

## Review

# Gene Expression Profiling in Non-Small Cell Lung Cancer: From Molecular Mechanisms to Clinical Application

Russell D. Petty,<sup>1,2,3</sup> Marianne C. Nicolson,<sup>1</sup>  
Keith M. Kerr,<sup>2</sup> Elaina Collie-Duguid,<sup>3</sup> and  
Graeme I. Murray<sup>2</sup>

Departments of <sup>1</sup>Oncology and <sup>2</sup>Pathology, Aberdeen Royal Infirmary, Aberdeen and <sup>3</sup>Oncology Research Group, Department of Medicine and Therapeutics, University of Aberdeen, Aberdeen, United Kingdom

## Abstract

Non-small cell lung cancer (NSCLC) is the most common cause of premature death from malignant disease in western countries. A better understanding of the molecular mechanisms underlying NSCLC etiology, pathogenesis, and therapeutics will lead to improved clinical outcomes. Recent technological advances in gene expression profiling (in particular, with cDNA and oligonucleotide microarrays) allow the simultaneous analysis of the expression of thousands of genes. In this review, the technology of global gene expression profiling is discussed, and the progress made thus far with it in NSCLC is reviewed. A new molecular classification of NSCLC has been developed, which has provided important insights into etiology and pathogenesis. Other studies have found potential biomarkers for NSCLC that may be of use in diagnosis, screening, and assessing the effectiveness of therapy. Finally, advances have been made in the understanding of the molecular mechanisms of NSCLC progression and the molecular mechanisms of action of currently used cytotoxic drugs. This may facilitate the improvement of current therapeutics and the identification of novel targets. Taken together, these advances hold the promise of an improved understanding of the molecular biology of NSCLC and its treatment, which in turn will lead to improved outcomes for this deadly disease.

## Introduction

Non-small cell lung cancer (NSCLC) is the most common human malignancy in western countries, but unfortunately, current 5-year survival is of the order of 8–12% (1). Improved

outcomes for NSCLC are therefore clearly needed. There are a number of different strategies currently under evaluation. However, a better understanding of the molecular mechanisms that determine clinical outcomes is likely to provide the basis for more effective therapeutic intervention. Transcriptome-wide gene expression profiling may provide the means to achieve this (Fig. 1).

At the molecular level, there are deficiencies in our current understanding of NSCLC etiology, pathogenesis, and therapeutics. In the following sections, we discuss the current state of understanding, progress, and problems in each of these areas and suggest how elucidation of the important underlying molecular mechanisms could lead to improved management of NSCLC.

**Etiology and Pathogenesis.** Tobacco smoke (including environmental tobacco smoke) is the major etiological factor (2). However, detailed information on the precise carcinogens in tobacco smoke, the molecular mechanisms of carcinogenesis, and individual predispositions to particular carcinogens has yet to be fully elucidated (2). Current pathological classification of NSCLC identifies three main histological subtypes: (a) adenocarcinoma (AC); (b) squamous cell carcinoma (SCC); and (c) large cell carcinoma. Whereas this classification identifies some differences in etiology, pathogenesis, and clinicopathological behavior, these differences are not sufficiently distinguishing to be useful in the clinical management of NSCLC. Therapeutic decisions are based on disease stage, performance status, and comorbidities, not on histological or molecular classification.

**Treatment.** Surgery is the main therapeutic option for curative treatment, but no more than 20% of patients have resectable disease (Stage I–IIIA) at presentation. Whereas surgery represents the best chance of cure in these patients, the outcomes after surgical resection are presently suboptimal. Radical radiotherapy alone can achieve long-term disease control and sometimes cure technically resectable tumors in patients who are medically unfit for surgery (3). In advanced disease, systemic chemotherapy can provide good symptom control with resultant improved quality of life in the majority of patients and also prolongs survival (4). Current data suggest that the combination of a platinum drug (carboplatin or cisplatin) with one of the “new generation” of cytotoxic agents (gemcitabine, docetaxel, and vinorelbine) provides optimal efficacy and tolerability of cytotoxic therapy (4, 5). Whereas adjuvant chemotherapy with current cytotoxic agents appears to have a minor role, the use of neoadjuvant chemotherapy in resectable NSCLC may improve cure rates and is currently under evaluation in several large randomized controlled trials (6). Recent data have demonstrated the efficacy and good tolerability of an oral targeted therapy, IRESSA, an epidermal growth factor receptor tyrosine kinase inhibitor (7, 8). Whereas this is undoubtedly an active drug, most recent trials combining IRESSA with chemotherapy have revealed that more detailed analysis and additional clinical

Received 10/30/03; revised 1/19/04; accepted 2/3/04.

**Grant support:** R. Petty's clinical research is supported by The Friends of the Aberdeen and Northern Centre for Oncology, Haematology, and Radiotherapy (ANCHOR), Tenovus Scotland, and the Grampian University Hospital NHS Trust Research and Development Office.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Requests for reprints:** Russell D. Petty, Oncology Group, Department of Medicine and Therapeutics, University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZN, United Kingdom. E-mail: r.d.petty@abdn.ac.uk.

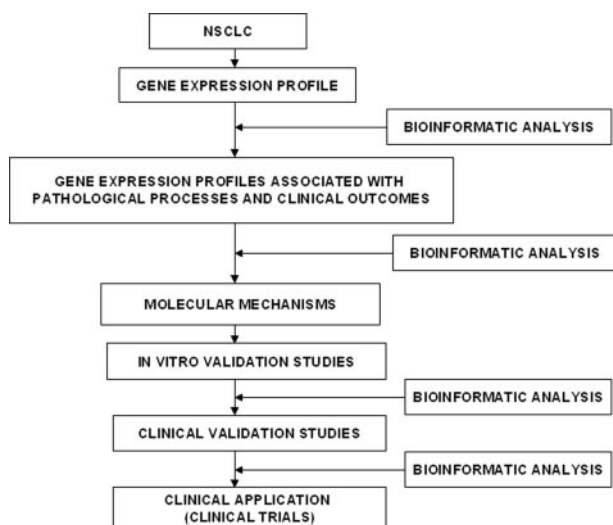


Fig. 1 Scheme for the use of gene expression profiling to elucidate the molecular mechanisms underlying important pathological processes and clinical outcomes with resulting clinical application and improved clinical outcome in non-small cell lung cancer (NSCLC).

trials will be necessary to evaluate the role of this new targeted therapy in the clinical management of NSCLC (9).

**From Molecular Mechanism to Clinical Application.**

A fuller understanding of NSCLC etiology, pathogenesis, and therapeutics would be afforded by elucidation of the molecular mechanisms involved. This would lead to more effective therapeutic intervention in NSCLC, resulting in improved clinical outcomes. For example, the clinical use of cytotoxic drugs in NSCLC, as in all common solid tumors, is limited due to inadequacies in sensitivity due to intrinsic or acquired resistance and poor selective toxicity resulting in significant and often unacceptable toxicity. These problems relate in part to a lack of understanding of the molecular mechanisms of action of cytotoxic drugs. This means that their use clinically is empirical, based on the best response rates and overall survival seen in large clinical trials. Such an approach requires the use of combination chemotherapy with increased toxicity and does not take into account the heterogeneity of chemosensitivity evident in patients. It is clear from clinical data that subsets of patients exist with differential sensitivities to different agents (5, 10, 11). At present, there is no means to identify these subsets before the instigation of therapy. Elucidation of the molecular mechanisms of cytotoxic drug action and the resultant identification of markers for individual patient chemosensitivity and prognosis would allow a paradigm change in the management of malignant disease by individualization of therapy. This would allow optimization of the chances of response and minimization of toxicity and would also allow novel therapeutic target identification. Fig. 2 summarizes this approach.

**The Potential of Transcriptome-Wide Gene Expression Profiling.**

It is clear from investigation to date that the molecular mechanisms concerned are extremely complex. Duplication, cross-talk, and redundancy of molecules and pathways contribute to this complexity, which far exceeds initial assumptions based on molecular amplification cascades within single

pathways (5, 10–16). Accordingly, whereas many workers have produced both clinical and preclinical data demonstrating the importance of particular molecules in the determination of particular phenotypes and clinical outcomes, none have proven sufficiently useful thus far to bring about alterations in the clinical management of NSCLC (10). The problem is magnified by data that illustrate the complexity and full extent of molecular abnormalities in cancer in general (17). The majority may represent the effects of neoplastic transformation rather than the cause, and many may be of minimal relevance in phenotypic determination. Nevertheless, their total represents a complex system that will behave in a complex manner in response to any therapeutic intervention. Overall, it is becoming increasingly clear that genome-, transcriptome-, or proteome-wide analysis will be necessary to explain the complex, clinically important phenotypes that will determine the success or failure of any intervention. In this review, the use of global transcriptome gene expression profiling in NSCLC is discussed.

