Translational Review

An Overview of Lung Cancer Genomics and Proteomics

Courtney A. Granville and Phillip A. Dennis

Cancer Therapeutics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland; and George Washington University Genetics Program, Washington, DC

Lung cancer is the cause of nearly 170,000 cancer deaths in the United States each year, accounting for nearly 25% of all deaths from cancer. The 5-yr survival rate for lung cancer is < 15% from the time of diagnosis. This is largely due to the late stage of diagnosis and the lack of effective treatments, reflecting the need for a better understanding of the mechanisms that underlie lung carcinogenesis. Unlike the study of a single gene, protein, or pathway, genomic and proteomic technologies enable a systematic overview that provides the potential to improve our understanding of this disease. Ultimately, this could improve the diagnosis, prognosis, and clinical management of patients with lung cancer. Here, we review studies that generated profiles of gene and protein expression in lung cancer specimens and relevant model systems, and make recommendations to facilitate the clinical application of these technologies.

Keywords: lung cancer; genomics; proteomics

Lung cancer accounts for 30% of cancer deaths in the United States every year, and it is the second most frequently diagnosed cancer in men and women (1). Among men the number of lung cancer deaths nearly triples the number of prostate cancer deaths, and among women, the number of lung cancer deaths is almost double the number of breast cancer deaths. Eighty percent of lung cancer cases are non–small cell lung cancer (NSCLC) and 20% are small cell lung cancer (SCLC). Regardless of subtype, the 5-yr survival rate for lung cancer is among the lowest of all cancers at 10–15% (1, 2). Although more than half of lung cancers are diagnosed at a late stage, when cure is unlikely, the survival of patients diagnosed with Stage I lung cancer is also surprisingly low (2). Thus, there is a great need to understand the molecular alterations that confer a poor prognosis and to use this information to improve diagnosis and patient management.

Genomics and proteomics were conceived to allow for the rapid, complete, and parallel analysis of the genes and proteins that are expressed in a given cell or tissue type. In the context of cancer, differential profiling can be used to determine if the genomic and/or proteomic profile of a set of cancers differs from a set of normal tissues or from one another. In addition, gene and protein expression profiles have potential to improve the clinical management of lung cancer by improving classification either via class prediction or class discovery, or by providing data to develop a diagnostic classifier. Gene and protein expression profiles could identify new molecular targets and improve patient care through the identification of profiles that predict responsiveness to therapy or prognosis.

The past decade has been marked by the rapid development of technologies for the high-throughput analysis of gene and protein expression. Microarray technology was initially described for the analysis of cDNAs by Schena and colleagues in 1995 (3), and was applied to the analysis of oligonucleotides in 1997 (4, 5) (Figure 1). It wasn't until 2001 that microarray technology was first applied to the global analysis of a large number of histologically diverse lung cancer specimens (6). The tools of proteomics, namely two-dimensional gel electrophoresis and mass spectrometry, have been developing over the past three decades, but it was not until 1995 that high-throughput analysis became possible due to the ability to use mass spectrometry for the quantitative analysis of complex mixtures (7, 8). The development of mass spectrometry was paralleled by the establishment of databases for the identification of peptides in such mixtures (9–13) (reviewed in Refs. 14 and 15). The first proteomic study of lung cancer was published in 2002 and has been followed by three additional studies in the past 2 yr. In contrast, several dozen microarray studies of lung cancer have been published since 1999. Although there are many platforms for the parallel analysis of gene expression, DNA microarrays are the most universally applied to the study of clinical lung cancer specimens, and are therefore the only platform included in this review. In addition, there are several approaches to analyzing the volumes of data that a single microarray experiment can generate, including hierarchical clustering, principal component analysis and training-testing, and nearest neighbor analysis. More detailed discussion of these methods are beyond the scope of this review, but the interested reader can refer to other publications on the topic (16–18).

THE OMICS OF LUNG CANCER

At the time of submission of this review, ~ 40 articles describing the application of microarray and proteomic technologies to the analysis of clinical lung cancer specimens have been published. These studies can be broadly categorized as: (*1*) defining categories or tumor subsets that may improve the diagnostic classification of lung tumors; (*2*) identifying specific genes or proteins that could serve as molecular targets for improved diagnosis and/or therapy; and (*3*) associating gene expression profiles with clinical outcome.

Classification of Tumors: Class Prediction and Class Discovery

Studies that used microarray analysis to classify NSCLC specimens based on gene or protein expression profiles are shown in Table 1 (6, 19–32). In many of these studies, a class discovery approach showed that gene expression profiles could not only group tumor samples consistent with classical histology, but also could identify subgroups within histologic subclasses. Moreover, many studies showed that profiles generated from one set of

(*Received in final form December 16, 2004*)

Correspondence and requests for reprints should be addressed to Phillip A. Dennis, National Cancer Institute, Building 8, Room 5101, 8901 Wisconsin Ave., Bethesda, MD 20889. E-mail: pdennis@nih.gov

Am J Respir Cell Mol Biol Vol 32. pp 169–176, 2005
DOI: 10.1165/rcmb.F290
Internet address: www.atsjournals.org

