Lung Cancer Genomic Signatures

M. Teresa Agulló-Ortuño, PhD, BD, Fernando López-Ríos, PhD, MD,† and Luis Paz-Ares, PhD, MD‡*

Background: Lung cancer (LC) is the dominant cause of death by cancer in the world, being responsible for more than a million deaths annually. It is a highly lethal common tumor that is frequently diagnosed in advanced stages for which effective alternative therapeutics do not exist. In view of this, there is an urgent need to improve the diagnostic, prognostic, and therapeutic classification systems, currently based on clinicopathological criteria that do not adequately translate the enormous biologic complexity of this disease.

Methods: The advent of the human genome sequencing project and the concurrent development of many genomic-based technologies have allowed scientists to explore the possibility of using expression profiles to identify homogenous tumor subtypes, new prognostic factors of human cancer, response to a particular treatment, etc. and thereby select the best possible therapies while decreasing the risk of toxicities for the patients. Therefore, it is becoming increasingly important to identify the complete catalog of genes that are altered in cancer and to discriminate tumors accurately on the basis of their genetic background.

Results and Discussion: In this article, we present some of the works that has applied high-throughput technologies to LC research. In addition, we will give an overview of recent results in the field of LC genomics, with their effect on patient care, and discuss challenges and the potential future developments of this area.

Key Words: Lung cancer, Genomic signatures, High-throughput technologies.

(*J Thorac Oncol.* 2010;5: 1673–1691)

Lung cancer (LC) constitutes an outstanding public health problem. The most recent World Health Organization (WHO) projections for the period 2002 to 2030 report that LC will take on a health problem of increasing importance in the coming decades.[1] In 2030, LC will occupy the sixth place and will account for 3.1% of global mortality. The small difference between the number of prevalent and incident cases reflects the high lethality of this tumor. Primary pulmonary neoplasias are classified into two large histologic subgroups with different prognoses and therapeutic approaches: the small cell lung carcinoma (SCLC) and the non-small cell lung carcinoma (NSCLC). The NSCLC constitutes 80% of the cases, and it is also subclassified into different morphologic mainstay varieties (squamous cell carcinoma [SCC], adenocarcinoma [AC], and large cell carcinoma). The fundamental treatment of early-stage NSCLC is surgical resection. Nevertheless, only 15 to 25% of cases are resectable on diagnosis, and of those, 30 to 70% eventually recurs after surgery. NSCLC is a heterogeneous disease; even in patients with similar clinical and pathologic features, the outcome varies: some are cured, whereas in others, the cancer recurs. Once recurred or metastasized, the disease is essentially incurable with survival rates at 5 years of less than 5%, and this has improved only marginally during the past 25 years.[2] The poor early detection of LC coupled with ineffective treatments for advanced disease is responsible for the low survival rate. In view of this, there is an urgent need to improve our diagnostic, prognostic, and therapeutic classification systems, currently based on crude clinicopathological criteria that do not adequately translate the enormous biologic complexity of this disease.

The complete sequencing of the human genome and the concurrent development of technologies that allow for high-throughput generation of genomic data have opened avenues for a systematic approach to understanding the complex biology of LC.[3,4] A greater knowledge of the molecular mechanisms involved in the genesis, progression, and dissemination of LC is essential for the development of diagnostic methods that allow an earlier detection of the disease and for the design of more suitable, individualized, and effective therapeutic strategies. Precise global analyses will be necessary, based on the genome, transcriptome, or proteome to explain the complexity of the clinically important phenotypes that determine the failure or the success of any therapeutic intervention. Staging systems for LC that are based on clinical and pathologic findings may have reached their limit of usefulness for predicting outcomes, but molecular methods add value. Gene-expression profiling with the use of microarrays or real-time reverse transcription polymerase chain reaction (RT-PCR) has been shown to estimate the prognosis for patients with LC accurately.

This review will focus on major recent advances in the genomic approaches to the study of LC biology. These advances have been facilitated by the development of molecular techniques and biomarkers for defining cancer risk,

*Servicio de Oncología, Hospital Universitario 12 de Octubre, Madrid; †Laboratorio de Dianas Terapéuticas, Centro Integral Oncológico Clara Campal, Hospital Universitario Madrid-Sanchinarro, Madrid; and ‡Servicio de Oncología, Hospital Universitario Virgen del Rocío, Sevilla, Spain.

Disclosure: The authors declare no conflicts of interest.

Address for correspondence: M. Teresa Agulló-Ortuño, PhD, BD, Hospital Universitario 12 de Octubre, Centro de Investigación, Laboratorio de Oncología, Avda de Córdoba S/N, Madrid 28041, Spain. E-mail: agullo@h12o.es

ISSN: 1556-0864/10/0510-1673

prognosis, and optimal therapy aimed at prevention and personalized treatment of LC.

GENE-EXPRESSION SIGNATURES

Gene expression is the technical term to describe how active a particular gene is, that is, how many times it is expressed, or transcribed, to produce the protein it encodes. Gene-expression signature is a specific gene-expression profile, often a subset of expressed genes usually associated with a specific phenotype.

High-throughput technologies, such as RNA microarray and RT-PCR, allow simultaneous counting of many gene transcriptions. This creates a snapshot of a tissue's global gene activity, called the transcriptome. Gene-expression measurements have been used to develop new biologic concepts, refine disease classification, improve diagnostic and prognostic accuracy, and identify new molecular targets for drugs, especially in cancer research. Results are commonly reported in the form of a list of genes that are differentially expressed between normal and diseased patients or that correlate with different prognoses or phenotypes. These lists are called gene-expression profiles or signatures.

The tumor-node-metastasis (TNM) staging system for LC is the standard for prediction of survival.[5] For NSCLC, TNM stage, age, sex, and histologic cell type are well-established prognostic factors. Nevertheless, these factors have reached their limit in the prognostic information they provide and do not explain the large outcome variation among patients with similar characteristics. Formidable obstacles to developing effective markers include tumor heterogeneity, the highly complex interplay between the environment and host and the complexity, multiplicity, and redundancy of tumor-cell signaling networks involving genetic, epigenetic, and microenvironmental effects. Tumor molecular heterogeneity is a major reason that patients with NSCLC with a similar clinical stage and tumor histology can have dramatically different clinical outcomes and responses to treatment. Microarray techniques that profile the expressions of tens of thousands of genes simultaneously can measure this tumor heterogeneity at a global level. The use of microarray data is not simply the measurement of the expression of individual genes. Rather, the power lies in the ability to assay many thousands of genes simultaneously and evaluate the multivariate patterns of change across subsets that characterize a physiological or clinical state. This complexity opens the way to powerful tools of statistical analysis, not merely simple measures of reproducibility but identification of complex patterns within the date that reflect biology. The power of genomic technology, generating data sets of enormous complexity, heralds the transformation of biology into a quantitative science.

Developing a Useful Signature

Investigations of transcript levels on a genomic scale using hybridization-based arrays have led to formidable advances in our understanding of the biology of many human illnesses. At the same time, these investigations have generated controversy because of the probabilistic nature of the conclusions and the surfacing of noticeable discrepancies between the results of studies addressing the same biologic question.[6] The studies generally vary with the platforms used, the tissues studied, and the populations being sampled. Nevertheless, because of cost and other practical limitations, most microarray studies have used a relatively small number of biologic samples. As a result, cross-referencing lists of genes found to be associated with disease phenotypes in two separate studies usually produce relatively few genes in common,[7] even when one restricts attention to genes measured in both experiments. Although an incomplete overlap is to be expected given the small number of samples typically used and the large number of comparisons made, discrepancies have generated skepticism for this type of investigation. In this scenario, three related statistical questions[6] are important for making progress toward an objective assessment of the worth of microarray analysis results: (1) reliability, that is whether different measuring techniques are capturing the same biologic variation, (2) validation, that is whether the conclusions of a study are supported by other similar studies, and (3) combination, that is whether more reliable conclusions can be reached by jointly analyzing multiple studies. Variation in measurements of gene expression includes "technological" variation, associated with limitations of the measuring technologies, and "biologic" variation, because of the phenotype or experimental condition being studied, as well as natural variation of levels of gene expression in different samples of the same type.[6]

Gene-expression microarrays have been analyzed using clustering algorithms that group genes and samples on the basis of expression profiles and statistical methods that score genes on the basis of their relevance to various clinical attributes. Despite the natural caution associated with the implementation of new technologies in the clinical field, the utility of the results of microarray analysis as an effective diagnostic tool at the point of care is already being assessed. Nevertheless, simply listing genes associated with a certain tumor type is far from identifying the biologic processes in which these genes are involved, and clustering genes with similar expression patterns does not identify the causal molecular mechanisms that regulate them.[8]

The main objectives of large-scale expression profiling are to identify homogenous tumor subtypes based on gene-expression patterns, to find genes that are differentially expressed in tumors with different characteristics, and to develop a rule on the basis of gene expression allowing the prediction of patient prognosis or response to a particular treatment.[9] There are a variety of statistical approaches used with expression profiling data to achieve these aims, including clustering to identify homogenous subgroups, rules to define statistical significance of differential expression of large number of genes,[6] various classification methods for developing prediction rules[10] and then evaluation of the performance of the classification rule, and, finally, replication of the results in an independent population.[9]

Meta-analysis is a broad area consisting of techniques for analyzing data obtained from different studies. Because of the availability of large LC data sets, several statistical research groups have performed meta-analysis in which they

searched for commonalities among the large expression studies. One of these studies[7] compared both complementary DNA (cDNA) and oligonucleotide array analyses of lung carcinoma,[11–13] whereas another study[14] compared two different oligonucleotide analyses of lung AC.[11,13] Both of these meta-analyses have identified a high degree of reproducibility among the primary data-gathering studies. Gene-expression levels across samples were found to correlate between the independent studies, even across the different experimental platforms. Each of the analyses was also able to identify genes across the data sets that predicted patient outcome in both data sets. Nevertheless, the predictive gene sets were not overlapping, suggesting that the true predictive gene sets may be large and might be discovered effectively by a method such as gene set enrichment analysis.[15]

Nevertheless, despite many publications reporting positive results, we have to take some issues into consideration: clustering is overused[10]; the choice of analysis methods should be made according to the objective of the study. Microarray study objectives are often categorized as class comparison (or gene finding), class prediction (prediction of clinical outcome), or class discovery (grouping samples or genes with similar expression profiles).[16] The sample size should be as large as possible and composed of patients representative of the set of patients for which the classifier might be used in the future.[17] Further, clinical samples are relatively heterogeneous and they contain a variable percentage of cancer cells and other infiltrating cells. In cancer studies, selecting a heterogeneous group of patients presenting with different stages of disease and receiving a variety of treatments usually leads to substantial difficulties in interpreting the results of outcome-related analyses. The main problem lies in the possibility of confounding patient outcome by stage and treatment.[17,18] Validation of mRNA expression differences using another technique such as quantitative RT-PCR (qRT-PCR) are worthwhile but only confirm the RNA expression differences and not whether the signatures are predictive or useful.[9] The test validation must come on samples other than those used to develop the prediction rule; and the use of the gene signature needs to give more information than already available (e.g., clinical parameters).[9,16,17] Similarly, we need to learn whether this approach is reproducible and robust, particularly with prior specification. The characterization of cancer processes in terms of transcriptional changes in genes or modules is only a step toward the goal of obtaining a detailed mechanistic model of the processes leading to malignancy.[8] A key limitation of such approaches is that many regulators are regulated posttranscriptionally, and their activity is undetectable in gene-expression data.

The use of microarray technology has generated great excitement for its potential to identify biomarkers for cancer outcomes, but the reproducibility and validity of findings based on microarray data have come under widespread challenge. A proposal of guidelines for statistical analysis and reporting for clinical microarray studies presented as a checklist of "Do's and Don'ts" is provided by Dupuy and Simon.[18]

Technologies Used for Gene-Signature Development

Broadly, experimental platforms for expression arrays include oligonucleotide arrays synthesized by photolithography, oligonucleotide arrays synthesized by ink-jet printing, spotted oligonucleotide arrays, and spotted cDNA arrays. After hybridizing a given sample on the array, signal intensities can be determined by scanning it. These signal intensities are directly correlated to the amount of a given transcript within the sample. The number of interrogated transcripts varies between 5000 and 20,000.

