

Molecular Signatures in Biopsy Specimens of Lung Cancer

Alain C. Borczuk, Lori Shah, Gregory D. N. Pearson, Kristin L. Walter, Liqun Wang, John H. M. Austin, Richard A. Friedman, and Charles A. Powell

Department of Pathology, Department of Medicine, Department of Radiology, Department of Biomedical Informatics, and Herbert Irving Comprehensive Cancer Center, Columbia University College of Physicians and Surgeons, New York, New York

Gene expression profiles of resected tumors may predict treatment response and outcome. We hypothesized that profiles derived from lung tumor biopsies would discriminate tumor-specific gene signatures and provide predictive information about outcome. Lung carcinoma specimens were obtained from 23 patients undergoing computed tomography-guided transthoracic biopsy or endobronchial brushing for undiagnosed nodules. Excess tissue was processed for gene profiling. We built class prediction models for lung cancer histology and for cancer outcome. The histology model used an F test to identify 99 genes that were differentially expressed among lung cancer subtypes. The histology validation set class prediction accuracy rate was 86%. The outcome model used the maximum difference subset algorithm to identify 42 genes associated with high risk for cancer death. The outcome training set class prediction accuracy rate was 87%. In conclusion, gene expression profiles of biopsy specimens of lung cancers identify unique tumoral signatures that provide information about tissue morphology and prognosis. The use of specimens acquired from lung biopsy procedures to identify biomarkers of clinical outcome may have application in the management of patients with lung cancer. The procedures are safe and feasible; the efficacy and utility of this strategy will ultimately be determined by prospective clinical trials.

Keywords: lung neoplasms; microarray analysis of gene expression; prognosis

Lung cancer is the leading cause of cancer death in the United States, with 187,000 cases and 165,000 deaths expected in 2004 (1). Despite innovations in diagnostic testing, surgical technique, and the development of new therapeutic agents, the five-year survival rate has remained about 13–15% throughout the past three decades. Factors contributing to the low lung cancer survival rate include the small proportion of patients presenting with resectable disease and chemotherapy response rates ranging from 13–42% in patients with advanced stage disease (2, 3). However, even for patients with resected Stage I lung carcinoma, up to 30% will succumb to their disease within five years. Research has been directed toward the identification of patients at high risk for death after resection or chemotherapy; these individuals could be candidates for adjuvant therapy or alternative management strategies. Other than clinical stage, there are no established cancer-specific clinical variables or biomarkers that reliably iden-

tify individuals at increased risk for death after either surgical resection for early-stage non-small cell carcinomas or chemotherapy and/or radiation therapy for advanced stage carcinomas.

Studies indicate that gene expression profiles of resected tumors can provide insights into lung carcinogenesis (4–6) and may predict risk for recurrence and death in early-stage lung carcinomas treated by surgical resection (7, 8). These studies suggest that prognostic information provided by molecular profiling of resected lung tumors may be useful in guiding adjuvant therapy or postresection surveillance strategies. However, because approximately only 20% of patients with lung cancer undergo surgical resection with curative intent (9), the applicability of this strategy may be limited. In contrast, biopsy specimens obtained by computed tomography (CT)-guided approaches or by fiberoptic bronchoscopy are available from patients with both resectable and unresectable disease (10). Therefore, approaches to examine gene expression profiles from lung cancer biopsies may identify clinically relevant signatures that offer the potential to be widely applicable to the management of patients with lung cancer.

We hypothesized that gene expression profiles derived from biopsies of lung tumors could discriminate tumor-specific gene signatures and provide predictive information about clinical outcome. Similar to other invasive diagnostic procedures, lung biopsies are safe but may be associated with complications that are infrequently medically significant. For transthoracic needle biopsies, pneumothorax rates are about 20%, and the incidence of hemoptysis varies from 5 to 15% (11–13). In addition, fiberoptic bronchoscopy, endobronchial biopsy, and brushing complications are uncommon (10). To eliminate the risk of complications from obtaining biopsy specimens specifically for gene profiling analysis, we utilized residual material obtained from diagnostic lung cancer biopsies. Thus, no additional biopsies were performed specifically for these studies. We show that biopsy molecular signatures identify genes associated with tumor histology and we show that a classifier set of 42 genes can predict risk for lung cancer death. Some of the results of these studies have been reported previously in the form of an abstract (14).