**Transcriptome-Wide Gene Expression Profiling**

**The Technology.**

Early techniques developed for gene expression profiling included differential display and serial analysis of gene expression. Although these techniques are useful, there are a number of drawbacks; in particular, they are time-consuming, and there are problems associated with sensitivity and reproducibility. In recent years, microarray technology has been developed. This has allowed automation and a degree of standardization that eliminates many of the problems encountered with earlier techniques. In contrast to serial analysis of gene expression and differential display, which are often described as being “open” because they do not require prior knowledge of the presence and sequence of genes to be analyzed, microarrays require the sequence of the genes to be known. In this way, they are sometimes described as being “closed” gene expression technologies. The fundamental principle of microarrays is complementary hybridization of cRNA

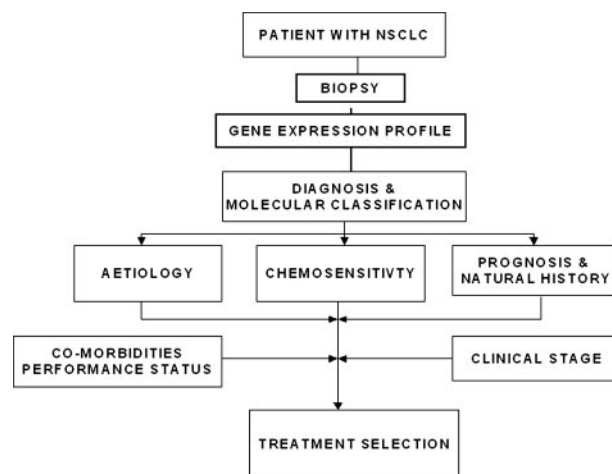


Fig. 2 Outline of scheme for therapy selection in non-small cell lung cancer (NSCLC) based on the use of gene expression profiling to provide a molecular classification of non-small cell lung cancer that forms the basis for an individualization of therapy.

or cDNA from the sample to probes representing the genes of interest. There are two types of microarrays used for gene expression profiling: spotted microarrays, which involve the use of robotics to spot customized cDNA, oligonucleotides, or PCR products on a glass slide; and the Affymetrix Genechip system, which uses prefabricated oligonucleotide microarrays. A detailed description of these technologies is not provided in this review. However, Fig. 3 provides an overview of the four main gene expression profiling techniques, and the reader is directed to the excellent reviews in Refs. 18–20 for further information.

**RNA Quality in Clinical Specimens.** The important issue of obtaining RNA of a sufficiently good quality for gene expression profiling studies from clinical specimens is worthy of separate discussion. The analysis of clinical specimens is a vital and essential part of research in this area, and good-quality RNA with minimal degradation is essential for valid profiling results and meaningful data analysis. The collection of samples must fit into current routines of clinical practice as far as possible. However, because many messenger RNAs have half-lives of less than 30 min, if appropriate measures are not taken rapidly after sample collection, then unacceptable degradation will occur.

Our own approach to this issue has been to establish a dedicated specimen collection service to allow the rapid collection of resected tumor specimens, their subsequent transportation to the pathology department, and the provision by the pathologist of representative biopsies from the specimen that are immediately snap frozen in liquid nitrogen. With this system working optimally, RNA of sufficiently good quality is obtained in virtually all cases. However, a degree of pragmatism is essential because the quality of RNA obtained from clinical specimens is unlikely ever to be as high as that obtained from cell culture or from animal tissues. In addition, there are a number of issues outside the control of the investigator that will influence RNA quality and that may be variable. For example, the time from the ligation of the arterial blood supply to the tumor (typically early in the procedure) until the resection and collection of the specimen is variable, as is the RNase activity in different tumors and different normal tissues from different

patients. The development of small sample preparation protocols may facilitate the use of clinical materials, and Lim *et al.* (21) recently reported in this journal gene expression profiling using material obtained from fine-needle aspirates of NSCLC.

RNAlater (Ambion Inc.) provides a means by which some of the issues concerning clinical sample collection can be addressed. The stability of RNA from human tissues collected into RNAlater has been well documented (for information, see the manufacturer's website<sup>4</sup>). Our own experience is that biopsy specimens can be easily collected in the clinical setting into RNAlater at room temperature, and RNA quality can be maintained for several hours. In addition, we have developed protocols that allow good-quality frozen sections to be obtained from specimens thus collected. This is essential to confirm histopathological diagnosis and collect other important histopathological information, such as tumor grade and tumor to normal cellularity. Similarly, a recent paper in this journal has addressed these issues regarding the collection of breast cancer biopsies into RNAlater and reported that good-quality RNA and frozen sections were obtained (22). In this study, satisfactory results from transcriptome profiling on a cDNA spotted microarray were obtained with samples thus collected and prepared.

The collection of specimens into RNAlater can easily be performed at the time of clinical procedures, and the specimen can be transported to the laboratory by routine means. Accordingly, this collection protocol fits well into current clinical practice, and RNAlater may prove to be suitable for sample collection in clinical trials and, ultimately, in clinical practice. The study by Ellis *et al.* (22) cited above is important because it recognizes the issues discussed here regarding clinical sample collection and successfully addresses them. More work in this area and the recognition of the importance of publication of such studies are needed to optimize clinical sample collection.

**Data Analysis and Bioinformatics.** Transcriptome-wide gene expression profiling generates large amounts of data. The analysis of this data is critical to the interpretation of experiments, and the magnitude of data means that this is a complex process. Accordingly, a new field of mathematics, bioinformatics, has emerged. In many instances, the optimal method of analysis is still a matter of debate, reflecting the early development of the discipline. However, a number of methodologies are becoming increasingly established. Whereas investigators do not need to have a comprehensive knowledge of the details of the mathematical algorithms involved, they do require an overall appreciation of the mathematic principles concerned and the uses and limitations of each method of analysis.

Initial data analysis involves production of absolute measurements of gene expression and comparative analysis. Subsequent, more extended analysis involves the use of this data to elucidate the molecular mechanisms underlying the important biological, pathological, and clinical processes. Further discussion of the bioinformatic analysis is not provided here, but a useful review with particular reference to cancer research is provided by Chung *et al.* (23), and the commonly used processes of data analysis and their use are summarized in Fig. 4.

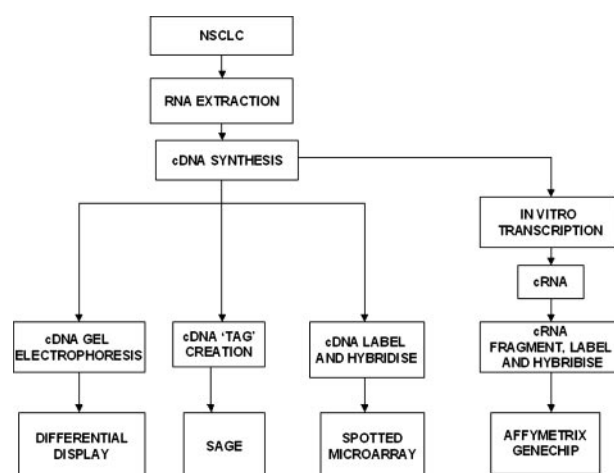


Fig. 3 Overview of technologies available for gene expression profiling.

<sup>4</sup> <http://www.ambion.com/>.

