ELSEVIER

European Journal of Cancer 40 (2004) 2560–2591

European Journal of Cancer

www.ejconline.com

Gene expression microarray technologies in the development of new therapeutic agents

Paul A. Clarke, Robert te Poele, Paul Workman *

Cancer Research UK Centre for Cancer Therapeutics, The Institute of Cancer Research, Haddow Laboratories, 15 Cotswold Road, Belmont, Sutton, Surrey SM2 5NG, UK

Received 24 March 2004; received in revised form 27 June 2004; accepted 28 July 2004
Available online 21 September 2004

Abstract

We review in detail how gene expression microarray technology is benefiting all phases of the discovery, development and subsequent use of new cancer therapeutics. Global gene expression profiling is valuable in cancer classification, elucidation of biochemical pathways and the identification of potential targets for novel molecular therapeutics. We exemplify the value in tissue culture and animal models of cancer, as well as in clinical studies. The power of expression profiling alongside gene knockout or knockdown methods such as RNA interference is illustrated. The use of basal or constitutive gene expression profiling to understand and predict drug sensitivity or resistance is described. The ability of expression profiling to define detailed molecular signatures of drug action is emphasised. The approach can identify on-target and off-target effects. It can be used to identify molecular biomarkers for proof of concept studies, pharmacodynamic endpoints and prognostic markers for predicting outcome and patient selection.

Keywords: Microarray; Drug development; Pharmacology; Cancer

1. Introduction

The ability to profile the expression of hundreds, thousands and eventually all genes encoded by the human genome using microarray analyses is transforming how biomedical research is carried out. The massively parallel nature of the technology empowers us to carry out analyses that could not be comprehended previously and is revolutionising the way we think about and conduct our science.

In this Review, we focus on how gene expression microarray technology is being used to improve and accelerate the development of new therapeutic agents. As we will show, the technology is impacting on all phases of the drug discovery process. We discuss in detail the use of gene expression profiling in the classification of cancer, the study of biochemical pathways and the identification of potential targets for novel therapeutics. We also show the complimentary value of microarray gene expression profiling in animal and tissue culture models of cancer. The power of using expression profiling alongside gene knockout methods such as RNA interference is illustrated. The use of basal or constitutive gene expression profiling to understand and predict drug sensitivity and resistance is described. Likewise, the value of determining changes in gene expression in response to drug treatment is illustrated. Examples are

Abbreviations: ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; CDK, cyclin-dependent protein kinase; DLCL, diffuse large cell lymphoma; ER, oestrogen receptor; FNA, fine-needle aspirate; HDAC, histone deacetylase; k-NN, K nearest neighbour; MLL, mixed lineage leukaemia; NCI, National Cancer Institute; RNAi, small interfering RNA; SAM, statistical analysis of microarrays; SOM, self-organising map.

* Corresponding author. Tel.: +44 20 8722 4301; fax: +44 20 8722 4324.
E-mail address: Paul.Workman@icr.ac.uk (P. Workman).

0959-8049/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2004.07.024

provided to illustrate how microarray gene expression profiling can be used to improve our understanding of the molecular pharmacology of established and novel drugs in cancer patients and we conclude with a brief summary and future perspective.

1.1. Cancer, the human genome and the development of targeted therapeutics

The last 25 years have witnessed previously undreamt of insights into the cellular and molecular pathology of cancer that have affected the aspirations of those involved in drug discovery. Arguably one of the most monumental contributory achievements was the publication of an initial working draft of the sequence of the human genome in February 2001 [1,2]. The human genome sequence was subsequently ''finished'' in 2003, appropriately in time for the 50th anniversary of the description of the double helical structure of DNA [3]. With the sequencing complete, the human genome is now being annotated on a chromosome-by-chromosome basis, a process requiring in depth manual review and curation of gene content. To date, seven chromosomes have been completed (Y, 6, 7, 14, 20, 21 and 22) revealing 7738 gene structures, with evidence that 5410 are functional genes, of which approximately half are previously undescribed [4].

Many genes are already known to play a role in malignancy and as a result of the output from the human genome project, it seems likely that most of the remaining cancer genes will be identified over the next five years or so [5]. From the current inventory of human genes, it has been estimated that there will be approximately 2–3 thousand genes encoding G-protein coupled receptors, nuclear receptors, ion channels, protein kinases and other enzymes currently considered to belong to 'druggable' classes of targets [6]. A number of these tractable targets are well known; however, the bulk of these candidates have yet to be implicated in the aetiology of any disease or to be shown to have relevant therapeutic potential. Nevertheless, a significant proportion of these genes are likely to have key roles in cancer biology and will also have potential as specific targets of novel rationally designed and targeted anticancer therapeutics.

There is now a consensus that drug development is moving away from generally cytotoxic agents to drugs that are designed to target genes causally involved in cancer [7]. Strong evidence from animal models of human cancer supports the potential therapeutic value of agents targeting particular cancer genes. For example, transgenic mice engineered to overexpress the C-*MYC* oncogene develop malignant osteosarcomas. Transient removal of C-MYC expression induces differentiation, while restoration of C-MYC expression does not result in reversion to malignant phenotype, but instead induces apoptosis [8]. Similarly, apoptosis and tumour regression are detected in HA-RAS-driven melanomas when HA-RAS expression is switched off [9]. Experiments of this type have stimulated the concept of 'oncogene addiction' [10]. According to this hypothesis, the multiple redundant signalling pathways in normal cells are lost in cancer cells through selection for critical oncogenic pathways during malignant progression. Importantly, this concept also supports the strategy of interfering with oncogene function by therapeutic intervention, particularly by acting on the specific pathways upon which cancer cells have become dependent.

Alongside experimental support from animal models of cancer, the molecular dissection of human oncogenesis and malignant progression has also been extremely important to the development of new molecular therapeutics. Although, our understanding of malignant progression is far from complete, we have now documented pathways and gene products that are causally involved in the initiation and progression of cancer. This provides significant numbers of novel and interesting molecular targets for rational therapeutic intervention. Several therapeutic agents targeting a specific genetic abnormality in cancer cells are already being tested and many more are in clinical trials [11]. 'Proof of principle' is now emerging that such agents show biological and therapeutic activity by the desired mechanism not only in pre-clinical models, but also in the cancer patient, and the success of this approach is exemplified by the regulatory approval of imatinib mesylate (Gleevec, Glivec), trastuzumab (Herceptin) and gefitinib (Iressa) [12–14].

1.2. Analysis of gene expression by microarrays

The gene expression profile of a cell determines its phenotype and response to the environment. The complement of genes expressed by a cell is very dynamic and can respond rapidly to external stimuli. Therefore, measurement of gene expression can potentially provide clues about common regulatory mechanisms, biochemical pathways and broader cellular functions. In addition, comparison of genes expressed in disease versus the normal counterpart tissue will further the understanding of disease pathology and identify points for potential therapeutic intervention.

Gene expression can be assessed by measuring the quantity of the end-product, protein, or the mRNA template used to synthesise protein. Transcription factors are frequently the targets of pathways that transduce environmental signals and numerous studies have identified remodelled gene expression, assessed by measuring mRNA, during many cellular processes ranging from cell division to differentiation to cell death. Significantly, in the context of this Review, there have been many observations of altered individual gene expression following drug exposure.

1.3. Analysis of mRNA by microarray technology

Gene expression microarray technology has been founded on a number of key developments in blotting methodologies, recombinant DNA technologies such as the polymerase chain reaction (PCR), and the accumulation of tens of thousands of sequenced cDNA clones from genome sequencing initiatives. There are a number of useful references that deal with the technological aspects of the fabrication and application of microarrays [15–19]. In general, the conduct of an array experiment can be divided into the following processes: (1) fabrication of the array; (2) RNA isolation and labelling; (3) application of the labelled sample to the array and measurement of hybridisation; and finally (4) data analysis and interpretation (Fig. 1; [15–19]).

There are two main types of array: those where oligonucleotides are synthesised on the array *in situ* using photolithographic or other techniques and arrays where PCR products, plasmids or oligonucleotides are robotically printed or deposited using pins or pens onto nitrocellulose, nylon, plastic or glass supports at densities of the order of 100 000 oligonucleotides or 10 000 PCR products per cm^2.

Labelled representations of cellular mRNA are generated using reverse transcriptase, an enzyme that generates a single-strand DNA copy of each mRNA. Depending on the labelling methodology, generally >1 μg total RNA or > 0.050 μg mRNA are required. In some circumstances, experiments will not yield sufficient RNA, in which case a commonly used approach is linear amplification. This method is based on cDNA synthesis by reverse transcription using primers that contain a bacteriophage RNA polymerase promoter sequence [15–19]. The cDNA is amplified by *in vitro* transcription using the appropriate bacteriophage RNA polymerase, and then labelled and used for microarray analysis. This approach is reproducible and allows the use of as few as 500–1000 cells or as little as 1–50 ng of total RNA.

Microarrays exploit a remarkable feature of nucleic acids that allows them to assemble into a duplex in a process that is reversible, but occurs with absolute fidelity [20]. In effect, each labelled cDNA searches out and pairs with its complementary sequence on the array. Hybridisation of radiolabel cDNAs can be detected and quantified by autoradiography or phosphorimaging. The hybridisation of fluorescently labelled cDNAs is detected using laser scanners that excite the fluorescent

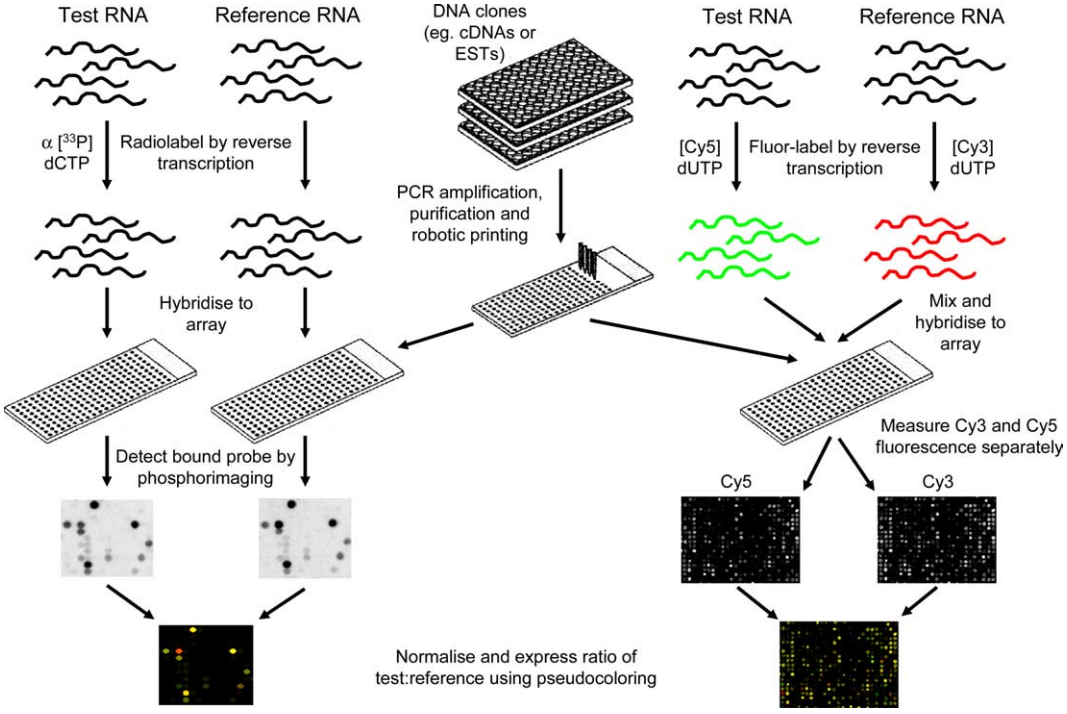

Fig. 1. Schematic of the various steps in a microarray experiment. Plasmid clones are propagated in bacteria, and the cloned inserts are amplified by polymerase chain reaction (PCR) and then purified. Next, the purified PCR products are robotically printed onto solid supports. Modifications of this approach 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. Samples can be analysed on separate arrays using radiolabelled cDNA or fluorescently labelled cDNA prepared from the test and reference samples. Alternatively, glass slide arrays can be hybridised simultaneously with Cy5 and Cy3 fluorescently labelled test and reference samples, respectively. Following stringency washes, hybridisation is detected by phosphorimaging or 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. ESTs, expression sequence tags; dUTP, deoxyuridine triphosphate.

dyes and collect their emission at the relevant wavelengths. The signal obtained following hybridisation to the arrays gives a measure of the number of molecules bound. 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 by overlaying and aligning a grid specifying spot location onto the array image. The background and hybridisation signal are calculated using algorithms that predict the expected position, size, and shape of the spot and then calculate local background in the vicinity of each spot.

Comparison of data from multiple arrays or multiple samples on a single array requires the data to be scaled or normalised, in order to ensure that the expression levels are directly comparable. One approach is to include ''housekeeping'' genes that are assumed to be constitutively expressed and relatively unchanged from experiment to experiment. Another method is to assume that the total amount of cellular mRNA is constant; or that the overall ratio of expression between test and control averaged over the ratios for every gene measured equals unity [21]. Locally weighted linear regression (lowess) analysis is also frequently used as a normalisation method [21,22]. Before proceeding with any kind of analysis, the data are frequently filtered as genes that do not change in any of the samples will not contribute to discrimination between samples. Most investigators typically apply an arbitrary global minimum threshold of a 1.5- to 4-fold change for differences in expression that might be considered biologically interesting. When possible, multiple replicates of the experiment are included, as this allows the elimination of false-positives through significance testing. A simple *t*-test can be used to determine significance between control and test data; however, in most cases, more than two conditions are tested and an analysis of variance is required [23]. In some cases, the data will not be normally distributed and in such instances non-parametric testing is necessary.

It is becoming increasingly clear that biological function incorporating features such as feed-back, feed-forward, error checking, and redundancy generally results from complex interactions between many components; for example, transcriptional regulation is more sophisticated than the traditional view of a simple on–off event [24,25]. Therefore, the acquisition of sufficiently large datasets to address such complex biological systems or genome-wide function is essential. Having established a suitable dataset, the process of mining the data for meaningful information is then initiated [24,26,27]. The easiest form of analysis is simply to list all of the genes that differ between the test and control 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 is relatively straightforward. However, these approaches fail to extract all the potential information in genome-scale experiments with many samples and more sophisticated analysis identifying non-random groups of genes associated with particular biological events are required.

One strategy is to use hierarchical clustering, an approach commonly employed in sequence/phylogeny experiments, coupled with outputs that facilitate visual examination of the data [28]. Several studies have noted transcriptional co-regulation 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 concept that genes with similar expression profiles are likely to be functionally related [20]. 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 [25]. 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.

However, there are some limitations to hierarchical clustering as well-separated data are required to give unambiguous results [24]. Other mathematical approaches commonly used for data mining are K means or self-organising maps (SOMs) that are useful for initial identification of expression patterns, but may still not be powerful enough to distinguish subtle differences in expression [29]. Classification of samples, for example, cancer sub-types, can be achieved using k nearest neighbour (k-NN) analysis that can be used on a relatively small dataset [30]. Another commonly used algorithm is significance of analysis of microarrays (SAM) that identifies statistically significant changes in expression by assimilating a set of gene-specific *t*-tests and repeated permutations of the data to identify a false detection rate [31]. If the number of samples is high, the more advanced classification methods of support vector machines and artificial neural networks can be employed [32]. A list of some of the software that is freely available for Microarray analysis is provided in Table 1.

1.4. Microarray analysis in contemporary drug discovery

In this Review, we will illustrate how gene expression microarrays are impacting on most stages of the cancer drug discovery process. An indication of the increasing application of this technology can be seen when searching PubMed for 'microarray' or 'microarray + cancer'. In 1998, these searches gave 21 and 6 citations, respectively; 5 years later, the same search resulted in 2394 and 704 hits, respectively. The current literature illustrates some principles of the application of gene expression profiling with chemotherapeutic agents currently used in the clinic. These examples are cited as paradigms for the application of this technology in the discovery of novel anticancer therapeutics.

Table 1
A selective list of freely available software for microarray analysis

Package	Internet address	Comments
The R project for statistical computing	http://www.r-project.org/	Number of packages run under R
Bioconductor	http://www.bioconductor.org/	Open source and development software project for the analysis of genomic data, runs under R
Jexpress	http://www.ii.uib.no/~bjarted/jexpress/index.html	Developed by Bjarte Dysvik and Inge Jonassen, Univ of Bergen, Norway. A java application for the analysis of gene expression data from microarray experiments
BRB arraytools	http://linus.nci.nih.gov/BRB-ArrayTools.html	Developed by: Richard Simon & Amy Peng, National Cancer Institute, USA. An integrated package for the visualisation and statistical analysis of DNA microarray gene expression data. Runs as an Excel plug-in
GeneCluster 2.0	http://www.broad.mit.edu/cancer/software/software.html	Implements the methodology used in Golub and colleagues [30]. Includes algorithms for building and testing supervised models using weighted voting and k-nearest neighbours
Analysis of variance	http://www.jax.org/research/churchill/software/anova/index.html	ANOVA programs for microarray data: from Gary Churchill's group at Jackson Laboratory, USA
Cluster/Treeview	http://rana.lbl.gov/EisenSoftware.htm	From Eisen lab at Lawrence Berkeley National Lab, USA [28]
Statistical analysis of microarrays	http://www-stat.stanford.edu/~tibs/SAM/index.html	From Tibshirani group at Stanford University, USA [31]. Runs as an Excel plug-in
Prediction analysis for microarrays	http://www-stat.stanford.edu/%7Etibs/PAM/	From Tibshirani group at Stanford University, USA. Class Prediction Software for Genomic Expression Data Mining. Runs as an Excel plug-in
BASE	http://base.thep.lu.se/	Comprehensive free web-based database solution for data generated by microarray analysis, supports Minimal Information About Microarray Experiment (MIAME) format
General sites		
European Bioinformatics Institute micorarray site	http://www.ebi.ac.uk/microarray/index.html	
Stanford University microarray site	http://genome-www5.stanford.edu/	

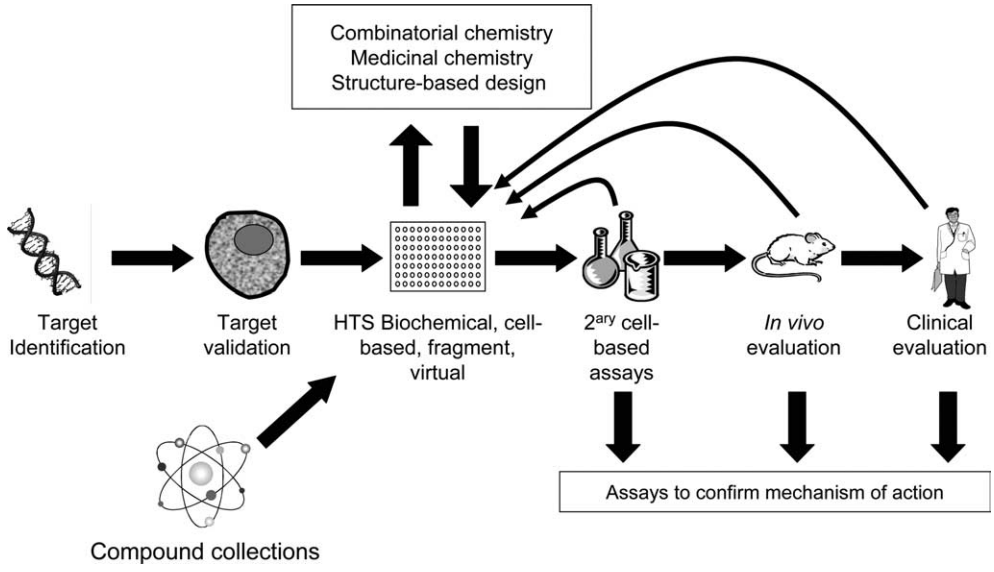

Fig. 2. The process of contempory drug discovery. Gene expression microarray analysis can be valuable at all stages of the process, including target identification and validation, mechanism of action studies and the identification of pharmacodynamic endpoints. Gene expression analysis can also be used as the basis for a cell-based high throughput screen (HTS) to identify compounds that alter gene expression patterns in a particular way. As shown, gene expression profiling *in vitro* cell systems, animal models and the clinical situation can all lead to identification of additional targets that feedback into earlier stages of the drug discovery process. For more details see the text. 2^{ary}, secondary.