METHODS

Subjects were recruited from a consecutive series of patients referred for transthoracic needle biopsy or bronchoscopy of an undiagnosed lung nodule or mass. An additional inclusion criterion was the diagnosis of a primary lung carcinoma. Tissue specimens were obtained from 26 patients undergoing CT-guided biopsy ($n = 23$) (Temno coaxial core biopsy system; Allegiance/Cardinal Health, McGaw Park, IL) or endobronchial brushing ($n = 3$) (Cellebriy endoscopic cytology brush; Microvasive/Boston Scientific, Watertown, MA) of undiagnosed pulmonary nodules. After needle biopsy and brushing specimens were collected for pathologic diagnosis, the needle or brush containing cells that would otherwise have been discarded was placed into 1 ml of RNA extraction buffer (RNeasy minikit; Qiagen, Valencia, CA). cRNA was generated by the modified Eberwine protocol (http://www.affymetrix.com/support/technical/technotes/smallv2_technote.pdf) (15). Compared with the standard amplification protocol,

(Received in original form January 15, 2004; accepted in final form April 12, 2004)

Supported by the National Institutes of Health (ES00354), the American Cancer Society (CRTG00058), and the Herbert and Florence Irving Scholar Fund.

Correspondence and requests for reprints should be addressed to Charles A. Powell, M.D., Department of Pulmonary, Allergy, and Critical Care Medicine, Columbia University College of Physicians and Surgeons, 630 W. 168th Street, Box 91, New York, NY 10032. E-mail: cap6@columbia.edu

This article has an online supplement, which is accessible from this issue's table of contents online at www.atsjournals.org

Am J Respir Crit Care Med Vol 170, pp 167–174, 2004

Originally Published in Press as DOI: 10.1164/rccm.200401-066OC on April 15, 2004

Internet address: www.atsjournals.org

the modified Eberwine procedure incorporates a second cycle of reverse transcription and a second cycle of *in vitro* transcription.

Biotinylated cRNA was hybridized to the Affymetrix (Santa Clara, CA) U95Av2 DNA array, which contains probes for about 12,600 human genes. Probe-level analysis and normalization to nonmalignant lung tissue were performed according to the robust multiarray algorithm (16) (Gene-Traffic; Iobion, La Jolla, CA). Affymetrix Microarray Suite 5.0 was used to designate each gene as present, absent, or marginal. We excluded from further analysis three arrays of poor quality as demonstrated by fewer than 35% of genes detected as present. Genes were filtered to remove those not present in at least two specimens and genes whose mean log ratio range was less than one. After filtering, 2,194 genes in 23 specimens were used for subsequent analyses. Analyses were performed with BRB ArrayTools (version 3.01; R. Simon and A. P. Lam, National Cancer Institute, Bethesda, MD) (17, 18) and with the Maximum Difference Subset (MDSS) algorithm (<http://bioinformatics.upmc.edu/GE2/GEDA.html>) (19) (microarray data available online at <http://hora.cpmc.columbia.edu/dept/pulmonary/5ResearchPages/Laboratories/Powell%20Lab.htm>).

It was not possible to perform cytologic analysis of specimens used for gene profiling because the residual specimens for research were immediately placed into lysis buffer. We examined the cellularity of four additional specimens acquired from transthoracic needle biopsies; these were collected by standard procedures but were not processed for gene expression analysis. We determined that 1,500–2,000 cells were present in residual specimens obtained from biopsy needles. Cells in the residual specimens were similar in morphology to the tumor cells in paraffin-embedded core biopsy tissues (see Figure E1 in the online supplement). RNA was not specifically quantitated. On the basis of cell counts and cRNA yields during processing for expression analysis, we estimate that needle biopsy specimens contained about 20–50 ng of total RNA. RNA yields from residual material on bronchoscopy brushings ranged from 500 to 600 ng.