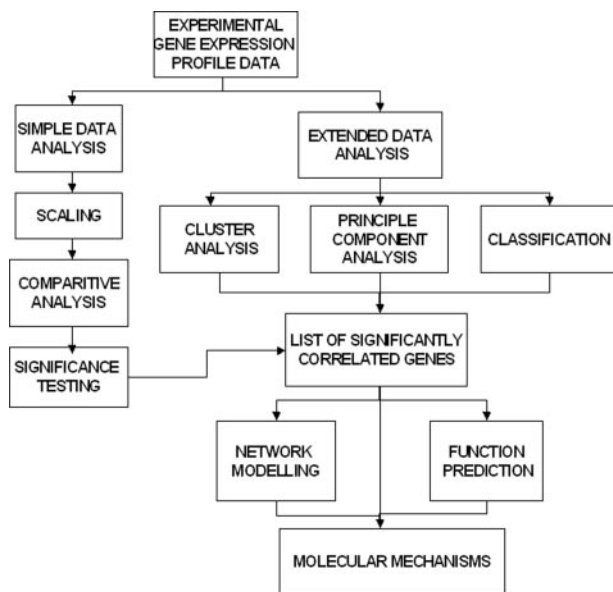


Fig. 4 Overview of bioinformatic analysis of gene expression profiling data. More details are given in the text.

Recent initiatives in the field have recognized the importance of standardization of reporting of data from microarray experiments. This will facilitate reproducibility, interpretation, and comparison of data produced by different groups. Criteria that define a minimum amount of information that must be provided have been established for gene expression data, and the editors of the highest impact journals have adopted these as a requirement for publication [minimum information about microarray experiment (MIAME)]. The Microarray Gene Expression Data Society and, within this, the Microarray and Gene Expression Group have been established as international bodies to facilitate the reporting of data in MIAME-compliant format and ensure its availability in the public domain. The availability of databases of gene expression data will provide an invaluable resource for studies on clinical material.

**Validation of Gene Expression Data.** Although the frequency of false positive and false negative expression data with most microarray technologies is very low, it is now accepted that validation of gene expression data must be performed using a different technique. Quantitative reverse transcription-PCR (RT-PCR) is accepted as the “gold standard.” Most investigators select a panel of candidate genes from their experiments for quantitative RT-PCR that represent the full range of expression seen. The majority of investigators also examine levels of protein expression with Western blots for the same selected candidates. Although this is not strictly required for validation of the gene expression data, it provides useful information with regard to the biology of any interesting candidate genes identified in the experiment and has evolved, for most investigators, into standard practice. However, the relationship between mRNA and protein levels is likely to be complex (24, 25).

The field of global proteomic profiling is developing rapidly. There are a number of potential advantages in proteomic profiling, for example, the ability to directly assay the protein as

effectors of cellular processes and the possibility of investigating different posttranslational modifications, all within the same experiment. However, at present (and likely for several years to come), the technology is inferior to that available for transcriptome profiling in terms of sensitivity, reproducibility, and the proportion of the proteome that can be analyzed within single experiments. Accordingly, investigations with transcriptome profiling are proceeding rapidly; recent advances in the field of RNA interference have suggested the potential for the development of a new class of drugs based on this mechanism and underscore the importance of transcriptome profiling research. In the longer term, it is likely that transcriptomic and proteomic global profiling technologies will be used in a complementary and cooperative fashion in the investigation of NSCLC. However, as already stated, the relationship between mRNA and protein levels (and function) is likely to be complex. Thus, Chen *et al.* (24) have reported a discordance between mRNA and protein levels for the majority of the 165 protein spots studied in lung adenocarcinomas, but Celis *et al.* (25) have suggested a good concordance between protein and microarray-determined transcript levels for 40 well-resolved abundant proteins in bladder cancer.

### Transcriptome-Wide Gene Expression Profiling in NSCLC

These studies can be divided into the five groups outlined below. In each section, the major findings and their implications are discussed along with representative individual studies. Finally, all of the studies are discussed in relation to their role in elucidating the molecular mechanisms that underlie important biological processes and clinical outcomes in NSCLC. The recent moves discussed above to standardize the methods of reporting and archiving of data from gene expression microarray experiments have facilitated the full interpretation and comparison of studies. Nevertheless, care must still be taken in interpretation with studies using different methodologies of profiling and data analysis. It remains controversial whether the most optimal analysis may be obtained from experiments with “whole tumors” [that is, tumor and surrounding stroma (and potentially contaminating normal tissues)] or microdissected tumor cells. Theoretical arguments exist supporting both positions; in particular, the importance of the tumor microenvironment and the interaction of tumor cells with this in determining tumor phenotype cannot be disregarded, but the potential confounding influence of tumor samples of low malignant to normal cellularity may also be important. A summary of the studies discussed is provided in Table 1.

**Molecular Phenotyping and Classification.** These studies have applied gene expression profiling to produce a classification of NSCLC beyond that provided by conventional histopathological means. This has provided insights and allowed hypothesis generation regarding the etiology and pathogenesis of NSCLC. Unique and characteristic gene expression profiles have been identified for each of the presently defined histological types of NSCLC (26–30).

The profiles that characterize SCC largely concern genes that are involved with cellular processes of detoxification and antioxidant proteins (*e.g.*, glutathione *S*-transferase, carboxyl

Table 1 Summary of studies discussed in text

This is not an exhaustive list, but studies have been selected as representative examples in each category.

Study and reference	Profiling platform	Specimen type (n)	Key findings	URLs
<b>Molecular phenotyping and classification</b>				
Battacharjee <i>et al.</i> (26)	Affymetrix U195A oligonucleotide microarray	Clinical lung cancer (186) Normal lung (17)	Clusters for histological subtype and normal lung Four AC <sup>c</sup> subclasses identified AC subclasses show differences in grade, smoking history, and survival	<a href="http://www.pnas.org/cgi/content/full/98/24/13790">http://www.pnas.org/cgi/content/full/98/24/13790</a> (paper) <a href="http://www.genome.wi.mit.edu/MPR/lung">http://www.genome.wi.mit.edu/MPR/lung</a> (raw data and methodological details)
Garber <i>et al.</i> (28)	Customized cDNA microarray	Clinical lung cancer specimens (67) Normal lung (5)	Clusters for histological subtype and normal lung Three AC subclasses identified AC subclasses show differences in grade and clinical stage Identified a molecular classifier for the above based on 918 cDNAs	<a href="http://www.pnas.org/cgi/content/full/98/24/13784">http://www.pnas.org/cgi/content/full/98/24/13784</a> (paper) <a href="http://www.pnas.org/">http://www.pnas.org/</a> (for supplementary information)
Beer <i>et al.</i> (31)	Affymetrix HuGeneFL oligonucleotide microarray	Clinical lung ACs (86)	Three AC subclasses identified AC subclasses show differences in grade, clinical stage, and smoking history 967 genes identified as being significantly different between stage I and stage II tumors	<a href="http://www.nature.com/cgi-taf/DynaPage.taf?file=/nm/journal/v8/n8/full/nm733.html">http://www.nature.com/cgi-taf/DynaPage.taf?file=/nm/journal/v8/n8/full/nm733.html</a> (paper) <a href="http://www.nature.com/nm/journal/v8/n8/extref/nm733-S1.pdf">http://www.nature.com/nm/journal/v8/n8/extref/nm733-S1.pdf</a> (Supplementary Fig. A) <a href="http://www.nature.com/nm/journal/v8/n8/extref/nm733-S2.pdf">http://www.nature.com/nm/journal/v8/n8/extref/nm733-S2.pdf</a> (supplementary methods) <a href="http://www.nature.com/nm/journal/v8/n8/extref/nm733-S3.pdf">http://www.nature.com/nm/journal/v8/n8/extref/nm733-S3.pdf</a> (supplementary table A)
Giordano <i>et al.</i> (32)	Affymetrix HuGeneFL oligonucleotide microarray	Clinical ACs (57 lung, 51 colon, 46 ovary)	Molecular classifier of 30 genes for tissue of AC origin	<a href="http://ajp.amjpathol.org/cgi/content/full/159/4/1231">http://ajp.amjpathol.org/cgi/content/full/159/4/1231</a> (paper)
Nacht <i>et al.</i> (27)	SAGE and Affymetrix U95A oligonucleotide microarray	SAGE-clinical lung cancer (4) Normal lung (1) A549 cell line Microarray-clinical lung cancer (32)	Clusters for histological subtype and normal lung	<a href="http://www.pnas.org/cgi/content/full/98/26/15203">http://www.pnas.org/cgi/content/full/98/26/15203</a> (paper)
McDoniels-Silvers <i>et al.</i> (29)	Competitive cDNA library screening	Clinical lung cancer (14) Normal lung (1)	Identified 68 genes underexpressed and 4 genes overexpressed in NSCLC	<a href="http://clincancerres.aacrjournals.org/cgi/content/full/8/4/1127">http://clincancerres.aacrjournals.org/cgi/content/full/8/4/1127</a> (paper)
<b>Gene expression profiling and prognosis</b>				
Battacharjee <i>et al.</i> (26)	Affymetrix U195A oligonucleotide microarray	Clinical lung cancer (186) Normal lung (17)	Four AC subclasses identified with significantly different survivals	<a href="http://www.pnas.org/cgi/content/full/98/24/13790">http://www.pnas.org/cgi/content/full/98/24/13790</a> (paper) <a href="http://www.genome.wi.mit.edu/MPR/lung">http://www.genome.wi.mit.edu/MPR/lung</a> (raw data and methodological details)
Garber <i>et al.</i> (28)	Customized cDNA microarray	Clinical lung cancer specimens (67) Normal lung (5)	Three AC subclasses identified with significantly different survivals	<a href="http://www.pnas.org/cgi/content/full/98/24/13784">http://www.pnas.org/cgi/content/full/98/24/13784</a> (paper) <a href="http://www.pnas.org/">http://www.pnas.org/</a> (for supplementary information)
Beer <i>et al.</i> (31)	Affymetrix HuGeneFL oligonucleotide microarray	Clinical lung ACs (86)	Identified a survival classifier of 50 genes Classifier validated on an independent set of 84 lung ACs from a different institution	<a href="http://www.nature.com/cgi-taf/DynaPage.taf?file=/nm/journal/v8/n8/full/nm733.html">http://www.nature.com/cgi-taf/DynaPage.taf?file=/nm/journal/v8/n8/full/nm733.html</a> (paper) <a href="http://www.nature.com/nm/journal/v8/n8/extref/nm733-S1.pdf">http://www.nature.com/nm/journal/v8/n8/extref/nm733-S1.pdf</a> (supplementary Fig. A) <a href="http://www.nature.com/nm/journal/v8/n8/extref/nm733-S2.pdf">http://www.nature.com/nm/journal/v8/n8/extref/nm733-S2.pdf</a> (supplementary methods) <a href="http://www.nature.com/nm/journal/v8/n8/extref/nm733-S3.pdf">http://www.nature.com/nm/journal/v8/n8/extref/nm733-S3.pdf</a> (supplementary Table A)
Wigle <i>et al.</i> (35)	Customized cDNA microarray	Clinical lung cancer (39)	Identified survival classifier of 22 genes	<a href="http://cancerres.aacrjournals.org/cgi/content/full/62/11/3005">http://cancerres.aacrjournals.org/cgi/content/full/62/11/3005</a> (paper)
<b>Gene expression profiling to identify Biomarkers</b>				
Iwao <i>et al.</i> (36)	mRNA differential display	Clinical lung cancer (35) Normal lung (5) Clinical cancer-non-lung (86) Clinical lymph nodes (16) NSCLC cell lines (5)	Identified a potential novel biomarker named LUNX for NSCLC	<a href="http://www3.interscience.wiley.com/cgi-bin/fulltext/76505817/HTMLSTART">http://www3.interscience.wiley.com/cgi-bin/fulltext/76505817/HTMLSTART</a> (paper)