Contemporary mechanism-based drug discovery programmes are usually directed towards achieving the essential profile of a clinical candidate that acts on the desired novel molecular target [33]. This defines the necessary potency, selectivity and therapeutic activity along with factors like the route and schedule of administration. A series of assays are assembled which form a biological test cascade (Fig. 2). At the top of the test cascade is the primary screen, usually a biochemical assay with recombinant target protein or a cell-based assay for the target pathway. Testing proceeds through the increasingly demanding levels of antitumour activity in cell culture, pharmacokinetic properties and therapeutic activity in an animal model. Throughout the pre-clinical development phase, and indeed in the early clinical trials, evidence that the drug effect occurs via the desired mechanism of action is extremely important. Furthermore, in the case of cell-based screens it is essential to have mechanistic assays available to help with the identification or 'deconvolution' of the cellular target. Such assays can also provide valuable molecular biomarkers and pharmacodynamic endpoints for use in animal studies and subsequent clinical trials.

Gene expression microarrays are now playing a valuable role at all phases of the drug discovery process, as shown in Fig. 2. Microarrays are providing new insights into the molecular pathology of human cancers and are helping to identify many new additional targets for drug discovery. For target identification, we can simply compare gene expression in the normal and disease tissue. The mechanism that cells employ to regulate the normal processes of signal transduction, cell division and cell death are frequently subverted during tumorigenesis. As mentioned earlier, it is also 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 [34]. Thus, 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 [35]. Changes in gene expression in cancer have been associated with cellular processes such as cell cycle progression, signal transduction and response to stress. Therefore gene expression profiling may allow the regulation and progression of these cellular processes to be followed and may facilitate the identification of genes associated with these processes. In the context of cancer and drug development, the application of microarray analysis to cellular processes, such as cell division or cell signalling, may identify genes involved in tumorigenesis that may be potential drug targets. Alternatively, in clinical classification studies, the products of genes associated with a poor outcome or with the subversion of oncogenic pathways could also be considered as potential targets.

Further along the discovery process, gene expression microarrays can be used to profile the pharmacological effects of lead compounds on a genome-wide basis. This facilitates the discovery of prognostic and pharmacodynamic markers of drug response, helps in the understanding of the molecular mechanism of action of drugs, aids the determination of 'on-target' and 'off-target' effects, and identifies undesirable expression signatures related to toxicity that can be dealt with during the chemical optimisation process. Microarrays are also being used to help identify genes and expression patterns that are associated with drug sensitivity and resistance using *in vitro* models.

The clinical trial is a critical stage of the drug development process. Studies of the molecular mechanism of action alongside the more established toxicity and pharmacokinetic evaluations need to be a strong component in the early clinical testing of agents acting on new molecular targets. Cytotoxic drugs generally have a relatively non-specific mechanism of action and therefore may have a relatively broad spectrum of activity in human cancers. In contrast, with novel drugs targeting a particular oncogenic target or pathway, considerable emphasis has to be placed on patient selection. Even in the case of targeting pathways that are activated in a large percentage of patients and across a variety of different cancers, there will be a considerable number of patients with tumours in which these pathways are not activated and do not contribute to tumour progression. Treating such patients with the molecularly targeted drug would not have any therapeutic benefit; on the contrary, this may cause toxicity and additionally prevent treatment with other drugs that are more likely to be active. Exclusion of patients that are unlikely to respond, but who might benefit from alternative treatments, also generates considerable pharmacoeconomic benefit. Expression profiling can be used to investigate the molecular mode of action of drugs in clinical trials, and to predict which patient is most likely to benefit from which particular drug, aiding the individualisation of cancer treatment. Thus, the ability to obtain quantitative information from a transcriptional profile represents an exceptionally powerful means to explore basic biology, diagnose the disease, facilitate drug development, tailor therapeutics to specific pathologies, and generate databases of biological processes or pathways [25]. These various specific applications of profiling gene expression by microarrays are reviewed in the following sections.

2. Classification of tumours or biochemical pathways and identification of potential targets for novel therapeutics

2.1. Link between diagnostics/prognostics and therapeutic targets

One of the current challenges of cancer treatment is to target specific therapies to pathogenetically distinct

tumour types, as therapeutics that work against one type of cancer are frequently inactive in other types, e.g., drugs active in breast cancer are generally inactive in lung or colon cancers. Moreover, cancers of an apparently similar type vary widely in response to treatment with the same drug. Improvements in cancer classification have been central to advances in cancer treatment; however, it is highly likely that many cancer subclasses are yet to be defined by molecular markers and gene expression profiling of clinical tumour samples by microarrays has the potential to contribute to improve cancer classification. Molecular classification studies will potentially benefit the application of novel agents by identifying:

- sub-groups likely to benefit from treatment;
- new disease targets by comparison of the tumour and its normal counterpart tissue;
- genes associated with a poor outcome or failure to respond that may be potential therapeutic targets.

Gene expression profiling of solid tumours is a considerable challenge as normal tissue content will influence the expression pattern and solid tumours are generally less well classified. In contrast, lymphoid tumours are extensively classified and it is relatively easy to obtain biopsy material. Taking advantage of these facts, key studies of leukaemia and lymphoma have examined a number of hypotheses concerning gene expression profiling. These classification in cancer and have demonstrated the principle of molecular characterisation and offering identified potential drug targets.

2.2. Haematopoietic cancers

Although the distinction between the acute leukaemias is established, there is no single test to distinguish between acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML). Distinguishing between these leukaemias is critical as their treatment regimes are different and mortality from treatment is still a factor. In one of the first cancer classification studies, Golub and colleagues [30] demonstrated that gene expression profiling by microarray coupled with k-NN could be used to develop a class predictor for the ALL/AML distinction. The predictor consisted of known cell surface markers, genes critical for S-phase progression, chromatin remodelling, cell adhesion and some known oncogenes e.g., *C-MYB*, *E2A*, *HOXA9*. Yeoh and colleagues [30,36] followed up this study in paediatric ALL and examined bone marrow biopsies from 6 major paediatric ALL sub-types. A χ^2 metric was used to rank statistically significant genes and identified those that could accurately classify the tumours. Some of the genes identified in this study were of therapeutic interest; for example *C-MER*, a receptor tyrosine kinase, was overexpressed in ALL associated with the *E2A-PBX1* chromosomal translocation and represents a potential therapeutic target for this sub-group.

Leukaemias with rearrangements of the mixed lineage leukaemia gene (*MLL*) are characterised by having a particularly poor outcome with current therapies. Microarray analysis determined that MLL was an entity distinct from ALL and AML [37]. Within the classification set of genes, *FLT3* was highly expressed in MLL, and as a receptor tyrosine kinase, represents an attractive novel target for rationale drug development. Stam and colleagues [38] profiled infant and childhood ALL and noted increased expression of *hENT1*, the nucleoside transporter used by the deoxycytidine analogue Ara-C, in MLL tumour cells. Increased expression of this nucleoside transporter correlated with sensitivity to Ara-C, suggesting a potential explanation for the sensitivity of these tumours to Ara-C treatment and a potential therapeutic strategy.

The T cell lineage of ALL has also been examined in detail as the molecular classification of this class of ALL remains obscure. Ferrando and colleagues [39] compared the expression profiles of T cell ALL to the expression of *HOX11*, *TAL1* and *LYL1* transcription factors using k-NN. The expression profiles were correlated with the degree of differentiation. In addition, increased expression of the anti-apoptotic *BCL2* and *BCL2A1* genes in the *LYL1* group and *TAL1* group, respectively, was consistent with the relative insensitivity of these sub-types to chemotherapy. Further analysis demonstrated that HOX11L2, structurally related to HOX11, was overexpressed in some T-ALLs and corresponded to a sub-group of *HOX11* tumours. This type of study is important as it contributes to the elucidation of pathways of transformation, the components of which may be attractive targets for new therapeutic approaches.

In one of the first significant microarray studies of lymphoma, Alizadeh and colleagues [40] asked whether expression profiling could be used to generate a molecular portrait of B cell malignancy and whether distinct B cell malignancies not recognised by current classification systems could be identified. Initial hierarchical clustering separated the malignancies by class. Further analysis identified 2 sub-types of diffuse large cell lymphoma (DLCL) that had a germinal centre expression pattern or an activated B cell gene expression pattern that was associated with a significantly poorer overall survival. Rosenwald and colleagues [41] followed up this study and constructed a molecular predictor of risk using genes with expression patterns that were associated with survival. Two gene expression sub-groups as described above were identified, together with an additional previously undescribed type 3 DLCL. The germinal-centre B cell-like sub-group had the highest five year survival of 60%, compared with a rate of 39% for patients with type 3 DLCL and 35% for those with activated B cell-like DLCL. A Cox proportional-hazards model was used

to identify individual genes, the expression of which correlated with outcome and a predictor of outcome was built. On the basis of their predictive scores, one quarter of the 240 patients in this study could be assigned to a risk group with a five-year survival rate of 15%. The activated B cell group had increased activation of the nuclear factor-κB (NF-κB) pathway that has been reported to be required for survival of activated B cell-like lymphoma cells [42]. This would potentially block apoptosis induced by chemotherapy and may account for the poor prognosis of this group. These observations suggested that DLCLs may consist of multiple entities and indicated that patients who have a profile indicating a poor prognosis might proceed directly to bone marrow transplantation rather than the potentially debilitating first-line regime. In addition, the results support the view that targeting the NF-κB pathway might be therapeutically beneficial for a subset of DLCL patients.

Shipp and colleagues [43] profiled tumour specimens from DLCL patients who received cyclophosphamide, doxorubicin, vincristine and prednisone-based chemotherapy. They applied a supervised learning prediction method and identified two categories of patients with very different five year overall survival rates (70% *versus* 12%) using a 13-gene predictor. The cyclic-adenosine monophosphate (cAMP)-specific phosphodiesterase *PDE4B*, that catalyses the degradation of cAMP and limits cAMP-dependent protein kinase signalling, was overexpressed in the poor prognostic group. It may be an attractive target as PDE4B inhibitors have been reported to induce apoptosis in B cell lymphoma models [44]. *Protein kinase C*-β was also overexpressed in the poor prognosis group; this protein kinase has a key role in B cell proliferation and inhibitors of protein kinase C-β have shown synergy with chemotherapeutic agents [45].

Follicular lymphoma frequently transforms into a more aggressive DLCL that is biologically distinct from sporadic DLCL. Expression profiling of a cohort of follicular lymphomas and their transformed counterparts yielded a classifier of genes altered in transformed DLCLs [46]. Genes included those encoding growth factor receptors, N-RAS, RAS-related genes and p38 mitogen-activated protein kinase (MAPK), an important mediator of cytokine- or stress-induced cellular responses. Treatment of a transformed follicular lymphoma cell line with a MEK inhibitor (U0126) had no effect. However, treatment with an inhibitor of p38 (SB203580) reduced viability, inhibited tumour growth *in vivo* and induced apoptosis [46]. Transformation was also associated with increased expression of a gene encoding phosphatidylinositol 3 (PI3)-kinase which is a potential therapeutic target, and the receptor tyrosine kinase platelet-derived growth factor receptor β (PDGFRβ) that is inhibited by imatinib mesylate.

These studies in easily accessible leukaemias and lymphoid tumours demonstrate the power of expression profiling using microarrays to classify tumours and in retrospective studies to predict outcome, but not initial response to treatment [30,40]. In addition, it is clear that these studies have also identified potential therapeutic targets. Some of these targets are acted upon by existing agents that have not been previously tested in haematopoietic malignancies, while others are novel targets against which new inhibitors can be rationally developed.

2.3. Breast cancer

Some of the most important breakthroughs and progress in the gene expression profiling of solid tumours have been made in studies of breast cancer. The ability to more accurately predict prognosis in breast cancer patients would improve the selection of patients that might benefit from adjuvant therapy. Chemotherapy or hormonal therapy reduces the risk of distant metastases by 30%; however, 70–80% of patients receive therapy they do not need. Many factors have been correlated with prognosis; but under multivariate analysis their prognostic value frequently disappears.

Perou and colleagues [47] detected a great variation in gene expression in breast tumour biopsies, but following hierarchical clustering, the majority of the pre- and post-treatment samples from individual patients clustered together, and likewise for the primary and lymph node tumours. This implied that the majority of tumour gene expression remains unchanged following either chemotherapeutic treatment or metastatic spread. The analysis also identified 4 tumour sub-groups with expression profiles characteristic of oestrogen receptor (ER)-positive/luminal epithelial cells, basal-like epithelial cells, ERBB2-positive cells and normal breast epithelial cells. Further expression profiling was able to sub-divide the tumours into distinct groups with characteristic expression profiles [48]. These were a low/no ER expression group, a basal-epithelium like group, a group that overexpressed ERBB2 and a normal breast-like group. The ER-positive tumours were sub-divided into 3 luminal subtypes A, B, C; luminal subtype C had an underlying basal-epithelium gene expression pattern. Mutation of *p53* was strongly associated with the basal-like and ERBB2 groups. The SAM algorithm was used to identify genes associated with survival. Kaplan–Meier plots demonstrated highly significant survival differences for tumours with the ERBB2/basal signature, which have the poorest overall and relapse-free survival rates. A follow-up study refined these observations and sub-divided the tumours as before [49].

Another study by van't Veer and colleagues [50] demonstrated that expression profiling could separate ER-positive and ER-negative breast tumours using unsupervised clustering. A leave-one out cross-validation strategy subsequently generated a 70-gene set that

could identify patients with poor prognosis as effectively as the current classification systems. This gene set included those encoding products involved in cell cycle regulation, invasion, metastases and angiogenesis, but did not include genes previously identified from single gene studies. The power of the 70-gene prognosis profile was confirmed in a large follow-up study of 295 consecutive patients with primary breast cancer, where the mean overall survival in patients with a poor prognosis profile was 54.6%, compared with 94.5% for those with a good prognosis signature [51]. Additionally, those in the high-risk group were more likely to develop metastases than those identified by conventional criteria. These observations suggested that the ability to spread is inherent within certain primary breast tumours rather than the clonal development model which proposes that metastatic potential is acquired late in tumorigenesis.

Sotiriou and colleagues [52] reported on a group of unselected breast cancer patients, and profiled 99 of the 700 patients consulted over a 2-year period. Expression profiles divided the tumours by ER-status and could separate the tumours by grade. The ER-negative tumours exhibited a basal-epithelium pattern and two sub-groups, one of which overexpressed members of the activator protein (AP-1) family of transcription factors. The ER-positive group expressed genes associated with ER activation and could be subdivided into three luminal groups with different clinical outcomes. Ninety-three genes could segregate the population into two groups with different survival; this gene set had some overlap with the 70 gene set reported by Van 't Veer and colleagues [50]. Genes overexpressed in the tumours with poor prognosis and that might be therapeutic targets included those encoding products of the glutathione *S*-transferase pathway, the melanoma tumour antigen *PRAME* and other gene products involved in cell cycle regulation.

2.4. Prostate cancer

An expression profiling study of prostate cancer demonstrated that benign conditions of the prostate clustered separately from malignant prostate cancer cell lines or tissues and that metastatic and localised prostate cancer formed distinct sub-groups [53]. The expression profiles were also ranked and identified genes with increased or decreased expression in the tumours. Expression of the *PTEN* tumour-suppressor gene that represses PI3-kinase signalling was decreased. In contrast expression of the proto-oncogene *C-MYC* was increased and co-transcriptionally regulated with the proto-oncogene *PIM1*, suggesting a potential synergistic oncogenic effect in prostate cancer. Expression of hepsin and PIM1 protein demonstrated significant correlation with measures of clinical outcome. PIM1 is an aurora-like serine/threonine protein kinase required for mitosis and hepsin is a transmembrane serine protease that has been implicated in cell growth and development. Luo and colleagues [54] also detected increased expression of *hepsin* and *C-MYC* in primary prostate cancer compared with benign prostatic hyperplasia. Using multidimensional scaling and weighted gene analysis, *hepsin* expression had the greatest ability to differentiate benign prostatic hyperplasia from cancer.

A study of normal prostate cells, transformed normal derivatives, prostate cancer cell lines, prostate stromal cells, prostate cancer biopsies and benign hyperplasias identified highly expressed genes that were potential pathological markers or therapeutic targets [55]. Among the highest scoring genes were the carcinoma-associated antigen *GA733-2* (a Ca^{2+}-independent cell adhesion molecule that is a potential immunotherapy target), *fatty acid synthase*, a secreted cytokine of the transforming growth factor-β (TGF-β) superfamily, *PLAB* and *hepsin*. A more recent study ranked gene expression using a variation of signal-to-noise metric and identified genes highly expressed in tumour or normal tissues [56]. No individual gene could predict outcome; however, a five gene model identified by k-NN could predict recurrence. The five genes were *chromogranin A* (a neuroendocrine secretory protein), the receptor tyrosine kinase *PDGFR*-β, a homeobox transcription factors *HOXC6*, *inositol triphosphate receptor 3* (an inositol-1,4,5-triphosphate regulated calcium channel) and glycosyltransferase family member *sialyltransferase-1*. Rhodes and colleagues [57] used meta-analysis of data from four studies and identified two high scoring genes, *fatty acid synthase* and *hepsin*. Metabolic pathway data were integrated with the transcriptional data and identified increased transcription of the polyamine pathway and purine biosynthesis pathway genes as potential targets. Fatty acid synthase may be of interest as previous studies have demonstrated that pharmacological inhibition of fatty acid synthase in prostate xenograft tumours results in dose-dependent growth inhibition [58]. These studies also suggested PI3-kinase, PIM1 and hepsin as potential targets.