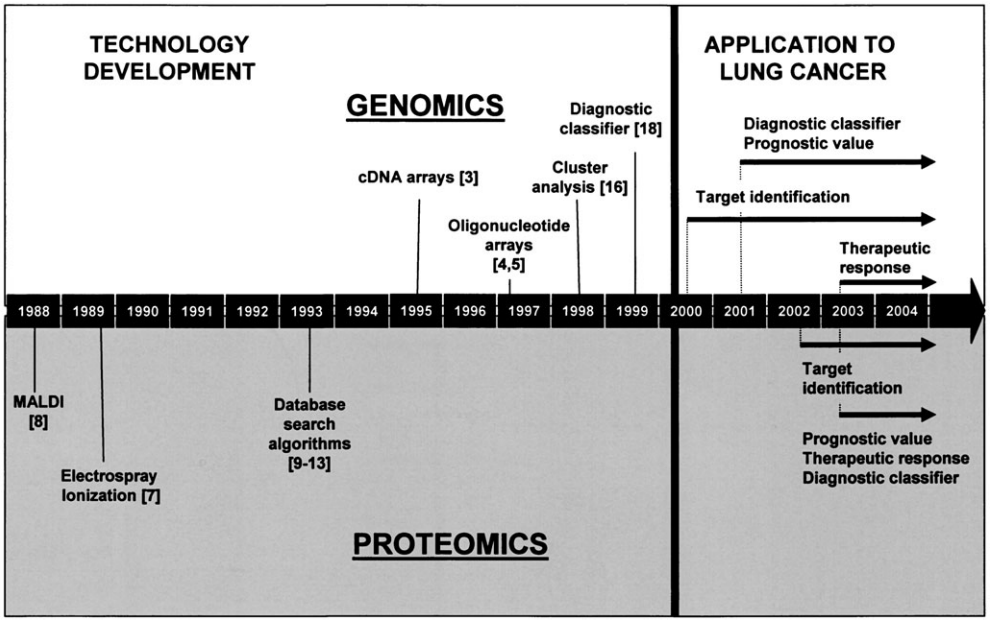

Figure 1. The development of genomics and proteomics and their application to the analysis of lung cancer. A timeline of the key advances in the development of microarray-based genomic and mass spectrometry-based proteomic technologies is shown on the *left*. The application of these technologies to the analysis of lung cancer specimens is shown on the *right*.

samples could then be applied to a test set of tumors to identify the histologic subclass in a class prediction approach. A recent retrospective analysis of the gene expression profiles generated by the initial studies of Bhattacharjee and coworkers (6), Beer and colleagues (31), and Garber and associates (25) showed that statistical methods could predict a gene-set that defines histology and survival (33). Despite these encouraging results, inconsistency in the number of classes observed has limited the clinical application of these findings.

Reproducibility and consistency are necessary for clinical applications. Although histologically similar tumors can have gene expression profiles that cluster, and in some cases can show groups that can be further broken down into distinct subclasses, the number of clusters is not consistent between groups (6, 24, 25, 30). For example, the analysis of 35 NSCLC specimens, including 18 adenocarcinomas (AC), by Borczuk and coworkers (24) revealed two distinct groups of AC; Bhattacharjee and associates (6) analyzed 186 primary lung tumors, including 139 AC, and identified four distinct AC subclasses, but only after excluding the other histologic subtypes of lung cancer; finally, Garber and colleagues (25) analyzed 67 lung tumors, 41 of which were AC, and revealed three AC subgroups. These differences in the number of subgroups of AC likely reflect the differences in the number and heterogeneity of the samples analyzed. However, the fact that the number of AC subgroups identified by microarray analysis correlates with the overall number of AC samples analyzed suggests that analysis of a larger group of AC would result in a larger number of subgroups. Do these molecular profiles reflect meaningful biologic differences that can be applied to the clinic, or do they simply reflect a heterogeneous disease that can be subclassified based on statistical analysis, experimental methods, or the particular chip that is used? Tomida and colleagues (30) identified two subclasses for squamous cell carcinoma (SCC) that associated with prognosis, suggesting that the identification of subgroups can have biologic or clinical meaning. However, this analysis only included 16 SCC specimens. Validation of methods and sharing of specimens, in addition to increasing sample size, are needed to answer this important question.

Genomic analysis has already provided clarification regarding important diagnostic distinctions. These include the identification of the cell of origin for SCLC (34), the distinction of histologically similar tumors such as malignant pleural mesothelioma (MPM) and AC (35), and the discrimination of primary from metastatic lung cancer (6, 27, 28).

The first lung cancer microarray study focused on understanding the cell of origin for SCLC (34). Morphologic characteristics and molecular markers suggest that SCLC originates from a neuroendocrine progenitor. However, the clinical course of SCLC is not like pure neuroendocrine carcinoid tumors, and a microarray analysis of SCLC specimens by Anbazhagan and coworkers provides support for the hypothesis that SCLC arises from lung epithelium (34). The authors assert that reclassification of these tumors as epithelial-derived tumors based on molecular changes rather than classification based on morphologic characteristics would be more accurate (34). Such a reclassification would be consistent with the exposure and transformation of lung epithelial cells by smoking, and the observation that SCLC is almost never observed in non-smokers. In contrast, other groups did identify several markers of neuroendocrine cells that were upregulated in SCLC specimens when the gene expression profiles of SCLC versus NSCLC tumors were compared (6, 25). It is not clear whether the differences between these studies are artifacts due to differences in the method of analysis (i.e., comparison of SCLC to normal bronchial epithelium and pulmonary carcinoids in the study by Anbazhagen and colleagues study versus comparison of SCLC to NSCLC in the other studies), or reflective of the small number of SCLC specimens in the studies by Bhattacharjee and associates (6) and Garber and coworkers (25). Because of the lower prevalence of SCLC, fewer genomic analyses have been performed on this type of lung cancer. Analysis of a larger number of SCLC specimens might better classify these tumors as derived from epithelial or neuroendocrine tissues.