Table 1 Continued

Study and reference	Profiling platform	Specimen type (n)	Key findings	URLs
Hibi <i>et al.</i> (37)	SAGE	Clinical lung cancer (2) Normal lung primary cell lines (2)	Two genes ( <i>PGP9.5</i> and <i>B-Myb</i> ) identified as potential biomarkers for NSCLC	
Gene expression profiling of specific pathways and processes				
Dong <i>et al.</i> (38)	mRNA differential display and cDNA microarray (NCI mouse array slides)	Murine squamous cell carcinoma cell lines (3) Parental primary cell line (1)	Identified 53 cDNAs by both techniques that were associated with the acquisition of an invasive and subsequent metastatic phenotype in this model system  10/22 genes identified by cDNA microarray were known targets of the NF- $\kappa$ B transcription factor	<a href="http://cancerres.aacrjournals.org/cgi/content/full/61/12/4797">http://cancerres.aacrjournals.org/cgi/content/full/61/12/4797</a> (paper)
Chen <i>et al.</i> (39)	Customized cDNA microarray	Human lung AC cell lines with increasing invasiveness (5)	271 genes identified with positive correlation and 312 genes with negative correlation with increasing invasiveness in an <i>in vitro</i> assay	<a href="http://cancerres.aacrjournals.org/cgi/content/full/61/13/5223">http://cancerres.aacrjournals.org/cgi/content/full/61/13/5223</a> (paper)
Gemma <i>et al.</i> (40)	cDNA macroarray (gene navigator cancer selected array)	Human AC cell line and highly metastatic ( <i>in vivo</i> assay) daughter cell lines	Identified 3 up-regulated and 9 down-regulated genes in metastatic cell lines as compared to parental cell line	
Tran <i>et al.</i> (41)	mRNA differential display	Clinical lung ACs with matched normal lung (2)	Isolated a novel tumor suppressor gene Tumor suppressor function confirmed by transfection studies	<a href="http://cancerres.aacrjournals.org/cgi/content/full/59/1/35">http://cancerres.aacrjournals.org/cgi/content/full/59/1/35</a> (paper)
Gene expression profiling in relation to therapy				
Kikuchi <i>et al.</i> (44)	Customized cDNA microarray	Clinical NSCLC (37)	Identified genes significantly correlated with <i>in vitro</i> chemosensitivity (collagen gel droplet assay) to docetaxel, paclitaxel, irinotecan, cisplatin, gemcitabine, and vinorelbine	<a href="http://www.nature.com/cgi-taf/DynaPage.taf?file=/onc/journal/v22/n14/full/1206288a.html">http://www.nature.com/cgi-taf/DynaPage.taf?file=/onc/journal/v22/n14/full/1206288a.html</a> (paper)
Kihara <i>et al.</i> (45)	Customized cDNA microarray	Clinical squamous cell carcinoma of esophagus treated with resection followed by adjuvant Cisplatin/5FU (20)	Identified 53 genes that correlated with long and short surviving groups after adjuvant therapy  These 53 genes were used to determine a Drug Response Score (DRS) that distinguished long and short survivors in a test set of 6 cases	<a href="http://cancerres.aacrjournals.org/cgi/content/full/61/17/6474">http://cancerres.aacrjournals.org/cgi/content/full/61/17/6474</a> (paper)
Natsume <i>et al.</i> (42)	cDNA microarray (Clontech human cancer cDNA expression array)	NSCLC cell line (1) and astrocyte cell line (1)	Time course experiments identified 86 genes with differential expression in NSCLC cell line after exposure to a novel spindle poison  Several proapoptotic genes were noted among these 86 genes, and all were down-regulated after drug exposure	<a href="http://content.kluweronline.com/article/338118/fulltext.pdf">http://content.kluweronline.com/article/338118/fulltext.pdf</a> (paper)
Ohira <i>et al.</i> (43)	Customized cDNA macroarray	Clinical NSCLC post neoadjuvant platinum-based chemotherapy and matched normal tissues (3)	Identified up-regulation of angiogenesis, invasion-related, and adhesion molecules following chemotherapy, for example, fibroblast growth factor 3 matrix metalloproteinase 15, 16 and 10, integrin beta 4, integrin alpha 9, endonexin, and several types of collagens	
Dan <i>et al.</i> (46)	Customized cDNA microarray	Cancer cell lines (39, including 7 NSCLC) treated with 55 anticancer drugs	Identified gene sets of approximately 50 genes each whose expression correlated with the sensitivity of all cell lines to each mechanistically similar group of drugs	
Korn <i>et al.</i> (48)	Customized cDNA microarray	Breast cancer specimens pre- and postneoadjuvant chemotherapy (39)	Developed a novel method for identifying genes with significantly differential expression in pre- and postchemotherapy-treated specimens, step down permutation <i>t</i> test  Identified 17 genes with significantly differential expression pre- and postchemotherapy using this method	<a href="http://cancerres.aacrjournals.org/cgi/content/full/62/4/1139">http://cancerres.aacrjournals.org/cgi/content/full/62/4/1139</a> (paper)