2.5. Lung cancer

Lung cancer is divided into four major histologic subtypes that define distinct treatment regimens. Of particular importance is the pathological difference between small cell and non-small cell lung cancer (NSCLC) that is further sub-divided into adenocarcinoma, squamous and large cell tumours. Similar to the breast cancer studies described earlier, expression profiling has divided lung cancers into sub-types [59]. The adenocarcinomas were sub-classified into three groups that exhibited different cumulative survivals. Another study classified adenocarcinomas into four subtypes, one containing poorly-differentiated tumours, two groups with moderately/well-differentiated tumours and a group of tumours with a neurondocrine-like gene expression pattern that

have significantly poorer outcome [60]. Expression profiling of stage I and III adenocarcinomas identified 3 clusters, including a normal lung tissue cluster, a cluster of well-differentiated tumours, and a cluster of stage 1 and III tumours with the highest number of poorly-differentiated tumours [61]. A 50-gene risk index could separate the stage 1 tumours by survival. Genes associated with poor survival included *ERBB2*, *REG1A*, *VEGF* and *CRK*. The *CRK* oncogene encodes an adaptor protein that interacts with C-JUN N-terminal kinase, a regulator of matrix metalloproteinase secretion and cell invasion. Another study has also used expression profiling to examine survival and noted that the receptor tyrosine kinase *FLT1* was among the genes predictive of outcome [62]. FLT1 is the receptor for the angiogenic growth factor vascular endothelial growth factor (VEGF) and a number of inhibitors of this receptor are currently under clinical development [63].

2.6. Ovarian cancers

Of the gynaecological tumours, epithelial ovarian tumours have the worst prognosis; survival with distant metastases is poor compared with patients that have localised tumours. Expression profiling identified genes overexpressed in tumours and underexpressed in normal tissue [64,65]. Genes consistently overexpressed in tumours that represent potential immunotherapeutic or pharmacological targets included a secreted protease inhibitor (*HE4*), a cell surface antigen (*CD24*), an anti-apoptotic protein with a proposed chaperone function (*clusterin*), *fatty acid synthase* and *14-3-3σ*.

2.7. Gastrointestinal tumours

A number of studies have compared normal colon and colorectal cancers. These have identified gene expression patterns associated with tumour or normal tissues and have also revealed genes associated with Dukes classification, as well as genes linked to disease progression, mainly from normal to early stage cancer [66–72]. The expression levels of genes classified within a given functional category showed remarkable correlation. Few genes encoding proteins involved in apoptosis and signal transduction had altered expression, rather most of the genes showing increased expression encoded products involved in nucleic acid metabolism, cell cycle regulation, translation, adhesion, or proteolysis, whereas those involved in membrane and protein trafficking and lipid metabolism, together with most of the kinases and phosphorylases, exhibited decreased expression. Interestingly, expression of *MGSA-α*, a chemokine frequently overexpressed in melanoma, was associated with progressive disease.

Although no investigations have addressed outcome in relation to expression profile in colorectal cancer, some studies have examined outcome in other gastrointestinal tumours. Kihara and colleagues [73] profiled oesophageal tumours from patients who were to receive cisplatin/5-fluorouracil (5FU) treatment and identified genes statistically correlated with sensitivity. Within this gene set, *GST-π* and *s-adenosylmethionine* were associated with drug resistance and were highly expressed in patients with <12 months survival. The gene encoding ornithine decarboxylase, the enzyme that regulates the rate-limiting step in polyamine synthesis, also showed increased expression.

2.8. Melanoma

The molecular pathology of melanoma remains unclear, the genetic changes responsible for progression are poorly characterised and the heterogeneity in response to treatment and clinical course is unexplained. Bittner and colleagues [74] profiled gene expression in melanoma biopsies that were passaged in culture or analysed straight from biopsy, but could not identify clinical or tumour cell characteristics specifically associated with the clustering pattern. Among the best predictive genes for determining different sub-classes of melanoma was *WNT5A* that is associated with increased motility and invasiveness [74–76]. Overexpression of *WNT5A* in melanoma cells resulted in increased invasive ability. Interfering with the WNT5A/frizzled-5 receptor signalling pathway using an antibody decreased signalling through protein kinase C and reduced motility, suggesting a potential therapeutic target. Another study in melanoma identified 32 genes apparently specific for metastases; included in this gene set were genes that regulate cytoskeletal organisation and cell migration [77]. Expression of one gene, *RHOC*, could stimulate motility and invasiness that was blocked by expression of a dominant-negative *RHOC*. RHO can activate protein kinase C-μ, suggesting that signalling through protein kinase C or RHO may be a potential therapeutic strategy [78].

2.9. Neuronal tumours

Beyond surgical resection, treatment options for patients with glioma are currently limited to DNA-damaging agents, such as carmustine or temozolomide. Expression profiling has detected gene expression patterns significantly associated with tumour grade including the overexpression of genes such as *epidermal growth factor receptor* (*EGFR*), *MDM2*, *cyclin-dependent kinase 4* (*CDK4*), *CD44* and *insulin-like growth factor binding protein 2* (*IGFBP2*), a number of which are targets of novel agents currently under clinical evaluation [63,79]. Another study of glial tumours has detected increased expression of *EGFR*, *AP2*, *VEGF*, *connective tissue growth factor*, *insulin-like growth factor 2* (*IGF2*), *IGFBP3*, *IGFBP5* and genes encoding extracellular matrix and cytoskeletal proteins [80].

Medulloblastomas exhibit a high degree of variability in response to chemotherapy or radiotherapy. There are no predictive molecular markers, but patients with localised disease have a better outcome. Pomeroy and colleagues [81] generated a molecular taxonomy that could separate gliomas from medulloblastomas and the medulloblastoma sub-types. A cluster of ribosomal protein genes was associated with poor outcome following unsupervised analysis; this signature was also correlated with an increased number of ribosomes on examination by electron microscope. k-NN and other classification methods could also accurately predict outcome and found additional genes associated with favourable or poor outcomes. A class prediction algorithm implicated PDGFR-α and the downstream RAS/MAPK pathway in progression from non-metastatic to metastatic medulloblastoma [82]. *In vitro* assays demonstrated that PDGF-α stimulated migration and activated downstream mediators of the MAPK pathway. Neutralising antibodies to PDGF-α or U0126, a specific MEK inhibitor, blocked migration.

2.10. Comments on expression profiling of human tumours

Gene expression profiling of normal and pathological tissue is now widely used as part of the identification and validation of biomarkers and new molecular targets. In addition to the published studies discussed above, a number of commercial enterprises have set up subscription databases of gene expression profiles of 100s to 1000s of tumour and normal tissue samples. However, much information on target identification through data-mining of these proprietary databases is not publicly available and, as such, cannot be dealt with here other than in principle. There are a number of caveats to be made regarding studies comparing tumour biopsies and normal tissue. Altered gene expression may not be the cause of the pathology, but could for example be a random event or a consequence of co-amplification with increased expression of another gene that is causally involved, or a result of altered physiology. Biopsies are generally not homogeneous since, depending on the biopsy type, there will be underlying gene expression patterns from normal tissue and components such as stroma, immune system and vasculature. However, despite these reservations, it is clear from the studies reviewed in this section that comparisons of normal and tumour biopsies can identify potential new targets for rational drug development that will require further validation.

The numerous examples given above also provide evidence that differences in expression profiles indicate tumour subclasses that are molecularly distinct, and have arisen from different transforming genetic aberrations. Although they share the same tissue of origin, distinct molecular pathways drive the malignant phenotype in these subclasses, and will likely need different agents targeted at the activated molecular pathways to inhibit tumour progression. Patients with tumours that show activation of a given pathway might be particularly suited for treatment with agents that inhibit this pathway. Generally, for the genes identified in the classifiers, it was not clear whether their gene products are mechanistically associated with outcome or are simply statistical associations. However, overexpressed genes associated with a poor outcome or disease pathology are clearly potential targets for new therapeutics which should be explored and validated.

Although the early studies reviewed here are very encouraging, the clinical usefulness of gene expression profiling has to be established in large prospective studies. It is significant that the European Organisation for Research and Treatment of Cancer (EORTC) has initiated a prospective study comparing breast cancer classification by expression profiling to the 'gold standard' classification parameters currently used in the clinic [83]. In this study, half the patients will be assigned to more aggressive therapy if the expression signature predicts high-risk, while the other half will be assigned to standard or aggressive therapy using conventional classification. The next step would be to use this technique to profile tumours before treatment, in order to determine which oncogenic pathways are active within individual cancers and thus to identify the most effective and logical treatment strategy.

2.11. Expression profiling of in vivo animal tumour models

Expression profiling of transgenic models of human cancer and of experimental metastatic cancers has furthered the understanding of the biological significance of the human profiling data. This is particularly important in the identification and validation of targets suitable for the development of novel agents.

Expression profiling of a transgenic mouse prostate cancer model restricting *C-MYC* expression to the prostate has confirmed potential targets for therapy identified in clinical studies [84]. Gene expression profiling could distinguish between normal and *C-MYC* transgenic prostates and generated a ranked list of genes overexpressed in the tumours. Highly ranked genes included *PIM1*, *L-MYC*, *TMPRSS2* (a serine protease related to hepsin), *SPARC* (an anti-adhesive protein differentially expressed during prostate cancer progression) and *EGF*. Comparison of the *C-MYC* transgenic mouse prostate tumour signature with published proliferation signatures did not reveal a significant overlap, but could classify the transgenic mouse prostate tumours into low-*MYC* and high-*MYC*. A classification gene set that could distinguish the low-*MYC* and high-*MYC* tumours was applied to three published prostate cancer databases and demonstrated co-transcriptional regulation of *PIM1* with *C-MYC*. As described earlier,

increased expression of *PIMl* mRNA or protein correlated significantly with measures of poor outcome in prostate cancer [53]. The *PIM1*/*C-MYC* association was also detected in breast and ovarian cancer datasets. This study illustrates that carefully designed transgenic mouse models can be used in combination with gene expression microarray profiling to reveal common features with human disease and identify significant targets such as PIM1 that may be exploited for drug discovery.

Comparison of hormone-sensitive and hormone-refractory human prostate cancer xenograft pairs by gene expression profiling identified the *androgen receptor* as the only gene differentially expressed in all the hormone-refractory tumours [85]. Expression profiling experiments of cell lines overexpressing the androgen receptor indicated that high expression of the androgen receptor resulted in the conversion of androgen receptor antagonists into agonists. The effect was weak as only a limited number of androgen-responsive genes were induced, suggesting that the receptor could not bind the full complement of transcriptional coactivators. These observations have implications for prostate treatment and perhaps for other hormone-dependent diseases and suggest that novel therapeutics that interfere with translocation of the receptor to the nucleus or its assembly into a transcription complex may have great potential.

Kang and colleagues [86] compared parental MDA-MB-231 breast cancer cells and sub-populations with enhanced metastatic abilities *in vivo*. The parental cell line had a gene expression profile similar to the poor prognostic group defined by the 70-gene predictor described earlier, consistent with the cell line being derived from a metastatic tumour [57,58]. The gene set specifically associated with bone metastases did not include genes in the 70-gene set, implying that the bone metastasis signature was superimposed on the bad prognosis gene set. Many of the genes encoded either secreted products or products located on the cell plasma membrane that would encourage growth in the bone environment. These included genes encoding an activator of osteoclast differentiation, *CXCR4* (a bone-homing chemokine receptor), angiogenesis factors, *MMP1* (promotes osteolysis), *IL11* (a potent inducer of osteoclast formation), and *osteopontin* (a factor that stimulates adhesion of osteoclast to the bone matrix). Cells engineered to overexpress *IL11* and *osteopontin* promoted bone metastases and cells expressing *IL11*, *osteopontin* and either *CXCR4* or *connective tissue growth factor* had greatly increased bone metastases. Dilutional cloning of the parental population *in vitro* followed by expression profiling identified a subset of cells that exhibited the bone metastatic gene expression profile and formed metastases *in vivo*. This observation demonstrated that cells existed within the population preprogrammed to express genes that would be required for bone metastases, an observation consistent with that proposed to occur in the clinic [50,51].

These studies show that gene expression profiling in *in vivo* animal models can be used to provide an experimental system in which data contained within human tumour databases can be overlaid and allows the re-evaluation of the clinical data. This approach can be used to clarify the role of individual genes within multi-gene predictors and, importantly, to validate novel targets suitable for rationale drug discovery. Examples include the identification of PIM1 in prostate cancer, a serine-threonine protein kinase that would be an attractive target for rationale drug discovery, or the overexpression of the androgen receptor in hormone-refractory models.

2.12. Expression profiling of in vitro models

As emphasised earlier, when developing agents targeting specific molecular targets and pathways it is imperative to define tumours that have become reliant on the activation of the target and its cognate pathways for maintaining the malignant phenotype. With the advance of microarray gene expression profiling, it is now possible to derive molecular signatures that are associated with activation of certain oncogenic pathways that are key in tumorigenesis. Not only will this help to identify tumours suitable for a particular treatment with drugs that act on these oncogenic pathways, but, in addition, genes induced by particular pathways associated with oncogenesis may also be potential therapeutic targets.

Later in this Review, we will discuss expression profiling following exposure to exisiting therapeutics to define molecular fingerprints that can be used to determine whether novel agents designed to target the pathway of interest are acting in the intended way. However, with the identification of novel targets, there will frequently be no small molecule inhibitors to establish an expression profile for comparison. In this instance, alternative strategies need to be employed. One approach is to overexpress the drug target or pathway regulator in a suitable model system and develop a gene predictor for the pathway. This approach may be especially valid for deconvoluting "hits" in cell-based drug discovery screens.

Passage of cells from G1 to S-phase of the cell cycle is a key pathway that is regulated by the Retinoblastoma (*RB*) tumour suppressor protein and is frequently deregulated in cancer. RB is phosphorylated by CDKs. Phosphorylation of RB releases sequestered E2F transcription factors that induce the expression of additional cell cycle regulators and proteins required for DNA replication. As such, this pathway is an attractive therapeutic target, for example, in the development of CDK inhibitors such as flavopiridol, UCN-01 and CYC202 that block RB phosphorylation ([63] and see below). The D-type cyclins are important regulators of the CDKs that phosphorylate RB. Gene expression

profiling has been used to understand the role and function of cyclin D1 [87]. Twenty-one genes were found to be upregulated >3-fold by expression of wild-type or mutant *cyclin D1* in MCF-7 breast cancer cells. This gene set did not resemble the gene set regulated by *E2F* transcription factors or induced by *cyclin D3*. The gene expression profile was compared with different databases of tumour gene expression using a mathematical approach known as Kolmogorov–Smirnov scanning. *Cyclin D1* was ranked at number 26 in the ordered list of all 16 063 genes measured and, more significantly, eight highly ranked genes in this list were also identified as markers of luminal epithelial tumours that frequently overexpress *cyclin D1*. These observations implied that this approach had captured an important biological function unique to cyclin D1 (when compared with another D-type cyclin) that was independent of CDK activation. Further analysis of other datasets suggested that *C/EBP*-β alone could mediate this effect. A dominant-negative C/EBP-β and a *C/EBP*-β knockout cell line gave the same profile as the mutant cyclin D1, suggesting that one function of cyclin D1 in tumorigenesis is to antagonise the transcriptional repressor function of C/EBP-β. Given that *cyclin D1* is frequently overexpressed or amplified in cancers, for example breast cancer and some lymphomas, gene products induced by the CDK-independent activity of cyclin D1 could be attractive therapeutic targets as the expression profiling data suggest that this novel mechanism is active in human tumours.

An alternative approach to identifying an RB pathway-regulated signature is to express *p16*, a non-phosphorylatable form of *RB* or different *E2F* isoforms in cell lines [88,89]. Genes regulated by this pathway included those encoding products involved in proliferation, differentiation and development. In addition, increased expression of pro-apoptotic genes, including *APAF1* and various caspases, were detected. The expression profile detected had some overlap with those patterns induced by serum stimulation, by receptor tyrosine kinases, p53, C-MYC or RAS.

Transcription factors are frequently downstream of signalling pathways that are subverted during oncogenesis and are responsive to new therapeutics that target receptor tyrosine kinase activity e.g., EGFR/gefitinib and PDGFR-β/imatinib mesylate. A number of studies have profiled gene expression following responses mediated by external signals. The response of normal human diploid fibroblasts to serum was rapid, with genes such as *C-FOS* and *JUNB* being induced within 15 min of exposure to serum [90]. The timing of gene expression was coincident with progression through the cell cycle, and distinct clusters of genes involved in cell division were identified. The induction of immediate early genes by receptor tyrosine kinase-activated signalling pathways has been examined in NIH3T3 cells expressing a fusion protein consisting of the cytoplasmic portion of PDGFR-β fused to the extracellular portion of the macrophage colony-stimulating receptor [91]. Sixty-six genes, including previously identified immediate early genes, were induced by more than 3-fold following a few hours exposure to macrophage colony-stimulating factor. Interestingly, a mutant form of the receptor that lacks binding sites for phospholipase C (PLC)-γ 1, PI3K, SHP2 and RASGAP adaptor proteins could still induce these genes. However, removal of a final site bound by Grb2 greatly reduced the induction of these genes. There was some evidence for specificity, as restoring a RasGAP binding site on the receptor resulted in the induction of genes regulated by interferon-γ. Thus, the gene expression microarray profiling suggested that the distinct pathways emanating from growth factor receptors exhibited a certain degree of functional redundancy for the induction of immediate early genes, but also indicated that diverse pathways could exert overlapping effects on their induction.

Downstream effectors of signalling pathways have also been examined by gene expression microarray profiling, for example, the protein kinase cascade downstream of receptor tyrosine kinase receptors that are also therapeutic targets [63]. As a case in point, induction of *C-RAF* expression in a spontaneously immortalised normal mammary epithelial cell line did not induce transformation, but protected cells from apoptosis induced by detachment from the growth substrate [92]. Gene expression analysis detected induction of immediate early genes such as *JUNB*, *C-FOS*, *cyclin D1* and soluble factors that can activate the EGF receptor and induce phosphorylation of ERK1/2 and C-AKT.

A number of immediate early genes regulated by the receptor tyrosine kinase-mediated signalling pathways are themselves transcription factors, for example *EGR1*, a gene frequently overexpressed in prostate tumours. Expression of EGR1 in a prostate cancer cell line results in increased expression of a number of genes involved in neuroendocrine differentiation and several growth factors (*PDGF*-α, *IGF-2* and *TGF*-β1) [93]. These observations suggested an early role for EGR1 in prostate malignancies. The JUN family of transcription factors are intermediate early genes that may have a key role in tumorigenesis. Expression of *C-JUN* and *JUNB* expression results in immortalised cells that exhibit anchorage-independent growth [94]. Expression of *JUND* did not alter morphology, but inhibited proliferation. Expression profiling identified genes altered following the expression of any JUN protein and others that exhibited JUN-specific regulated expression [94].