Regarding differentiation between tumors with similar histology, Gordon and colleagues showed that MPM and AC can be accurately distinguished 99% of the time using a set of three gene pairs that were inversely correlated with MPM and AC (35). Using specific gene pair ratios may be more clinically applicable because it could be done with little sample preparation and

TABLE 1. GENOMIC AND PROTEOMIC ANALYSES OF CLINICAL LUNG CANCER SPECIMENS

	Genomic Analysis				
Year	Author (Ref.)	Test Samples	Reference Samples	Clinical Contribution	Data[†]
1999	Anbazhagan (34)	2 Primary SCLC*	2 Primary carcinoids; 2 primary brain tumors; 2 SCLC & 2 NHBE* cell lines	Class prediction: expression profile (ep) of SCLC more like EP of epithelial tumor than EP of neuroendocrine tumor	Yes[‡]
2000	Wang (52)	4 Primary SCC* 4 Primary AC*	1 Primary colon tumor; normal tissues	Target ID: 17 genes differentially expressed in SCC	No
2001	Bhattacharjee (6)	186 Primary lung tumors (125 AC w/clinical data)	15 Normal lung	Class prediction Class identification: (AC subclasses identified) Target ID Outcome predictor	Yes[§]
2001	Garber (25)	67 Lung tumors	5 Normal lung; reference RNA	Class prediction Class identification: (ac subclasses identified) Survival index	Yes[¶]
2001	Giordano (26)	57 Primary lung AC 51 Colon & 46 ovarian	51 Colon & 46 ovarian	Class prediction	No
2002	Bangur (54)	2 Primary SCLC 6 SCLC cell lines	Normal lung	Target ID: 209 genes overexpressed in SCLC	No
2002	Beer (31)	86 Primary lung AC	10 Normal lung 84 Lung AC validation set	Class prediction Survival index	Yes[‖]
2002	Gordon (35)	31 MPM* 150 Primary AC		Diagnotic classifier: ratios of 6 genes discriminate MPM from AC (99% accuracy)	No
2002	Heighway (46)	26 Tumor w/13 matched tissues from chemo trial	Matched normal tissue	Target ID	No
2002	Miura (32)	19 Primary lung AC		Class identification: smokers versus nonsmokers Survival index	No
2002	Virtanen (19)	44 Primary lung tumor	41 Cell lines	Class prediction	Yes[¶]
2002	Wigle (38)	39 NSCLC*		Outcome predictor: relapse group ID	Yes**
2002	Wikman (53)	14 NSCLC	Matched normal tissue	Target ID	No
2003	Borczuk (24)	32 Microdissected NSCLC	5 Normal lung; murine expression data	Class prediction Target ID	Yes[††]
2003	Creighton (45)	Reanalysis of Beer (2002)		Target ID	Yes[‡‡]
2003	Gordon (39)	Reanalysis of Bhattacharjee (2001), Beer (2002), and Garber (2001)		Diagnostic classifier: correctly classifies 74% tumors Survival index	Yes[§§]
2003	Kikuchi (29)	37 Microdissected NSCLC		Class prediction Prognostic indicator: 40 genes predict lymph node metastasis	No
2003	Lim (20)	46 Fine needle aspirates	17 Matched tumor/normal	Class prediction	Yes[¶¶]
2003	Miller (48)	86 Primary lung AC	10 Normal lung	Target ID	No
2003	Nakamura (49)	7 Primary lung AC 3 Primary SCC	Matched normal tissue	Target ID	No
2003	Powell (42)	6 Smoker-lung tumors 6 Non-smoker lung tumor	Case-matched normal tissue	Target ID	Yes[‖‖]
2003	Singhal (50, 51)	15 Lung AC	5 Non-neoplastic tissue	Target ID	Yes***
2003	Yamagata (27)	50 Resected NSCLC	3 Normal lung; pooled mrna	Diagnostic classifier Prognostic indicator	Yes[†††]
2004	Borczuk (21)	26 Primary tissue specimens		Class prediction: 86% accuracy survival index: 87% accuracy	Yes[‖‖]
2004	Bonner (22)	17 Primary AC 6 Primary LCC*	Murine normal lung, adenomas, and AC	Class prediction Class ID: (AC subclasses identified) target ID	No
2004	Diederichs (37)	14 Primary NSCLC; partial reanalysis of Bhattacharjee (2001)	Gene expression profiles from non-metastatic tumors	Target ID Prognostic indicator	Yes[‡‡‡]
2004	Erez (44)	26 NSCLC; reanalysis of Bhattacharjee (2001)	10 Normal lung	Target validation	No
2004	Kettunen (47)	13 SCC	13 AC; normal lung	Target ID	No
2004	Spira (43)	93 Primary bronchial airway brushings		Target ID	Yes[§§§]
2004	Tomida (30)	50 NSCLC		Class ID: SCC subclasses identified Survival index: 82% accuracy for 5-yr survival; 90% accuracy if class specific	No

(*Continued*)

TABLE 1. (*Continued*)