Table 1 Continued

Study and reference	Profiling platform	Specimen type (n)	Key findings	URLs
Staunton <i>et al.</i> (14)	Affymetrix Hu6800 oligonucleotide microarray	NCI-60 cell line panel treated with 6084 drugs	232 drugs identified with variable activity across the cell lines Classifiers consisting of 5–200 genes for each drug developed Classifiers applied to independent test sets for each drug with overall 75% accuracy in prediction of cell line chemosensitivity Proof of principle that this approach can be used to develop profiles predictive of response to chemotherapy	<a href="http://www.pnas.org/cgi/content/full/98/19/10787">http://www.pnas.org/cgi/content/full/98/19/10787</a> (paper) <a href="http://www.genome.wi.mit.edu/MPR/NC160/NC160.html">http://www.genome.wi.mit.edu/MPR/NC160/NC160.html</a> (supplementary data) <a href="http://dtp.nci.nih.gov/">http://dtp.nci.nih.gov/</a> (supplementary methods) <a href="http://www.genome.wi.mit.edu/MPR">http://www.genome.wi.mit.edu/MPR</a> (supplementary methods)
Chang <i>et al.</i> (12)	Customized cDNA microarray	HCT116 human colon cancer cell line and p53 and p21 homozygous knockout mutants	Examined changes in gene expression associated with doxorubicin-induced terminal growth arrest and subsequent accelerated cell senescence Characterized 68 up-regulated genes and 29 down-regulated genes Overall changes paralleled the differences seen in normal vs. cancerous cell and suggested a normalization of cancer cells Changes also similar to those seen in replicative senescence as part of normal aging, drug-induced accelerated senescence is a similar process p53 function is not critical for drug-induced senescence, but without it, the process is delayed Overall, the molecular basis of a novel drug-induced cell death pathway was characterized and potentially useful molecular markers of senescence were identified	
Kudoh <i>et al.</i> (13)	Customized cDNA microarray	MCF-7 breast cancer cell line and doxorubicin and cisplatin-resistant daughter cell lines	Time course experiments performed on each cell line after drug exposure A far smaller number of genes showed changed expression in the resistant daughter cell lines compared to the sensitive parental cell line, indicating that the response to cytotoxic drug-induced lesions is the critical determinant of response In the sensitive parental cell line, down-regulation of a novel group of zinc finger transcription factors was seen, and expression changes were consistent with an apoptotic response Only minimal overlap between cisplatin- and doxorubicin-resistant cell lines was seen on drug exposure, indicating different mechanisms of resistance in operation	<a href="http://cancerres.aacrjournals.org/cgi/content/full/60/15/4161">http://cancerres.aacrjournals.org/cgi/content/full/60/15/4161</a> (paper)
Zembustu <i>et al.</i> (15)	Customized cDNA microarray	Human tumor xenografts (85, including 11 NSCLCs) treated with cytotoxic drugs	Identified 333 genes whose expression correlated with chemosensitivity to 2 or more drugs and 200 genes whose expression that correlated with chemosensitivity to all drugs Identified 15 genes whose expression correlated with expression only in specific tumor types Therefore, drug mechanism and tumor type-specific chemosensitivity related genes were identified	<a href="http://cancerres.aacrjournals.org/cgi/content/full/62/2/518">http://cancerres.aacrjournals.org/cgi/content/full/62/2/518</a> (paper)
Qin <i>et al.</i> (47)	cDNA microarray (Clontech human apoptosis array)	Hepatoma cell line (1)	Identified 43 genes significantly changed after exposure to cisplatin	

<sup>a</sup> AC, adenocarcinoma; SAGE, serial analysis of gene expression; NSCLC, non-small cell lung cancer; NCI, National Cancer Institute; NF- $\kappa$ B, nuclear factor  $\kappa$ B; 5FU, 5-fluorouracil.

<http://www.pnas.org/cgi/content/full/99/1/389> (paper).

<http://www.pnas.org/> (supplementary data).

esterase, and peroxiredoxin 1). This implies that SCC represents a cellular response of the bronchial epithelium to ongoing insult from environmental (likely predominantly tobacco-related) carcinogens (27, 29). From here, one might postulate that population polymorphisms in these genes may be important determinants of lung SCC predisposition and, as such, useful information to incorporate into preventative or screening strategies. Cytokeratin and cytokeratin-regulatory genes are also overexpressed in SCC (27–29).

Lung ACs are characterized and distinguished by abnormal expression of surfactant-related and small airway-associated genes [e.g., surfactants A2 and B, mucin 1, and pronapsin A (26, 28–32)]. This indicates that these tumors likely originate from the small airway epithelial cell type II pneumocytes and Clara cells (27). The overexpression of such genes may also be implicated in the pathogenesis of NSCLC; for example, mucin 1 is known to be overexpressed in breast ACs, and activation by tyrosine phosphorylation here has been shown to result in activation of the mitogen-activated protein kinase cell signaling pathway, thereby promoting cell division and propagation of the malignant cell phenotype (33).

Large cell lung cancer has a gene expression profile suggestive of an epithelial to mesenchymal transition [e.g., overexpression of HMG(Y), tissue plasminogen activator, and FOS-related antigen 1 and underexpression of E-cadherin, PAX-8, and  $\gamma$ -catenin (28)].

AC is by far the most heterogeneous type, based on gene expression profiling. Subclassification based on gene expression profiles into at least three subtypes is possible (including a neuroendocrine subtype and a subtype with similarity to large cell lung cancer). Garber *et al.* (28) used a cDNA array (23,200 cDNA clones of 17,108 functionally annotated genes) and hierarchical cluster analysis to identify 3 subgroups of ACs in 38 patients. Surfactant-related genes were highly expressed in groups 1 and 2 but poorly expressed in group 3, which overexpressed many genes involved in tissue remodeling, specifically plasminogen activator urokinase receptor and cathepsin L, which are involved in extracellular proteolysis, and vascular endothelial growth factor C and other angiogenic factors. This finding is significant because many of the tumors in group 3 were metastatic. Furthermore, the gene expression profiles of two intrapulmonary metastases from group 1 were similar and, in fact, clustered with group 3. Overall, this suggests that the gene expression profile of group 3 may be related to progression and/or acquirement of metastatic potential and appears to identify a subgroup of ACs with aggressive behavior.

In another study, Bhattacharjee *et al.* (26) used the Affymetrix Genechip system with human U95A microarrays and hierarchical clustering to identify four subgroups in 125 ACs. The four subgroups include a neuroendocrine subgroup, a group with high expression of genes associated with cell division and proliferation, a group expressing surfactant-related genes, and a group with high expression of type II pneumocyte markers. Two of the subgroups (the surfactant-expressing group and the type II pneumocyte-expressing group) have counterparts in the data of Garber *et al.* (28); however, the other findings were distinct. These differences may relate to the different gene expression profile tools used but most likely reflect the heterogeneity within each of the two samples, which, it appears, are probably

both too small to generate a classification scheme that will fully represent the total complexity in ACs in particular and NSCLC in general.

These studies have also identified gene expression profiles that correlate independently with histological grade and differentiation, tumor stage, and, interestingly, smoking history (26, 31, 34). It is possible to classify NSCLC in smokers and non-smokers on the basis of these profiles, perhaps reflecting a distinct pathogenesis. Interestingly, genes involved in gemcitabine activation (hence sensitivity) were discriminatory between smokers and nonsmokers in one study (34).

In summary, these gene expression profiles may provide the basis for a new classification of NSCLC, which will allow treatment decisions to be based on molecular pathology. Realization of this will require additional, larger studies to investigate in detail the relationship between the profiles and NSCLC etiology, pathogenesis, and biological and clinical behavior. Such studies may also prove a useful means to identify novel therapeutic targets. These points are discussed further in “Discussion of Studies and Conclusions.”

**Gene Expression Profiling and Prognosis.** The aim of these studies was to specifically identify gene expression profiles from global analysis that predict disease-free survival and overall survival, which would be useful in clinical diagnostics. Accordingly, the aim was that they should outperform current clinicopathological prognosticators (*i.e.*, distinguish differing prognoses in patients with the same histological type and clinical stage of tumor).