Two basic strategies have been described for the analysis of microarray data. One involves the discovery of structure in a given data set without regard for prior knowledge of the underlying biology. This approach often referred to as "unsupervised analysis" uses the gene-expression data to find structure in the data that can then be used to infer biologically meaningful structure. This approach can be an effective tool in classifying biologic samples into categories that were not previously known to exist. By contrast, "supervised analysis" strategies do consider existing information and, indeed, use it to guide the analysis of the gene-expression data. The power of the supervised analysis lies in the ability to specifically drive the analysis to the phenotype of interest, taking advantage of the relevant information as a guide.[19]

Systematic microarray gene-expression profiling has proven to be a powerful and versatile tool for analysis of cancer classification. Analysis of purified RNA samples, frozen cell pellets, and frozen tumor tissues across different institutions and even different experimental platforms has revealed a high degree of correlation between institutions, and hierarchical clustering likewise correctly classified the samples across institutions.[7,15,20] Nevertheless, although we can often identify function associated with some of the genes in a signature, the challenge is to put this information into perspective with respect to the entire genomic profile.

qRT-PCR has emerged as a preferred method for independent validation of microarray-based results because it has equivalent or superior technical characteristics. The qRT-PCR assay is convenient in terms of laboratory work load and applicable for large-scale routine use, making it a viable alternative to more complex microarrays. qRT-PCR is a molecular biology technique that allows amplification and quantification in real time of defined RNA molecules from specific specimens. In brief, in the first step, DNA copies of the investigated RNA molecules are obtained by a process called reverse transcription, and DNA amplification is then obtained by using PCR. The quantification of the target RNA molecule is based on the analysis of the accumulation curve of the cDNA, as measured by the fluorescence detected at each cycle of the reaction. Two common methods of quantification are the use of fluorescent dyes that intercalate with double-stranded DNA and modified DNA oligonucleotide probes that fluoresce when hybridized with a cDNA. An example of these technologies is shown in Figure 1.

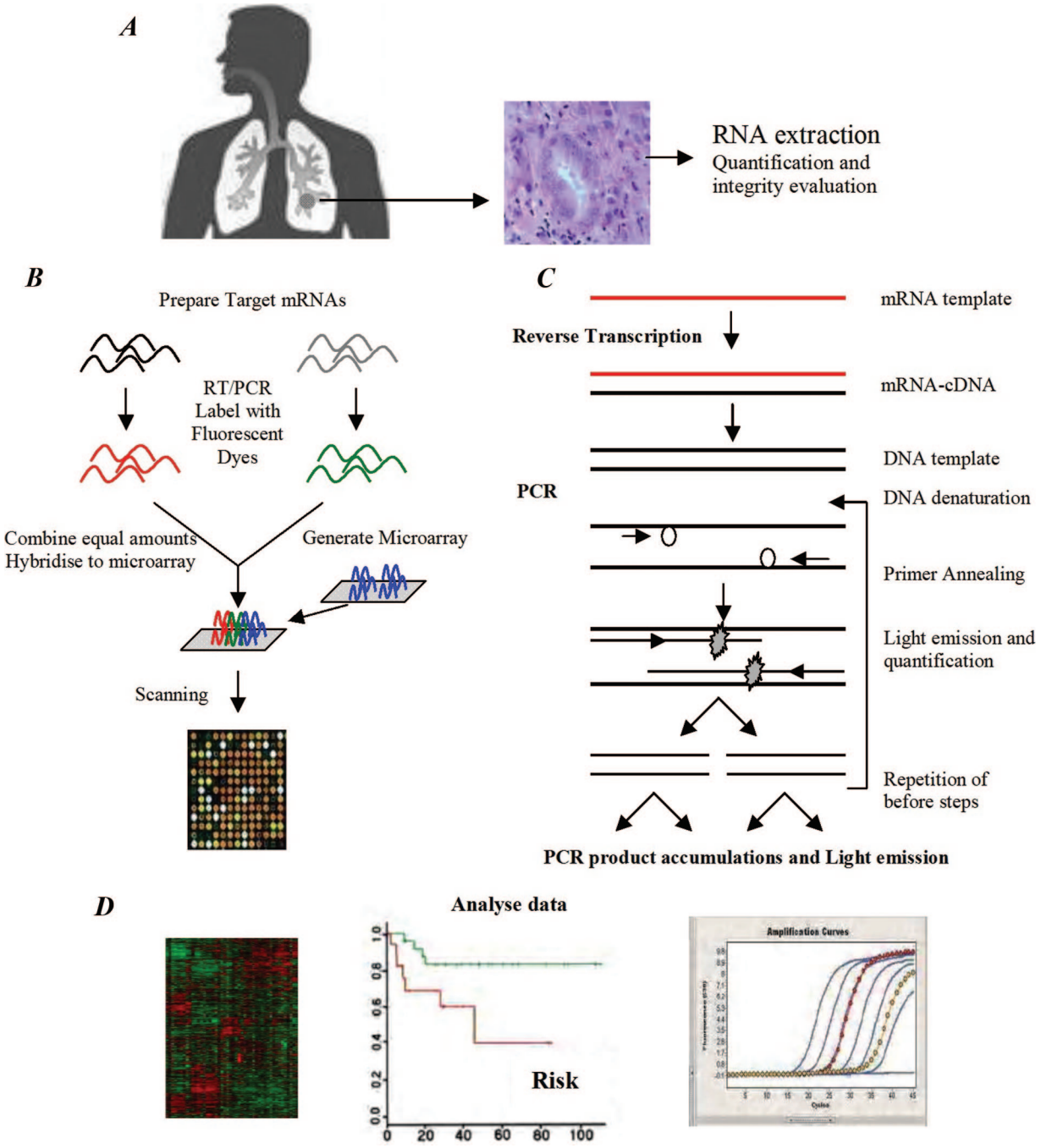

FIGURE 1. Microarrays and quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) are the molecular biology techniques enabling gene-expression analysis. *A*, Lung cancer tumors are sampled and shipped to the laboratory, where pathologic review is done to assess cancer cell contents, followed by RNA preparation and integrity evaluation. *B*, RNA is labeled with fluorescent dye and hybridized against thousands of different nucleotide sequences corresponding to different genes and arrayed on a solid surface. On hybridization, fluorescence emitted by single locations on the microarray is used to estimate gene-expression levels. *C*, qRT-PCR is based on reverse transcription of a specific mRNA into the complementary DNA (cDNA) molecule, which is used as a template in PCR. The production of double-stranded DNA is accompanied by emission of light, which is recorded throughout the process and correlates to the amount of the initial amount of RNA in the sample. *D*, Gene-expression levels are mathematically transformed into indexes predicting.

TABLE 1. Signatures Define Distinct Histological Subtypes of Lung Cancer

References	Disease	No. of Genes	Methods of Analysis	No. of LC Samples	Findings
Bhattacharjee et al.[11]	LC	18	Microarray	186	Signature define distinct subclasses of LC
Garber et al.[12]	LC	30	Microarray	67	Subclassification of AC into subgroups
Nacht et al.[22]	NSCLC	115	SAGE Microarray	9 SAGE libraries	Distinctive signature among types of LC
Beer et al.[13]	AC	50	Microarray	86	Signature differ between stage I and stage III tumors
Sugita et al.[23]	LC	20	Microarray RT-PCR	4 cell lines	Cancer/testis antigens as biomarkers in LC
Fujii et al.[24]	NSCLC		SAGE	9 SAGE libraries	Identification of novel genetic changes among NSCLC
Wang et al.[25]	NSCLC	6	RDA	5	Signature represents novel candidate tumor biomarker genes for NSCLC and its histological subtypes.
Pedersen et al.[26]	SCLC	Several	Microarray RT-PCR Xenografts	21 cell lines	Genes with differential expression between variants of SCLC
Yamagata et al.[27]	NSCLC	Several	Microarray	31 6 cell lines	Clustering identify histological subgroups of NSCLC
Kikuchi et al.[28]	NSCLC	Several	Microarray	37	Clustering distinguish two major histological types of NSCLC
Jiang et al.[14]	AC	10–13	Microarray Meta-analysis		Expression patterns differentiate diseased from normal samples
Takeuchi et al.[29]	NSCLC	30	Microarray GO analysis	149	Expression profile define two major types of AC
Hayes et al.[30]	AC	Several	Microarray Meta-analysis	231	Tumor subtypes correlate with clinically relevant covariates
Motoi et al.[31]	AC	Several	Microarray	100	Gene profiling clusters correlate with AC subtypes and EGFR mutations
Angulo et al.[32]	NSCLC	23	Microarray IH	69	Cluster analysis segregate tumors by histology and the presence of *EGFR* mutations
Kuner et al.[33]	NSCLC	30	Microarray RT-PCR	60	AC and SCC are characterized by distinct sets of cell adhesion molecules

LC, lung cancer; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; AC, adenocarcinoma; RT-PCR, real-time reverse transcriptase polymerase chain reaction; SAGE, serial analysis of gene expression; RDA, representational difference analysis technique; GO, gene ontology; IH, immunohistochemistry; EGFR, epidermal growth factor receptor.

LC-RELEVANT EXPRESSION SIGNATURES

Correlation between Lung Carcinoma Expression Profiles and Histopathology

The current staging system for LC has remained largely unchanged and continues to be based on histopathology and extent of disease at presentation. Thus, tumor classification systems provide the foundation for tumor diagnosis and, more importantly yet, a critical basis for patient therapy election. The heterogeneity of LC patients at each disease stage with respect to outcome and treatment response suggests that additional subclassification and substaging remain possible.[21] In the field of LC, genetic analysis by independent investigators has demonstrated a wide variety of potentially clinically important uses, including the ability to distinguish morphologic variants reliably, which might affect the treatment election and predict prognosis. The first results in this area were the findings of close correlations between gene-expression patterns and known histologic categories of lung carcinoma (Table 1). Thus, the histopathologic phenotypes of LC correspond to particular gene-expression patterns. Furthermore, novel subtypes had been identified within the histologic subtype of AC. This is particularly important because this category comprises patients with markedly differing outcomes, for example, the bronchioloalveolar carcinoma (BAC) subtype, which is histopathologically difficult to determine but is characterized by a more favorable prognosis. Nevertheless, to date, just a few systematic studies have been reported comparing detailed histopathologic examination, including immunohistochemistry of selected markers, with subclasses of carcinoma-defined gene-expression profiles.

Interestingly, in the study by Garber et al.,[12] the AC samples formed subclusters, whereas SCCs and SCLC samples clustered tightly together within their respective groups. Some degree of overlap in these subcategories with those identified by Bhattacharjee et al.[11] underline the reproducibility of the results across the two different experimental platforms used, namely, oligonucleotide arrays in the study by Bhattacharjee et al. and cDNA arrays in the study by Garber et al. The hypothesis tested in the work by Hayes et al.[30] is that lung AC subtypes defined by gene array analysis are reproducible and clinically relevant. Tumor subtypes were named according to overall similarity of gene-expression

patterns across hundreds or thousands of genes to easily recognizable morphologic LC variants. This naming choice emphasizes the view that the tumor subtypes are not dependent on identification of a fixed set of genes, specific analytic method, or microarray platform and allows future investigators to establish a common reference point lacking in this heterogeneous disease. Although unrecognized technical artifacts can drive clustering patterns in a single data set, it is unlikely that similar effects would be present in multiple cohorts using different assay platforms. The main focus of the analysis is the validation of AC subtypes derived from clustering of expression profiles, but they do not exclude the possibility that additional tumor subtypes might be described if the sample set were larger or of a different composition. In attempt to address AC subclassification comprehensively in the context of genetic data, Motoi et al.[31] report an analysis of 100 lung AC with correlation between detailed histologic subclassification and epidermal growth factor receptor (*EGFR*) mutation, chromogenic in situ hybridization, and protein expression and cDNA gene-expression profiles. They found several distinctive clinical, pathologic, and molecular subsets of lung AC using the 2004 WHO histologic classification and further classifying the mixed subtype AC according to the major histologic subtype. Their comprehensive approach examining all histologic subtypes has allowed them for identification of additional genetic-histologic correlation.

