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. Oncogene 1999; 18:3666–72.
[68] Kudoh K, Ramanna M, Ravatn R, Elkahoun AG, Bittner EL, Meltzer PS, Trent JM, Dalton WS, Chin KV. Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. Cancer Res 2000;60:4161–6.
[69] Levenson VV, Davidovich IA, Roninson IB. Pleiotropic resistance to DNA-interactive drugs is associated with increased expression of genes involved in DNA replication, repair and stress response. Cancer Res 2000;60:5027–30.
[70] Brachat A, Pierrat B, Brungger A, Heim J. Comparative microarray analysis of gene expression during apoptosis-induction by growth factor deprivation or protein kinase C inhibition. Oncogene 2000;19:5073–82.
[71] Kelland LR, Sharp SY, Rogers PM, Myers TG, Workman P. DT-diaphorase expression and tumor cell sensitivity to 17-allylamino, 17-demethoxygeldanamycin, an inhibitor of heat shock protein 90. J Natl Cancer Inst 1999;91:1940–9.
[72] Hostein I, Robertson D, DiStefano F, Workman P, Clarke PA. Inhibition of signal transduction by the Hsp90 inhibitor 17-allylamino-17-demethoxygeldanamycin results in cytostasis and apoptosis. Cancer Res 2001;61:4003–9.
[73] Workman P. New drug targets for genomic cancer therapy: success, limitations, opportunities and future challenges. Curr Cancer Drug Targets 2001;1:33–47.
[74] Workman P. Scoring a bullseye against cancer genome targets. Curr Opin Pharmacol 2001;1:342–52.
[75] Workman P. Changing times: developing cancer drugs in genomeland. Curr Opin Invest Drugs 2001;2:1128–35.
[76] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286: 531–7.
[77] Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt L. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2001;403:503–11.
[78] Harris NL, Jaffe ES, Stein H, Banks PM, Chan JKC, Cleary ML, Delsol G, De Wolf-Peeters C, Falini B, Gatter KC, Grogan TM, Isaacson PG, Knowles DM, Mason DY, Muller-Hermelink H-K, Pileri SA, Piris MA, Ralfkiaer E, Warnke RA. A Revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. Blood 1994;84:1361–92.
[79] Hedenfalk MS, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallionemi O-P, Wilfond B, Borg A, Trent J. Gene-expression profiles in hereditary breast cancer. New Engl J Med 2001;344:539–48.
[80] Welsh JB, Zarrinkar PP, Sapinosa LM, Kern SG, Behling CA, Monk BJ, Lockhardt DJ, Burger RA, Hampton GM. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Cancer Res 2001;98:1176–81.
[81] Ono K, Tanaka T, Tsunoda T, Kitahara O, Kihara C, Okamoto A, Ochiai K, Takagi T, Nakamura Y. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. Cancer Res 2000;60: 5007–11.
[82] Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhani Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V, Hayward N, Trent J. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 2000;406:536–40.
[83] Ellis PA, Smith IE, McCarthy K, Detre S, Salter J, Dowsett M. Preoperative chemotherapy induces apoptosis in early breast cancer. Lancet 1997;349:849.
[84] Knight J. When the chips are down. Nature 2001;410:860–1.
[85] Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavat G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesiser. Nat Biotechnol 2001;19:342–7.
[86] Penn SG, Rank DR, Hanzel DK, Barker DL. Mining the human genome using microarrays of open reading frames. Nat Genet 2000; 26:315–8.
[87] Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engele P, McDonagh PD, Loerch PM, Leonardson A, Lum PY, Cavet G, Wu LF, Altschuler SJ, Edwards S, King J, Tsang JS, Schimmack G, Schelter JM, Koch J, Ziman M, Marton MJ, Li B, Cundiff P, Ward T, Castle J, Krolewski M, Meyer MR, Mao M, Burchard J, Kidd MJ, Dai H, Phillips JW, Linsley PS, Stoughton R, Scherer S, Boguski MS. Experimental annotation of the human genome using microarray technology. Nature 2001;409:922–7.
[88] van de Rijke F, Zijlmans H, Li S, Vail T, Raap AK, Niedbala RS, Tanke HJ. Up-converting phosphor reporters for nucleic acid microarrays. Nat Biotechnol 2001;19:273–6.
[89] Blohm DH, Guiseppi-Elie A. New developments in microarray technology. Curr Opin Biotechnol 2001;12:41–7.
[90] Jenssen T-K, Laegreid A, Komorowski J, Hovig E. A literature network of human genes for high-throughput analysis of gene expression. Nat Genet 2001;28:21–8.
[91] Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlet DR, Aebersold R, Hood L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001;292:929–34.
[92] Pradet-Balade B, Boulme F, Beug H, Mullner EW, Garcia-Sanz JA. Translational control: bridging the gap between genomics and proteomics. Trends Biochem Sci 2001;26:225–9.
[93] Turton NJ, Judah DJ, Riley J, Davies R, Lipson D, Styles JA, Smith AG, Gant TW. Gene expression and amplification in breast carcinoma cells with intrinsic and acquired doxorubicin resistance. Oncogene 2001;20:1300–6.
[94] Robert J. Resistance to cytotoxic agents. Curr Opin Pharmacol 2001; 1:353–7.
[95] Gwynne P, Heebner G. DNA chips and microarrays, part 1. Science 2001;292:949–73.