Several groups have been able to identify gene expression profiles involving between 22 and 50 genes that can separate otherwise identical tumors in terms of disease-free survival and overall survival after potentially curative resection (26, 28, 31, 35). These profiles could form the basis for individualized therapeutic selection and follow-up involving the rational use of adjuvant or neoadjuvant systemic therapies. All studies to date are on relatively small samples and involve retrospective analysis. Larger scale and prospective studies are necessary (5, 10, 11). Some overlap was seen between profiles produced by different groups, but there were also some distinct differences. These can be explained on the basis of differences in microarray technique, samples analyzed, and statistical methods applied, but the differences observed are probably of biological relevance themselves.

A good example of this type of study is the work of Beer *et al.* (31). Beer *et al.* (31) took a sample of 86 lung ACs (67 stage I and 19 stage III ACs) from patients who underwent potentially curative resection. They used the Affymetrix HuGeneFL oligonucleotide microarray (7129 genes) to generate gene expression profiles in these patients with mature follow-up data. A gene set to classify survival was then built by dividing all tumors into a training set and a testing set ( $n = 43 \times 2$  with equal stage I and III tumors in each set) and then correlating genes with documented survival and using the top 10, 20, 50, and 75 genes as profiles based on cutoff at the 50<sup>th</sup>, 60<sup>th</sup>, and 70<sup>th</sup> percentile for survival in the training set. They found that using top 50 genes provided the most accurate classification in each case. This 50-gene classifier was then applied to the test set, using the 60<sup>th</sup> percentile as the cutoff for low- and high-risk groups; this produced two groups with significantly different



survival ( $P = 0.024$ ). Importantly, 11 stage I tumors were placed in the high-risk group and thus identified as having a poor prognosis. When the classifier was applied only to stage I tumors in the test set, two groups with significantly different survival ( $P = 0.028$ ) were separated. Applying the classifier to the full set of stage I tumors ( $n = 67$ ) identified 46 “low-risk” and 21 “high-risk” tumors with significantly different survival ( $P = 0.005$ ). The classifier was validated using a “leave one out cross-validation” (see Ref. 21) on the full set and was also validated on an independent test set of 84 ACs from a different institution.

Other workers have attempted to identify a gene expression profile that identifies tumors with “high metastatic potential” and thereby identify patients with a high probability of micro-metastatic disease. The approach here has been to examine the differences between the gene expression profiles in tumors of indistinguishable histological type and clinical stage that do and do not metastasize (28).

#### Gene Expression Profiling to Identify Biomarkers.

These studies have used gene expression profiling to identify genes and their protein products that are specific to lung tissue and that are markedly overexpressed in NSCLCs. In this way, they have potential for clinical use as biomarkers for diagnosis (and possibly screening), response to therapy, and follow-up/early detection of metastases.

Several genes/proteins have been identified that are lung specific and overexpressed in NSCLCs and not other tumor types tested: they are mostly novel genes/proteins (36, 37). For example, Iwao *et al.* (36), using the differential display technique, found a novel gene/protein named LUNX specifically overexpressed in normal lung but not in other normal tissues. Subsequently, they used RT-PCR for LUNX and found 100% expression in 35 NSCLCs, <5% expression in 86 non-lung tumors, and no expression in 16 normal lymph node specimens. They proceeded to analyze by RT-PCR 104 lymph nodes taken from 20 patients at the time of resection for NSCLC. They found in 25% of the lymph nodes that histology was reported negative, but LUNX RT-PCR was positive, thereby upstaging these patients by the detection of micrometastases. LUNX may be useful as a NSCLC biomarker for micrometastasis detection, but it may also be a clinically useful biomarker for diagnosis, response assessment, and follow-up if the high levels of expression in NSCLC lead to high levels of protein detectable in serum, sputum, or urine.

#### Gene Expression Profiling of Specific Pathways and Processes.

These studies have used gene expression profiling as a means to investigate the molecular mechanisms of specific processes involved in carcinogenesis (usually invasion and metastases) or to identify and investigate particular gene/pathways of interest and importance in the pathogenesis of NSCLC. They include studies on clinical specimens but more commonly on cell lines that have been characterized for the process of interest in a well-defined model system.

Many groups have successfully used model systems that provide assays of invasive or metastatic ability to identify gene expression profiles that correlate with invasiveness or metastatic ability (38–40). These profiles can be used for the foundation of future work to identify the molecular mechanisms underlying the acquisition of invasive and metastatic ability. For example,

using a mouse model and a cDNA array comprising 400 cancer-associated genes (National Cancer Institute Mouse array slides), Dong *et al.* (38) have successfully identified 287 genes with >2-fold increased expression in tumor cell lines, with 34 genes showing differential expression associated with tumor progression as defined by the acquisition of metastatic ability within the model. Twenty-two were known genes. Ten of the 22 genes identified were known targets of the nuclear factor- $\kappa$ B transcription factor, thereby illustrating the importance of activation of this pathway in tumor progression and the potential importance of this pathway as a target for novel therapeutics.

Several other workers have used gene expression profiling techniques to identify novel molecules, including tumor suppressor genes involved in the pathogenesis of NSCLC (41).

#### Gene Expression Profiling in Relation to Therapy.

There are few studies that have specifically examined gene expression profiling in relation to therapy in NSCLC (42–44). However, there are many other studies that have investigated gene expression profiling in relation to cytotoxic therapy in malignant disease in general, and these are of relevance to NSCLC (12–15, 45–47).

This group of studies has been undertaken with three inter-related aims: (a) firstly, to use gene expression profiling to elucidate the mechanisms of action and resistance of specific cytotoxic drugs; (b) secondly, to examine whether gene expression profiles can be identified that can be predictive of sensitivity/resistance to particular agents that would be useful for individual patients; and (c) thirdly, to identify novel therapeutic targets and investigate their potential usefulness in clinical treatment including their integration with conventional cytotoxic therapy.

However, few studies have looked at clinical response to cytotoxic chemotherapy. This is due in part to problems with the precise assessment of clinical response to therapy; radiological evaluations may not accurately represent the degree of tumor cell death, and “harder” end points such as time to tumor progression/relapse or overall survival (although important clinically) are dependent on many (and thereby confounding) other patient-related variables. The exception to this is pathological assessment of response, which can be assessed in the setting of neoadjuvant chemotherapy; such studies have been performed in breast cancer using pre- and posttreatment samples and provided profiles of altered gene expression after chemotherapy, although no data have yet been published on the relationship of these to response, posttreatment tumor morphology, or cytotoxic mechanism of action (48). In NSCLC, we were only able to find one small study involving neoadjuvant chemotherapy. Ohira *et al.* (43) used a cDNA macroarray (568 cancer-related genes) to profile three NSCLC specimens and paired normal tissue from three patients who underwent platinum-based chemotherapy before surgery. All patients had a partial response to therapy. Significantly, they found that in postchemotherapy tissues, there was a marked up-regulation of angiogenesis-, invasion-, and adhesion-related genes, and cluster analysis revealed the close association of angiogenesis genes, which in turn could be separated into three distinct groups. These findings have important clinical implications that need to be confirmed in larger numbers of patients. They suggest that angiogenesis-related genes are important therapeutic targets in NSCLC and, in particular, that antiangiogenesis therapies may be effective against chemotherapy-

pretreated tumors. This small study illustrates the ability of gene expression profiling to answer clinically important questions and provide details regarding the molecular mechanisms involved.

Large databases of gene expression profiles that determine the sensitivity or resistance to particular cytotoxic drugs have been developed, based on studies of a large number of cytotoxic drugs in a large number of cancer cell lines (14, 15, 46). These profiles appear to be specific for the mechanism of action of the drugs, such that drugs with similar mechanisms cluster together with the same profile (46). With a novel cytotoxic drug, its mechanism of action can be predicted from the gene expression profile that its sensitivity/resistance correlates with in the panel of cell lines tested (46). The gene expression profiles for each cytotoxic drug can be applied to predict the chemosensitivity or resistance of a particular cell line to that drug with a high degree of accuracy (14). The details of the individual genes in the profiles have provided useful insights into molecular mechanisms of action (15, 46). The most comprehensive database has been generated by Staunton *et al.* using Affymetrix Hu6800 oligonucleotide microarray applied to the National Cancer Institute 60-cell line panel tested with 6,084 drugs (14). The 6,084 drugs were cut down to 232 with variable activity (SD, 1.6) across the cell lines. Subsequently, for each drug, training and testing sets were made. Using the training sets and the weighted voting classifier method, gene expression classifiers for each drug were formulated and validated using leave one out cross-validation. Classifiers contained 5–200 genes. Classifiers were then applied to test sets for each drug, and 75% accuracy in prediction of cell line sensitivity was achieved. This study provides a proof of principle that gene expression profiles can be developed that can predict response to chemotherapy. A similar approach specifically in NSCLC using clinical tumor samples has been adopted by Kikuchi *et al.* (44). They used microdissected cancer cells from NSCLCs from 37 patients and performed expression profiling using a custom-made cDNA array (23,040 genes). *In vitro* chemosensitivity was determined using the collagen gel droplet embedded culture drug sensitivity test. By calculating a Pearson's correlation coefficient with cutoffs of  $P < 0.01$  and slope of the regression line  $> 1.5$ , profiles of between 29 and 92 genes were found that correlated with *in vitro* chemosensitivity to six drugs (docetaxel, irinotecan, paclitaxel, cisplatin, vinorelbine and gemcitabine). No attempt to develop predictive profiles was made.