Biopsy histologic diagnosis was acquired from the medical record. Permanent sections were reviewed by a second pathologist, who concurred with the original diagnosis in each instance. The histology was classified according to the World Health Organization lung tumor classification scheme for small cell and non-small cell carcinoma (20). In biopsy and brushing specimens, a diagnosis of adenocarcinoma or squamous cell carcinoma was rendered when there were features associated with differentiation (e.g., gland formation or mucin droplets for adenocarcinoma; keratin or intercellular bridges for squamous carcinoma).

If the carcinoma was poorly differentiated, a designation of “non-small cell carcinoma” was assigned. Clinical information for the subjects was obtained from the medical record and from patients’ physicians (Table 1). All procedures were approved by the Columbia University Medical Center (New York, NY) Institutional Review Board and informed consent was obtained from participants.

For validation of the histology class prediction model, an independent set of 29 lung carcinoma resection specimens was microdissected and processed for microarray analysis according to standard protocols, as reported previously (6). For validation of the outcome class prediction model, gene expression and clinical data from a Massachusetts-based independent cohort of 109 patients with lung adenocarcinoma were accessed from <http://www-genome.wi.mit.edu/mpr/lung/>. Hu95Av2 CEL files from Massachusetts-based Dataset A(7) were imported into Gene-Traffic and processed as described above. For the Mantel–Haenszel test for survivorship data (log rank test) (21), specimens were classified as high expression or low expression on the basis of gene expression relative to the median across all specimens. Statistical analyses of survival (22) were performed with SPSS 11.0 (SPSS, Chicago, IL).

The following data sets were used for analysis: histology training set (n = 19 biopsies of adenocarcinoma and squamous, and small cell carcinoma), histology validation set (n = 29 microdissected primary lung carcinoma specimens), outcome training set (n = 23 biopsies), and outcome validation set (n = 109 patients with lung adenocarcinoma from the Massachusetts-based cohort).

Immunohistochemistry

Immunohistochemical staining was performed with antibodies for cyclin B1 (clone GN5a; Neomarkers, Fremont, CA) and FHL2 (Santa Cruz Biotechnology, Santa Cruz, CA). Formalin-fixed, paraffin-embedded biopsy tissue blocks were sectioned at a thickness of 5 μ m, dewaxed in xylene, rehydrated through a graded ethanol series, and washed with phosphate-buffered saline. For FHL2, antigen retrieval was achieved by heat treatment in a steamer for 40 minutes in citrate buffer (pH 6.0, 10 mmol/L); secondary antibody was rabbit anti-goat diluted 1:200 (Vector Laboratories, Burlingame, CA). For cyclin B1, antigen retrieval was achieved with Protease XXV (Neomarkers) at 1 mg/ml for 10 minutes at 37°C; secondary antibody was horse anti-mouse diluted 1:200 (Vector Laboratories). Before staining the sections, endogenous peroxidase was quenched; for both antibodies, primary antibody incubation was for 1 hour at 37°C (FHL2, diluted 1:100; cyclin B1, diluted 1:50).