	Proteomic Analysis				
Year	Author (Ref.)	Test Samples	Reference Samples	Clinical Contribution	Data†
2002	Chen (55)	93 Primary lung AC	10 Normal lung	Target ID Single case where overexpression MRNA correlates with overexpression of corresponding protein	No
2003	Campa (23)	10 Lung tumors	Matched normal lung	Class identification Target identification	No
2003	Chen (36)	90 Primary lung AC	10 Normal lung	Target identification Prognostic indicator	No
2003	Yanagisawa (28)	79 Primary lung tumors and secondary lung metastases	14 Normal lung	Class prediction Target ID Prognostic indicator: predicts lymph node involvement Survival index: 15 peptides	No

Definition of abbreviations: AC, adenocarcinoma; ID, identification; LCC, large cell carcinoma; MPM, malignant pleural mesothelioma; NHBE, normal human bronchial epithelium; NSCLC, non–small cell lung cancer; SCC, squamous cell carcinoma; SCLC, small cell lung cancer.

The types of samples analyzed, the reference sample(s) used, the area(s) in which each study contributes to an improved understanding of lung cancer biology, and the availability of data on the world wide web are provided for each study.

† Data online at:

‡ http://128.220.85.49/genomics.

§ http://www-genome.wi.mit.edu/cgi-bin/cancer/datasets.cgi.

¶ www.pnas.org.

‖ www.nature.com/nm.

** http://www.cs.utoronto.ca/~juris/publicationsData.html.

†† http://lungtranscriptome.bwh.Harvard.edu/.

‡‡ http://dot.ped.med.umich.edu:2000/pub/diff/index.htm.

§§ http://www.chestsurg.org/microarray.htm.

¶¶ http://www.omniarray.com/lungFNA.htm.

‖‖ http://hora.cpmc.Columbia.edu/dept/pulmonary/5ResearchPages/Laboratories/Powell%20Lab.htm.

*** www.uphs.upenn.edu/lungctr/academic_programs/pulmonary/research/labs/albelda.

††† http://array.mc.vanderbilt.edu/supl/yamagata/index.html.

‡‡‡ http://www.klinikum.uni-muenster.de/institute/meda/research/index.htm.

§§§ http://pulm.bumc.bu.edu/aged.

manipulation, and because it does not require the use of a reference sample.

Since the publication of Gordon and associates (35), two other studies have suggested that the expression profiles of smaller groups of genes could be used as a classifier to discriminate between the histologic types of lung cancer and to distinguish primary from metastatic lung cancer (27, 28). Yamagata and colleagues (27) showed that a classifier based on sets of gene expression values could be used on blinded samples to differentiate NSCLC from normal lung, as well as primary lung tumors from lung metastases from tumors originating at other sites. However, this classifier was less accurate at differentiating between subtypes of NSCLC in a blinded analysis—AC was correctly identified, but 1 of 4 blinded SCC were incorrectly classified and large cell carcinoma (LCC) could not be differentiated from other histologic subtypes of NSCLC (27). These limitations may partially reflect the small size of the study. Proteomic studies have also used a subset of markers to discern pathologic differences. Yanagisawa and coworkers (28) showed that a set of 82 MS peaks could discriminate between tumor and normal tissue in the training set with 100% accuracy. When applied to the blinded test set, 42/43 samples were correctly classified as tumor or normal (28). Collectively, these studies show that a small number of genes or mass spectral peaks combined in a formal classifier have the potential to further define subsets of lung cancer.

Correlation of Gene and Protein Expression Profiles with Prognosis

Microarray analysis has the potential to predict the survival of patients with NSCLC. For example, although cluster analysis by Beer and colleagues did not perfectly segregate stage I from stage III lung AC, the authors point out that the stage I tumors that clustered with stage III samples came from patients that exhibited worse survival (31). This study shows that the expression of genes that confer a poor prognosis is independent of stage of disease at the time of diagnosis. Because microarray analysis can identify tumor subsets that share molecular alterations important for cancer progression, incorporating gene expression profiles might provide added therapeutic and prognostic value when combined with traditional staging and histologic analysis.

Other studies have also correlated gene or protein expression profiles with prognosis (6, 21, 25, 27–32, 36–39). In these studies, subsets of genes or protein peaks differentially expressed in tumors could predict survival differences among patients with AC. In an extension of their study on the use of multiple gene pairs to classify lung cancer based on histology, Gordon and associates found that sets of three gene expression ratios could accurately predict prognosis more than 90% of the time (39). The authors then used the most predictive gene pairs to analyze data from the earlier study of Bhattacharjee and coworkers (6). In this analysis, the set of gene expression ratios could correctly predict survival of patients with stage I AC tumors 64% of the time in low-tumor-content samples, and 74% of the time for high-tumor-content samples (39).

The development of prognostic identifiers could guide clinical decision-making, but such application requires validation in prospective clinical trials. The inclusion of genomic and proteomic analyses into future clinical trials, and the analysis of specimens from lung cancer patients receiving standard of care therapy could provide information that might ultimately lead to more accurate, individualized clinical decisions.

Correlation of Gene Expression Profiles with Smoking Status and Lung Cancer

Smoking accounts for 85–90% of lung cancer cases. However, the remaining cases occur among never-smokers. This raises the hypothesis that lung cancers in smokers and never-smokers might arise through different mechanisms. Support for this hypothesis is manifest by the selective expression of EGFR mutations in NSCLC specimens from never-smokers (40, 41).