Time course studies with cell lines have allowed evaluation of the changes in gene expression that occur after cytotoxic drug exposure and provided useful insights into molecular mechanisms of action. Comparison of these profiles in drug-sensitive and drug-resistant daughter cell lines has provided useful insights on the mechanisms of development of resistance and the molecular basis of drug resistance. A good example of this type of work and its potential can be seen in the work of Kudoh *et al.* (13). They performed time course experiments after exposure to doxorubicin or cisplatin using the MCF-7 breast cancer cell line and MCF-7 doxorubicin- and cisplatin-resistant daughter cell lines. A customized cDNA microarray was used. They found 500 genes with altered expression in the parent cell line (2–30-fold). This began at 3 h after exposure and continued to 15 h for up-regulated genes and began within 1 h and continued to 5 h for down-regulated genes. Overall, a global down-regulation of

transcription factors was seen, but in particular, a marked down-regulation of a novel group of Zinc finger transcription factors was demonstrated. Up-regulation of cytochrome *c* and down-regulation of *bcl-2* consistent with an apoptotic response were seen. Proteasome genes and genes consistent with a  $G_2$  arrest were also up-regulated. In the doxorubicin-resistant cell line, much less gene expression change was seen after drug exposure, and 18 of the genes that were induced in the parent cell line were found to be constitutively overexpressed. In the cisplatin-resistant cell line, only minimal overlap in gene expression patterns was seen, suggesting different mechanisms of resistance. This study illustrates the complexity of the molecular response and the molecular mechanisms of action and resistance of cytotoxic drugs, but it also demonstrates the ability of transcriptome-wide gene expression profiling to investigate and begin to elucidate the mechanisms.

Another time course study is worthy of specific mention because it has provided evidence and some preliminary details about the molecular mechanisms of a novel form of cell death following cytotoxic therapy. Chang *et al.* (12) examined the gene expression changes associated with doxorubicin-induced terminal growth arrest and subsequent accelerated cell senescence. The HCT116 human colonic cancer cell line and p53 and p21 homozygous knockout HCT116 mutants were used. A spotted cDNA microarray comprising more than 9000 sequence verified genes and expressed sequence tags was used (Human Unigen 2.0; Incyte Genomics). Exposure to doxorubicin led to growth arrest and accelerated senescence in a proportion of wild-type HCT116 cells that were isolated by fluorescence-activated cell-sorting profiles of fluorescence of PKH2 (a lipophilic fluorophore that stably incorporates into the plasma membrane and evenly distributes between daughter cells). Cells that died after doxorubicin treatment were identified and excluded using propidium iodide staining. In the doxorubicin-induced senescent wild-type HCT116 cells, 68 genes were found to be up-regulated ( $\geq 2$ -fold), and 29 were found to be down-regulated ( $\geq 2$ -fold). Those up-regulated were growth-inhibitory genes; many were secreted proteins, therefore suggesting a possible paracrine or autocrine mechanism. Cell proliferation and DNA repair genes were down-regulated. Overall, the gene expression changes paralleled the differences seen in normal *versus* cancerous cells and therefore suggested a "normalization" of cancer cells occurring in the process of drug-induced accelerated cell senescence. The changes seen were also similar to those seen in the processes of replicative senescence and also as part of normal aging, suggesting that drug-induced accelerated senescence is a related process. Also up-regulated were some growth-promoting and potentially pathogenic genes, including mitogens, antiapoptotic proteins, proteases, and angiogenic factors, many with paracrine functions. Again, this is similar to replicative senescence and normal aging. The studies with the p21-null cell line revealed that the majority of these genes were under the control of this transcription factor. Studies with the p53-null cell line revealed that the same gene expression patterns were seen, but the time course was delayed, thus the senescence is dependent in part, but not entirely, on p53 function. Overall, this study has provided information on the molecular basis of drug-induced accelerated cell senescence. The genes identified may be useful targets for novel drug development that show particular promise because any drugs used are likely to

have limited cytotoxic effects and thus toxicities. Importantly, this study has also identified genes and determined gene expression profiles that could act as senescence-specific markers in future studies.

## Discussion of Studies and Conclusions

Transcriptome-wide gene expression profiling in NSCLC has provided insights into the molecular aspects of the etiology, pathogenesis, treatment, and natural history of this disease. Both studies using clinical specimens and those using model systems (*in vivo* animal studies or studies with cell lines) have proven instructive. Important findings include the ability of transcriptome-wide gene expression profiling to identify and define subgroups of NSCLC both as part of the current histological classification (thereby defining the molecular basis) and also beyond this, in the case of lung AC, to identify novel subgroups. The details of the defining gene expression profiles provide insights regarding etiology and pathogenesis, and it is of particular note that smokers and nonsmokers can be defined at the molecular level as distinct entities. Other studies have identified potential biomarkers for NSCLC and the important molecular determinants and networks that control invasiveness and metastasis and response to cytotoxic drug treatment. The proof of principle that gene expression profiles can predict probability of response to cytotoxic treatment and probability of relapse after resection is an exciting and important advance.

According to the schema proposed at the start of this review (Fig. 1), this work is at the stage of demonstrating correlative relationships between biological features and clinical outcomes of interest and importance, and these have in turn have been applied to propose molecular classifiers. The work to date therefore provides the foundation for a molecular classification of NSCLC that will integrate molecular abnormalities with the heterogeneity that is observed in etiology, pathogenesis, response to therapy, and subsequent natural history of NSCLC. Such a classification will allow innovative approaches to screening, prevention, diagnosis, therapy, and clinical follow-up of patients with NSCLC. This will be applicable within the existing framework of clinical management by rationalizing the use of currently available modalities to individual patients or subgroups of patients. This will require innovative clinical trial design in terms of how the data from gene expression profiling studies are integrated into current clinical management and also in terms of the most appropriate clinical efficacy measurements that should be used. In this area, there are many issues that remain to be resolved. For example, it is not yet clear whether predictive or prognostic profiles will take the form of pattern recognition within a global expression profile or alternatively be determined by the expression of a defined “panel” of genes. In the era of targeted therapy for NSCLC, it seems reasonable that, as far as prediction of response is concerned, identification of the functionality and importance to individual tumors of the “targets” of particular therapies as defined by a “panel” of genes will be applicable. Alternatively, the current state of our understanding of NSCLC pathogenesis means that prognostication might be more accurately determined by means of pattern recognition within a more global expression profile. Translational research and clinical trials should address these issues, but

practical issues are likely to dominate discussions, at least initially. We have already discussed optimizing clinical specimen collection, but it remains to be determined whether the genes identified by gene expression profiling research will be analyzed by means of the global profiling platforms discussed here or, alternatively, whether their products will be assayed by more conventional and widely available techniques such as immunohistochemistry. The practical and organizational issues of different pathology departments and clinicians from a number of different institutions providing specimens in an agreed upon, standardized way will also provide challenges for the designers of clinical trials. However, in this regard, considerable experience already exists from previous investigations that have involved tissue collection, and the practical experiences within these studies are informative and illustrate that these issues, although often difficult, are not insurmountable. The progression of gene expression profiling research into this area also requires a number of ethical and political issues related to genetic testing to be addressed. At present, this has been mostly a matter for individual institutional research ethics boards. A critical distinction in much of the work in this area is the gene expression profiling of patients’ tumors as opposed to the patients themselves. This is usually recognized by ethical boards and generally accepted by patients themselves. The gene expression profiling of patients themselves does represent a more complex issue and brings the research into an area where there is more widespread public awareness, debate, and legislation. Clinical studies will not proceed without addressing this issue. However, in light of the advances already made by gene expression profiling and the potential for significant breakthroughs in the management of what is a major public health issue, ethical issues and objections are unlikely to represent a major obstacle.