In recent years, there has been a great expectation of the potential of microRNAs (miRNAs). miRNA form a class of endogenously expressed, small noncoding RNA gene products about 22-nt long, with a recently established key role in the posttranscriptional regulation of gene expression.[34,35] To date, more than 900 human miRNAs have been experimentally identified (available at: http://www.sanger.ac.uk/Software/Rfam/mirna/), and it has been estimated they regulate more than one third of cellular messenger RNAs,[36] but only a handful of specific targets have been experimentally validated.[37,38] From several studies, it is now clear that many miRNAs are associated with primary human tumors.[39–43] More than 50% of human miRNAs genes are located in cancer-associated genomic regions, such as common breakpoint regions and fragile sites, in minimal regions of loss of heterozygosity, or minimal regions of amplification,[44–46] suggesting that miRNAs may play an important role in the pathogenesis of human cancers. Similar to mRNA-encoding genes, several miRNA-encoding genes have been meanwhile classified as oncogenic ("oncomiRs")[47,48] or tumor-suppressive genes according to their function in cellular transformation and expression in tumors.[37,49] Furthermore, tumor cells seem to undergo a general loss of miRNA expression, and forced reduction of global miRNA expression promotes transformation.[50] Recent advances in classification of NSCLC have identified differences in miRNA expression between SCC and AC.[51–54] Lebanony et al.,[54] have identified hsa-miR-205 as a biomarker for SCC that exhibit the most significant and strongest difference in expression between SCC and AC. A previous study identified hsa-miR-205 as one of a set of six miRNAs, which were differentially expressed in SCC compared with lung AC[51] (Table 2). Similarly, Yanaihara et al.[51] studied more than 100 patient-matched pairs of primary malignant and normal adjacent lung tissue and found that the expression of 43 miRNAs was significantly different in the tumor tissues compared with the normal adjacent tissues. Of these miRNAs, 28 were down-regulated, and 15 were up-regulated in the malignant tissue. Although miR-34b/c is mainly expressed in lung tissues, the expression level of miR-34b is decreased by more than 90% in LC cells.[53,63] Raponi et al.[62] have identified 15 miRNAs that were differentially expressed between normal lung and SCC, including members of the mirR-17-92 cluster and its paralogs, miR-155 and let-7, which had previously been shown to have prognostic value in AC. Also, miR-146b alone was found to have

TABLE 2. Lung Cancer-Related miRNAs

miRNAs	Events	References
hsa-miRNA let-7 family	Lung cancer development	Takamizawa et al.[39]
	Cell growth	Johnson et al.[40]
	Survival	Calin and Croce[55]
	Prognosis	Yanaihara et al.[51]
	Resistance to cytotoxic therapy	Johnson et al.[56]
		Brueckner et al.[57]
		Weidhaas et al.[58]
		Yu et al.[59]
hsa-miR-17–92 cluster	Cell growth	Hayashita et al.[60]
	Overexpression	He et al.[41]
		O'Donnell et al.[61]
		Raponi et al.[62]
hsa-miR-34 family	p53 targets	Bommer et al.[63]
		Raver-Shapira et al.[64]
		Tarasov et al.[65]
hsa-miR-34 b/c	Diagnostic	Bommer et al.[63]
		Liang[53]
hsa-miR-449	Diagnostic	Liang[53]
hsa-miR-205	Subclassification of NSCLC	Yanaihara et al.[51]
		Lebanony et al.[54]
hsa-miR-155	Prognosis and Diagnosis	Yanaihara et al.[51]
		Volinia et al.[66]
		Raponi et al.[62]
hsa-miR-21	Subclassification	Volinia et al.[66]
		Lebanony et al.[54]
hsa-miR-99b	Subclassification	Yanaihara et al.[51]
has-miR-202		
hsa-miR-203		
hsa-miR-102		
hsa-miR-204		
hsa-miR-191	Differential expression	Volinia et al.[66]
hsa-miR-128b		
hsa-miR-199a-1		
hsa-miR-196	Survival	Hu et al.[67]
hsa-miR-106 family	Cell cycle progression	Ivanovska et al.[68]
hsa-miR-221	Survival and relapse	Yu et al.[59]
hsa-miR-137		
hsa-miR-372		
hsa-miR182		

miRNA, micro RNA.

the strongest prediction accuracy for stratifying prognostic groups at 78%.[62]

The importance of tumor subtyping is clear even in the absence of a complete biologic understanding. Thus, correlations of molecular profiles from individual tumor samples to clinical outcome data hold the promise of better classification of LC and subsequently improved diagnostic and prognostic information for patient management. Taken together, gene-expression analyses have led to important insights into LC, such as the existence of molecularly defined subclasses of lung AC. In an ongoing collaborative study, these subgroups will hopefully be characterized in more detail.

Signatures that Predict Disease-Free and Overall Survival

Recent developments in Genomics, Proteomics, and Bioinformatics have conditioned the approach to the discovery of prognostic and predictor factors of human cancers. Global gene-expression profiling using high-throughput technologies has identified potential biomarkers and gene signatures for classifying patients with significantly different survival outcomes (Table 3). Although morphologic features and clinical stage based on the TNM system can roughly stratify patients for prognosis, it is often difficult to predict either which surgically managed patients are at risk for early relapse or which rare advanced-stage patients may experience prolonged survival. Patients whose early-stage tumors contain signatures predicting short survival times would benefit from the aggressive therapies currently given only to those with later-stage cancers.[71,94,103] Unfortunately, current methods of classification and staging are not completely reliable or sufficiently precise,[104] and no reliable markers exist to predict the outcome in patients with LC. Additionally, using clinical covariates together with the gene-expression data improved outcome prediction compared with using gene-expression data alone.[79,81,82,91,94]

Several studies have identified expression signatures that partition patients into prognostic groups. Furthermore, cross-study analyses of the data sets using different statistical approaches have generated additional prognostic gene sets. Nevertheless, Lau et al.[84] have shown a Venn diagram with 158 candidate prognostic genes from several studies showing minimal overlaps, but prognostic genes from some of these studies seem to share partially common protein-protein interaction pathways.[84] This discordance has been attributed to insufficiently powered studies[105] and variability in patient cohorts, arbitrary selection of a different time point for performance evaluation, expression profiling platforms, or statistical methodologies.

Beer et al.[13] have identified a gene-expression signature predictive of patients at high risk for poor overall survival. They built a 50-gene predictor using leave-one-out cross-validation to predict the survival of patients. Notably, in this study, some of the gene-expression results were confirmed by Northern blot and immunohistochemistry, validating the results of microarray analysis and representing a first step to the development of a clinically feasible test for predicting survival. Hopefully, in the near future, the results of all gene-expression studies in LC will translate into a clinically widely applicable test (e.g., immunohistochemistry or PCR) to allow for early identification of patients at high risk. Guo et al.[80] present a model system to identify important marker genes, which could improve the prognosis for individual patients with lung AC. They used several standard feature selection algorithms, random forests, correlation-based feature selection, and gain ratio attribute evaluation to identify novel molecular signatures with respect to the interactions among genes. Chen et al.[83] derived their five-gene signature from a global analysis of major subtypes of NSCLC (both SCC and AC) and identified only five genes (*STAT1*, *DUSP6*, *ERBB3*, *MMD,* and *LCK*) that were found to be capable of separating patients into two distinct prognostic groups. Lau et al.[84] have identified a three-gene mRNA expression-based classifier (*STX1A*, *HIF1A,* and *CCR7*) that can partition patients with early-stage NSCLC into subgroups with significantly different prognoses, and they show that gene-expression patterns are both independent and additive to the predictive ability of clinical parameters such as stage and histology. Nevertheless, Raz et al.,[93] show a four-gene model that predicts mortality better than clinical stage or tumor size (*WNT3a*, *RND3*, *LCK,* and *ERBB3* genes). Two genes of them overlap with the signature from Chen et al. Guo et al.[106] have used published microarray data sets to evaluate their previously identified cancer prognostic gene signature,[80] and their results indicate that the signature is an accurate predictor of survival in NSCLC. This signature was compared with the five-gene signature from Chen et al.[83] and the 133-gene signature from Potti et al.[79] The signature from Chen et al. was not validated in any of the three validation cohorts used, and they suggested that the gene signature from Potti et al contains correlated and/or redundant biomarkers. Shedden et al.[94] conducted a large retrospective, training-testing, multisite, blinded validation study to characterize the performance of several prognostic models based on gene expression for 442 lung AC, the specific type of LC that is increasing in incidence. The hypotheses proposed examined whether microarray measurements of gene expression either alone or combined with basic clinical covariates (stage, age, and sex) could be used to predict overall survival in LC subjects. The consideration of clinical covariates is highly relevant as gene-expression-based prediction is important in practice only if it provides more information than these measures. They show that using clinical covariates together with the gene-expression data improved outcome prediction compared with using gene-expression data alone, and thus, lung AC can be divided into groups with different survival rates. Sun et al.[91] described two gene signatures for AC and SCC suggesting that a prognostic signature may not be cell type specific, and a universal signature reflecting tumor aggressiveness and subsequent clinical outcome may exist across histologic cell types. This agrees with the studies by Raponi et al.[82] and, more recently, Roepman et al.[97] This would be clinically important because a unified gene signature would dramatically simplify the outcome evaluation process for different or unspecified types of carcinoma. Nevertheless, although unique prognostic signatures are more attractive because of

TABLE 3. Signatures Predicting Disease-Free, Overall Survival, and Metastatic Status

References	Disease	No. of Genes	Methods of Analysis	No. of LC Samples	Findings	Accuracy
Chen et al.[69]	AC	Several	Microarray Northern-blot Flow cytometry	4 cell lines	Clusters correlated with invasiveness	
Beer et al.[13]	AC	50	Microarray Northern-blot IH	86	Signature predict survival	
Wigle et al.[21]	NSCLC	16	Microarray	39	Signature correlated with disease free survival	
Ramaswamy et al.[70]	AC	17	Microarray	64	Signature predictive of metastasis	
Kikuchi et al.[28]	AC	40	Microarray	18	Molecular marker predict metastasis	
Endoh et al.[71]	AC	8	qRT-PCR	85	Stratification of patients according to their prognosis	
Sun et al.[72]	SCC	27	Microarray	15	Difference between high and low aggressiveness	80%
Takada et al.[73]	SCC AC	23 43	Microarray	92	Stratification of patients	100% 94%
Tomida et al.[74]	NSCLC	40	Microarray	50	Subclasses of SCC with different prognosis and invasiveness	82%
Diederichs et al.[75]	NSCLC	39	Microarray qRT-PCR	14	Expression of 2 proteins is associated with metastasis and predicts survival	
Müller-Tidow et al.[76]	NSCLC	10	qRT-PCR	70	RTKs predict metastasis and survival	
Xi et al.[77]	AC	50	Microarray	Meta-analysis	Expression profiles from primary tumor may predict lymph node status	88%
Inamura et al.[78]	SCC	50	Microarray	48	Two groups of SCC with different survival to 6 yr	68%
Potti et al.[79]	NSCLC	133	Microarray	Lung metagene model	Signature predict the risk of recurrence	79%
Guo et al.[80]	AC	37 12 18	Microarray Computacional model system	86	Signature predict tumor stage and differentiation	96% 94% 84%
Lu et al.[81]	NSCLC	64	qRT-PCR Tisuee microarray	Meta-analysis	Signature segregated patients into high- and low-risk groups	90%
Raponi et al.[82]	NSCLC	50	Microarray qRT-PCR IH	129	Stratification of high- and low-risk patients	71%
Chen et al.[83]	NSCLC	5	Microarray qRT-PCR	125	Predictor of relapse-free and overall survival	96%
Lau et al.[84]	NSCLC	3	qRT-PCR	147	Stratification stage I and II patients	
Larsen et al.[85]	SCC	71 79	Microarray	51	Signature predict outcome	72%
Larsen et al.[86]	AC	54	Microarray	48	Signature predict recurrence	
Bianchi et al.[87]	AC	10	qRT-PCR	Meta-analysis	Signature predict survival	75%
Rosell et al.[88]	NSCLC	9	qRT-PCR	126	Overexpression of BRCA1 mRNA associated with poor survival	
Seike et al.[89]	AC	11	qRT-PCR	80	Signature classified patients according to risk of recurrence	75%
Skrzypski et al.[90]	SCC	3	qRT-PCR	66	Signature associate with prognosis	70%
Sun et al.[91]	NSCLC	50	Microarray	Meta-analysis	Two signatures predict survival regardless of histologic cell type	
Landi et al.[92]	AC	12	Microarray qRT-PCR	135	Smoking-associated gene-expression signature	
Raz et al.[93]	AC	4	qRT-PCR	107	Prediction of overall and disease-free survival	
Shedden et al.[94]	AC	Several	Microarray	256	Several classifiers stratifying subjects according to risk	
Vicent et al.[95]	NSCLC	4	Microarray	Xenograft model	Signature associated with bone metastasis	
Landemaine et al.[96]	Breast cancer	6	Microarray qRT-PCR	72	Signature predict breast cancer lung metastasis	
Roepman et al.[97]	NSCLC	72	Microarray	103	Signature predict risk of recurrence	
Boutros et al.[98]	NSCLC	6	qRT-PCR	147	Signature stratify patients into groups with different prognosis	
Bröet et al.[99]	NSCLC		BAC array-CGH Tech.	85	Copy number alterations linked to relapse-free survival	
Hsu et al.[100]	NSCLC	4	Microarray databases qRT-PCR	9 NCI-60 cell lines	Identification of invasion-associated genes	
Showe et al.[101]	NSCLC	29	Peripheral blood samples	228	Signature in peripheral blood mononuclear cells identified presence of disease	
Tomida et al.[102]	AC	46	Microarray	60	Signature associated with relapse and death.	