Three groups have applied technologies for tumor profiling to dissect the molecular mechanisms underlying smoking-induced and smoking-independent lung cancer (Table 1; Refs. 32, 42, 43). In the study by Miura and colleagues, laser-capture microdissection was used to isolate tumor tissue before analysis (32). The analysis revealed 45 genes that were differentially expressed in the tumors of smokers versus nonsmokers. Powell and associates (42) compared the gene expression profiles of tumor and matched normal tissues for six smokers and six nonsmokers. Hierarchical clustering using all data did not separate tumors from smokers versus nonsmokers, but it did separate tumor and nontumor tissue and grouped the nontumor tissues from nonsmokers together. Otherwise, gene expression profiles did not differentiate between each of the groups of tissues unless the analysis was confined to data from genes with the most divergent expression between the groups. This implies that although the underlying "normal" tissues from smokers versus nonsmokers are quite different, their cancers cannot be distinguished using microarray technology. Although similar mechanisms may mediate the development of lung cancer in smokers and nonsmokers alike, the selective expression of EGFR mutations in nonsmokers raises the possibility that differences between smokers and nonsmokers may be most apparent by identifying mutations in relevant oncogenes and tumor suppressor genes. Alternatively, the similarity of tumors from smokers and nonsmokers may simply reflect the small size of this study and illustrates the need for further investigation of this issue. In the third study correlating changes in gene expression with smoking, Spira and colleagues compared the gene expression profiles of bronchial epithelial cells isolated from current, former, and never-smokers (43). Their analysis showed that several gene expression changes that are observed in current smokers persist in former smokers. This is important because over 50% of lung cancer cases are diagnosed in former smokers, and these persistent changes in gene expression may underlie the increased risk of these patients.

Identification of Novel Molecular Targets by Gene and Protein Expression Profiling

Genomic and proteomic technologies produce a wealth of data that can be mined for specific mechanistic information, in addition to their usefulness in detecting patterns. Many studies have identified and validated specific genes whose expression differs between tumor and normal tissue, or throughout the histologic progression of cancer (23, 37, 44–54). Pending validation, these genes may serve as biomarkers for diagnosis or targets for therapy. For example, two groups found that the insulinoma-associated 1 gene is overexpressed in SCLC (6, 25). In addition, gene expression profiles from the most clinically aggressive AC samples showed increased expression of genes that play a role in angiogenesis such as VEGF and PPARγ, as well as degradation of extracellular proteins such as cathepsin L and plasminogen activator of urokinase. The analysis of Yamagata and coworkers identified one gene, *ACTN4*, whose expression level was correlated with poor survival, and therefore could serve as a marker for survival (27). Other gene expression profiling studies have shown that phosphoglycerate kinase 1 (PGK1) is overexpressed in lung tumors and correlates with poor survival, as well as 14 other genes that are important in the glycolytic pathway (31, 55).

By comparing gene expression profiles in tumors from smokers and never-smokers, Powell and colleagues identified several genes whose expression levels either positively or negatively correlated with the number of pack-years of smoking (42). Although it is tempting to view these genes as possible preventative or therapeutic targets, it is also possible that the data are confounded by the fact that the authors did not collect data on gene expression changes associated with smoking in the absence of cancer (42). Interestingly, the study by Spira and associates (43) showed that the expression of several tumor suppressor genes and oncogenes are permanently altered in bronchial epithelial cells from former and current smokers. In contrast, alterations in genes associated with metabolism and antioxidation pathways were reversible depending on the number of years of abstinence (43). The genes that remain altered in former smokers may be better targets for the prevention and treatment of smoking-induced lung cancer.

Proteomic studies also identified proteins that could serve as therapeutic targets. For example, Chen and coworkers (36) showed that four proteins that regulate gluconeogenesis and glycolysis are associated with survival; phosphoglycerate kinase 1, phosphoglycerate mutase, α enolase, and pyruvate kinase M1. Combined with their gene expression analyses described earlier, this study provides further evidence for the idea that patients whose tumors exhibit increased glycolysis have a poorer prognosis (36). In addition to the markers identified by Chen and colleagues, Yanagisawa and associates (28) identified three proteins that could serve as tumor markers; small ubiquitin-related modifier-2 (SUMO-2), thymosin-β4, and ubiquitin. Thymosin-β4 has been previously associated with proliferation in cancerous tissue (56), but SUMO represents a potential new protein for analysis of its role in carcinogenesis (28). All of these potential markers require validation in prospective trials.

Analysis of Relevant Murine Models

Although very few mouse models of lung cancer exist and their relevance to human lung cancer is often questioned, they have been used to study lung tumorigenesis in a defined genetic background. Gene expression studies have begun to correlate changes in gene expression with murine lung tumorigenesis. For example, Gariboldi and coworkers (57) used microarray analysis to compare the gene expression profiles of normal mouse lungs from strains that varied in their susceptibility and resistance to lung cancer. They identified 89 genes whose expression profile correlated with each particular strain's tumor susceptibility. Of these 89 genes, 26 were previously identified for their involvement in cancer (57). Consistent with their hypothesis, this study showed that the gene expression profile of normal murine lung could predict tumor susceptibility for each strain. Although differences in gene expression may reflect "unconnected strain differences," the authors suggest that this study provides evidence for the existence of genes whose expression in normal lung predisposes for lung cancer in mice. These studies are consistent with the study by Ramaswamy and colleagues, who showed that the development of metastasis is strain-dependent (58). In another study, Bonner and colleagues (22) compared the gene expression profiles of murine lung adenomas and adenocarcinomas from strain A/J × 129/SvJ F_1 mice and identified genes that could differentiate between the stages, as well as genes whose expression consistently increased or decreased in the human tumors independent of histology. Finally, this group also identified gene expression changes in the tumors that were either consistent with, or opposite to, the expression in the developing lung of this strain (22). The results of these studies have refined our

knowledge of murine lung tumorigenesis and could facilitate the creation of novel mouse models that might more closely resemble human lung cancer.