A more recent approach to classify biochemical pathways is to generate a 'metagene', a combination of individual genes that describes a particular pathway or gene activity. Huang and colleagues [95] expressed *HA-RAS*, *C-MYC*, *E2F-1*, *E2F-2* or *E2F-3* in quiescent fibroblasts

and identified metagenes that could distinguish cells transformed by *C-MYA*, *HA-RAS* or the different *E2F's* from quiescent cells. Metagene analysis following serum stimulation of mouse embryo fibroblasts correctly predicted early activity of C-MYC and HA-RAS, whereas E2F activity was evident at later time-points. The metagenes were also tested *in vivo* by profiling of mammary tumours from transgenic mice expressing *HA-RAS* or *C-MYC* and comparing with the results obtained for tumours from *ERBB2* transgenic mice and normal mammary tissue. The *C-MYC* metagene predicted activity with a high degree of confidence and separated C-MYC-driven tumours from the other cancers.

The above studies demonstrate that it is possible to identify a particular molecular signature associated with activation of defined biochemical pathways or transformation associated with a particular oncogene. This approach can also identify novel functions of established proteins, for example, the CDK-independent activity cyclin D1 [87]. Significantly, the identification of genes regulated by these pathways may in turn be potential targets of novel anticancer therapeutics.

The identification of a molecular signature can also be utilised in cell-based chemical biology screening in which compounds could be identified that altered gene expression in a particular, therapeutically advantageous way. For example, in a project looking for novel agents that specifically reversed C-MYC transformation, it would be possible to screen for compounds that interfere with the expression of the C-MYC metagene mentioned above. Expression profiling by microarray may not be sufficiently amenable for high throughput approaches; however, strategies such as pooling compounds or reducing the molecular signature or classifier to a multiplexed RT-PCR-based analysis can improve throughput.

2.13. Knockout strategies and expression profiling

One drawback of studies discussed in the last section is that they are frequently restricted to primary cell lines such as fibroblasts. This may not accurately reflect cancer cell biology and/or may rely on abnormally high levels of expression that may obscure subtle functional differences by overwhelming normal regulatory mechanisms. An alternative approach is to pursue an inhibition or knockout approach where expression of the target or pathway is reduced. This strategy has the advantage of specificity and can potentially provide a 'clean' profile of inhibition that is unlikely to be achieved currently with a small molecule inhibitor. The purest approach within this strategy is to use genetic knockouts. The yeast, *S. cerevisae* is an ideal model system to investigate the value of this approach as the entire gene complement of this organism has been established and has been subjected to extensive mutational analysis. Knockout approaches can give functional information regarding the consequences of inhibiting the activity of a particular drug target; alternatively, one can compare the genetic knockout to pharmacological inhibition of the target and gain information regarding the specificity of the inhibitor by determining the extent to which the genetic and chemical knockouts are similar. However, it should be noted that removal of a component of a multi-protein complex using knockout approaches may result in different effects from simply inhibiting functional activity.

Studies in *S. cerevisae* have focused on mutations of specific sets of pathways or processes. Roberts and colleagues [96] profiled gene expression during the pheromone response of wild-type and mutants of the different mitogen-activated protein kinases involved in a number of signal transduction pathways. They found subsets of co-expressed genes that reflected the activity, cross-talk and overlap of the different signalling pathways, particularly two distinct mitogen-activated protein kinase mutants that revealed overlap between filamentous growth and mating responses. Cell cycle-dependent mRNA fluctuation has been described in numerous settings and has been profiled at regular intervals following release of *CDC28* and *CDC15* mutants from their cell cycle block. This identified genes with periodic cycling consistent with cell cycle phase-specific expression [97]. Gene expression changes in the *CDC28* mutant have also been compared with those induced by the CDK inhibitors flavopiridol or a trisubstituted purine. These agents have some selectivity towards CDK2/cyclin complexes and inhibit the activity of CDC28 [98]. A number of genes showed altered expression both in the *CDC28* mutant and the small molecule-treated cells. However, another set of genes were altered only in the *CDC28* mutant and were not affected by the CDK inhibitors and *vice versa*. The differences suggest that the small molecule CDK inhibitors may have off-target effects or, as discussed above, that there may be an intrinsic difference between chemical and genetic manipulation of a target. The inclusion of an inactive analogue of the CDK inhibitor in these experiments was important as few changes were detected, suggesting that chemotype effects were minimal and that the gene expression changes that were detected resulted from kinase inhibition.

Marton and colleagues [99] also used *S. cerevisae* as a model organism to compare mutations and known inhibitors of the same pathway. They examined the calcineurin-signalling pathway, which can be inhibited by FK506 or cyclosporin A, and HIS3 which is inhibited by 3-aminotriazole. In both cases, there was a correlation in the gene expression profile between drug inhibition of the target and inactivation of the target by mutation. A more sophisticated approach using a 'decoder' strategy was also developed. Initially the gene

expression profile of drug-treated wild-type cells was compared with a panel of mutant strains. The mutant strains with the most similar expression profile are selected and treated with the drug. For a 'perfect' drug with absolute single target specificity, treatment of a mutant strain that lacks the drug target should not alter the gene expression profile. However, in reality no drug is perfect in this way and hence any changes in expression profile detected in the drug-treated mutant will give clues to off-target effects. Treatment with FK506 gave a very similar profile to the *calcineurin* mutants; subsequent treatment of these mutant strains with FK506 altered gene expression in a manner corresponding to an off-target effect. This profile corresponded to an effect dependent on the GCN4 transcriptional activator, thereby identifying another potential locus of action.

Hughes and colleagues [100] extended this approach using a "compendium" of expression profiles to demonstrate that the expression profile of a mutation serves as a molecular phenotype that predicts phenotype defined by more conventional biochemical assays. They achieved this by creating a reference database of expression profiles from full-genome expression experiments of *S. cerevisae* mutated in 276 characterised and uncharacterised genes using a single growth condition. Several classes of co-regulated genes were identified, e.g., those involved in mitochondrial respiration. A general observation was that different mutants that affected the same pathway frequently exhibited similar expression profiles. Additionally, cells treated with a small molecule inhibitor showed a similar profile to the mutation of the target. For example, inhibition of HMG-CoA reductase by lovastatin gave a similar profile to an *HMG2* mutant. This study also demonstrated that the method can be used to assign a putative function to genes and several examples were confirmed by biochemical analysis. An example was dycyclone, an anaesthetic, which gave an expression profile that resembled mutations affecting the egosterol pathway. Biochemical analysis confirmed that this pathway was affected and one mutant, *ERG2*, was hypersensitive to the drug, while overexpression of *ERG2* resulted in decreased drug sensitivity. 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 genes.

Application of knockout approaches in mammalian cells is more challenging. One major challenge will be the production of 'targetless' cells that are either permanent knockouts or are conditional knockouts that can be switched on and off. Examples of using the approach are to mimic the pharmacological modulation of the target in mammalian cells by the construction of stable knockouts using homologous recombination, transfection of dominant-negative constructs, the use of anti-sense constructs, anti-sense oligonucleotides, ribozymes or small interfering RNA (RNAi), or the microinjection of dominant-negative proteins or neutralising antibodies.

The metagene strategy discussed earlier has also been described in the knockout setting using *RB* and *p107*/*p130* knockout mouse embryo cell lines [101]. These proteins are key regulators of E2F transcription factors. Genes characterising a *RB*-null state included those required for DNA replication and cell cycle regulatory proteins. In contrast, genes characterising cells lacking *p107*/*p130* included those involved in regulating proliferation and a second group that maintain the in interaction between the cell and its extracellular environment. The observations were consistent with RB controlling cell cycle progression versus a role for p107/p130 in controlling entry or exit from a quiescent state. The metagene was then applied to an $RB^{+/-}$ mouse model in which animals develop pituitary and thyroid tumours. These tumours were compared with pituitary tumours that develop in mice expressing a truncated isoform of the fibroblast growth factor 4 receptor. The *RB* metagene, but not the *p107*/*p130* metagene distinguished between tumours driven by *RB* loss and those driven by fibroblast growth factor signalling. This study demonstrated that it is possible to develop a molecular signature associated with a knockout that could subsequently be used to compare with a novel inhibitor or for deconvoluting the mechanism of action of a compound with an unknown mechanism or target.

An alternative to the stable genetic knockout approach is to use strategies such as treatment with anti-sense oligonucleotides that sterically hinder the translation apparatus or induce RNase H-mediated cleavage of the target mRNA. This approach has the advantage of being relatively quick to establish compared with the time taken to isolate a recombinant knockout clone. Moreover, in some ways, the transient nature of a knockout by treatment with an anti-sense oligonucleotide mimics inhibition by treatment with a small molecule. However, care has to be taken, as there are known challenges with design, uptake and specificity associated with the use of anti-sense oligonucleotides.

We have used rationally designed anti-sense oligonucleotides to explore the role of *KI-RAS* in colorectal cancer, as we had previously demonstrated that acquisition of mutated *KI-RAS* significantly increases the risk of recurrence and reduces overall survival in this disease [102,103]. Specific inhibition of *KI-RAS* expression by anti-sense oligonucleotides decreased constitutive phosphorylation of ERK1/2, but not C-AKT, and decreased VEGF secretion. Gene expression profiling by microarray analysis detected altered expression of a number of genes following anti-sense treatment, including genes

encoding products involved in protein trafficking and cell adhesion. The results suggested that, at least in this model, activation of KI-RAS may contribute to malignancy predominantly through effects on angiogenesis, invasion and metastasis, and that therapies directed at KI-RAS may have particular utility through inhibiting these processes.

An alternative to employing anti-sense oligonucleotides involves the use of RNAi that can be delivered as double-stranded RNA oligonucleotides or expressed as hairpin RNAs following gene transfer. The RNAi strategy is based on a conserved biological mechanism that silences homologous genes in response to small double-stranded RNA. Although we are only just beginning to appreciate the mechanistic complexity of this response, RNAi has already been developed as a means to manipulate gene expression experimentally and to probe gene function on a whole-genome scale, as in *C. elegans* [104].

Two recent studies have examined the potential of coupling RNAi with gene expression profiling in mammalian cells. The first study explored the principle of RNAi and its potential specificity [105]. Cells expressing green fluorescence protein from jellyfish (*GFP*) were either transfected with two RNAi oligonucleotides specifically designed to target *GFP* expression or with two control RNAi sequences. One oligonucleotide reduced *GFP* expression, but did not significantly alter the expression of other genes when the data were analysed by SAM. A second study targeted *RB*, *C-AKT* and *PLK1* expression in NSLC carcinoma cells [106]. A concentration of 100 nM RNAi resulted in the non-specific induction of large number of genes that were common to all the oligonucleotides. However, transfection at a lower concentration of RNAi regained specificity. For each individual gene, transfection with all five different RNAi sequences resulted in target-specific gene expression profiles. Gene expression patterns unique for each target were identified; however, even at the lower doses of RNAi, 14 genes were commonly altered by all the RNAi oligonucleotides, suggesting that an underlying background non-specific effect was still evident.

The powerful combination of expression profiling coupled with RNAi technology may help with the identification and characterisation of new targets for therapy and expression profiles associated with their inhibition may provide molecular signatures for use in the rationale design of inhibitors. To help accomplish this, Cancer Research UK and the Netherlands Cancer Institute, have started an initiative to probe gene function genome-wide using RNAi [107]. However, as with most technologies, there are some challenges and we have to bear in mind that this technology may not be the predicted panacea. Depending on the study, 30–50% of the RNAi tested are inactive, the inhibition is frequently incomplete, 12–24 h is required for the knockdown to be achieved and there are instances where different RNAi species targeting the same gene can alter biological effects such as proliferation or cell death to different extents. Similar to small molecules, there may be off-target effects associated with RNAi that could result from cross-hybridisation or aptamer binding to proteins. Careful experimental design is therefore essential, involving the use of 3–5 individual RNAi and accompanied by at least one appropriate control mismatch or scrambled RNAi sequence together with appropriate transfection controls.

A good example of the combined use of gene expression profiling and RNAi to identify and validate a potential cancer drug target is the study of Verambally and colleagues [108]. Using a cDNA microarray, 55 genes were found to be significantly overexpressed in clinical metastatic prostate cancer relative to localised prostate cancer. The most significantly overexpressed gene was *EZH2*. This gene encodes a member of the polycomb family of proteins that act in large protein complexes as chromatin-associated transcriptional regulators and contains a highly conserved SET motif which is found in methyltransferases. Increased EZH2 protein was also detected with prostate cancer progression using a tissue array. Disruption of EZH2 expression using RNAi resulted in reduced proliferation of prostate cancer cells in tissue culture. Microarray analysis of cells transfected with EZH2 detected only repression of gene expression, an effect that required the SET domain of EZH2. These observations suggested that EZH2 may be an excellent target for drug development in prostate cancer.

3. Basal gene expression profiling and response to therapeutics

With the development of targeted molecular therapeutics, additional emphasis has to be placed on defining factors that confer sensitivity and resistance. It will be critical to be able to predict which drugs will be effective in sub-groups of individual patients. As described above, we have seen how microarrays can be used to predict clinical response and to describe the activity of biochemical pathways. Herein, we describe how these approaches are being used to classify existing chemotherapeutics for which the precise molecular mechanisms of action are sometimes unclear, as well as to study the mode of action of novel agents. In particular, the use of gene expression microarrays for identifying genes involved in drug sensitivity or resistance is illustrated.

One strategy is to establish basal expression profiles for a panel of cell lines, and also to measure cytotoxicity or another cellular parameter for compounds in the same panel. It is then possible to compare the biological activity of the compound to the basal expression profile.

This type of study has the potential to classify the mechanisms of action of known and unknown compounds and to identify factors that govern sensitivity. This allows one to not only identify resistance/sensitivity factors, but also to identify potential on- and off-target mechanisms.

In one of the first examples of this type of study, Ross and colleagues [109] profiled gene expression of the US National Cancer Institute (NCI) 60 human tumour cell line panel and Sherf and colleagues [110] analysed these data to search for relationships to drug sensitivity. The gene expression data were correlated with cytotoxicity following 48 h drug exposure to the compounds. The clustering pattern of 1376 genes in relation to the activity of 118 agents was analysed and this separated the agents by putative mechanism of action. The antifols, purine analogues and pyrimidine analogues were separated, while the alkylating agents were separated into guanine N-7 reactive nitrogen mustards as distinct from ethylamines, nitrosoureas and alkyl alkane sulfonates that alkylate DNA in a different way from the mustards. The topoisomerase inhibitors were arranged in a manner that revealed mechanistic differences among the subclasses of compounds. The topoisomerase I inhibitors clustered by whether or not they required activation, while the topoisomerase II inhibitors separated into an anthracycline node and another node that contained mitoxantrone and the bioreductive alkylating agents. This study also used a clustered image map to visualise the data and to summarise the relationship between drug activity and gene expression. Two significant examples of drug/gene expression correlations, 5-FU and *dihydropyrimidine dehydrogenase* expression and L-asparaginase and *asparagine synthase* expression, were also cited [110]. Observations of these known correlations provided validation for the microarray approach.

A more recent critical analysis has examined three datasets corresponding to the NCI 60 human tumour cell line panel, including the above dataset, and the relationship of these to the corresponding cytotoxicity data [111]. There was substantial variation between the three datasets obtained with different microarray platforms; however, 376 genes survived the stringent filtering. Gene-agent relationships were suggested by the microarray data, where individual gene expression was correlated with cytoxicity. An assumption was made that this relationship was caused by direct gene–agent interaction. Based on this assumption, ligand–target crystallographic data from Protein Data Bank were used to compare the structure of known targets and ligands to the structure of compounds deposited in the NCI database. Eleven interactions were identified using this approach. These included a relationship between *calcium/calmodulin-dependent kinase 1* and staurosporine, a known ligand of this kinase. Renal cancer cells were the most sensitive cell type to staurosporine and expressed high levels of *cAMP-dependent protein kinase-α*, a kinase structurally related to calcium/calmodulin-dependent kinase 1, which also binds staurosporine. MAP2K4, an activator of stress-activated kinases that has structural homology with the two kinases, had a relationship with a novel agent that although structurally distinct from staurosporine had some similar fused-ring features suggesting a common pharmacophore. Other relationships between sensitivity to novel compounds and expression of *PDGF-α*, *3-hydroxybutyrate dehydrogenase*, *protein kinase C-β1*, *cytochrome b-5 reductase*, *protein tyrosine phosphatase receptor type C*, *matrix metalloproteinase 1*, *apolipoprotein D*, *alcohol dehydrogenase 5* and *cathepsin H* were also identified in this study.

The expression profile of the NCI 60 human tumour cell line panel was also used to ask whether constitutive gene expression patterns were sufficient to predict sensitivity or resistance to 232 compounds [112]. The constitutive gene expression data were divided into two groups: a training set used to develop a classifier using a weighted voting scheme and a test set to evaluate the accuracy of the classifier. To avoid developing a classifier that identified tissue type rather than drug sensitivity, the training set included a sensitive and resistant cell line from each tissue type. In general, the classifiers were complex and difficult to interpret. However, this analysis did identify some mechanistically understandable drug-target relationships, for example, the classifier for cytochalasin D, which binds to and interferes with actin polymerisation, contained genes encoding products related to the cytoskeleton or extracellular matrix. This suggested that these types of studies could reveal information on factors governing drug resistance/sensitivity, in addition to providing information on the potential target.

Dan and colleagues [113] used microarrays to determine the gene expression pattern in 39 human cancer cell lines and integrated the data with the chemosensitivity profile to 55 agents, including commonly used anticancer drugs. Some genes, e.g., *aldolase reductase*, which encodes an enzyme that catalyses the conversion of glucose to sorbitol, and also *damage-specific DNA binding protein* 2, which is involved in DNA repair, showed a positive correlation with sensitivity to many drugs, indicating they could be common markers of chemosensitivity. Other genes, such as that encoding LIM domain protein kinase 2, a protein that regulates the actin cytoskeleton, exhibited a negative correlation with responsiveness to drugs and may represent common markers of chemoresistance. Correlations were also observed with specific classes of drugs. For example, expression of *survivin* and *inhibitor of apoptosis 1* were correlated with resistance to 5FU derivatives and expression levels of the 3 *aldo-keto reductase* family members, 2 *aldehyde dehydrogenases* and *galectin 4* were associated with sensitivity to fluoropyrimidines. Interestingly, *cathepsin H*, one of the 11 genes of interest identified by Wallqvist

and colleagues [111], and also *14-3-3σ* were associated with resistance to anthracyclines, bleomycin derivatives and topoisomerase I inhibitors. There were also some expected relationships; for example, a negative correlation was observed between *thymidylate synthetase* expression and responsiveness to inhibitors of this enzyme and a positive correlation was seen between sensitivity to the DNA topoisomerase II inhibitor etoposide and *topoisomerase IIA* and *IIB* expression.