LC, lung cancer; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; AC, adenocarcinoma; SCC, squamous cell carcinoma; qRT-PCR, real-time quantitative reverse transcriptase polymerase chain reaction; RTK, receptor tyrosine kinases; IH, immunohistochemistry.

their wider applicability, using independent prognostic signatures for different carcinomas may be more biologically significant and less influenced by genetic heterogeneity.[87] Boutros et al.[98] use a nonlinear algorithm that learned patient grouping (i.e., a semisupervised algorithm). The six-gene signature identified by this algorithm was validated in multiple testing data sets and with permutation analysis. This permutation analysis suggests a rationale for the number and diversity of distinct NSCLC prognostic markers identified. The nonoverlapping yet equally predictive gene signatures suggest the possibility that multiple sets of gene-expression biomarkers may exist in tumors that could be useful for outcome prediction. These genes may participate in similar molecular processes related to tumor aggressiveness. This may explain some of the heterogeneity of NSCLC gene-expression profiles observed to date in the literature.

Yu et al.[59] have shown that a five-miRNA signature (Table 2) can distinguish high-risk versus low-risk patients within stage and histologic subgroups of patients with NSCLC. Moreover, Hu et al.[67] provide evidence that common SNPs in miRNAs might play an important role in prediction of NSCLC survival. This finding may potentially identify and select high-risk patients for effective adjuvant therapy in addition to standard surgery, to improve the treatment outcome of NSCLC. Takamizawa et al.,[39] observed that the expression levels of let-7 were frequently reduced in both in vitro and in vivo LC studies, and reduced let-7 expression was significantly associated with shortened postoperative survival, independent of disease stage.[39,51] They also observed that overexpression of miRNA let-7 in A549 lung AC cell lines inhibited cancer cell growth. let-7 negatively regulates the expression of RAS and MYC by targeting their mRNAs for translation repression.[40] On the other hand, let-7g levels are up-regulated after irradiation in LC cell lines.[58] These findings suggest that miRNA let-7 may be a tumor suppressor gene.[40,56] Moreover, in humans, *let-7* is located at a chromosome region that is usually deleted in human cancers.[44] The *miR-17-92* cluster also enhances LC cell growth.[60] It has been demonstrated recently that the members of the miR-34 family are direct TP53 targets, which induce apoptosis, cell cycle arrest, and senescence.[63–65]

Based on the reports so far, it is possible that multiple LC genes classifiers provide similar prognostic capabilities, especially when they include genes that belong to the commonly deregulated pathways in lung carcinogenesis. For example, prognostic genes identified in four[71,83,84,88] qRT-PCR-based studies interact with proteins that are often implicated in cancer such as TP53, ERBB3, BRCA1, and EGFR.

Signatures in Metastatic Status Prediction

Metastasis is a complicated multistep process that involves interactions between cancer cells and their surrounding microenvironments. It is the principal event leading to death in individuals with cancer, but its molecular basis is poorly understood. The metastatic phenotype includes the ability to migrate from the primary tumor, survive in blood or lymphatic circulation, invade distant tissues, induct angiogenesis, and proliferate establishing distant metastatic nodules.[107]

The ability to detect postoperative residual tumor cells, occult metastases, or early tumor recurrence potentially may improve survival by early aggressive adjuvant therapy. In these situations, the availability of cancer cell-specific biomarkers is essential for the development of effective screening modalities. Recently, gene-expression profiling studies in several cancer tissue types have reported molecular signatures that are associated with metastasis (Table 3) and have a potential use for diagnostic purposes. No reliable clinical or molecular predictors of recurrent disease are currently available. Because of heterogeneity in recurrence rates among patients with the same stage of cancer, it is critical to isolate a reliable molecular signature in tumors that could be used to identify those who are likely to develop recurrent disease. Nevertheless, early-stage tumors have better clinical outcome and tumor staging aids treatment planning, but there are instances where patients unexpectedly develop recurrent disease, illustrating the limitations of current clinical staging techniques in accurately predicting tumor recurrence.[85] It is unclear whether the metastasis potential of individual tumors develops over time or whether the basic genetic program of the primary tumor predetermines the metastasis capability.[76] Although both concepts seem reasonable, recent data indicate that a metastatic program is inherent to tumors that do metastasize early.[70,75,76,89,102]

Potti et al.[79] identified groups of metagenes that can stratify patients with stages I to III NSCLC based on their risk of recurrence using a decision tree model that incorporates clinical data. The model seemed to be predictive in stage IA patients and was validated in two independent patient cohorts. Nevertheless, the precise components of the metagenes were not provided, and the histologic influences of lung tumors were not considered. Seike et al.[89] describe a novel approach to identify metastasis-related genes and their potential use for diagnostic purposes based on the knowledge that a few receptor tyrosine kinases are known to play an important role in solid tumor metastasis. In addition to the known metastasis-associated genes *EGFR* and *ERBB2*, several receptor tyrosine kinases previously not known to be associated with the metastatic process were identified as strong predictors for the development of metastasis in early-stage NSCLC.

Patients whose early-stage tumors contain signatures predicting short survival may benefit from more aggressive therapies and assign less aggressive treatments to patients at low risk for recurrence. Moreover, identification of genes critical for development of metastasis could lead to advances in therapeutics.

Signatures of Oncogenic Signaling Pathways

The study of oncogenic signaling pathways has progressed remarkably over the past few decades, resulting in the identification of a large collection of activated receptors, receptor-coupled activators, kinases, phosphatases, transcription factors, and various negative regulators of these activities. By using gene-expression signatures rather than specific gene mutations, we may detect the consequence of the mutation in the form of pathway deregulation. Various studies have also demonstrated the potential for using gene-expression profiles for the analysis of oncogenic pathways.[32,108–115]

Combining signature-based predictions across several pathways identifies coordinated patterns of pathway deregulation that distinguish between specific cancers and tumor subtypes. Linking pathway deregulation with sensitivity to therapeutics that target components of the pathway provides an opportunity to make use of these oncogenic pathway signatures to guide the use of targeted therapeutics.

A 2006 study by Bild et al.[109] describes a novel approach to targeted therapeutics in LC, relying on the principle that assay of gene-expression profiles provides a measure of the consequence of the oncogenic process, irrespective of how the pathway might have been altered. Thus, even if the known oncogene is not mutated but another component of the pathway is altered, the gene-expression profile will still detect the alteration. These authors used gene-expression profiling to investigate the coordination of these oncogenic pathways in driving tumor phenotype through examining the resulting probabilities of a tumor having deregulation of the signaling pathways. Angulo et al.[32] have shown that *EGFR*-mutant tumors clustered together in an unsupervised analysis, indicating that the presence of *EGFR* mutations confers very specific patterns of gene expression. This statement agrees with Motoi et al.[31] Thus, it is likely that *EGFR*-mutant lung tumors constitute a very closely defined disease entity. Also, studies involving the combination of mutation and gene-expression data[116] will hopefully help to model in more detail the oncogenic pathways that are active in the different subtypes of LC.

The RAS proteins are pivotal regulators of cellular proliferation, differentiation, motility, and apoptosis. K-*ras* oncogenes are frequently detected in mouse lung tumors. To analyze profiling of genes regulated by K-*ras* oncogene, Lee et al.[117] generated K-*ras*^{G12D} Tg mice expressing mutant type K-*ras* gene in lung tissue by the regulation of SPC promoter sequence. Gene-expression profiles of normal lungs and AC showed a distinct pattern in hierarchical clustering. These studies suggest that genes related to cancer development and inflammation were up-regulated, whereas genes related to the tumor suppression were down-regulated by K-*ras*, resulting in the tumor growth. By exploring the gene-expression correlates of a mouse model for lung AC, dependent of K-ras2 activation, Sweet-Cordero et al.[118] were able to identify a signature of *KRAS2* mutation in human lung AC. The *KRAS2* mutation-associated genes were validated by qRT-PCR analysis in the setting of RNA interference directed against K-*ras2*.

The signaling pathway mediated by transforming growth factor-β (TGF-β) participates in various biologic processes, including cell growth, differentiation, angiogenesis, apoptosis, and extracellular matrix remodeling. In the context of cancer, TGF-β signaling can inhibit tumor growth in early-stage tumors. Nevertheless, in late-stage tumors, the very same pathway promotes tumor invasiveness and metastasis. Because a major role for TGF-β has been established in several pathologic conditions, this pathway is a very attractive target for therapeutic intervention.[119–122] Ranganathan et al.[119] found that 267 genes were regulated in several LC cell lines in a similar manner but with different kinetics. Most actions of TGF-β are brought about by regulation of gene expression. The genes that are regulated and the way they are regulated are largely dependent on the cell type under consideration. They have shown that signaling pathways such as MAP kinase, focal adhesion, Wnt signaling, and Integrin αV are regulated by TGF-β. Borczuk et al.[123] identified a lung AC signature that segregated tumors into three subtypes distinguished by histologic invasiveness. Among the genes, differentially expressed was the type II TGF-β receptor (*TGFβRII*), which was lower in AC-mixed and solid invasive tumors compared with BAC. This finding, which suggested that *TGFβRII* repression was required for lung AC invasion, was confirmed using qRT-PCR and immnunohistochemistry, and by in vitro studies indicating that *TGFβRII* expression was inversely correlated with LC invasion. They used a tumor cell invasion system to identify and characterize downstream mediator in *TGFβRII*-repressed cells important for lung AC invasion. Candidate targets were identified using DNA microarray gene-expression signatures of AC tumor specimens and of *TGFβRII* knockdown cell in vitro. Afterward, they have focused on characterizing the molecular mechanisms important for invasion. In lung AC, loss of *TGFβRII* expression with concomitant altered TGF-β signaling is an important initiating event of invasion. To determine these events in lung AC tumor cells, they used genomics and in vitro-based invasion assay. Among the genes identified was *CCL5,* which encodes the CC chemokine RANTES.[124] Microarray data indicating *CCL5* expression was increased in *TGFβRII*-deficient cells were confirmed by qRT-PCR and by enzyme-linked immunosorbent assay. These studies provide insights into the molecular pathways that mediate progression of AC from noninvasive BAC to invasive AC and, thus, are of high clinical significance.

Understanding the molecular underpinnings of cancer may be of critical importance to the development of targeted intervention strategies. Identification of such targets, however, is notoriously difficult and unpredictable. Malignant cell transformation requires the cooperation of a few oncogenic mutations that cause substantial reorganization of many cell features and induce complex changes in gene-expression patterns.[107] Critical genes to this multifaceted cellular phenotype have, therefore, only been identified after signaling pathway analysis. Synergistic control of gene expression by oncogenic mutations thus emerges as an underlying key to malignancy and provides an attractive rationale for identifying intervention targets in gene networks downstream of oncogenic gain- and loss-of-function mutations. On the other hand, current efforts are concentrated on developing drugs that specifically target abnormal regulatory pathways of cancer cells. Alterations in these cell processes are driven by the activation/inactivation of key genes, essential point controllers, which constitute the molecular targets for the design of such specific therapies.