Analysis of Lung Cancer Cells *In Vitro*

Gene expression studies have also evaluated the effects of reconstituting tumor suppressor genes that are commonly lost in lung cancer. Two noteworthy *in vitro* studies have been published in lung cancer cellular systems. Hong and colleagues studied the effect of the phosphatase and tensin homology deleted on chromosome 10 (PTEN) tumor suppressor gene on the overall gene expression profile of a human lung cancer cell line CL_{1-5} (59), and Russo and coworkers used microarray analysis to look for targets of the retinoblastoma tumor suppressor gene (pRb/p300) in the human lung AC cell line H23 (60). In both studies, wild-type tumor suppressor genes of interest were transfected into the parental cells and the gene expression profiles were compared with the profiles generated from the parental cell transfected with the control vector. Overexpression of PTEN led to decreased expression of 19 genes and 4 ESTs, including 3 that had previously been described to play a role in cell invasiveness, such as laminin β_3, integrin α_6, and Akt2. PTEN overexpression also increased expression of 17 genes and 2 ESTs, including protein phosphatase 2A1B, an enzyme that was previously shown to maintain integrity of the cytoskeleton (59). In the studies on pRb, Russo and colleagues showed that 40 genes were downregulated in the pRb/p300 overexpressing cells compared with control. Many of these genes are important in cancer, including cyclin D1, Raf-1, and B-Myb (60). Together, these studies have contributed to a better understanding of how known tumor suppressor genes function in lung cancer cells.

DISCUSSION

The studies reviewed here reveal that genomics and proteomics are powerful tools for classification of tumor subtypes and identification of genes or proteins that may serve as diagnostic, predictive, or prognostic markers. The fact that molecular fingerprints generated by genomic and proteomic analysis can recapitulate the histologic variation between NSCLC subtypes is good proof of principle, but one has to ask: how does this improve upon what we already knew? Several issues will need to be addressed before the potential of these powerful technologies is realized and integrated into the clinical setting.

In many of these studies, the results of classification using cluster analysis of gene expression profiles did not match the results of classification based on tumor histology (24, 26, 27). In one analysis, immunohistochemistry was used to assess samples whose expression profile did not cluster with histologic subclass (24). Borczuk and associates (21) showed that segregation of these tumors was likely due to the presence of nontumor cells that explain the segregation into groups of cancers with another subtype. Similarly, the reanalysis of "incorrectly classified" samples by Giordano and coworkers and Yamagata and colleagues revealed that their original pathologic classification might have been incorrect (26, 27). Thus, although the molecular profile was an accurate diagnostic method, these studies highlight the fact that array technology relies on the molecular information that is given to it for its power to predict. One possible solution to eliminate "contaminating" nontumor cells from tumor specimens is to employ laser capture microdissection (61). Laser capture dissection can be used to separate tumor from normal tissue and can alleviate problems due to contamination of tumor tissue from normal. However, exclusion of surrounding stroma also eliminates cells that contribute to the tumor microenvironment.

Validation of changes in gene expression at the protein level would improve many of these studies, because changes in protein expression provide a more functional readout. Although evaluating protein expression may not be necessary to develop fingerprints for diagnostic use, it is necessary if markers that are identified by this method are to be used in their own right as diagnostic markers or as targets for therapy. Thus, despite the fact that many genes were identified as potential biomarkers, all of them require further validation. Even though proteomics provides a more functional approach than genomics, proteomics alone might provide an insufficient readout because post-translational modifications such as phosphorylation control the stability and function of many proteins. The development of sophisticated proteomic techniques such as phospho-proteomics will add another level of complexity to these systematic analyses (62).

The genomic and proteomic studies of lung cancer reviewed here have advanced the biologic understanding of lung cancer and have demonstrated the potential of these technologies to improve the classification and diagnosis of lung cancer. Numerous genes and proteins were identified as biomarkers, but the mere identification of markers does not provide any information about whether the changes in expression of genes or proteins are causal for biologic behavior. Although focusing on known genes and proteins has already yielded new information, unknown markers may also lend insight into the biology of lung cancer. These nascent studies using new and rapidly developing technologies have yielded exciting results. Improving sample size, reproducibility, and the ability to discriminate tissue or tumor subtypes will facilitate application of these new technologies to the clinic.

Conflict of Interest Statement: C.A.G. has no declared conflicts of interest; and P.A.D. has no declared conflicts of interest.

References

1. Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ, Thun MJ. Cancer statistics, 2004. *CA Cancer J Clin* 2004;54:8–29.
2. Fry WA, Phillips JL, Menck HR. Ten-year survey of lung cancer treatment and survival in hospitals in the United States: a national cancer data base report. *Cancer* 1999;86:1867–1876.
3. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–470.
4. Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Biotechnol* 1997;15:1359–1367.
5. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci USA* 1997; 94:13057–13062.
6. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 2001;98:13790–13795.
7. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989; 246:64–71.
8. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988;60:2299–2301.
9. Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci USA* 1993;90:5011–5015.
10. James P, Quadroni M, Carafoli E, Gonnet G. Protein identification by mass profile fingerprinting. *Biochem Biophys Res Commun* 1993;195: 58–64.
11. Mann M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* 1993;22:338–345.
12. Pappin DJ, Hojrup P, Bleasby AJ. Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* 1993;3:327–332.