A similar approach examined the NCI 60 human tumour cell database of gene expression and identified correlations sensitivity to 171 drugs with known and unknown mechanisms of action: a powerful statistical procedure known as partial least squares modelling was used [114]. Six gene products appeared to influence response to topoisomerase inhibitors, RNA/DNA antimetabolites and alkylating agents, but not to the antimitotic agents. *Methylenetetrahydrofolate dehydrogenase-2*, the product of which is required for formylmethionyl-tRNA synthesis in mitochondria, was associated with sensitivity to RNA/DNA antimetabolites, as was *SP2* which encodes a splicing protein that is localised in mitochondria. This observation suggested mitochondrial transcription as a potential target for RNA/DNA antimetabolites.

Blower and colleagues [115] have used a 'structure–activity–target strategy' that incorporates structural features that are associated with observed correlations between gene expression and growth inhibition. Sensitivity to two compound classes, benzothiophenediones and indolonaphthaquinones, were well correlated with several genes expressed in melanoma and leukaemia cells, respectively. Sensitivity to dihydrobenzothiophenedione-4,8-diones was highly correlated with *RAB7* expression and expression of *L-plastin* was linked to sensitivity to indolo-1,4-naphthaquinones. The benzothiophenediones were COMPARE negative in the NCI standard agent database, suggesting a novel mechanism of action. Sensitivity to benzothiophenediones with electron donating substituents exhibited a high correlation with genes such as *RAB7*, suggesting that these compounds may disrupt an essential redox process. RAB7 has a key regulatory role for aggregation and fusion of late endocytic lysosomes: expression of dominant-negative RAB7 results in dispersed lysosomes with a considerably higher pH due to disruption of the vacuolar proton pump. Five other genes encoding lysosomal products correlated with sensitivity to the benzothiophenediones, and included adenosine triphosphate (ATP)ases involved in regulating lysosome pH. This study demonstrated that the structure–activity–target strategy can be used to prioritise genes or candidate compounds to follow-up as potential targets or therapeutics.

A different analytical approach is to employ relevance networks that have nodes with varying degrees of cross-connectivity; this is in contrast to phylogenetic tree approaches that can only link one feature to another. Phylogenetic trees cannot easily display associations between gene expression and anticancer activity, whereas relevance networks can capture negative correlations and positive correlations and display them on a single network. Butte and colleagues [116] applied this approach to the NCI 60 human tumour cell line panel data and found the majority of the networks linked anticancer agents exclusively; these generally linked one agent to a structural derivative of that agent, while the larger networks linked agents with similar mechanisms of action, e.g., alkylating agents. Nine of the networks showed evidence of biological association; for example, one network linked *keratin 8* and *18* that are known to be co-expressed and function together. One single network demonstrated an association between decreased sensitivity to a thiazolidinecarboxylic acid derivative and increased expression of *L-plastin*, an actin binding protein. In addition, three other agents were linked in this same network. This type of analysis has also been performed with data from eight hepatoma cell lines treated with eight commonly used anticancer agents [117]. For most of the associations generated, the biological significance was unclear; however, this study did detect a relationship between increased *topoisomerase IIβ* expression and resistance to two anthracyclines, a previously reported association between *transporter associated with antigen processing-1* expression and mitoxantrone resistance, and a correlation between *superoxide dismutase-2* expression and resistance to the nitrosourea nimustine.

A single study has attempted the correlative approach *in vivo* by comparing sensitivity with constitutive gene expression profile determined by genome wide cDNA microarray screening in 85 tumour xenografts treated with nine commonly used anticancer agents [118]. Pearson correlation analysis found certain associations between gene expression and anticancer agent sensitivity that were biologically relevant or interesting; for example, an association was seen between increased *topoisomerase IIα* expression and doxorubicin resistance, a negative correlation was observed between *thymidylate synthetase* expression and 5FU sensitivity, and also a negative correlation was identified between *aldehyde dehydrogense 1* and camptothecin sensitivity. Other correlations were noted between sensitivity to some agents and the expression of genes such as the multidrug resistance/antigen transporter family, *major vault protein*, *cytochrome P450 subfamily 3A* and *galectin 4*. In addition, reduced expression of 2 regulators of G2M cell cycle progression (*cyclin B1* and *BUB 1β*) was associated with chemoresistance. This observation could be biologically significant as many anticancer agents induce a G2M arrest.

It should be noted that the majority of the gene expression-drug sensitivity relationships discussed above, frequently derived from the NCI 60 human

tumour cell line panel, involve mostly the traditional cytotoxic drugs rather than the newer types of molecular therapeutics. This may be because the databases are dominated by the traditional agents. We are currently exploring novel agent–gene expression relationships for new drugs currently under development at the Cancer Research UK Centre for Cancer Therapeutics. Forty cell lines with different tissue of origin are being profiled using an in-house platform with 36 000 cDNA elements (www.crcdmf.icr.ac.uk/). The IC_{50} (concentration causing 50% growth inhibition) data for drug or compound sensitivity are used to separate the cell lines into sensitive or resistant populations. Nearest neighbour analysis and support vector machine analysis with leave-one-out cross validation are being used with the aim of building a molecular predictor based on the classification sensitive and resistant. These analysis tools will select genes that are most significantly correlated with drug sensitivity or resistance, and should predict whether a tumour cell will be sensitive or resistant according to the expression of the genes in the classifier.

Gene expression-drug sensitivity correlation investigations can be seen as analogous to the clinical gene expression profiling studies discussed earlier in this Review that look for predictors of classification or outcome. However, there are several potential limitations as: (1) the cell lines have been selected for growth in culture and should really be considered as surrogates for clinical tumours; (2) the database is frequently generated from a single end-point of short-term growth inhibition and/or cytotoxicity in tissue culture; (3) the sensitivity is assumed to be determined at the transcriptional level; (4) genes can exhibit transcriptional co-regulation if their products function in the same complex or if they are all targets of a particular transcription factor, raising the possibility that although only one gene in the cluster may be important for drug sensitivity, co-clustering will result in all of them being detected as significant; and (5) finally, the relationship between drug activity and expression is correlative and not necessarily causal as expression following drug exposure has not been measured. This strategy requires extensive prior information as the correlations described above were identified using background information to which the gene array data provided extra value. In addition, the correlations can be difficult to interpret, mainly due to the paucity of 'correct answers' in the literature which means that each hypothesis requires prospective testing. The significant correlations between gene-drug pairs frequently do not make biological sense and this raises the questions: what are these data telling us and does there have to be a biological significance behind the association in order for it to be useful? In some respects, biological sense may be irrelevant on one level as we are simply seeking statistically significant and robust predictive markers of response. Possible explanations could be that the reference drugs are working by mechanisms other than their predicted mechanism of action, that the genes could be acting downstream of the drug target, that the genes may affect drug metabolism or transport, or that the function of the significant gene products is not fully understood. There is ample precedent in the literature to support the last point, for example, the unexpected role of cytochrome *c* in apoptosis, emphasising the point that we have a substantial amount to learn regarding protein function. However, the benefit of an established database of known compounds with defined mechanisms of action is that it will allow comparison with novel compounds that have unknown mechanisms of action.

3.1. Expression profiling of drug resistance

Intrinsic and acquired resistance to drugs seriously limits the efficacy of cancer treatment and is a major problem in cancer therapy. Determination of the molecular mechanisms involved provides a possible approach to overcome this problem. Molecular mechanisms of acquired drug resistance and downstream mediators of drug action can be examined *in vitro* by continuous exposure of cell lines to a particular agent until a subclone becomes resistant to the drug and is selected out. The resistant clone can subsequently be compared with its parental line by gene expression microarray analysis.

Such an approach was used to investigate the molecular mechanisms of acquired resistance to camptothecin, a topoisomerase I inhibitor, analogues of which are used in the treatment of lung, breast, ovarian and colon cancers. The prostate cancer cell line DU145 and a selected 9-nitro-camptothecin resistant sub-line were compared by microarray analysis [119]. Using the statistical method of a stratum-adjusted Kruskal–Wallis test, the expression of 181 genes were found to be significantly altered by >1.5-fold in the resistant compared with the parent line. Several functional groups of genes were significantly over-represented, among them being those involved in NF-$\kappa\beta$ signalling and apoptosis. The resistant subline was cross-resistant to a variety of cellular stresses, an observation consistent with the gene changes involving a more general anti-apoptotic phenotype. However, several of the gene changes observed, especially those in the NF-κB and transforming growth factor β pathways, were contrary to expectation. This led the authors to suggest a 2-step mechanism for the development of drug resistance. The first step would involve gene expression changes that lead directly to resistance to apoptosis, in this case apoptosis regulators of the BCL2 family and pathways signalling through C-AKT. This would in turn allow the emergence of secondary genetic events that stimulate cell cycle progression, but would normally be constrained as they induce apoptosis, for example E2F1 and C-MYC expression or, in the case of this particular study,

components of the NF-kB pathway. The proposed model is analogous to one suggested for the development of malignant cancers in which acquisition of resistance to apoptosis allows subsequent deregulation of signalling pathways leading to increased proliferation that can now occur without cell death as a consequence.

The melanoma cell line RPMI8226 was selected for resistance to different concentrations of doxorubicin and developed a MDR1 phenotype [120]. However, this was not the sole factor as resistant clones grown in the absence of drug were found to lose their MDR1 phenotype, but retain some degree of resistance. The clone with the highest resistance to doxorubicin was also resistant to fas-ligand stimulation and treatment with staurosporine. Expression profiling detected the differential expression of 380 genes, including the loss of expression of genes encoding proteins associated with the fas death receptor, members of the *FOS* family and the pro-apoptotic *BAD* gene. In another study of drug resistance in melanoma, the differences in expression profile of parental MeWo melanoma cells and sub-clones made resistant to cisplatin, fotemustine and etoposide were less clear-cut [121]. *MPP1*, a membrane-associated guanylate kinase homologue that may have a role in signal transduction, and also *crystalline-αB*, a small heat-shock protein and antagonist of caspase 3-mediated apoptosis, were strong candidates for resistance to DNA damage. In addition, a group of six interferon-regulated genes were increased in expression in the cisplatin- and etoposide-resistant lines.

Resistance to DNA-interacting agents has also been examined using an alternative method of transfecting cell lines with cDNAs and isolating clones that exhibit resistance. Levenson and colleagues [122] transfected cell lines with short biologically active cDNAs and isolated a fibrosarcoma cell line that was resistant to a diverse range of DNA-interacting agents. Four transfected cDNAs were recovered from this cell line: an expressed sequence tag of unknown function, *mitochondrial subunit 3 of cytochrome c oxidase*, *ring-3-related gene* (*ORFX*; a mitogen-activated nuclear protein kinase that shares substrate specificity with myosin light chain protein kinase and cAMP-dependent kinase) and *widely interspersed zinc finger protein* (*WIZ*). The original cell line and a cell line transfected with all four of these cDNAs were expression profiled and exhibited altered expression of genes involved in DNA repair and replication, signal transduction, cell cycle control, and transcription.

Antimetabolites such as 5FU can induce DNA damage or affect RNA metabolism, depending on the dose and route of administration. Wang and colleagues [123] examined the expression profile of cells made resistant to 5FU or the more specific thymidylate synthetase inhibitor, raltitrexed, and identified *thymidylate synthetase* as the sole gene in expression of which was consistently associated with resistance to these agents. This observation was significant as thymidylate synthetase is a target of both 5FU and raltitrexed treatment and negative correlations have been detected between sensitivity to 5FU and its derivatives and *thymidylate synthetase* expression [113,118]. A similar approach was taken with paclitaxel, a microtubule stabilising chemotherapeutic used to treat ovarian, breast and non-small cell lung cancers. Lamendola and colleagues [124] developed SKOV-3 ovarian carcinoma cells with different degrees of resistance to paclitaxel and profiled expression as the cells acquired resistance. Generally, the expression pattern associated with resistance was complex. Using SOMs the authors detected increased *MDR1* expression and loss of pro-apoptotic *BAK* expression with increased resistance. Early paclitaxel resistance was associated with increased expression of genes encoding proteins involved in inflammatory responses. Intermediate resistance was associated with increased expression of genes encoding products involved in the G1–S transition of the cell cycle. Finally, late resistance was associated with expression of genes encoding plasma membrane proteins.

One general observation from most of the studies described above is the complexity of the gene expression changes acquired during the development of resistance. It is not entirely clear whether all the changes acquired during resistance are required or whether only one or two genes are key to resistance and the others showed altered expression purely by coincidence, for example, as a result of co-transcriptional regulation. In addition, a series of these studies suggest that the acquisition of resistance may involve a series of stages that each require a different class of gene product.

4. Profiling gene expression alterations in response to drug treatment

4.1. The value of profiling drug-induced gene expression changes

Demonstrating 'proof of concept' is extremely important, both in the pre-clinical discovery phase and in early clinical trials of new molecular therapeutics. The overall challenge is to develop new molecular therapeutics as expediently and effectively as possible [5,7,33]. Identifying molecular pharmacodynamic markers that can be used to assess whether the drug is actually modulating the intended molecular target and the biochemical pathways and biological processes in which it operates is critically important. Indeed, it has been proposed that such molecular biomarkers are essential to allow the construction of a pharmacological 'audit trail' that links the status and expression of the molecular target and the pharmacokinetic and pharmacodynamic effects of the drug to the clinical outcome of treatment [125]. Gene expression profiling of cells in response to drug

treatment provides a valuable means to identify potential pharmacodynamic markers. Furthermore, determination of expression profile using microarrays can also be used to investigate the detailed cellular mechanisms of action of both novel and established drugs, revealing molecular signatures of both on- and off-target effects. In this section, we discuss examples of the types of studies that have been carried out using the expression profiling approach.

4.2. Cytotoxic agents

A number of investigations have examined gene expression responses following treatment with commonly used DNA damaging drugs or irradiation. Zhou and colleagues [126] measured the transcriptional response of HCT116 colon carcinoma cells, synchronised in S phase, to two different concentrations of the topoisomerase I inhibitor, camptothecin. With the lower concentration, a reversible G2 arrest of the cell cycle was observed, whereas at the higher concentration an irreversible arrest in the G2 phase was detected. Different patterns of gene expression according to drug concentration were observed; for example, a group of genes were upregulated only in response to high concentrations. The genes in this group were known DNA damage-inducible genes and also genes associated with cell cycle arrest and apoptosis, e.g., *DDB2*, *cyclin dependent kinase inhibitor p21* and *fas-receptor*. Some genes in this group were genes known to be transcriptionally activated by the p53 DNA damage response pathway. These transcription profiles suggest that there is a fundamental difference in the response to mild DNA damage resulting in a reversible G2 arrest compared with the permanent G2 arrest following extensive DNA damage.

The response to DNA damage has also been assessed following exposure of human myeloid cells to methanosulfonate, γ- or ultraviolet (UV)-irradiation [127]. Genes expected to be induced by these treatments, including *ATF3* and *FOS*, as well as those regulated by p53, were detected; in addition, the expression of several other genes previously unassociated with DNA damage were shown to be responsive to genotoxic stress. Also of significance was the observation that response to these DNA damaging modalities varied widely between cell types, indicating that cellular context played an important part in response. The role of p53 in response to DNA damage was also examined for topotecan, a semi-synthetic water-soluble derivative of camptothecin used in first-line therapy of cisplatin-refractory ovarian cancer and second-line in small cell lung cancer [128]. In this case, a recombinant p53 knockout HCT116 colon carcinoma cell line was compared with its parental p53 wild-type line. Treatment of the wild-type cell line altered the expression of many more genes (approximately 10% of those measured) compared with the knockout cell line. P53 activation was associated with a coherent pattern of gene expression leading to cell cycle arrest and apoptosis. Genes induced independently of p53 included the anti-apoptotic protein encoded by *cFLIP* and genes encoding stress signalling proteins in the mitogen-activated protein kinase and FOS/JUN pathways. Other studies have also examined the effect of DNA damaging agents. Kudoh and colleagues [129] compared doxorubicin-sensitive and -resistant MCF-7 breast cancer cells. Many genes showed altered expression following exposure of the sensitive cells to doxorubicin; in contrast, the resistant line exhibited fewer changes, although the genes that exhibited altered expression were also changed in the sensitive line.

Cells treated with cytotoxic agents, such as drugs that result in DNA damage, undergo an irreversible cell growth arrest that may result in cell death or a permanent growth arrest that has features of cell senescence described earlier. Following treatment with doxorubicin, Chang and colleagues [130] used staining with PHK2, a lipophilic fluorophore that incorporates into the plasma membrane, to separate senescent growth-arrested cells from cells able to proliferate. In the growth arrested cells, 25/68 genes encoding products with a role in the G2M phase of the cell cycle exhibited decreased expression. Other genes showing increased expression in this population included those encoding secreted mitogenic, anti-apoptotic factors, angiogenic factors and transcription factors such as the AP-1 family, *C-JUN* and *XBP-1*, together with *ATF3* and *ELF-1* whose gene products interact with AP-1. Both *p53* and *p21* were induced and there was some overlap between genes induced in the senescent population and those regulated by p53 and p21, although increased expression of many of these genes did not require p53 or p21. Interestingly, agents that induce the expression of genes which would cause secretion of proliferative, anti-apoptotic and angiogenic factors would actually benefit any cancer cells that survive the treatment.

The effect of specific biochemical pathways on response can also be assessed following treatment with a DNA damaging agent. For example, the effect of *C-MYC* expression on treatment with VP-16, a topoisomerase II inhibitor that induces double-strand breaks, has been examined [131]. VP-16 treatment of the *C-MYC*-expressing cell line resulted in high levels of apoptosis that occurred rapidly, as opposed to the corresponding *C-MYC*-null line for which apoptosis was reduced and delayed to later time-points. Groups of genes dependent or independent of *C-MYC* status were altered by VP-16 treatment. The *C-MYC* independent group included members of RAS/MAPK pathway. Genes that were altered in a time-dependent manner included p53-regulated genes and were also associated with the onset of apoptosis. C-MYC-dependent genes included *C-JUN*,

fos-related antigen 2 and other genes previously reported to regulate apoptosis. *C-JUN* and *FOS* can be induced by the stress-activated kinase, p38, and pre-treatment with the p38 inhibitor SB203580 blocked their induction. The degree of apoptosis in the *C-MYC*-overexpressing line was reduced by SB203580 pre-treatment, was slightly decreased in the parental line, but was unchanged in the *C-MYC*-null line. This suggested a model in which p53 induction triggers a basal rate of apoptosis independently of the presence of C-MYC, but C-MYC is required to engage other death effector pathways to gain maximal effect [131]. Perhaps significantly, *in vitro* and *in vivo* studies have suggested that tumours with elevated C-MYC and wild-type p53 are particularly sensitive to 5FU [132].