These issues notwithstanding, it will remain important to continue with the progression of this type of research at the scientific end of the translational research spectrum. Resolution of the clinical trial issues discussed will facilitate the advancement of this work. Here the aim should be to use the correlative patterns and molecular classifiers already elucidated to propose the underlying molecular mechanisms. This will require both innovative experimental design and data analysis but holds the promise of the identification of novel therapeutic and management strategies based on molecular mechanisms. Such interventions will be specifically targeted to individual tumors and individual patients based on a comprehensive understanding of the important molecular abnormalities present.

## References

1. Stat bite: global cancer incidence, 2000. *J Natl Cancer Inst* (Bethesda) 2003;95:847.
2. Hecht SS. Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *Lancet Oncol* 2002;3:461–9.
3. Bogart JA, Scalzetti E, Dexter E. Early stage medically inoperable non-small cell lung cancer. *Curr Treat Options Oncol* 2003;4:81–8.
4. Bunn PA Jr. Chemotherapy for advanced non-small-cell lung cancer: who, what, when, why? *J Clin Oncol* 2002; 20(Suppl):23S–33S.
5. Rosell R, Fossella F, Milas L. Molecular markers and targeted therapy with novel agents: prospects in the treatment of non-small cell lung cancer. *Lung Cancer* 2002;38(Suppl):43–9.
6. Felip E, Rosell R. Neoadjuvant chemotherapy in non-small cell lung cancer. *Curr Med Chem* 2002;9:893–8.

7. Cella D. Impact of ZD1839 on non-small cell lung cancer-related symptoms as measured by the functional assessment of cancer therapy-lung scale. *Semin Oncol* 2003;30(Suppl 1):39–48.
8. Herbst RS. Dose-comparative monotherapy trials of ZD1839 in previously treated non-small cell lung cancer patients. *Semin Oncol* 2003;30(Suppl 1):30–8.
9. Ranson M, Thatcher N. Commentary on ZD1839 (Iressa) in non small cell lung cancer. *Lung Cancer* 2003;40:77–8.
10. Rosell R, Monzo M, O'Brate A, Taron M. Translational oncogenomics: toward rational therapeutic decision-making. *Curr Opin Oncol* 2002;14:171–9.
11. Sarries C, Haura EB, Roig B, et al. Pharmacogenomic strategies for developing customized chemotherapy in non-small cell lung cancer. *Pharmacogenomics* 2002;3:763–80.
12. Chang BD, Swift ME, Shen M, et al. Molecular determinants of terminal growth arrest induced in tumor cells by a chemotherapeutic agent. *Proc Natl Acad Sci USA* 2002;99:389–94.
13. 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;60:4161–6.
14. Staunton JE, Slonim DK, Collier HA, et al. Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci USA* 2001;98:10787–92.
15. Zembutsu H, Ohnishi Y, Tsunoda T, 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;62:518–27.
16. Leist M, Jaattela M. Four deaths and a funeral: from caspases to alternative mechanisms. *Nat Rev Mol Cell Biol* 2001;2:589–98.
17. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Genet* 2003; 33(Suppl):238–44.
18. Heller MJ. DNA microarray technology: devices, systems, and applications. *Annu Rev Biomed Eng* 2002;4:129–53.
19. Yamamoto M, Wakatsuki T, Hada A, Ryo A. Use of serial analysis of gene expression (SAGE) technology. *J Immunol Methods* 2001;250:45–66.
20. Baugh LR, Hill AA, Brown EL, Hunter CP. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* 2001; 29:E29.
21. Lim EH, Aggarwal A, Agasthian T, et al. Feasibility of using low-volume tissue samples for gene expression profiling of advanced non-small cell lung cancers. *Clin Cancer Res* 2003;9:5980–7.
22. Ellis M, Davis N, Coop A, et al. Development and validation of a method for using breast core needle biopsies for gene expression microarray analyses. *Clin Cancer Res* 2002;8:1155–66.
23. Chung CH, Bernard PS, Perou CM. Molecular portraits and the family tree of cancer. *Nat Genet* 2002;32(Suppl):533–40.
24. Chen G, Gharib TG, Huang CC, et al. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* 2002; 1:304–13.
25. Celis JE, Kruhoffer M, Gromova I, et al. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett* 2000;480:2–16.
26. 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;98:13790–5.
27. Nacht M, Dracheva T, Gao Y, et al. Molecular characteristics of non-small cell lung cancer. *Proc Natl Acad Sci USA* 2001;98:15203–8.
28. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 2001;98:13784–9.
29. McDoniels-Silvers AL, Nimri CF, Stoner GD, Lubet RA, You M. Differential gene expression in human lung adenocarcinomas and squamous cell carcinomas. *Clin Cancer Res* 2002;8:1127–38.
30. McDoniels-Silvers AL, Stoner GD, Lubet RA, You M. Differential expression of critical cellular genes in human lung adenocarcinomas and squamous cell carcinomas in comparison to normal lung tissues. *Neoplasia* 2002;4:141–50.
31. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002; 8:816–24.
32. Giordano TJ, Shedden KA, Schwartz DR, et al. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am J Pathol* 2001;159:1231–8.
33. Meerzaman D, Shapiro PS, Kim KC. Involvement of the MAP kinase ERK2 in MUC1 mucin signaling. *Am J Physiol Lung Cell Mol Physiol* 2001;281:L86–91.
34. Miura K, Bowman ED, Simon R, et al. Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res* 2002;62:3244–50.
35. 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;62:3005–8.
36. Iwao K, Watanabe T, Fujiwara Y, et al. Isolation of a novel human lung-specific gene, LUNX, a potential molecular marker for detection of micrometastasis in non-small-cell lung cancer. *Int J Cancer* 2001;91:433–7.
37. Hibi K, Liu Q, Beaudry GA, et al. Serial analysis of gene expression in non-small cell lung cancer. *Cancer Res* 1998;58:5690–4.
38. Dong G, Loukinova E, Chen Z, et al. Molecular profiling of transformed and metastatic murine squamous carcinoma cells by differential display and cDNA microarray reveals altered expression of multiple genes related to growth, apoptosis, angiogenesis, and the NF-kappaB signal pathway. *Cancer Res* 2001;61:4797–808.
39. Chen JJ, Peck K, Hong TM, et al. Global analysis of gene expression in invasion by a lung cancer model. *Cancer Res* 2001;61:5223–30.
40. Gemma A, Takenaka K, Hosoya Y, et al. Altered expression of several genes in highly metastatic subpopulations of a human pulmonary adenocarcinoma cell line. *Eur J Cancer* 2001;37:1554–61.
41. Tran YK, Bogler O, Gorse KM, et al. A novel member of the NF2/ERM/4.1 superfamily with growth suppressing properties in lung cancer. *Cancer Res* 1999;59:35–43.
42. Natsume T, Nakamura T, Koh Y, et al. Gene expression profiling of exposure to TZT-1027, a novel microtubule-interfering agent, in non-small cell lung cancer PC-14 cells and astrocytes. *Investig New Drugs* 2001;19:293–302.
43. Ohira T, Akutagawa S, Usuda J, et al. Up-regulated gene expression of angiogenesis factors in post-chemotherapeutic lung cancer tissues determined by cDNA macroarray. *Oncol Rep* 2002;9:723–8.
44. Kikuchi T, Daigo Y, Katagiri T, et al. Expression profiles of non-small cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. *Oncogene* 2003;22:2192–205.
45. 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;61:6474–9.
46. 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;62:1139–47.
47. Qin LF, Lee TK, Ng IO. Gene expression profiling by cDNA array in human hepatoma cell line in response to cisplatin treatment. *Life Sci* 2002;70:1677–90.
48. Korn EL, McShane LM, Troendle JF, Rosenwald A, Simon R. Identifying pre-post chemotherapy differences in gene expression in breast tumours: a statistical method appropriate for this aim. *Br J Cancer* 2002;86:1093–6.