13. Yates JR III, Speicher S, Griffin PR, Hunkapiller T. Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem* 1993;214:397–408.
14. Hirsch J, Hansen KC, Burlingame AL, Matthay MA. Proteomics: current techniques and potential applications to lung disease. *Am J Physiol Lung Cell Mol Physiol* 2004;287:L1–L23.
15. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nat Genet* 2003;33:311–323.
16. 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.
17. Quackenbush J. Computational analysis of microarray data. *Nat Rev Genet* 2001;2:418–427.
18. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–537.
19. Virtanen C, Ishikawa Y, Honjoh D, Kimura M, Shimane M, Miyoshi T, Nomura H, Jones MH. Integrated classification of lung tumors and cell lines by expression profiling. *Proc Natl Acad Sci USA* 2002;99: 12357–12362.
20. Lim EH, Aggarwal A, Agasthian T, Wong PS, Tan C, Sim E, Tan L, Goh PS, Wang SC, Khoo KL, *et al.* Feasibility of using low-volume tissue samples for gene expression profiling of advanced non-small cell lung cancers. *Clin Cancer Res* 2003;9:5980–5987.
21. Borczuk AC, Shah L, Pearson GD, Walter KL, Wang L, Austin JH, Friedman RA, Powell CA. Molecular signatures in biopsy specimens of lung cancer. *Am J Respir Crit Care Med* 2004;170:167–174.
22. Bonner AE, Lemon WJ, Devereux TR, Lubet RA, You M. Molecular profiling of mouse lung tumors: association with tumor progression, lung development, and human lung adenocarcinomas. *Oncogene* 2004; 23:1166–1176.
23. Campa MJ, Wang MZ, Howard B, Fitzgerald MC, Patz EF Jr. Protein expression profiling identifies macrophage migration inhibitory factor and cyclophilin a as potential molecular targets in non-small cell lung cancer. *Cancer Res* 2003;63:1652–1656.
24. Borczuk AC, Gorenstein L, Walter KL, Assaad AA, Wang L, Powell CA. Non-small-cell lung cancer molecular signatures recapitulate lung developmental pathways. *Am J Pathol* 2003;163:1949–1960.
25. Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 2001;98:13784–13789.
26. Giordano TJ, Shedden KA, Schwartz DR, Kuick R, Taylor JM, Lee N, Misek DE, Greenson JK, Kardia SL, Beer DG, *et al.* Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am J Pathol* 2001;159:1231–1238.
27. Yamagata N, Shyr Y, Yanagisawa K, Edgerton M, Dang TP, Gonzalez A, Nadaf S, Larsen P, Roberts JR, Nesbitt JC, *et al.* A training-testing approach to the molecular classification of resected non-small cell lung cancer. *Clin Cancer Res* 2003;9:4695–4704.
28. Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, Roberts JR, Edgerton M, Gonzalez A, Nadaf S, *et al.* Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 2003; 362:433–439.
29. Kikuchi T, Daigo Y, Katagiri T, Tsunoda T, Okada K, Kakiuchi S, Zembutsu H, Furukawa Y, Kawamura M, Kobayashi K, *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;22:2192–2205.
30. Tomida S, Koshikawa K, Yatabe Y, Harano T, Ogura N, Mitsudomi T, Some M, Yanagisawa K, Takahashi T, Osada H. Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. *Oncogene* 2004;23:5360–5370.
31. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002;8:816–824.
32. Miura K, Bowman ED, Simon R, Peng AC, Robles AI, Jones RT, Katagiri T, He P, Mizukami H, Charboneau L, *et al.* Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res* 2002;62:3244–3250.
33. Parmigiani G, Garrett-Mayer ES, Anbazhagan R, Gabrielson E. A cross-study comparison of gene expression studies for the molecular classification of lung cancer. *Clin Cancer Res* 2004;10:2922–2927.
34. Anbazhagan R, Tihan T, Bornman DM, Johnston JC, Saltz JH, Weigering A, Piantadosi S, Gabrielson E. Classification of small cell lung cancer and pulmonary carcinoid by gene expression profiles. *Cancer Res* 1999; 59:5119–5122.
35. Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Ramaswamy S, Richards WG, Sugarbaker DJ, Bueno R. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Res* 2002; 62:4963–4967.
36. Chen G, Gharib TG, Wang H, Huang CC, Kuick R, Thomas DG, Shedden KA, Misek DE, Taylor JM, Giordano TJ, *et al.* Protein profiles associated with survival in lung adenocarcinoma. *Proc Natl Acad Sci USA* 2003;100:13537–13542.
37. Diederichs S, Bulk E, Steffen B, Ji P, Tickenbrock L, Lang K, Zanker KS, Metzger R, Schneider PM, Gerke V, *et al.* S100 family members and trypsinogens are predictors of distant metastasis and survival in early-stage non-small cell lung cancer. *Cancer Res* 2004;64:5564–5569.
38. Wigle DA, Jurisica I, Radulovich N, Pintilie M, Rossant J, Liu N, Lu C, Woodgett J, Seiden I, Johnston M, *et al.* Molecular profiling of non-small cell lung cancer and correlation with disease-free survival. *Cancer Res* 2002;62:3005–3008.
39. Gordon GJ, Richards WG, Sugarbaker DJ, Jaklitsch MT, Bueno R. A prognostic test for adenocarcinoma of the lung from gene expression profiling data. *Cancer Epidemiol Biomarkers Prev* 2003;12:905–910.
40. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–1500.
41. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–2139.
42. Powell CA, Spira A, Derti A, DeLisi C, Liu G, Borczuk A, Busch S, Sahasrabudhe S, Chen Y, Sugarbaker D, *et al.* Gene expression in lung adenocarcinomas of smokers and nonsmokers. *Am J Respir Cell Mol Biol* 2003;29:157–162.
43. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, Palma J, Brody JS. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci USA* 2004;101:10143–10148.
44. Erez A, Perelman M, Hewitt SM, Cojacaru G, Goldberg I, Shahar I, Yaron P, Muler I, Campaner S, Amariglio N, *et al.* Sil overexpression in lung cancer characterizes tumors with increased mitotic activity. *Oncogene* 2004;23:5371–5377.
45. Creighton C, Hanash S, Beer D. Gene expression patterns define pathways correlated with loss of differentiation in lung adenocarcinomas. *FEBS Lett* 2003;540:167–170.
46. Heighway J, Knapp T, Boyce L, Brennand S, Field JK, Betticher DC, Ratschiller D, Gugger M, Donovan M, Lasek A, *et al.* Expression profiling of primary non-small cell lung cancer for target identification. *Oncogene* 2002;21:7749–7763.
47. Kettunen E, Anttila S, Seppanen JK, Karjalainen A, Edgren H, Lindstrom I, Salovaara R, Nissen AM, Salo J, Mattson K, *et al.* Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. *Cancer Genet Cytogenet* 2004;149:98–106.
48. Miller CT, Chen G, Gharib TG, Wang H, Thomas DG, Misek DE, Giordano TJ, Yee J, Orringer MB, Hanash SM, *et al.* Increased C–CRK proto-oncogene expression is associated with an aggressive phenotype in lung adenocarcinomas. *Oncogene* 2003;22:7950–7957.
49. Nakamura H, Saji H, Ogata A, Hosaka M, Hagiwara M, Saijo T, Kawasaki N, Kato H. cDNA microarray analysis of gene expression in pathologic Stage IA nonsmall cell lung carcinomas. *Cancer* 2003;97: 2798–2805.
50. Singhal S, Amin KM, Kruklitis R, DeLong P, Friscia ME, Litzky LA, Putt ME, Kaiser LR, Albelda SM. Alterations in cell cycle genes in early stage lung adenocarcinoma identified by expression profiling. *Cancer Biol Ther* 2003;2:291–298.
51. Singhal S, Amin KM, Kruklitis R, Marshall MB, Kucharczuk JC, DeLong P, Litzky LA, Kaiser LR, Albelda SM. Differentially expressed apoptotic genes in early stage lung adenocarcinoma predicted by expression profiling. *Cancer Biol Ther* 2003;2:566–571.
52. Wang T, Hopkins D, Schmidt C, Silva S, Houghton R, Takita H, Repasky E, Reed SG. Identification of genes differentially over-expressed in lung squamous cell carcinoma using combination of cDNA subtraction and microarray analysis. *Oncogene* 2000;19:1519–1528.
53. Wikman H, Kettunen E, Seppanen JK, Karjalainen A, Hollmen J, Anttila