The gene expression response to treatment with other classes of commonly used anticancer agents, such as paclitaxel or 5FU, has been profiled. Treatment of lung carcinoma or melanoma cells with paclitaxel induced the expression of *MGSA*-α, a gene frequently overexpressed in melanoma, colorectal cancer and other tumours [66,67,133]. Treatment with the MEK1/2 inhibitor, U0126, suppressed the induction of *MGSA*-α by paclitaxel and induced apoptosis. Perhaps significantly, melanomas are frequently insensitive to paclitaxel treatment, although the role of MGSA in resistance to paclitaxel in the clinical treatment of melanoma has not yet been determined. In the lung cancer cell line, *topoisomerase III*β expression was induced by paclitaxel, and suppressed by the combination of paclitaxel and U0126. This gene was overexpressed in lung tumours and its expression has been reported to have a role in cell growth and to reduce basal cell death; as such it is a potential therapeutic target. Treatment of MCF-7 breast and H630 colon cancer cells with 5FU, raltitrexed or oxaliplatin resulted in the induction of p53 [134]. A number of genes showed increased expression following 5FU treatment; these included the *fas-receptor* and genes encoding products involved in signal transduction, cell cycle regulation, polyamine metabolism, and other diverse functions, none of which had been previously described as altered by 5FU. Inactivation of p53 by expression of the viral E6 protein caused a decrease in the induction of a number of these genes, whereas the response of others was unaltered. Five of the genes induced by 5FU also had higher basal levels in a 5FU-resistant cell line that overexpressed *thymidylate synthetase*.

4.3. New chemotherapeutic agents

Natural products and their derivatives frequently exhibit potent anticancer activity *in vitro* and *in vivo*. However, one challenge with this class of agent is identifying their molecular targets. Expression profiling has contributed to the elucidation of the putative mechanism of action of ecteinascidin-743 (ET-743), a DNA binding alkaloid natural product derived from a marine tunicate that is currently undergoing phase III clinical studies. Friedman and colleagues [135] used expression profiling to demonstrate that ET-743 reduces transcription of genes that are induced following treatment with trichostatin A, an histone deacetylase (HDAC) inhibitor. This effect was not apparent with ET-743 alone and suggested that ET-743 acts by inhibiting the expression of genes, the promoters of which are activated following histone deacetylation.

Non-steroidal anti-inflammatory drugs not only have a potential role in cancer prevention, but also induce growth arrest and apoptosis of cancer cells. The anti-inflammatory action of these agents is via inhibition of the cyclooxygenases (COX) that are required for the synthesis of prostaglandins. Sulindac sulphide inhibits both COX1 and COX2, and induces apoptosis of colon cancer cells. Expression profiling demonstrated that sulindac sulphide treatment significantly alters gene expression in SW480 colon carcinoma cells that express COX1, but little COX2; these included induction of *PLAB*, *C/EBP-b* and *ATF3* [136]. Eleven genes were picked for further analysis and generally showed altered expression to a greater degree in COX1- and COX2-deficient HCT116 cells. Equimolar concentrations of sulindac, sulindac sulphone or treatment with prostaglandin E2 did not alter the expression of the 11 genes; however, treatment with high concentrations of indomethacin did induce the changes detected with sulindac sulphide. This study demonstrated COX-independent regulation of anti-tumorigenic and pro-apoptotic gene expression by sulindac sulphide that may be linked to its anti-tumour mechanism of action. However, the mechanism is still unclear and is potentially complicated. Studies in our own laboratory have also examined the effects of COX inhibitors on gene expression of a human colon carcinoma cell panel and compared sulindac sulphide, the COX2 inhibitor celecoxib, the COX1/2 inhibitor indomethacin and sulindac sulphone, a structurally related compound that does not inhibit COX 1/2 activity [137]. Indamethacin and celecoxcib exhibited a similar gene expression response, in contrast to sulindac sulphide. Genes exhibiting altered expression included those involved in the regulation of the cell cycle, growth arrest, cell death and signal transduction, particularly those involved in interferon-regulated signalling.

There is considerable interest in targeting specific PI3-kinase isoforms according to the therapeutic indication. However, the PI3-kinase family comprises a large number of members, and there are indications that certain isoforms may be involved in insulin signalling, whereas others are more responsible for tumour progression [138]. Profiling such inhibitors using gene expression microarrays may help to identify compounds that do not inhibit insulin signalling, but retain the

anti-cancer effects, including apoptosis and growth inhibition. We are currently using this approach to profile novel inhibitors that may have differential effects on the specific PI3-kinase isoforms. SOMs identified a number of clusters of genes that are affected downstream of PI3-kinase inhibition following treatment with the broad spectrum PI3-kinase inhibitor LY294002 (te Poele, R and Workman, P data not shown). These included a cluster of mitotic genes that appear to be co-regulated by the PI3-kinase pathway, two of which have previously been implicated in malignancy.

Enzymes that affect chromatin structure and hence alter gene expression are another novel class of target now being investigated [139]. Microarrays have been used to identify a set of common genes regulated by HDAC inhibitors. Glaser and colleagues [140] compared the gene expression profiles of cell lines treated with hydroxamic acid derivatives suberoylanilide hydroxamic acid together with trichostatin A, and the structurally novel HDAC inhibitor MS-27-275. All three agents maximally induced p21 after a 24 h exposure and caused a robust hyperacetylation of histone H4. The gene expression profiles of the three HDAC inhibitors were generally similar and distinct from those produced with structurally-related, but inactive, analogues, indicating that the changes observed were mechanism-related and not due to the chemical backbone. The correlation of gene expression changes induced by the 2 hydroxamic acid-based compounds was higher than the correlation between either of these and the MS-27-275 compound, consistent with their different effects on cells. As treatment with HDAC inhibitors is believed to increase gene expression, it was surprising that as many genes were downregulated as were upregulated after HDAC inhibitor treatment. All three inhibitors consistently altered the expression of only thirteen genes across all three cell lines, of which most genes were involved in the regulation of the cell cycle, apoptosis and DNA synthesis.

DNA methylases directly modify DNA and regulate gene expression. As many tumour suppressors are silenced by methylation, abnormal methylation could be a potential target in cancer. The nucleoside 5-aza-2′-deoxycytidine, one of the most potent inhibitors of methylation, forms a covalent complex with DNA methyltransferases following its incorporation into DNA. This nucleoside is a powerful inducer of differentiation and is undergoing clinical trials in myeloid dysplasia syndromes and leukaemias. Liang and colleagues [141] profiled gene expression in human fibroblasts and T24 human bladder carcinoma cells following treatment with 5-aza-2′-deoxycytidine. Overall, a greater number of genes were induced following treatment of the T24 cells and these genes included those regulated by interferon and cytokines. This observation may be of therapeutic significance as 5-aza-2′-deoxycytidine sensitised colon cancer cells to treatment with interferon, an agent that has been used in the past to treat colon cancer in combination with 5FU. Treatment of prostate cancer cells with 5-aza-2′-deoxycytidine induced the expression of a novel LIM domain protein *ZNF185*, a gene the expression of which is repressed during progression from benign hyperplasia to metastatic carcinoma [142]. Additional genes altered by 5-aza-2′-deoxycytidine treatment included *BPAG1*, a gene that encodes a hemidesmosomal protein involved in adhesion to basement membranes, and *PSP94* which is a prostate-specific gene involved in inducing apoptosis.

Expression profiling has revealed that the proposed CDK inhibitor flavopiridol has a different cellular mechanism of action to two other chemically dissimilar CDK inhibitors studied, namely roscovitine and 9-nitopaullone [143]. Flavopiridol appeared to inhibit gene expression in a global manner causing an expression profile similar to the effects seen with general transcription inhibitors such as actinomycin D. Measurement of mRNA turnover following flavopiridol exposure revealed that different functional classes of genes had distinct distributions of turnover rates. Several apoptosis genes and key cell cycle regulators had very short half-lives, suggesting that flavopiridol may be particularly effective in cancers dependent on genes encoding mRNAs with high turnover rates, such as *C-MYC*. In our own gene expression studies with the tri-substituted aminopurine CDK inhibitor CYC202 (*R*-roscovitine) an interesting pattern of gene expression was observed in colon cancer cells that included reduced expression of genes encoding products that regulate the mitotic phase of the cell cycle [144].

Our laboratory has also examined alterations in gene expression pattern following exposure to 17AAG, an ansamycin antibiotic derivative and novel HSP90 molecular chaperone inhibitor that has great potential as the first in a new class of anticancer agent [145]. The response of different human colon adenocarcinoma cell lines to this agent varied widely at the gene expression level, again indicating that cellular context has an important role in response to anticancer agents. However, a number of interesting consistent changes were detected; these included induction of the HSP90 drug target in cell lines that had reduced sensitivity to 17AAG, contrasting with low HSP90 expression in cell lines particularly sensitive to 17AAG. Other changes included induction of HSP70 expression in all cell lines and differential effects on the expression of cytoskeletal and signalling genes. The induction of HSP70 along with a number of HSP90 client proteins is being used as a pharmacodynamic marker in clinical studies of 17AAG at our institution.

The examples above show how expression profiling following treatment with anticancer therapeutics with predicted mechanisms of action can identify mechanism-based gene expression changes, reveal clues to

on- and off-target effects, identify molecular pharmacodynamic markers, and, in some instances, suggest new targets for therapy. However, there are some issues that need to be considered very carefully. The choice of cell line to be studied is important. As discussed earlier, different cell lines can behave quite differently when exposed to the same pharmacological agent and this effect can be amplified at the level of gene expression profile. These differences can be beneficial in some instances, where clues to gene products affecting sensitivity can be obtained by comparing different cell lines. However, such differences can be confusing when different cell lines give non-identical expression profiles. It is usually sensible to choose a cell line where the pathway being targeted has been confirmed to be active. However, where possible, it is useful to profile the effects of drugs, particularly those with an unknown or unclear mechanism of action, in several cell lines. Another issue is the concentration of drug used to treat the cell line. In our laboratory, we generally use pharmacologically-relevant concentrations of between 1× and 5× the cell growth inhibitory IC_{50}. Furthermore, it is also very useful to determine gene expression changes over multiple time-points. This can be particularly useful to help dissect the issue of cause and effect: *ie* to answer the question is a particular gene expression pattern a cause of the phenotypic effect on the cell or a consequence of it? Analysis of early time-points prior to phenotypic changes, such as cell cycle arrest or induction of cell death, are, on the whole, preferable as they are more likely to define mechanism-related gene expression changes as opposed to gene expression alterations resulting from an altered phenotype. Finally, the inclusion of an inactive or much less active analogue of the drug under study can also be used to discriminate between on- and off-target effects.

4.4. Expression profiling in drug discovery

Microarray expression profiling can play a role in studies of structure–activity relationships during development of novel agents and has a potentially important application in prioritising lead compounds early in preclinical drug development. We have used a microarray strategy to compare inactive and active derivatives of 17AAG as well as to profile radicicol, a structurally dissimilar natural product HSP90 inhibitor and a structurally novel, synthetic HSP90 inhibitor CCT018159 that was discovered in our centre [146; Maloney, A., Clarke, P.A. and Workman, P. data not shown]. Expression profiling by microarray identified signatures associated with HSP90 inhibition that would be consistent with their mechanism of action. These changes were independent of chemotype, although additional changes in gene expression specific to the particular chemotype could also be detected.

Although not a cancer study, Gunther and colleagues [147] demonstrated the potential value of this approach in the development of psychoactive drugs. The absence of clear understanding of psychosis and depression has prevented the development of cell-based models for these disorders such that "anti-depression" compounds are usually tested *in vivo*. The authors profiled human primary neurones following 24 h treatment with multiple classes of psychoactive drugs. Using a classification tree and random forest supervised classification, it was possible to predict functional categories with good accuracy. This study clearly demonstrated the capacity to predict efficacy on the basis of gene expression signature.

When microarray expression profiling has been used in screening of novel compounds, it tends to be at later stages of lead optimisation when the throughput is more manageable. Owa and colleagues [148] examined seven sulphonamide antitumour agents, based on two different pharmacophores, against HCT116 cells. One class of pharmacophore were COMPARE similar to vincristine, paclitaxel, and navelbine, induced a G2M arrest and were potent inhibitors of tubulin polymerisation. Few genes showed altered expression by this class of agent, the most significant change being decreased expression of α-*tubulin* that was associated with a feedback regulatory loop in which accumulation of monomeric-tubulin represses tubulin transcription. The other class of pharmacophore was not COMPARE similar to inhibitors of tubulin polymerisation, induced limited effects at G2M, showed some reduction in S phase and altered the expression of more genes, including those encoding products involved in mitochondrial function and energy metabolism. One of the class of tubulin polymerisation inhibitors, E7010 which binds the colchicine binding site of β-tubulin, has undergone phase I clinical studies and demonstrated activity in 2 of 16 patients. Further studies have demonstrated that this compound repressed tubulin gene expression and induced the expression of *C-JUN* and *FRA-1* [149]. A compound from the other class, E7070, exhibited some activity in phase I and is now under phase II clinical study. This agent altered the expression of more genes than did E7010. However, the primary target of E7070 remains unknown.

DF-203 (2-(4-amino-3-methylphenyl)benzothiazole) is a potent anticancer agent that is COMPARE negative within the NCI database and may therefore have a novel mechanism of action [150]. The current lead candidate 5F-203 (2-(4-amino-3-methylphenyl) 5-fluorobenzothiazole) has good potency and pharmaceutical properties and requires activation by CYP1A1, although the active metabolite remains unidentified. Expression profiling was used to compare sensitive MCF-7 cells and insensitive MDA-MB-435 cells 6 and 24 h treatment with 5F-203. No significant changes in gene expression were detected in the insensitive MDA-MB-435 cells. However, treatment of MCF-7 cells resulted in a number of

changes, including increased expression of cytochrome P450 isoform *CYP1B1* (dioxin-inducible), *CYP1A1* (aromatic-inducible) and *PLAB*, which were the only changes common to both time-points studied. Additional changes included increased expression of *ATF3*, *fas-receptor*, *fos-related antigen 1* and *DDB2*. This profile was consistent with studies of other agents that induce DNA damage [121,127,130,131,134]. Treatment also resulted in p53 induction and there was evidence for induction of *p21* and *PIG3*, but not for other p53-regulated genes such as *MDM2* or *BAX*. Thus, the results are to some extent consistent with the induction of p53 protein and suggest that cytochrome P450 metabolism of this compound liberates a DNA damaging agent.

It may become increasingly frequent that genes with unknown function will be identified from global expression profiling studies. Transfection or RNAi can be used to validate their importance to the disease pathology without knowledge of function. This leaves the investigator in the position of having a potential drug target with an unknown function. The next logical step is to determine the protein function and then to set up biochemical or cell-based high throughput screens. This can be a major rate-limiting step in drug discovery for genes of unknown function as it requires extensive characterisation of the protein prior to assay development. While it is obviously preferable to be understand the function of the target to which drugs are being developed, there are strategies that may avoid the time-consuming assay development. One approach is to profile gene expression, including the unknown target, within a drug screen to measure transcriptional responses to exposure to diverse compound libraries. For screens involving large numbers of compounds, it is generally not feasible to use microarrays; an alternative is to select genes from the initial microarray analysis and to use a multiplex RT-PCR approach that can be adapted for 96-well assays. Johnson and colleagues [151] developed a multiplex assay for six genes described in the literature as being overexpressed in prostate cancer. The assay involved measurement of a homeodomain transcription factor *HOXB-13*, an *ETS* transcription factor family member *hPSE*/*PDEF*, *survivin*, a PDZ-domain protein *INA D* and a novel gene *BLX-33*. They also measured 5 genes indicative of stress or DNA damage plus 2-housekeeping control genes. Eighty compounds representative of different mechanistic classes were screened against the human prostate cancer cell line PC-3. Two compounds, digitoxin and ouabain, representative of cardiac glycosides that inhibit the Na^+K^+-ATPase sodium pump, were potent inhibitors of the expression of the 3 transcription factors (*HOXB13*, *hPSE* and *HNF-3A*). A further 9000 compounds were tested in pools of 10. A pool containing another cardiac glycoside was identified by this altered expression profile and an additional 34 hits had a similar profile. The authors identified an active steroid structure that inhibited the expression of the prostate target genes and induced apoptosis in PC-3 human prostate cancer cells, but not adherent cells derived from a benign prostatic hyperplasia. A structure–activity relationship was defined that was consistent with sodium pump inhibition. Within this group were two bufalins previously demonstrated to be shown to be active apoptosis inducers in other cancer cell lines.

A similar approach is also being developed by other groups and the challenges involved have been discussed, particularly with respect to data handling and informatics [152]. Hits obtained in a 24-gene screen have been characterised more thoroughly against 30 000 genes by microarray analysis. In this way the true 'core genes' that represent the gene expression signature of the hit compound can be identified and medicinal chemistry optimisation can be directed by the response of these genes.

4.5. Gene expression profiling in response to or during treatment in the clinic

As described earlier, there are many examples that demonstrate that basal gene expression profiling of cancer cells can classify and predict long-term clinical outcome. However, so far, these studies have not predicted initial response to chemotherapy. One possibility is that profiling genes that show altered expression in response to chemotherapy will be more revealing. This type of study will also allow the measurement of pharmacodynamic responses that are important for confirming the mechanism of action of existing and novel anticancer agents; such responses form a critical part of the pharmacological audit trail mentioned earlier [125]. In addition, profiling gene expression in response to a single course of chemotherapy may allow prediction of whether the tumour will respond in the long-term and this may provide an opportunity for directed therapy for individual patients. Clinical gene expression profiling studies of this type have to accommodate logistical and ethical considerations, particularly those associated with the need for repeat tumour biopsy. However, several initial studies have demonstrated both the feasibility and utility of gene expression profiling as a valuable hypothesis-generating tool for studying the molecular pharmacology of cancer drugs in the clinical setting.

A number of studies have explored the possibility of using repeat fine-needle aspirates (FNA) of breast tumours as a source of tumour tissue for this type of investigation. Gene expression profiling of RNA from FNAs was inefficient, with only 4 of 27 giving adequate data for analysis [153]. Importantly, for the four FNAs that were successful in generating data, the expression profile closely resembled that of their matched tumour at surgical resection. Buchholz and colleagues [154] examined

core biopsies using direct radiolabelling and were able to obtain adequate data from 25 of 56 available biopsies. In total, five patients could be analysed before and 24 and 48 h after doxorubicin/docetaxel, single agent docetaxel or 5FU/doxorubicin/cyclophosphamide. The greatest effects were seen with genes encoding products required for nucleoside/nucleotide metabolism, protein metabolism, signal transduction and cell cycle. A follow-up study demonstrated that expression profiling of core biopsies from a group of 38 patients could identify gene expression signatures specific to ER status [155]. Similarly, Chang and colleagues [156] expression-profiled core biopsies of 24 patients prior to neoadjuvant treatment with docetaxel. Ninety-two genes could discriminate between sensitive and resistant tumours. These genes encoded products required for apoptosis, adhesion, cytoskeleton, protein transport, signal transduction, RNA transcription and splicing, cell cycle and protein synthesis; for example, the pro-apoptotic BAX gene showed increased expression in docetaxel-sensitive tumours. However, this study looked only at constitutive expression and not at treatment-induced changes.