Signatures that Predict Responses to Chemotherapy

Various gene signatures and sequence alteration in target genes have been obtained for prediction of drug response in patients (Table 4) with remarkable clinical suc-

TABLE 4. Use of Gene-Expression Signatures to Predict Response to Chemotherapy

References	Disease	No. of Genes	Drugs	Methods of Analysis	No. of LC or Tumor Lines	Accuracy
Kikuchi et al.[28]	NSCLC	29–92	6 drugs	CD-DST	37 samples	
Szakacs et al.[125]		48	Human ABC transporters	qRT-PCR	NCI-60	
Oshita et al.[126]	NSCLC	9	Multiple	Microarray	29 samples	
	SCLC				18 samples	
Kakiuchi et al.[127]	NSCLC	12	Gefitinib	Microarray sqRT-PCR IH	33 biopsy samples	
Yauch et al.[128]	NSCLC	Several	Erlotinib	Microarray	42 cell lines	
Gemma et al.[129]	LC	Several	8 drugs	Microarray	29 cell lines	
Potti et al.[108]		50	Docetaxel and others drugs	Microarray	NCI-60 panel	>80%
Coldren et al.[130]	NSCLC	8	Gefitinib	Microarray qRT-PCR Flow cytom. IH	11 cell lines	
Gÿorffy et al.[131]		42–297	11 drugs	Microarray qRT-PCR	30 cell lines	
Balko et al.[132]	NSCLC	180 50	EGFR TKIs	Microarray	10 cell lines	90%
McDermott et al.[133]	LC	Several	14 TKIs	Microarray	7 cell lines	
Hsu et al.[134]	NSCLC	46 95	Cisplatin Pemetrexed	Microarrays	17 cell lines	83%
Miyanaga et al.[135]	NSCLC	9	HDAC inhibitors	Microarray	16 cell lines	

LC, lung cancer; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; CD-DST, collagen gel droplet embedded culture-drug sensitivity test; qRT-PCR, real-time quantitative reverse transcriptase polymerase chain reaction; TKI, tyrosine kinases inhibitors; IH, immunohistochemistry; HDAC, histone deacetylase; EGFR, epidermal growth factor receptor.

cesses. Improving our ability to manage the disease by optimizing the use of existing drugs and/or developing new agents is essential in this endeavor. To this end, individualizing treatments by identifying patients who will or will not respond to specific agents will potentially increase the overall effectiveness of these drugs and limit the incidence and severity of toxicities that impair the functional status of patients and their ability to tolerate further therapies.[134] The therapeutic improvement resulting from the new generation of cytotoxic chemotherapeutic agents seems to have reached a plateau because the mortality associated with metastatic LC has not changed for the past two decades. The major cause of failure of successful cancer treatment is resistance to currently available antineoplastic agents. Resistance can occur to individual anticancer drugs or more broadly to multiple drugs with different chemical structures and different mechanisms of action. From the clinician's point of view, the aim of drug resistance research is to improve treatment outcome by devising strategies that are able to circumvent primary drug resistance or to prevent the development of secondary antineoplastic drug resistance. Moreover, the detailed knowledge about the drug resistance status of a given patient with cancer can provide the basis for an individual patient-tailored chemotherapy regiment in the future. To achieve this aim, an exact prediction of the resistance status of a tumor patient is necessary.[131]

Of greatest importance will be to learn whether the human tumor cell lines can be used to generate signatures predictive (at least in part) of what goes on in the patient. If this is true then their use in preclinical studies will be greatly validated and increased; if not then great emphasis should be placed on study of tumor samples directly from patients.[136] Nevertheless, the use of cultured cell lines has the advantage to minimize the influences of sampling methods, although cell lines differ from tumors and should, therefore, be considered as surrogates that may contain information on the molecular cell biology and molecular pharmacology of cancer.[135] Many groups have been using preclinical models that make use of human tumor cell lines and/or xenografts to investigate gene-expression profiles associated with in vitro sensitivity (drug response phenotypes) to hundreds or even thousands of drugs. This approach was pioneered by John Weinstein and his team at the National Cancer Institute (NCI; Bethesda, MD) using data on the panel of 60 human tumor cell lines of various tissue origins (NCI-60 panel), which have been tested for sensitivity to more than 100,000 agents, and they correlated these drug response phenotypes with their gene-expression profiles.[137,138] Miyanaga et al.[135] with a view toward developing predictive markers for determining response to histone deacetylase inhibitor treatment in the context of individualized therapy for NSCLC related the cytotoxic activity of trichostatin A (TSA) and suberoylanilide hydroxamic acid (vorinostat) to corresponding gene-expression patterns using a modified NCI program. In this study, these two histone deacetylase inhibitors had distinct and differential activities in the panel of NSCLC cell lines tested. These results suggested that clinical studies in selected pa-

tients with NSCLC would be required for a more refined evaluation of these drugs.

Using in vitro drug sensitivity data coupled with microarray data, Potti el al.[108] developed gene-expression signatures that predict sensitivity to individual chemotherapeutic drugs. Each signature was validated with response data from an independent set of cell line studies. They show that many of these signatures can accurately predict clinical response in individuals treated with the drugs tested and with combination of those. Hsu et al.[134] described a novel approach to rationalized drug therapy in NSCLC, by developing predictors of cisplatin (a first-line agent) and pemetrexed (a second-line agent) sensitivity and demonstrating the clinical value of identifying the most appropriate drug on the basis of sensitivity profile for the treatment regimen of each individual patients, thus moving beyond empirical therapeutic choices that are now in current practice. They have made use of in vitro drug sensitivity data in cancer cell lines, coupled with microarray expression data, to develop gene-expression signatures reflecting sensitivity to cisplatin and pemetrexed. The capacity of these signatures to predict response in independent sets of cell lines and patient studies begins to define a strategy that addresses the potential to identify cytotoxic agents that best match individual patients with advanced NSCLC. In addition, it can potentially be applied to patients with early-stage NSCLC to predict who may benefit from adjuvant cisplatin-based therapy.

Kinase inhibitors constitute an important new class of cancer drugs, whose selective efficacy is largely determined by underlying tumor cell genetics.[133] Increased focus surrounds identifying patients with advanced NSCLC who will benefit from treatment with EGFR tyrosine kinase inhibitors (TKIs). Nevertheless, individual markers do not encompass all potential responders because of high levels of interpatient and intertumor variability. A multivariate predictor of EGFR TKI sensitivity based on gene-expression data would offer a clinically useful method of accounting for the increased variability inherent in predicting response to EGFR TKI and for elucidation of mechanisms of aberrant EGFR signaling.[127,128,130,132,139] Signatures which predict response to kinase inhibitors provide an important preclinical model to guide early clinical applications of novel targeted inhibitors.[133] Indeed, the development of targeted therapies approaches has experienced an increase in the last years.[140–146]

The goal of developing new prognostic and predictive signatures is to improve adjuvant treatment. Recent advances in biologically directed therapies for NSCLC require more accurate subclassification, as treatment may be dictated by histologic subtype. Differences in *EGFR* mutation rates and response to specific TKIs between histologic subtypes and serious hemorrhagic complications with vascular endothelial growth factor (VEGF) inhibitors in SCC may affect therapeutic choices. Hence, the importance of a single genetic profile determination in the treatment election. In conclusion, the development of signatures of drug sensitivity provides an opportunity to optimize therapy for patients with LC and perhaps other patients with advanced cancer where some drug-based therapy is considered the standard of care.

Analysis of Gene Expression in a Developmental Context

In recent years, the molecular underpinnings of the long-observed resemblance between neoplastic and immature tissue have begun to emerge. Genome-wide transcriptional profiling has revealed similar gene-expression signatures in several tumor types and early developmental stages of their tissue of origin. Nevertheless, it remains unclear whether such a relationship is a universal feature of malignancy, whether heterogeneities exist in the developmental component of different tumor types, and to which degree the resemblance between cancer and development is a tissue-specific phenomenon.[147] The similarities between cancer and development are evident on many levels of observation: microscopically, cancerous tissues appear as undifferentiated masses, with some tumor types even exhibiting embryonic tissue organization. The recapitulation of embryonal pulmonary gene expression in LC has been reported in several articles. Progresses have been made in our understanding of the embryological development of the lung and in the characterization of the putative LC stem cells. Nevertheless, in contrast with hematology, a hierarchical distribution of lung epithelial cells in lineages according to their degree of "stemness," commitment and differentiation have not been defined.

Naxerova et al.[147] have demonstrated a clear imprint of developmental gene expression in a wide range of tumors and with respect to different, even noncognate developmental backgrounds. They identified a set of genes that are up-regulated in most cancers and shown that this signature is active in early development. Genes that are active in lung AC are preferentially expressed in early lung development; the pattern is inversed for down-regulated genes, meaning that genes that are characteristic for the mature, differentiated stage of the lung are suppressed in LC. Hassan et al.[148] showed that an increased expression of the embryonic stem cells (ESCs) gene set and a decreased expression of the polycomb target gene set identify poorly differentiated lung AC. In addition, this gene-expression signature was associated with markers of poor prognosis and worse overall survival in lung AC. Nevertheless, there was no correlation between this ESC gene signature and any histologic or clinical variable assessed in lung SCC. Furthermore, lung AC that share a common gene-expression pattern with normal human ESCs were associated with decreased survival, increase biologic complexity, and increased likelihood of resistance to cisplatin.[149] In the study by Borczuk et al.,[150] the authors compared gene-expression analysis between LCs and a mouse study of early development. They found that genes expressed at highest levels in large cell carcinoma were generally correlated with early stages of lung development, in particular the pseudoglandular and canalicular stages of development. These include a variety of proliferation-related genes such as *CDK4*, *PCNA*, and *E2F3*. In contrast, the genes expressed at highest levels in lung AC were associated with later stages of normal lung development, in particular with terminal branching and the formation of

alveolar structures such as surfactant protein genes and their major regulatory factor, thyroid transcription factor 1. It should be noted that these same marker genes for alveolar differentiation have also been noted as the hallmarks of specific AC subtypes, in particular, gene-expression classes that correlate with the presence of bronchioalveolar carcinoma features.[11,12]

Expression studies suggest that miRNA let-7 may control a variety of processes both during development and in the maintenance of adult tissue homeostasis or at least play a critical role in the pathogenesis of LC.[56] In contrast to miRNA let-7, the expression of miRNA cluster miR-17-92 is remarkably increased in LC, especially in the most aggressive form, SCLC,[60] when compared with miRNA expression in normal cells. More detailed understanding of the stages and cell types during lung development, and their associated expression profiles, will be valuable to identify more completely the relationship between LC cell types and lung development.[15]

CHALLENGES OF USING EXPRESSION SIGNATURES

The application of high-throughput technologies to identify genes that are expressed differentially in tumor cells and normal tissues seems to be promising. Gene-expression profiling is considered a standard procedure for these analyses. Nevertheless, the levels of mRNA do not always reflect accurately the levels of the corresponding proteins nor do they reveal changes in epigenetic posttranscriptional modulation of proteins (e.g., phosphorylation, acetylation, methylation, ubiquitination, ADP ribosylation, glycosylation, and myristoylation) or changes in protein degradation rates. Therefore, analysis of differences in protein levels or modifications is important to complement studies on mRNA expression.[151] The reported lack of correlation between mRNA and protein expression has highlighted the importance of conducting parallel proteomics studies to complement cDNA or oligonucleotide array data. On the other hand, a change in protein level could be the result of increased gene expression or stabilization of mRNA or protein. Also, discrepancy could be due to posttranslational changes in proteins or differences between mRNA and protein analysis.

Signature has been variably formulated as a simple ratio or as an index, normalized to different sets of genes or standardized with calibration RNA, and stratified by using thresholds optimized within each study. The consequences of threshold decisions on the interpretation of data obtained during microarray studies have not been elucidated.[152] Threshold choice and, consequently, the number of genes in genetic signatures, also dramatically affect the gene function categories represented nonrandomly in signatures. There is a need for routine assessment of the robustness of microarray-based biologic conclusions by evaluation of the conclusion's statistical validity under a range of threshold parameters.[152] Pan et al.[152] argue that conclusions about nonrandom representation of certain biologic processes and cellular components in gene signatures identified by microarray analysis can depend significantly on the signal-to-noise threshold used to select these genes, and the relationship between category representation and threshold choice is neither linear nor predicable. This effect may result from possible nonlinearity of the ratio between the total number of genes identified and the number of genes in a particular category.

Attempts to integrate and cross-study validate the results of various gene-expression profiling projects are complicated by the use of diverse microarray platforms, sample set, protocols, and analytical approaches. Nevertheless, Parmigiani et al.[7] have developed a practical analysis for cross-study comparison, validation, and integration of cancer molecular classification studies using public data. They evaluated genes for cross-platform consistency of expression patterns, using integrative correlations, which quantify cross-study reproducibility without relying on direct assimilation of expression measurements across platforms.[11–13] Cross-study comparison revealed significant, albeit incomplete, agreement of gene-expression patterns related to LC biology and identified genes that reproducibly predict outcomes. With respect to the question of integrating knowledge across studies, the comparative analysis of three data sets provides encouragement that there is a significant level of consistency with respect to genes that distinguish well-defined classes of LC (e.g., SCC versus AC) and genes that are associated with LC patient survival, but there was also significant scatter in the comparative correlations of gene-expression levels with relationship to classification of outcome. The integrative correlation analysis indicated that there remains a substantial component of unexplained variability across studies.[7]

It is unknown whether gene-expression profiles are more or less likely than traditional biomarkers to be generalizable across populations with varying genetic background. Gene-expression patterns have also been associated with specific genetic mutations (e.g., *EGFR*), indicating that specific DNA mutations or polymorphisms may affect the performance of a signature.