S, Knuutila S. Identification of differentially expressed genes in pulmonary adenocarcinoma by using cDNA array. *Oncogene* 2002;21:5804–5813.

54. Bangur CS, Switzer A, Fan L, Marton MJ, Meyer MR, Wang T. Identification of genes over-expressed in small cell lung carcinoma using suppression subtractive hybridization and cDNA microarray expression analysis. *Oncogene* 2002;21:3814–3825.
55. Chen G, Gharib TG, Huang CC, Thomas DG, Shedden KA, Taylor JM, Kardia SL, Misek DE, Giordano TJ, Iannettoni MD, *et al.* Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in tumors. *Clin Cancer Res* 2002;8:2298–2305.
56. Hall AK. Differential expression of thymosin genes in human tumors and in the developing human kidney. *Int J Cancer* 1991;48:672–677.
57. Gariboldi M, Spinola M, Milani S, Pignatiello C, Kadota K, Bono H, Hayashizaki Y, Dragani TA, Okazaki Y. Gene expression profile of normal lungs predicts genetic predisposition to lung cancer in mice. *Carcinogenesis* 2003;24:1819–1826.
58. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003;33:49–54.
59. Hong TM, Yang PC, Peck K, Chen JJ, Yang SC, Chen YC, Wu CW. Profiling the downstream genes of tumor suppressor PTEN in lung cancer cells by complementary DNA microarray. *Am J Respir Cell Mol Biol* 2000;23:355–363.
60. Russo G, Claudio PP, Fu Y, Stiegler P, Yu Z, Macaluso M, Giordano A. pRB2/p130 target genes in non-small lung cancer cells identified by microarray analysis. *Oncogene* 2003;22:6959–6969.
61. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA. Laser capture microdissection. *Science* 1996;274:998–1001.
62. Blagoev B, Ong SE, Kratchmarova I, Mann M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 2004;22:1139–1145.