Although the limited RNA yield obtained from small biopsies was a significant factor, the above studies suggested that the approach was feasible. Moreover, improved labelling technology and linear amplification methods should ensure that RNA yields will become less of an issue. Sotiriou and colleagues [157] incorporated a linear amplification step into their expression profiling analysis of FNAs and obtained good quality constitutive expression profiles from 12 of 14 patients. FNA pairs from the same patient clustered closely together compared with FNAs from other patients. A group of 10 patients had 2 FNAs prior to neo-adjuvant treatment with doxorubicin/cyclophosphamide followed by another at the end of the first chemotherapy cycle on day 21. Sixteen genes were identified for which the expression change during treatment could significantly differentiate between the responders and non-responders. The expression of *CDK9* and *dCMP deaminase* were increased in poor responders after one chemotherapy cycle.

One of the most promising new molecular therapeutic agents is the c-ABL kinase inhibitor imatinib mesylate [14]. High response rates to imatinib mesylate have been achieved for the treatment of chronic myelogenous leukaemia that is characterised by deregulated c-ABL kinase activity arising from a chromosomal translocation involving the *BCR* and C-*ABL* genes. However, resistance to imatinib mesylate is a common clinical occurrence, suggesting pre-existing mechanisms of resistance or the development of resistance during treatment. Hofmann and colleagues [158] profiled bone marrow aspirates of 19 patients with BCR-ABL-positive ALL to assess the possibility of predicting resistance to imatinib mesylate using nearest neighbour class prediction. Gene sets that could discriminate between sensitive and primary resistant samples and that were predictive for secondary resistance while on imatinib mesylate treatment were identified. There was no apparent functional association between the genes and mechanisms of resistance, although the expression patterns correctly classified samples as sensitive or resistant.

Fifty-six genes were also identified by direct comparison of sensitive and secondary resistant cells characterised by acquired resistance to imatinib mesylate [158]. Several could be organised into functional groups, some of which had previously been implicated in mechanisms of resistance or altered signalling in resistant cells. There was no increased expression of BCR-ABL protein or overexpression of multi-drug resistance genes. Of note, was overexpression of the gene encoding Bruton's tyrosine kinase (*BTK*) that is regulated by the PI3-kinase and SRC family tyrosine kinases. Increased expression of *BTK* could overcome inhibition of the ABL kinase function by phosphorylating downstream effectors of the ABL pathway. Genes decreased in expression in the resistant cells included the pro-apoptotic *BAK* gene and members of the AP1 transcription factor family, *JUNB*, *JUND* and *C-FOS*. The results of this small study are encouraging as not only was it possible to predict sensitivity and resistance to imatinib mesylate, which if confirmed in prospective studies should be of great clinical potential benefit, but possible mechanisms of resistance were identified, some of which may be amenable to therapeutic intervention.

Cheok and colleagues [159] collected ALL biopsies before and 24 h after single or combined treatment of patients with 6-mercaptopurine, an antimetabolite that is incorporated into RNA and DNA, or methotrexate, an inhibitor of dihydrofolate reductase. Unsupervised hierarchical clustering of the post-treatment samples clustered the tumours by the underlying characteristics of the ALL blasts and not according to treatment. Different analytical algorithms were all able to identify expression patterns associated with the different types of treatment using the ratios of pre- to post-treatment gene expression. Genes overrepresented in this set included those that regulate apoptosis, stress response, DNA and RNA metabolism, cell cycle control, signal transduction, metabolism and transport. Expression of pro-apoptotic *BAX* and cell cycle checkpoint regulator *ATM* were induced by methotrexate alone, but not by the other treatments. Changes induced by mercaptopurine alone included increased expression of the *E2F5* transcription factor and a splice variant of DNA mismatch repair *MLH1*. *C-FOS* showed decreased expression and expression of *DNA damage inducible transcript 3* was increased by all treatments. Interestingly, the combination treatments had a limited overlap with the single treatments. Significantly, treatment of 2 ALL cell lines *in vitro* failed to recapitulate the gene

expression patterns detected *in vivo*. This suggested that there were either fundamental differences between the response of the cancer cells *in vitro* and the primary malignant cells *in vivo*, for example, disparities in gene methylation, or that there were differences in the *in vitro* and *in vivo* milieu.

Using a collection of serial samples, taken pre- and during-treatment, from a consecutive series of patients with locally advanced rectal cancer, we explored the potential for the application of gene expression profiling by microarray to the measurement of global molecular pharmacodynamic responses after treatment with a single dose of mitomycin C and during a continuous infusion of 5FU [71] (Fig. 3). We demonstrated that the approach was feasible, and detected a novel molecular response that would not have been predicted from *in vitro* studies and that would have otherwise been missed by conventional approaches. 5FU treatment resulted in the accumulation of the inactive ternary complex of thymidylate synthetase:FdUMP:5–10, consistent with the inhibition of thymidylate synthetase. SAM analysis identified 247 individual genes encoding products involved in RNA synthesis, protein synthesis and metabolism that exhibited significantly decreased expression by >2-fold during treatment; of these, 80 had been reported previously as C-MYC-regulated genes. Thus, the treatment regimen was potentially reverting transcriptional changes associated with the expression of a known oncogene. Although only single patients, a similar gene expression pattern to that seen for 5FU was also detected following treatment with the thymidylate synthetase specific inhibitor raltitrexed, but not the topoisomerase I inhibitor irinotecan. The treatment-induced effects were observed, irrespective of response, suggesting that they were not a consequence of treatment outcome, but were more likely a consequence of drug action at its target or a secondary downstream event after target modulation. The nature of the mechanistic link between 5FU treatment and C-MYC *in vivo* was not clear. However, there is evidence that patients with colorectal cancers that have amplified C-MYC have significantly improved outcome after adjuvant treatment with 5FU/ levamisole [132]. Interestingly, although we detected evidence for thymidylate synthetase inhibition in the tumour tissue, a mechanism consistent with the prolonged infusional delivery of 5FU, we also detected decreased expression of genes required for RNA synthesis and processing. This suggests that prolonged infusional 5FU could exert effects on RNA metabolism by a mechanism distinct from the RNA-dependent effects induced by incorporation of 5FUTP into RNA. We have also compared the *in vivo* expression profile to that seen in colon carcinoma cell lines treated with 5FU. We observed little correlation between the *in vivo* and *in vitro* expression profiles (Fig. 3). This was similar to the observation of Cheok and colleagues [159] and emphasizes the point that although cell lines are an invaluable resource for biological and pharmacological studies and there is a need to study pharmacological responses and interactions in the setting of a clinical trial.

Taken together, our own observations [71] and those of others detailed here show that gene expression profiling in response to treatment could greatly benefit clinical studies of conventional or novel chemotherapeutic agents and provide encouragement for additional applications of the technology in the molecular pharmacol-

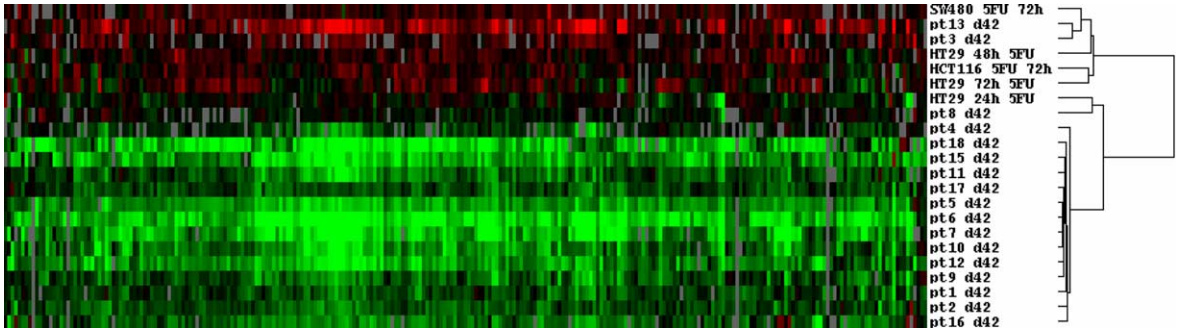

Fig. 3. Gene expression profiling of patients treated with mitomycin/5-flurouracil (5FU) and colon carcinoma cell lines (HT29, HCT116, SW480) treated *in vitro* with 5× IC_{50} (concentration causing 50% growth inhibition) 5FU. Statistical analysis of microarrays (SAM) identified 247 genes significantly altered by >2-fold during treatment of rectal cancer patients [71]. With the exception of patient 3 and patient 13 (treated with irinotecan), all the patients show decreased expression of these 247 genes during treatment (relative to pre-treatment controls). This decrease was not observed in the cell lines treated with 5FU, where there was either no change or even increased expression of these 247 genes (relative to untreated controls). Red indicates increase by treatment; green, a decrease; black, no change; and gray, a missing value. A major conclusion was that treatment of patients with locally advanced rectal cancer with 5FU decreases the expression of genes involved in RNA synthesis, protein synthesis and metabolism that are positively regulated by C-MYC. Thus this study identified previously unknown mechanisms of action of 5FU in which the the drug reverses the gene expression profile associated with a known oncogene. The expression profile of SW480 (72 h), HCT116 (72 h) and HT29 (24,48 and 72 h) colorectal cancer cells with 5× IC_{50} 5FU was different from that detected in the 5FU-treated patients. This highlights the potential differences between preclinical models and the clinical situation and emphasises the importance of carrying out expression profiling studies in the setting of clinical pharmacology.

ogy of cancer therapy. With judicious study design, such as the inclusion of samples to examine the effects of treatment on gene expression in normal tissue and also the collection of earlier time-points to detect more transient responses to drug exposure, global gene expression profiling has the potential to enhance the development of both established and novel anticancer agents, and also to facilitate the understanding of their mode of action in the clinic.

5. Conclusions and future prospects

It should be clear from this Review that microarray-based gene expression profiling has, over the last few years, begun to have a major impact on the development of cancer therapeutics. The most obvious benefits have been in our enhanced understanding of the global regulatory networks of normal and tumour cells – greatly empowering the 'systems biology' approach – and also the improved classification of human cancers and our increased ability to predict the outcome of treatment for individual patients. Related to this, gene expression profiling of cancer versus normal tissues has now become a standard approach to the identification and validation of new molecular targets for therapeutic intervention. As new drugs are developed, gene expression profiling is increasingly used to investigate mechanism of action and to determine on-target *versus* off-target effects. The comparison of gene expression changes induced by the test therapeutic with that produced by knockout or knockdown of the target – increasingly by the use of RNA interference – is proving to be exceptionally valuable. Transcriptional profiling is being used to improve lead optimisation and to characterise clinical development candidates. Furthermore, there are already several examples of the use of microarrays to determine global genome expression changes that are induced in cancer tissues by drug treatment in cancer patients. In some cases, the changes are unexpected, as in our study in which the cytotoxic drug 5FU was found to reverse transcriptional changes induced by the oncoprotein C-MYC in rectal cancer [71].

Where next? We can expect increasing usage of microarrays as the value of the technology continues to be exemplified, the costs decrease, accessibility increases and genome-wide coverage is imminent [160]. As the technology matures, there is demand for greater sensitivity, reproducibility, robustness and user-friendliness. Bioinformatics is increasingly seen by the user community as the critical bottleneck. Algorithms that incorporate measures of statistical confidence are now widely available and improved platforms for sharing and storing microarray data are being developed (see Table 1). Validation of observations by alternative techniques remains important, but as the robustness and reproducibility of expression profiles continues to improve, it is increasingly likely that these molecular signatures will be accepted as stand-alone measures of biological function and pharmacological action. Progress to this goal will be helped by the development of guidelines for microarray analysis, including reproducibility and statistical aspects.

However, it is important, to emphasise that microarray gene expression profiling is not a technology that is used in isolation. The microarray platform is being rolled out to allow a wide range of analytical readouts. In cancer, DNA loss and amplification, together with sequence variation, can be measured routinely. Single nucleotide polymorphism analysis will allow DNA polymorphisms to be determined on a genome-wide scale. Microarray analysis of proteins is developing rapidly and the use of tissue arrays for higher throughput biochemical analysis is now widespread. Gene expression microarrays and related global technologies are rapidly taking their place alongside other technologies as important tools in the discovery, development and use of new cancer therapeutics. The development of safe and effective drugs remains challenging. Nevertheless, these new technologies will improve the rate at which novel molecular therapeutics are developed and evaluated, and they increase the likelihood that the future of cancer medicine will involve individualised therapies targeted to the genomic and molecular abnormalities of individual patients and their cancers.

Conflict of interest

Paul Workman is a consultant to Avalon Pharmaceuticals.

Acknowledgements

We thank our colleagues and collaborators for valuable interaction and discussions. The authors' work is supported primarily by Cancer Research UK, of which PW is a Life-Fellow.

References

1. Lander ES, Linton RM, Birrer B, *et al*. 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**, 1304–1351.
3. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953, **171**, 737–738.
4. Grimwood J, Schmutz J. Genomics: six is seventh. *Nature* 2003, **425**, 775.
5. Workman P. Overview: changing times: developing cancer drugs in genomeland. *Curr Opin Investig Drugs* 2001, **2**, 1128–1135.

6. van Duin M, Woolson H, Mallinson D, Black D. Genomics in target and drug discovery. *Biochem Soc Trans* 2003, **31**, 429–432.
7. Workman P, Kaye SB. Translating basic research into new cancer therapeutics. *Trends Mol Med* 2001, **8**, S1–9.
8. Jain M, Arvanitis C, Chu K, *et al*. Sustained loss of a neoplastic phenotype by brief inactivation of MYC. *Science* 2002, **297**, 102–104.
9. Chin L, Tam A, Pomerantz J, *et al*. Essential role for oncogenic Ras in tumour maintenance. *Nature* 1999, **400**, 468–472.
10. Weinstein IB. Cancer. Addiction to oncogenes–the Achilles heal of cancer. *Science* 2002, **297**, 63–64.
11. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000, **100**, 57–70.
12. Baselga J, Averbuch SD. ZD1839 ('Iressa') (1,2) as an anticancer agent. *Drugs* 2000, **60**, 33–40.
13. Slamon D, Pegram M. Rationale for trastuzumab (Herceptin) in adjuvant breast cancer trials. *Semin Oncol* 2001, **28**, 13–19.
14. Druker BJ. STI571 (Gleevec) as a paradigm for cancer therapy. *Trends Mol Med* 2002, **8**, S14–18.
15. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Gen* 1999, **21**, 10–14.
16. Heyde P, Qi R, Abernathy K, *et al*. A concise guide to microaray analysis. *Biotechniques* 2000, **29**, 548–562.
17. The chipping forecast. *Nat Gen* 1999; (Suppl. 21):1–160.
18. The chipping forecast II. *Nat Gen* 2002; (Suppl. 32):461–552.
19. Eisen MB, Brown PO. DNA arrays for analysis of gene expression. *Methods Enzymol* 1999, **303**, 179–205.
20. Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 2000, **405**, 827–836.
21. Quackenbush J. Microarray data normalisation and transformation. *Nat Gen*(Suppl. 32), 496–501.
22. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979, **74**, 829–836.
23. Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. *Bioinformatics* 2001, **17**, 509–519.
24. Bittner M, Meltzer P, Trent J. Data analysis and integration: of steps and arrows. *Nat Gen* 1999, **22**, 213–215.
25. Young RA. Biomedical discovery with DNA arrays. *Cell* 2000, **102**, 9–15.
26. Bassett DE, Eisen MB, Boguski MS. Gene expression informatics - it's all in your mine. *Nat Gen*(Suppl. 21), 51–55.
27. Ermolaeva O, Rastogi M, Pruitt KD, *et al*. Data management and analysis for gene arrays. *Nat Gen* 1998, **20**, 19–23.
28. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998, **95**, 14863–14868.
29. Tamayo P, Slonim D, Mesirov J, *et al*. Interpretting the patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999, **96**, 2907–2912.
30. Golub TR, Slonim DK, Tamayo P, *et al*. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, **286**, 531–537.
31. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001, **98**, 5116–5121.
32. Khan J, Wei JS, Ringner M, *et al*. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001, **7**, 673–679.
33. Workman P. New drug targets for genomic cancer therapy: successes, limitations, opportunities and future challenges. Curr. *Cancer Drug Targets* 2001, **1**, 33–47.
34. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Gen* 2003, **33**, 238–244.
35. Szallasi Z. Gene expression patterns and cancer. *Nat Biotech* 1998, **16**, 1292–1293.
36. Yeoh EJ, Ross ME, Shurtleff SA, *et al*. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002, **1**, 133–143.
37. Armstrong SA, Staunton JE, Silverman LB, *et al*. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukaemia. *Nat Gen* 2002, **30**, 41–47.
38. Stam RW, den Boer ML, Meijerink JPP, *et al*. Differential mRNA expression of Ara-C-metabolising enzymes explains Ara-C sensitivity in MLL gene-rearranged infant acute lymphoblastic leukemia. *Blood* 2003, **101**, 1270–1276.
39. Ferrando AA, Neuberg DS, Staunton J, *et al*. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 2002, **1**, 75–87.
40. Alizadeh AA, Eisen MB, Davis RE, *et al*. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000, **403**, 503–511.
41. Rosenwald A, Wright G, Chan WC, *et al*. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New Engl J Med* 2002, **346**, 1937–1947.
42. Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor κB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med* 2001, **194**, 1861–1874.
43. Shipp MA, Ross KN, Tamayo P, *et al*. Diffuse large B-cell lymphoma outcome prediction by gene expression profiling and supervised machine learning. *Nat Gen* 2002, **8**, 68–74.
44. Lerner A, Kim D, Lee R. The cAMP signaling pathway as a therapeutic target in lymphoid malignancies. *Leuk Lymphoma* 2000, **37**, 39–51.
45. Teicher BA, Alvarez E, Mendelsohn LG, Ara G, Menon K, Ways DK. Enzymatic rationale and preclinical support for a potent protein kinase C beta inhibitor in cancer therapy. *Adv Enzyme Regul* 1999, **39**, 313–327.
46. Elenitoba-Johnson KS, Jenson SD, Abbott RT, *et al*. Involvement of multiple signaling pathways in follicular lymphoma transformation: p38-mitogen-activated protein kinase as a target for therapy. *Proc Natl Acad Sci USA* 2003, **100**, 7259–7264.
47. Perou CM, Serlie T, Eisen MB, *et al*. Molecular portraits of human breast tumours. *Nature* 2000, **406**, 747–752.
48. Sorlie T, Perou CM, Tibshirani R, *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001, **98**, 10869–10874.
49. Sorlie T, Tibshirani R, Parker J, *et al*. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003, **100**, 8418–8423.
50. van 't Veer LJ, Dai H, van de Vijver MJ, *et al*. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002, **415**, 530–536.
51. van de Vijver MJ, He YD, van't Veer LJ, *et al*. A gene-expression signature as a predictor of survival in breast cancer. *New Engl J Med* 2002, **347**, 1999–2009.
52. Sotiriou C, Neo S-Y, McShane LM, *et al*. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci USA* 2003, **100**, 10393–10398.
53. Dhanasekaran SM, Barrette TR, Ghosh D, *et al*. Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001, **412**, 822–826.
54. Luo J, Duggan DJ, Chen Y, *et al*. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. *Cancer Res* 2001, **61**, 4683–4688.