Although several genome-wide expression microarray-based prognostic models of LC have been reported, such array-based technology may be suboptimal for clinical use because of the need for specialized laboratory facilities and complex statistical analysis. It is also limited by the large number of genes in the analysis, lack of reproducibility and independent validation of the results, and the need of fresh-frozen tissues. Prognostic models based on gene expression of a limited number of genes using qRT-PCR may be more clinically practical. A limited qRT-PCR-based model has the advantage of being more reproducible and more feasible in the clinical setting and requires smaller quantities of tumor tissue than microarray-based models.[71]

In recent years, there has been a significant interest in developing tools and protocols that enable mRNA profiling from more readily available formalin fixation and paraffin embedding (FFPE) tissues. There is a huge resource of FFPE tissues specimens held in histopathology departments around the world. These samples provide an invaluable resource for studying the molecular basis of disease, making it possible to perform large retrospective studies correlating molecular features with therapeutic response

and clinical outcome. To date, gene-expression profiling from FFPE tissues has been problematic, as the retrieval of RNA from FFPE material is challenging.[153] Although tissue architecture and proteins are preserved with paraffin embedding, this method does not preserve nucleic acids very well, resulting in RNA that is often significantly degraded. Moreover, formalin fixation causes cross-linkage between nucleic acids and proteins[154,155] and modifies RNA covalently by the addition of monomethylol groups to the bases, making subsequent RNA extraction, reverse transcription, and quantitation problematic. Consequently, significant efforts to improve extraction of RNA from formalin-fixed tissue have been made by introducing various modifications to the extraction steps. These modifications have been implemented in almost all commercially available FFPE extraction methods today (from Nugen Technologies Inc., Ilumina Inc., Almac Diagnostic, etc.).

Successful extraction and amplification of RNA from FFPE tissue have been reported since the late 1980s.[154,156,157] Preliminary findings of Linton et al.[158] suggest that reliable microarray data could be generated using FFPE tissue and that this could be used for the identification of prognostic genes. The authors predicted that the sensitivity (about 50%) would be improved in the future by using improved mRNA extraction and amplification methods and by the use of new microarray platforms specifically designed for the interrogation of FFPE tissue. On the other hand, Specht et al.[159] used qRT-PCR to analyze the expression of a panel of cancer-relevant genes in matched frozen and FFPE xenograft sections. The authors reported no significant differences between gene-expression levels obtained with both approaches, and neither fixation time nor grade adversely affected these results. In fact, miRNA are well preserved in FFPE tissue, as has been demonstrated by Li et al.[160] and Lebanony et al.[54] in their studies, making them ideal candidates for molecular markers for use in routinely processed material.

Research published in recent years has taken us from a point where mRNA profiling from FFPE was not possible to a point where it is now accepted that gene-expression measurements can be performed from FFPE tissue using both qRT-PCR methods and by microarray profiling. It has become clear that the implementation of a standardized approach to fixation and storage of FFPE tissues and the further improvements in technology and techniques to interrogate this tissue are required if we are to fully embrace the utility of this resource.

FUTURE DIRECTIONS

Global molecular profiling of cancers has shown broad utility in delineating pathways and processes underlying disease, in predicting prognosis and response to therapy, and in suggesting novel treatments. To gain further insights from such data, Rhodes et al.[161] have integrated and analyzed a comprehensive collection of "molecular concepts" representing more than 2500 cancer-related gene-expression signatures from Oncomine and revision of the literature, drug treatment signatures from the Connectivity Map, target gene sets from genome-scale regulatory motif analyses, and reference gene sets from several gene and protein annotation databases. The Oncomine database represents a concerted effort to integrate and analyze such data. Another genome scale analysis, the Connectivity Map, examined hundreds of compound treatment gene-expression profiles and showed that such profiles could be used in a screen to identify compounds capable of reversing a gene-expression program active in disease.[19,162]

Although gene-expression studies are the predominant type of genome-scale molecular analyses to date, other high-throughput experimental modalities include proteomic profiling, transcription factor binding analysis, epigenetic profiling, and sequence-based analyses. In addition, several systematic annotation efforts have provided a variety of valuable genome-scale characterizations. Several tools are available to compare a query gene list to a reference set of gene lists. For example, Gene Set Enrichment Analysis allows one to compare a query signature with a variety of gene sets based on pathways, Gene Ontology terms, regulatory motifs, chromosomal regions, and perturbation experiments.[163] Databases with complete data on each patient are needed, including all analysis and procedures used to produce a risk estimate from a tumor sample. Such databases could renew and expand the currently limited pool of validation databases. It is necessary to do comparative effectiveness studies, oversight of test development, or research funding should encourage contrasts and combinations with existing expression-based predictors.

The molecular origins of LC lie in complex interactions between the environment and host genetic susceptibility. LC then evolves through genetic and epigenetic changes, including deregulated signaling pathways, which are potential targets for chemoprevention and therapy. So far, all the work that have dealt with diagnostic and prognostic markers of LC have focused on genes expressed in tumors. Nevertheless, a key challenge with LC is the early detection of the primary tumor or of its recurrence. Clearly, the tumor interacts with its microenvironment through a variety of autocrine and paracrine mechanisms, and there undoubtedly will be biomarkers of tumor response that focus on the microenvironment that will need to be developed because that microenvironment may properly be the therapeutic target (e.g., the tumor vasculature, which is targeted with the anti-VEGF monoclonal antibody bevacizumab). The gene-expression patterns in surrogate tissues are also another source of information, although few data are available to support the use of microarrays to identify surrogate markers in peripheral tissues.[89] Emerging techniques for genomic, gene expression, epigenetic, and proteomic profiling[164–166] could revolutionize clinical approaches across the spectrum of LC types and subtypes. Genome-wide and other molecular assessments are helping elucidate germ-line variations that may contribute to LC risk,[167–169] prognosis[170] (Tables 1–3), treatment sensitivity[171,172] (Table 4), and somatic genetic alterations that occur in lung carcinomas[116,173] and in high-risk lung tissue associated with tumors.[89]

CONCLUSIONS

New gene-expression signatures based on functional criteria may provide complementary information that will

help to refine a patient's prognosis and inform therapeutic choices. Also, gene-expression profiles that can predict response to commonly used cytotoxic agents provide opportunities to better use these drugs, including using them in combination with existing targeted therapies. The promise of effective targeted therapy for LC requires rigorous identification of potential targets combined with intensive discovery and development efforts aimed at developing effective drugs for these targets. We now recognize that getting the right drug to the right target in the right patient is more complicated than one could have imagined a decade ago.[174] To move toward the goal of individualizing or personalizing the approach to LC treatment, molecular profiling of treatment-sensitive and treatment-resistant tumor molecular subtypes at the time of diagnosis, during treatment, and at relapse will be essential.[175]

The introduction of gene-expression tests has ushered in a new era in which many conventional clinical markers may be seen merely as surrogates for more fundamental genetic and physiologic processes that can be measured with these tests. The multidimensional nature of these predictors' demands that large numbers of clinically homogeneous patients be used in the validation process and that exceptional rigor and discipline are be applied in evaluation. Every study provides an opportunity to modify a genetic signature, but the most important is to find the right balance between speed of innovation and development of reliable tools. It will be important to preserve genetic and clinical information from tested patients to facilitate further evaluation and innovation in current populations. Although these tests show great promise to improve predictions of prognosis and treatment benefit for patients with early-stage LC, more must be learned about the extent of that improvement, in which it is most improved, and how the tests are best incorporated into decision making about current LC treatment.

Lack of overlap in the composition of LC prognostic signatures is not unexpected and likely reflects the fact that numerous gene-expression signatures may be capable of predicting outcome. The reconstruction of the molecular mechanisms that underlie a complex process, such as tumorigenesis, is a formidable challenge. This challenge arises in part from difficulties associated with microarray assays, including noise in the data and limited reproducibility across platforms and researches. Thus, there is a strong possibility that sample collection methods, processing protocols, single-institution subject cohorts, small sample sizes, and peculiarities of the different microarray platforms are contributing significantly to the results. On the other hand, human LC has extensive alterations of miRNA expression that may deregulate cancer-related genes, and alteration in miRNA expression may play a critical role in tumorigenesis and cancer progression. The development of new miRNA markers in the near future will represent one of the main goals in molecular medicine. The variety of genes found useful for classification suggests that several mechanisms contribute to the clinical progression of LC and that several classifiers may be equally effective. Therefore, an integrated approach using gene expression together with associated clinical, pathologic, and other information may be more promising for future work. Moreover, most analyses implicitly treat mRNA expression as a surrogate for protein activity level, an assumption that does not account for processes such as mRNA stability, protein degradation, and posttranslational modification. Although genomic approaches are prevalent in cancer research, we are still far from reconstructing molecular mechanisms in human cancer. Both regulatory and signaling networks are larger and more elaborate, and the control of many genes and processes involves undefined epigenetic mechanisms, a higher degree of combinatorial regulation and multiple signaling pathways. Furthermore, many interactions are context specific, as different components of the molecular network are active in different cellular states and phenotypes. Therefore, the critical test of prognostic signatures is validation in independent data sets and with different assays. Only a few profiles have been validated rigorously in independent cohorts of patients using the qRT-PCR technique that is considered the gold standard for mRNA expression analysis.[176] On the other hand, it remains to be systematically determined which algorithms are best suited for selecting and validating stable prognostic gene list.

Given this complexity inherent to LC, it would be surprising if a single gene-expression pattern could effectively describe and ultimately predict the clinical course of the disease for all patients. Recognizing the importance of addressing this difficulty, it is very important to integrate various forms of data, including clinical variables and multiple gene-expression profiles, to build robust predictive models for the individual patient. The published information about gene-expression patterns that are predictive of survival in LC, the identification of new diagnostic markers, and the new and emerging preliminary data on gene-expression patterns in surrogate tissues and cells suggests that the quality of information will drive the clinical application. The ability to generate multiple relevant descriptions of gene-expression data that predict oncogenic pathway activation creates a unique opportunity to better match individual patient characteristics with what may be the most appropriate therapeutic option in advanced-stage LC. If we want to make progress, we will need to develop better techniques for assessing the benefits of targeted drugs (particularly combinations of them), as well as more complete methods for molecularly profiling each patient's cancer. To move toward a goal of individualized medicine with rationally targeted therapies, it is imperative to do more molecular profiling of treatment-sensitive and treatment-resistant tumors, to better understand the underlying biology of specific molecular subtypes.[177]

One important difficulty of the signatures with high number of genes is that the methodology used is not directly transferable to the clinical setting. Thus, there is need to develop strategies aimed at the identification of small signatures that can be easily analyzed in the clinical laboratory. Nevertheless, Shedden et al.[94] in their multisite study argues that the variety of probe sets showing some predictive capacity suggests that information about lung AC outcomes may not be concentrated in just a few exceptional genes and the variety of genes found useful for AC classification into groups with different survival rates suggests that several

mechanisms contribute to the clinical progression of lung AC and that several classifiers may be equally effective. So, how many genes are enough? Many molecular markers that predict patient survival independent of the TNM staging system have been reported in the literature. These include oncogene (*KRAS, BCL2, ERRB2*, and *EGFR*), tumor suppressor genes (*TP53, RB, TP16,* and *P27KIP1*), cell cycle modulators (cyclins), molecules related to tumor invasion and metastasis (CDRR, cathepsin B, and matrix metalloproteinase), telomerase, molecules involved in tumor angiogenesis (VEGF and VEGF receptor), cyclo-oxygenase 2, etc. Nevertheless, for the moment there is no single biomarker available that can be routinely used for prediction of prognosis of LC. This may be quite reasonable considering that cancer is a complex multigene disease, and each new signature generated from distinct models, platforms, or mathematical methods has the potential of adding prognostic information. Molecular profiling of the type described in this review has begun in clinical trials[79,143,178] and promises to select patients who are most likely to benefit from therapy and to guide the development of more effective agents that will personalize standard medicine for LC.[179]

Although some of the modular approaches outlined earlier in the text enhance our ability to analyze disease process-relevant signatures, we are still far from understanding the role that these signatures have in cancer. We may be able to derive a more comprehensive perspective on cancer processes by integrating existing assays with histopathologic, clinical, and environmental information on the one hand, and with measurements of genetic variation, such as SNPs or DNA copy-number changes, on the other. Although there still remain significant challenges to the use of gene-expression-based classifiers in the clinical setting, the potential that these tools can improve patient care and increase survival provides a strong impetus to continue to refine these approaches for eventual clinical use.