55. Welsh JB, Sapinoso LM, Su AI, *et al*. Analysis of gene expression identifies candidate markers and pharmalogical targets in prostate cancer. *Cancer Res* 2001, **61**, 5974–5978.
56. Singh D, Febbo PG, Ross K, *et al*. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 2002, **1**, 203–209.
57. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 2002, **62**, 4427–4433.
58. Pizer ES, Pflug BR, Bova GS, Han WF, Udan MS, Nelson JB. Increased fatty acid synthase as a therapeutic in androgen-independent prostate cancer progression. *Prostate* 2001, **47**, 102–110.
59. Garber ME, Troyanskaya OG, Schluens K, *et al*. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 2001, **98**, 13784–13789.
60. Bhattacharjee A, Richards WG, Staunton J, *et al*. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 2001, **98**, 13790–13795.
61. Beer DG, Kardia SLR, Huang C-C, *et al*. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002, **8**, 816–824.
62. Wigle DA, Jurisica I, Radulovich N, *et al*. Molecular profiling of non-small cell lung cancer and correlation with disease-free survival. *Cancer Res* 2002, **62**, 3005–3008.
63. Baird R, Workman P. Emerging molecular therapies drugs interfering with signal transduction pathways. In Bronchud M, Foote MA, Giaccone G, Olopade O, Workman P, eds. *Principles of molecular oncology*. Totowa, NJ, USA, Humana Press Inc., 2003. pp. 569–606.
64. Welsh JB, Zarrinker PP, Sapinoso LM, *et al*. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc Natl Acad Sci USA* 2001, **98**, 1176–1181.
65. Ono K, Tanaka T, Tsunoda T, *et al*. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. *Cancer Res* 2000, **60**, 5007–5011.
66. Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* 2001, **61**, 3124–3130.
67. Alon U, Barkai N, Notterman DA, *et al*. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 1999, **96**, 6745–6750.
68. Kithara O, Furukawa Y, Tanaka T, *et al*. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res* 2001, **61**, 3544–3549.
69. Zou T-T, Selaru FM, Xu Y, *et al*. Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. *Oncogene* 2002, **21**, 4855–4862.
70. Lin Y-M, Furukawa Y, Tsunoda T, Yue C-T, Yang K-C, Nakamura Y. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. *Oncogene* 2002, **21**, 4120–4128.
71. Clarke PA, George ML, Easdale S, *et al*. Molecular pharmacology of cancer therapy in human colorectal cancer by gene expression profiling. *Cancer Res* 2003, **63**, 6855–6863.
72. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, *et al*. Gene expression in colorectal cancer. *Cancer Res* 2002, **62**, 4352–4363.
73. Kihara C, Tsunoda T, Tanaka T, *et al*. Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res* 2001, **61**, 6474–6479.
74. Bittner M, Meltzer P, Chen Y, *et al*. Molecular classification of cutanteous malignant melanoma by gene expression profiling. *Nature* 2000, **406**, 536–540.
75. Carr KM, Bittner M, Trent JM. Gene-expression profiling in human cutaneous melanoma. *Oncogene* 2003, **22**, 3076–3080.
76. Weeraratna AT, Jiang Y, Hostetter G, *et al*. Wnt5a signaling directly affects cell motility and invasion of metastatic melanoma. *Cancer Cell* 2002, **1**, 279–288.
77. Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 2000, **406**, 532–535.
78. Sahai E, Marshall CJ. RHO-GTPases and cancer. *Nat Rev Cancer* 2002, **2**, 133–142.
79. Ljubimova JY, Lakhter AJ, Loksh A, *et al*. Overexpression of a4 chain-containing laminins in human glial tumors identified by gene microarray analysis. *Cancer Res* 2001, **61**, 5601–5610.
80. Rickman DS, Bobek MP, Misek DE, *et al*. Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res* 2001, **61**, 6885–6891.
81. Pomeroy SL, Tamayo P, Gaasenbeek M, *et al*. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002, **415**, 436–442.
82. MacDonald TJ, Brown KM, LaFleur B, *et al*. Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat Gen* 2001, **29**, 143–152.
83. Branca M. Putting gene arrays to the test. *Science* 2003, **300**, 238.
84. Ellwood-Yen K, Graeber TG, Wongvipat J, *et al*. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* 2003, **4**, 223–238.
85. Chen CD, Welsbie DS, Tran C, *et al*. Molecular determinants of resistance to antiandrogen therapy. *Nat Med* 2004, **10**, 33–39.
86. Kang Y, Siegel PM, Shu W, *et al*. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003, **3**, 537–549.
87. Lamb J, Ramaswarmy S, Ford HL, *et al*. A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell* 2003, **114**, 3230–3340.
88. Vernell R, Helin K, Muller H. Identification of target genes of the p16INK4A-pRb-E2F pathway. *J Biol Chem* 2003, **278**, 46124–46137.
89. Muller H, Bracken AP, Vernell R, *et al*. E2Fs regulate the expression of genes involved in differention, development proliferation, and apoptosis. *Genes Dev* 2001, **15**, 267–285.
90. Iyer VR, Eisen MB, Ross DT, *et al*. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999, **283**, 83–87.
91. 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**, 727–741.
92. Schulze A, Lehmann K, Jefferies HBJ, McMahon M, Downward J. Analysis of the transcriptional program induced by Raf in epithelial cells. *Genes Dev* 2001, **15**, 981–994.
93. Svaren J, Ehrig T, Abdulkadir SA, Ehrengruber MU, Watson J, Milbrandt J. EGR1 target genes in prostate carcinoma cells identified by microarray analysis. *J Biol Chem* 2000, **275**, 38531–385240.
94. Leaner VD, Kinoshita I, Birrer MJ. AP-1 complexes containing cJun and JunB cause cellular transformation of Rat1a fibroblasts and share transcriptional targets. *Oncogene* 2003, **22**, 5616–5629.
95. Huang E, Ishida S, Pittman J, *et al*. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat Gen* 2003, **34**, 226–230.

96. Roberts CJ, Nelson B, Marton MJ, *et al*. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* 2000, **287**, 873–880.
97. Cho RJ, Campbell MJ, Winzeler EA, *et al*. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* 1998, **2**, 65–73.
98. Gray NS, Wodicka L, Thunnissen A-MWH, *et al*. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* 1998, **281**, 533–538.
99. Marton MJ, DeRisi JL, Bennett HA, *et al*. Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat Med* 1998, **4**, 1293–1301.
100. Hughes TR, Marton MJ, Jones AR, *et al*. Functional discovery via a compendium of expression profiles. *Cell* 2000, **102**, 109–126.
101. Black EP, Huang E, Dressman H, *et al*. Distinct gene expression phenotypes of cells lacking Rb and Rb family members. *Cancer Res* 2003, **63**, 3716–3723.
102. Andreyev HJ, Norman AR, Cunningham D, Oates JR, Clarke PA. Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *J Natl Cancer Inst* 1998, **90**, 675–684.
103. Ross PJ, George M, Cunningham D, *et al*. Inhibition of Kirsten-ras expression in human colorectal cancer using rationally selected Kirsten-ras antisense oligonucleotides. *Mol Cancer Ther* 2001, **1**, 29–41.
104. Kamath RS, Fraser AG, Dong Y, *et al*. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* 2003, **421**, 231–237.
105. Chi JT, Chang HY, Wang NN, Chang DS, Dunphy N, Brown PO. Genomewide view of gene silencing by small interfering RNAs. *Proc Natl Acad Sci USA* 2003, **100**, 6343–6346.
106. Semizarov D, Frost L, Sarthy A, Kroeger P, Halbert DN, Fesik SW. Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci USA* 2003, **100**, 6347–6352.
107. Berns K, Hijmans EM, Mullenders J, *et al*. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 2004, **428**, 431–437.
108. Varambally S, Dhanasekaran SM, Zhou M, *et al*. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002, **419**, 624–629.
109. Ross DT, Scherf U, Eisen MB, *et al*. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Gen* 2000, **24**, 227–235.
110. Scherf U, Ross DT, Waltham M, *et al*. A gene expression database for the molecular pharmacology of cancer. *Nat Gen* 2000, **24**, 236–244.
111. Wallqvist A, Rabow AA, Shoemaker RH, Sausville EA, Covell DG. Establishing connections between microarray expression data and chemotherapeutic cancer pharmacology. *Mol Cancer Ther* 2002, **1**, 311–320.
112. Staunton JE, Slonim DK, Coller HA, *et al*. Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci USA* 2001, **98**, 10787–10792.
113. Dan S, Tsunoda T, Kitahara O, *et al*. An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines. *Cancer Res* 2002, **62**, 1139–1147.
114. Musumarra G, Condorelli DF, Scire S, Costa AS. Shortcuts in genome-scale cancer pharmacology research from multivariate analysis of the national cancer institute gene expression database. *Biochem Pharmacol* 2001, **62**, 547–553.
115. Blower PE, Yang C, Fligner MA, *et al*. Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. *Pharmacogenom J* 2002, **2**, 259–271.
116. Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci USA* 2002, **97**, 12182–12186.
117. Moriyama M, Hoshida Y, Otsuka M, *et al*. Relevance network between chemosensitivity and transcriptome in human hepatoma cells. *Mol Cancer Ther* 2003, **2**, 199–205.
118. Zembutsu H, Ohnsihi Y, Furukawa Y, *et al*. Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Res* 2002, **62**, 518–527.
119. Reinhold WC, Kouros-Mehr H, Kohn KW, *et al*. Apoptotic susceptibility of cancer cells selected for camptothecin resistance: gene expression profiling, functional analysis, and molecular interaction mapping. *Cancer Res* 2003, **63**, 1000–1011.
120. Watts GS, Futscher BW, Isett R, Gleason-Guzman M, Kunkel SE, Salmon SE. cDNA microarray analysis of multidrug resistance: doxorubicin selection produces multiple defects in apoptosis signaling pathways. *J Pharmacol Exp Ther* 2001, **299**, 434–441.
121. Wittig R, Nessling M, Will RD. Candidate genes for cross-resistance against DNA-damaging drugs. *Cancer Res* 2002, **62**, 6698–6705.
122. 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**, 5027–5030.
123. Wang W, Marsh S, Cassidy J, McLeod HL. Pharmacogenomic dissection of resistance to thymidylate synthetase inhibitors. *Cancer Res* 2001, **61**, 5505–5510.
124. Lamendola DE, Duan Z, Yusef RZ, Seiden MV. Molecular description of evolving paclitaxel resistance in the SKOV-3 human ovarian carcinoma cell line. *Cancer Res* 2003, **63**, 2200–2205.
125. Workman P. Challenges of PK/PD measurements in modern drug development. *Eur J Cancer* 2003, **38**, 2189–2193.
126. Zhou Y, Gwadry FG, Reinhold WC, *et al*. Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependent effects. *Cancer Res* 2002, **62**, 1688–1695.
127. Amundson SA, Bittner M, Chen Y, Trent J, Meltzer P, Fornace AJ. Fluorescent cDNA microarray hybridisation reveals complexity and heterogeneity of cellular genotoxic stress responses. *Oncogene* 1999, **18**, 3666–3672.
128. Daoud SS, Munson PJ, Reinhold W, *et al*. Impact of p53 knockout and topotecan treatment on gene expression profiles in human colon carcinoma cells: a pharmacogenomic study. *Cancer Res* 2003, **63**, 2782–2793.
129. Kudoh K, Ramanna M, Ravatn R, *et al*. Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. *Cancer Res* 2000, **60**, 4161–4166.
130. Chang B-D, Swift ME, Shen M, Fang J, Broude EV, Roninson IB. Molecular determinants of terminal growth arrest induced in tumor cells by a chemotherapeutic agent. *Proc Natl Acad Sci USA* 2002, **99**, 389–394.
131. Yu Q, He M, Lee NH, Liu ET. Identification of Myc-mediated death response pathways by microarray analysis. *J Biol Chem* 2002, **277**, 13059–13066.
132. Arango D, Corner GA, Wadler S, Catalano PJ, Augenlicht LH. c-myc/p53 interaction determines sensitivity of human colon carcinoma cells to 5-fluorouracil *in vitro* and *in vivo*. *Cancer Res* 2001, **61**, 4910–4915.
133. Taxman DJ, Mackinaw JP, Clements C, Bergstralh DT, Ting JP-Y. Transcriptional profiling of targets for combination therapy of lung carcinoma with paclitaxel and Mitogen-activated protein/extracellular signal-related kinase kinase inhibitor. *Cancer Res* 2003, **63**, 5095–5104.

134. Maxwell PJ, Longley DB, Latif T, *et al.* Identfication of 5-fluorouracil-inducible target genes using cDNA microarray profiling. *Cancer Res* 2003, **63**, 4602–4606.
135. Friedman D, Hu Z, Kolb EA, Gorfajn B, Scotto KW. Ecteinascidin-743 inhibits activated but not constitutive transcription. *Cancer Res* 2002, **62**, 3377–3381.
136. Bottone jr FG, Martinez JM, Collins JB, Afshari CA, Eling TE. Gene modulation by the cyclooxgenase inhibitor, sulindac sulphide, in human colorectal carcinoma cells. *J Biol Chem* 2003, **278**, 25790–25801.
137. Easdale S, Clarke PA, Titley J, *et al.* Genes expression and cell cycle analysis of HT29 human colon cancer cells, exposed to celecoxib, indomethacin, sulindac sulphide and sulindac sulphone. *Proc Am Assoc Cancer Res* 2003, **44**, 914.
138. Vivanco I, Sawyers CL. The phosphatidylinositol 3-kinase AKT pathway in human cancer. *Nat Rev Cancer* 2002, **2**, 489–501.
139. Kristeleit R, Stimson L, Workman P, Aherne W. Histone modification enzymes: novel targets for cancer drugs. *Expert Opin Emerg Drugs* 2004, **9**, 135–154.
140. Glaser KB, Staver MJ, Waring JF, Stender J, Ulrich RG, Davidsen SK. Gene expression profiling of multiple histone deacetylase (hdac) inhibitors: defining a common gene set produced by HDAC inhibition in T24 and MDA carcinoma cell lines. *Mol Cancer Ther* 2003, **2**, 151–163.
141. Liang G, Gonzales FA, Jones PA, Orntoft TF, Thykjaer T. Analysis of gene induction in human fibroblasts and bladder cancer cells exposed to the methylation inhibitor. *Cancer Res* 2002, **62**, 961–966.
142. Vanaja DV, Cheville JC, Iturria SJ, Young CYF. Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. *Cancer Res* 2003, **63**, 3877–3882.
143. Lam LT, Pickeral OK, Peng AC, *et al.* Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anticancer drug flavopiridol. *Genome Biol* 2001, **2**., RESEARCH 0041.
144. Whittaker SR, te Poele R, Walton MI, *et al.* Gene expression profiling of the cyclin-dependent kinase inhibitor CYC202 (R-roscovitine). *Proc Am Assoc Cancer Res* 2003, **44**, 797.
145. Clarke PA, Hostein I, Banerji U, *et al.* 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**, 4125–4133.
146. Aherne W, Maloney A, Sharp S, *et al.* Discovery of a novel synthetic inhibitor of the Hsp90 molecular chaperone. *Proc Am Assoc Cancer Res* 2003, **44**, 915.
147. Gunther EC, Stone DJ, Gerwien RW, Bento P, Heyes MP. Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. *Proc Natl Acad Sci USA* 2003, **100**, 9608–9613.
148. Owa T, Yokoi A, Yamazaki K, Yoshimatsu K, Yamori T, Nagasu T. Array-based structure and gene expression relationship study of antitumor sulfonamides including N-[2-[(4-hydroxyphenoyl)amino]-3-pyridinyl]-4-methoxybenzenesulfonamide and N-(3-chloro-7-indolyl)-1,4-benzenedisulfonamide. *J Med Chem* 2002, **45**, 4913–4922.
149. Yokoi A, Kuromitsu J, Kawai T, *et al.* Profiling novel sulfonamide antitumour agents with cell-based phenotypic screens and array-based gene expression analysis. *Mol Cancer Ther* 2002, **1**, 275–286.
150. Monks A, Harris E, Hose C, Connelly J, Sausville EA. Genotoxic profiling of MCF-7 breast cancer cell line elucidates gene expression modifications underlying toxicity of the anticancer drug 2-(4-amino-3-methylphenyl)-5-fluorobenzothiazole. *Mol Pharmacol* 2003, **63**, 766–772.
151. Johnson PH, Walker RP, Jones SW, *et al.* Multiplex gene expression analysis for high-throughput drug discovery: screening and analysis of compounds affecting genes overexpressed in cancer cells. *Mol Cancer Ther* 2002, **1**, 1293–1304.
152. Bol DK, Soppert DR, Glodek A. The changing landscape of drug discovery screening. *Curr Drug Discovery* 2003, **3**, 17–21.
153. Assersohn L, Gangi L, Zhao Y, *et al.* The feasibility of using fine needle aspiration from primary breast cancer for cDNA microarray analysis. *Clin Cancer Res* 2002, **8**, 794–801.
154. Buchholz TA, Stivers DN, Stec J, *et al.* Global gene expression changes during neoadjuvant chemotherapy for human breast cancer. *Cancer J* 2002, **8**, 461–468.
155. Pusztai L, Ayers M, Stec J, *et al.* Gene expression profiles obtained from fine-needle aspirations of breast cancer reliably identify routine prognostic markers and reveal large-scale molecular differences between estrogen-negative and estrogen-positive tumors. *Clin Cancer Res* 2003, **9**, 2406–2415.
156. Chang JC, Wooten EC, Tsimeizon A, *et al.* Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003, **362**, 362–369.
157. Sotiriou C, Powles TJ, Dowsett M, *et al.* Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res* 2002, **4**, R3.
158. Hofmann WK, de Vos S, Elashoff D, *et al.* Relation between resistance of Philadelphia-chromosome-positive acute lymphoblastic leukaemia to the tyrosine kinase inhibitor STI571 and gene-expression profiles: a gene-expression study. *Lancet* 2002, **359**, 481–486.
159. Cheok MH, Yang W, Pui C-H, Downing JR, Cheng C, Naeve CW, *et al.* Treatment-specific changes in gene expression discriminate *in vivo* drug response in human leukemia cells. *Nat Genet* 2003, **34**, 85–90.
160. Stears RL, Martinsky T, Schena M. Trends in microarray analysis. *Nat Med* 2003, **9**, 140–145.