REFERENCES

1. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3:e442.
2. Jemal A, Siegel R, Ward E, et al. Cancer statistics 2009. *CA Cancer J Clin* 2009;59:225–249.
3. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
4. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291:1304–1351.
5. Dressman HK, Bild A, Garst J, et al. Genomic signatures in non–small-cell lung cancer: targeting the targeted therapies. *Curr Oncol Rep* 2006;8:252–257.
6. Garrett-Mayer E, Parmigiani G, Zhong X, et al. Cross-study validation and combined analysis of gene expression microarray data. *Biostatistics* 2008;9:333–354.
7. Parmigiani G, Garrett-Mayer ES, Anbazhagan R, et al. A cross-study comparison of gene expression studies for the molecular classification of lung cancer. *Clin Cancer Res* 2004;10:2922–2927.
8. Segal E, Friedman N, Kaminski N, et al. From signatures to models: understanding cancer using microarrays. *Nat Genet* 2005;37(Suppl): S38–S45.
9. Michiels S, Koscielny S, Hill C. Interpretation of microarray data in cancer. *Br J Cancer* 2007;96:1155–1158.
10. Allison DB, Cui X, Page GP, et al. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 2006;7:55–65.
11. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 2001;98:13790–13795.
12. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 2001;98:13784–13789.
13. Beer DG, Kardia SLR, Huang C-C, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Med* 2002;8:816–824.
14. Jiang H, Deng Y, Chen HS, et al. Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC Bioinformatics* 2004;5:81. Available at: http://www.biomedcentral.com/1471–2105/5/81.
15. Meyerson M, Carbone D. Genomic and proteomic profiling of lung cancers: lung cancer classification in the age of targeted therapy. *J Clin Oncol* 2005;23:3219–3226.
16. Pawitan Y, Michiels S, Koscielny S, et al. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics* 2005;21:3017–3024.
17. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;365: 488–492.
18. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 2007;99:147–157.
19. Nevins JR, Potti A. Mining gene expression profiles: expression signatures as cancer phenotypes. *Nat Rev Genet* 2007;8:601–609.
20. Dobbin KK, Beer DG, Meyerson M, et al. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin Cancer Res* 2005;11:565–572.
21. Wigle DA, Jurisica I, Radulovich N, et al. Molecular profiling of non-small cell lung cancer and correlation with disease-free survival1. *Cancer Res* 2002;62:3005–3008.
22. Nacht M, Dracheva T, Gao Y, et al. Molecular characteristics of non-small cell lung cancer. *Proc Natl Acad Sci USA* 2001;98:15203–15208.
23. Sugita M, Geraci M, Gao B, et al. Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma. *Cancer Res* 2002;62:3971–3979.
24. Fujii T, Dracheva T, Player A, et al. A preliminary transcriptome map of non-small cell lung cancer. *Cancer Res* 2002;62:3340–3346.
25. Wang K-k, Liu N, Radulovich N, et al. Novel candidate tumor marker genes for lung adenocarcinoma. *Oncogene* 2002;21:7598–7604.
26. Pedersen N, Mortensen S, Sorensen SB, et al. Transcriptional gene expression profiling of small cell lung cancer cells. *Cancer Res* 2003; 63:1943–1953.
27. Yamagata N, Shyr Y, Yanagisawa K, 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. 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;22:2192–2205.
29. Takeuchi T, Tomida S, Yatabe Y, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. *J Clin Oncol* 2006;24:1679–1688.
30. Hayes DN, Monti S, Parmigiani G, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J Clin Oncol* 2006;24:5079–5090.
31. Motoi N, Szoke J, Riely GJ, et al. Lung adenocarcinoma: modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. *Am J Surg Pathol* 2008;32:810–827.
32. Angulo B, Suarez-Gauthier A, Lopez-Rios F, et al. Expression signatures in lung cancer reveal a profile for EGFR-mutant tumours and identify selective PIK3CA overexpression by gene amplification. *J Pathol* 2008;214:347–356.

33. Kuner R, Muley T, Meister M, et al. Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. *Lung Cancer* 2009;63:32–38.
34. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–297.
35. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 2007;8:93–103.
36. Lim LP, Lau NC, Garrett-Engele P, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005;433:769–773.
37. Liu Z, Sall A, Yang D. MicroRNA: an emerging therapeutic target and intervention tool. *Int J Mol Sci* 2008;9:978–999.
38. Backes C, Meese E, Lenhof HP, et al. A dictionary on microRNAs and their putative target pathways. *Nucleic Acids Res* 2010;38:4476–4486.
39. Takamizawa J, Konishi H, Yanagisawa K, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* 2004;64:3753–3756.
40. Johnson SM, Grosshans H, Shingara J, et al. RAS is regulated by the let-7 microRNA family. *Cell* 2005;120:635–647.
41. He L, Thomson JM, Hemann MT, et al. A microRNA polycistron as a potential human oncogene. *Nature* 2005;435:828–833.
42. Sassen S, Miska EA, Caldas C. MicroRNA: implications for cancer. *Virchows Arch* 2008;452:1–10.
43. Bartels CL, Tsongalis GJ. MicroRNAs: novel biomarkers for human cancer. *Clin Chem* 2009;55:623–631.
44. Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci USA* 2004;101:2999–3004.
45. Calin GA, Croce CM. Chromosomal rearrangements and microRNAs: a new cancer link with clinical implications. *J Clin Invest* 2007;117:2059–2066.
46. Sevignani C, Calin GA, Nnadi SC, et al. MicroRNA genes are frequently located near mouse cancer susceptibility loci. *Proc Natl Acad Sci USA* 2007;104:8017–8022.
47. Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. *Nat Rev Cancer* 2006;6:259–269.
48. Manikandan J, Aarthi JJ, Kumar SD, et al. Oncomirs: the potential role of non-coding microRNAs in understanding cancer. *Bioinformation* 2008;2:330–334.
49. Zhang B, Pan X, Cobb GP, et al. microRNAs as oncogenes and tumor suppressors. *Dev Biol* 2007;302:1–12.
50. Kumar MS, Lu J, Mercer KL, et al. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet* 2007;39:673–677.
51. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 2006;9:189–198.
52. Rosenfeld N, Aharonov R, Meiri E, et al. MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* 2008;26:462–469.
53. Liang Y. An expression meta-analysis of predicted microRNA targets identifies a diagnostic signature for lung cancer. *BMC Med Genomics* 2008;1:61. Available at: http://www.biomedcentral.com/1755–8794/1/61.
54. Lebanony D, Benjamin H, Gilad S, et al. Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from non-squamous non-small-cell lung carcinoma. *J Clin Oncol* 2009;27:2030–2037.
55. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006;6:857–866.
56. Johnson CD, Esquela-Kerscher A, Stefani G, et al. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res* 2007;67:7713–7722.
57. Brueckner B, Stresemann C, Kuner R, et al. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res* 2007;67:1419–1423.
58. Weidhaas JB, Babar I, Nallur SM, et al. MicroRNAs as potential agents to alter resistance to cytotoxic anticancer therapy. *Cancer Res* 2007;67:11111–11116.
59. Yu SL, Chen HY, Chang GC, et al. MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell* 2008;13:48–57.
60. Hayashita Y, Osada H, Tatematsu Y, et al. A polycistronic microRNA cluster, miR-17–92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res* 2005;65:9628–9632.
61. O'Donnell KA, Wentzel EA, Zeller KI, et al. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 2005;435:839–843.
62. Raponi M, Dossey L, Jatkoe T, et al. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. *Cancer Res* 2009;69:5776–5783.
63. Bommer GT, Gerin I, Feng Y, et al. p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Curr Biol* 2007;17:1298–1307.
64. Raver-Shapira N, Marciano E, Meiri E, et al. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol Cell* 2007;26:731–743.
65. Tarasov V, Jung P, Verdoodt B, et al. Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. *Cell Cycle* 2007;6:1586–1593.
66. Volinia S, Calin GA, Liu CG, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci USA* 2006;103:2257–2261.
67. Hu Z, Chen J, Tian T, et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. *J Clin Invest* 2008;118:2600–2608.
68. Ivanovska I, Ball AS, Diaz RL, et al. MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Mol Cell Biol* 2008;28:2167–2174.
69. Chen JJW, Peck K, Hong T-M, et al. Global analysis of gene expression in invasion by a lung cancer model. *Cancer Res* 2001;61:5223–5230.
70. Ramaswamy S, Ross KN, Lander ES, et al. A molecular signature of metastasis in primary solid tumors. *Nature Genet* 2003;33:49–54.
71. Endoh H, Tomida S, Yatabe Y, et al. Prognostic model of pulmonary adenocarcinoma by expression profiling of eight genes as determined by quantitative real-time reverse transcriptase polymerase chain reaction. *J Clin Oncol* 2004;22:811–819.
72. Sun Z, Yang P, Aubry MC, et al. Can gene expression profiling predict survival for patients with squamous cell carcinoma of the lung? *Mol Cancer* 2004;3:35. Available at: http://www.molecular-cancer.com/content/3/1/35.
73. Takada M, Tada M, Tamoto E, et al. Prediction of lymph node metastasis by analysis of gene expression profiles in non-small cell lung cancer. *J Surg Res* 2004;122:61–69.
74. Tomida S, Koshikawa K, Yatabe Y, et al. Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. *Oncogene* 2004;23:5360–5370.
75. Diederichs S, Bulk E, Steffen B, 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.
76. Muller-Tidow C, Diederichs S, Bulk E, et al. Identification of metastasis-associated receptor tyrosine kinases in non-small cell lung cancer. *Cancer Res* 2005;65:1778–1782.
77. Xi L, Lyons-Weiler J, Coello MC, et al. Prediction of lymph node metastasis by analysis of gene expression profiles in primary lung adenocarcinomas. *Clin Cancer Res* 2005;11:4128–4135.
78. Inamura K, Fujiwara T, Hoshida Y, et al. Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization. *Oncogene* 2005;24:7105–7113.
79. Potti A, Mukherjee S, Petersen R, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 2006;355:570–580.
80. Guo L, Ma Y, Ward R, et al. Constructing molecular classifiers for the accurate prognosis of lung adenocarcinoma. *Clin Cancer Res* 2006;12:3344–3354.
81. Lu Y, Lemon W, Liu PY, et al. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS Med* 2006;3:e467.
82. Raponi M, Zhang Y, Yu J, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* 2006;66:7466–7472.
83. Chen H-Y, Yu S-L, Chen C-H, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 2007;356:11–20.
84. Lau SK, Boutros PC, Pintilie M, et al. Three-gene prognostic classifier for early-stage non–small-cell lung cancer. *J Clin Oncol* 2007;25:5562–5569.

85. Larsen JE, Pavey SJ, Passmore LH, et al. Expression profiling defines a recurrence signature in lung squamous cell carcinoma. *Carcinogenesis* 2007;28:760–766.
86. Larsen JE, Pavey SJ, Passmore LH, et al. Gene expression signature predicts recurrence in lung adenocarcinoma. *Clin Cancer Res* 2007;13: 2946–2954.
87. Bianchi F, Nuciforo P, Vecchi M, et al. Survival prediction of stage I lung adenocarcinomas by expression of 10 genes. *J Clin Invest* 2007; 117:3436–3444.
88. Rosell R, Skrzypski M, Jassem E, et al. BRCA1: a novel prognostic factor in resected non-small-cell lung cancer. *PLoS One* 2007;2:e1129.
89. Seike M, Yanaihara N, Bowman ED, et al. Use of a cytokine gene expression signature in lung adenocarcinoma and the surrounding tissue as a prognostic classifier. *J Natl Cancer Inst* 2007;99:1257–1269.
90. Skrzypski M, Jassem E, Taron M, et al. Three-gene expression signature predicts survival in early-stage squamous cell carcinoma of the lung. *Clin Cancer Res* 2008;14:4794–4799.
91. Sun Z, Wigle DA, Yang P. Non-overlapping and non-cell-type-specific gene expression signatures predict lung cancer survival. *J Clin Oncol* 2008;26:877–883.
92. Landi MT, Dracheva T, Rotunno M, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* 2008;3:e1651.
93. Raz DJ, Ray MR, Kim JY, et al. A multigene assay is prognostic of survival in patients with early-stage lung adenocarcinoma. *Clin Cancer Res* 2008;14:5565–5570.
94. Shedden K, Taylor JM, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;14:822–827.
95. Vicent S, Luis-Ravelo D, Anton I, et al. A novel lung cancer signature mediates metastatic bone colonization by a dual mechanism. *Cancer Res* 2008;68:2275–2285.
96. Landemaine T, Jackson A, Bellahcene A, et al. A six-gene signature predicting breast cancer lung metastasis. *Cancer Res* 2008;68:6092–6099.
97. Roepman P, Jassem J, Smit EF, et al. An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. *Clin Cancer Res* 2009;15:284–290.
98. Boutros PC, Lau SK, Pintilie M, et al. Prognostic gene signatures for non-small-cell lung cancer. *Proc Natl Acad Sci USA* 2009;106:2824–2828.
99. Broet P, Camilleri-Broet S, Zhang S, et al. Prediction of clinical outcome in multiple lung cancer cohorts by integrative genomics: implications for chemotherapy selection. *Cancer Res* 2009;69:1055–1062.
100. Hsu YC, Yuan S, Chen HY, et al. A four-gene signature from NCI-60 cell line for survival prediction in non-small cell lung cancer. *Clin Cancer Res*2009;15:7309–7315.
101. Showe MK, Vachani A, Kossenkov AV, et al. Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. *Cancer Res* 2009;69:9202–9210.
102. Tomida S, Takeuchi T, Shimada Y, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *J Clin Oncol* 2009;27:2793–2799.
103. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. *J Comput Biol* 2002;9:505–511.
104. Bach PB, Kelley MJ, Tate RC, et al. Screening for lung cancer: a review of the current literature. *Chest* 2003;123(Suppl 1):72S–82S.
105. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci USA* 2006;103:5923–5928.
106. Guo NL, Wan YW, Tosun K, et al. Confirmation of gene expression-based prediction of survival in non-small cell lung cancer. *Clin Cancer Res* 2008;14:8213–8220.
107. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100: 57–70.
108. Potti A, Dressman HK, Bild A, et al. Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 2006;12:1294–1300.
109. Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006;439:353–357.
110. Lamb J, Ramaswamy S, Ford HL, et al. A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell* 2003;114:323–334.
111. Huang ES, Black EP, Dressman H, et al. Gene expression phenotypes of oncogenic signaling pathways. *Cell Cycle* 2003;2:415–417.
112. Black EP, Huang E, Dressman H, et al. Distinct gene expression phenotypes of cells lacking Rb and Rb family members. *Cancer Res* 2003;63:3716–3723.
113. Segal E, Friedman N, Koller D, et al. A module map showing conditional activity of expression modules in cancer. *Nat Genet* 2004; 36:1090–1098.
114. Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci USA* 2004;101: 9309–9314.
115. McMurray HR, Sampson ER, Compitello G, et al. Synergistic response to oncogenic mutations defines gene class critical to cancer phenotype. *Nature* 2008;453:1112–1116.
116. Ding L, Getz G, Wheeler DA, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;455:1069–1075.
117. Lee S, Kang J, Cho M, et al. Profiling of transcripts and proteins modulated by K-ras oncogene in the lung tissues of K-ras transgenic mice by omics approaches. *Int J Oncol* 2009;34:161–172.
118. Sweet-Cordero A, Mukherjee S, Subramanian A, et al. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* 2005;37:48–55.
119. Ranganathan P, Agrawal A, Bhushan R, et al. Expression profiling of genes regulated by TGF-beta: differential regulation in normal and tumour cells. *BMC Genomics* 2007;8:98. Available at: http://www.biomedcentral.com/1471–2164/8/98.
120. Jeon HS, Dracheva T, Yang SH, et al. SMAD6 contributes to patient survival in non-small cell lung cancer and its knockdown reestablishes TGF-beta homeostasis in lung cancer cells. *Cancer Res* 2008;68:9686–9692.
121. Jeon HS, Jen J. TGF-beta signaling and the role of inhibitory Smads in non-small cell lung cancer. *J Thorac Oncol* 2010;5:417–419.
122. Korpal M, Kang Y. Targeting the transforming growth factor-beta signalling pathway in metastatic cancer. *Eur J Cancer* 2010;46:1232–1240.
123. Borczuk AC, Kim HK, Yegen HA, et al. Lung adenocarcinoma global profiling identifies type II transforming growth factor-beta receptor as a repressor of invasiveness. *Am J Respir Crit Care Med* 2005;172:729–737.
124. Borczuk AC, Papanikolaou N, Toonkel RL, et al. Lung adenocarcinoma invasion in TGFbetaRII-deficient cells is mediated by CCL5/RANTES. *Oncogene* 2008;27:557–564.
125. Szakacs G, Annereau JP, Lababidi S, et al. Predicting drug sensitivity and resistance: profiling ABC transporter genes in cancer cells. *Cancer Cell* 2004;6:129–137.
126. Oshita F, Ikehara M, Sekiyama A, et al. Genomic-wide cDNA microarray screening to correlate gene expression profile with chemoresistance in patients with advanced lung cancer. *J Exp Ther Oncol* 2004;4:155–160.
127. Kakiuchi S, Daigo Y, Ishikawa N, et al. Prediction of sensitivity of advanced non-small cell lung cancers to gefitinib (Iressa, ZD1839). *Hum Mol Genet* 2004;13:3029–3043.
128. Yauch RL, Januario T, Eberhard DA, et al. Epithelial versus mesenchymal phenotype determines in vitro sensitivity and predicts clinical activity of erlotinib in lung cancer patients. *Clin Cancer Res* 2005;11: 8686–8698.
129. Gemma A, Li C, Sugiyama Y, et al. Anticancer drug clustering in lung cancer based on gene expression profiles and sensitivity database. *BMC Cancer* 2006;6:174. Available at: http://www.biomedcentral.com/1471–2407/6/174.
130. Coldren CD, Helfrich BA, Witta SE, et al. Baseline gene expression predicts sensitivity to gefitinib in non-small cell lung cancer cell lines. *Mol Cancer Res* 2006;4:521–528.
131. Gyorffy B, Surowiak P, Kiesslich O, et al. Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. *Int J Cancer* 2006;118:1699–1712.
132. Balko JM, Potti A, Saunders C, et al. Gene expression patterns that predict sensitivity to epidermal growth factor receptor tyrosine kinase inhibitors in lung cancer cell lines and human lung tumors. *BMC*

Genomics 2006;7:289. Available at: http://www.biomedcentral.com/1471–2164/7/289.
133. McDermott U, Sharma SV, Dowell L, et al. Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc Natl Acad Sci USA* 2007; 104:19936–19941.
134. Hsu DS, Balakumaran BS, Acharya CR, et al. Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *J Clin Oncol* 2007;25:4350–4357.
135. Miyanaga A, Gemma A, Noro R, et al. Antitumor activity of histone deacetylase inhibitors in non-small cell lung cancer cells: development of a molecular predictive model. *Mol Cancer Ther* 2008;7:1923–1930.
136. Minna JD, Girard L, Xie Y. Tumor mRNA expression profiles predict responses to chemotherapy. *J Clin Oncol* 2007;25:4329–4336.
137. Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24:236–244.
138. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;6:813–823.
139. Ercan D, Zejnullahu K, Yonesaka K, et al. Amplification of EGFR T790M causes resistance to an irreversible EGFR inhibitor. *Oncogene* 2010;29:2346–2356.
140. Rosell R, Felip E, Paz-Ares L. How could pharmacogenomics help improve patient survival? *Lung Cancer* 2007;57(Suppl 2):S35–S41.
141. Engelman JA, Zejnullahu K, Mitsudomi T, et al. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* 2007;316:1039–1043.
142. Press MF, Lenz HJ. EGFR, HER2 and VEGF pathways: validated targets for cancer treatment. *Drugs* 2007;67:2045–2075.
143. Cobo M, Isla D, Massuti B, et al. Customizing cisplatin based on quantitative excision repair cross-complementing 1 mRNA expression: a phase III trial in non-small-cell lung cancer. *J Clin Oncol* 2007;25: 2747–2754.
144. Engelman JA, Janne PA. Mechanisms of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer. *Clin Cancer Res* 2008;14:2895–2899.
145. Felip E, Rojo F, Reck M, et al. A phase II pharmacodynamic study of erlotinib in patients with advanced non-small cell lung cancer previously treated with platinum-based chemotherapy. *Clin Cancer Res* 2008;14:3867–3874.
146. Sharma SV, Settleman J. ErbBs in lung cancer. *Exp Cell Res* 2009; 315:557–571.
147. Naxerova K, Bult CJ, Peaston A, et al. Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers. *Genome Biol* 2008;9:R108. Available at: http://genomebiology.com/2008/9/7/R108.
148. Hassan KA, Chen G, Kalemkerian GP, et al. An embryonic stem cell-like signature identifies poorly differentiated lung adenocarcinoma but not squamous cell carcinoma. *Clin Cancer Res* 2009;15:6386–6390.
149. Stevenson M, Mostertz W, Acharya C, et al. Characterizing the Clinical Relevance of an Embryonic Stem Cell Phenotype in Lung Adenocarcinoma. *Clin Cancer Res* 2009;15:7553–7561.
150. Borczuk AC, Gorenstein L, Walter KL, et al. Non-small-cell lung cancer molecular signatures recapitulate lung developmental pathways. *Am J Pathol* 2003;163:1949–1960.
151. Shen J, Behrens C, Wistuba, II, et al. Identification and validation of differences in protein levels in normal, premalignant, and malignant lung cells and tissues using high-throughput Western Array and immunohistochemistry. *Cancer Res* 2006;66:11194–11206.
152. Pan KH, Lih CJ, Cohen SN. Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. *Proc Natl Acad Sci USA* 2005;102:8961–8965.
153. Krafft AE, Duncan BW, Bijwaard KE, et al. Optimization of the isolation and amplification of rna from formalin-fixed, paraffin-embedded tissue: the Armed Forces Institute of Pathology experience and literature review. *Mol Diagn* 1997;2:217–230.
154. Finke J, Fritzen R, Ternes P, et al. An improved strategy and a useful housekeeping gene for RNA analysis from formalin-fixed, paraffin-embedded tissues by PCR. *Biotechniques* 1993;14:448–453.
155. Park YN, Abe K, Li H, et al. Detection of hepatitis C virus RNA using ligation-dependent polymerase chain reaction in formalin-fixed, paraffin-embedded liver tissues. *Am J Pathol* 1996;149:1485–1491.
156. Jackson DP, Quirke P, Lewis F, et al. Detection of measles virus RNA in paraffin-embedded tissue. *Lancet* 1989;1:1391.
157. Stanta G, Schneider C. RNA extracted from paraffin-embedded human tissues is amenable to analysis by PCR amplification. *Biotechniques* 1991;11:304, 306, 308.
158. Linton KM, Hey Y, Saunders E, et al. Acquisition of biologically relevant gene expression data by Affymetrix microarray analysis of archival formalin-fixed paraffin-embedded tumours. *Br J Cancer* 2008; 98:1403–1414.
159. Specht K, Richter T, Muller U, et al. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. *Am J Pathol* 2001;158:419–429.
160. Li J, Smyth P, Flavin R, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC Biotechnol* 2007;7:36.
161. Rhodes DR, Kalyana-Sundaram S, Tomlins SA, et al. Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* 2007;9:443–454.
162. Lamb J, Crawford ED, Peck D, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;313:1929–1935.
163. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102:15545–15550.
164. Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007;450:893–898.
165. Esteller M. Epigenetics in cancer. *N Engl J Med* 2008;358:1148–1159.
166. Tsou JA, Galler JS, Siegmund KD, et al. Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma. *Mol Cancer* 2007;6:70.
167. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008;40:616–622.
168. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008;452:633–763.
169. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008;452:638–642.
170. Heist RS, Zhai R, Liu G, et al. VEGF polymorphisms and survival in early-stage non-small-cell lung cancer. *J Clin Oncol* 2008;26: 856–862.
171. Gregorc V, Hidalgo M, Spreafico A, et al. Germline polymorphisms in EGFR and survival in patients with lung cancer receiving gefitinib. *Clin Pharmacol Ther* 2008;83:477–484.
172. Liu G, Gurubhagavatula S, Zhou W, et al. Epidermal growth factor receptor polymorphisms and clinical outcomes in non-small-cell lung cancer patients treated with gefitinib. *Pharmacogenomics J* 2008;8: 129–138.
173. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 2007; 448:561–566.
174. Lynch TJ Jr, Blumenschein GR Jr, Engelman JA, et al. Summary statement novel agents in the treatment of lung cancer: Fifth Cambridge Conference assessing opportunities for combination therapy. *J Thorac Oncol* 2008;3:S107–S112.
175. Anguiano A, Nevins JR, Potti A. Toward the individualization of lung cancer therapy. *Cancer* 2008;113(Suppl 7):1760–1767.
176. Canales RD, Luo Y, Willey JC, et al. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* 2006;24:1115–1122.
177. Lynch TJ, Bonomi PD, Butts C, et al. Novel agents in the treatment of lung cancer: Fourth Cambridge Conference. *Clin Cancer Res* 2007;13: s4583–s4588.
178. Simon G, Sharma A, Li X, et al. Feasibility and efficacy of molecular analysis-directed individualized therapy in advanced non-small-cell lung cancer. *J Clin Oncol* 2007;25:2741–2746.
179. Herbst RS, Lippman SM. Molecular signatures of lung cancer-toward personalized therapy. *N Engl J Med* 2007;356:76–78.