TABLE 1. PATIENT CHARACTERISTICS

Sample	Age (yr)	Sex	Pathology	Source	Tumor Size (cm)	Stage	Cancer Death	Follow-Up (d)
1	62	M	Adenocarcinoma	ttn	5.1	IV	No	432
2*	88	M	Adenocarcinoma	ttn	4	IB	No	502
3	63	M	Adenocarcinoma	ttn	2.6	IIIA	No	379
4	67	F	Adenocarcinoma	ttn	4.3	IV	No	389
5	80	F	Adenocarcinoma	ttn	2.5	IB	No	108
6	70	F	Adenocarcinoma	ttn	2.5	IV	No	230
7	61	F	Squamous	Brush	2.9	IA	No	248
8	77	F	Squamous	ttn	2.4	IIIA	No	341
9	56	M	Squamous	ttn	9.3	IIIA	No	59
10	56	M	Squamous	ttn	6.7	IIIA	No	281
11	69	M	Squamous	ttn	4.5	Ila	No	328
12	55	F	Non-small cell	ttn	10.5	IIB	Yes	102
13	66	M	Squamous	Brush	4.5	IIIA	Yes	259
14	65	F	Adenocarcinoma	ttn	1.2	IIIA	No	437
15	89	M	Non-small cell	ttn	10	IV	Yes	54
16*	77	M	Adenocarcinoma	ttn	2.6	IB	No	355
17	85	F	Adenocarcinoma	ttn	3.8	IV	Yes	442
18	72	M	Squamous	ttn	5.2	IIA	Yes	58
19	64	M	Non-small cell	ttn	4.8	IV	Yes	265
20	40	F	Non-small cell	Brush	2.5	IIIB	No	270
21	55	M	Adenocarcinoma	ttn	8.1	IV	No	275
22	74	M	Small cell	ttn	8	E	No	400
23	72	F	Small cell	ttn	3.7	E	Yes	346

Definition of abbreviations: brush = bronchoscopy brushing; E = extensive stage; ttn = transthoracic needle biopsy.

* Resected tumor available for gene expression analysis.

RESULTS

Biopsy specimens were adequate for gene expression profiling analysis in 23 of 26 cases. Because our procedures utilized residual material from clinically indicated biopsies, there were no patient complications attributable to the research procedures. A limitation of gene expression profiling of small specimens obtained in this manner is that the number of cells captured does not provide an adequate quantity of total RNA for analysis on Affymetrix oligonucleotide arrays, using standard amplification protocols. We therefore instituted the modified Eberwine procedure, which is an established modification designed to uniformly amplify RNA obtained from small samples for analysis on microarrays.

We examined two potential sources of variability in gene profiling of small specimens obtained from diagnostic biopsies: nucleic acid amplification and cellular heterogeneity. To examine the variability introduced by the additional round of amplification in the modified Eberwine procedure, we compared gene expression data of tumor RNA (2 μ g) processed by standard procedures with that of diluted tumor RNA (200 ng) from the same specimen, but processed by the modified Eberwine protocol. Examination of scatter plots and correlation coefficients shows that gene signal intensities were highly similar between the two methods of amplification, as has been shown by other researchers (23–25) (Figure 1A).

To examine variability introduced by the admixture of cells present in the diagnostic specimens, we compared gene expression data of biopsy material with that of diluted microdissected tumor RNA from the same patient. The results indicate that the

gene expression intensities are similar, but that there is more heterogeneity than in the comparison of amplification protocols (Figure 1B). Because both specimens were processed according to the modified Eberwine procedure, the variability was likely attributable to the presence of cellular heterogeneity in biopsy specimens. Compared with microdissected resected tumors, which contain more than 90% tumor cells, the biopsy specimens often contain cells from normal lung, pleura, muscle, and skin; inflammatory cells; and blood leukocytes in addition to tumor cells. Despite this heterogeneity, we hypothesized that unique tumor-specific molecular signatures (i.e., histology classifiers) could be detected in these specimens.

Histology

Previous work demonstrates that lung tumor histologic subtypes can be distinguished by gene expression profiles (6, 7). To determine whether gene expression profiles of lung biopsies could identify specific tumor signatures, we performed class comparison, using an F test (26) within BRB ArrayTools to identify 99 genes that were differentially expressed among the histologic classes ($p < 0.01$; Table 2). To address the problem of multiple comparisons in statistical testing, class labels were randomly permuted 1,000 times and a permutation p value less than 0.01 was associated with each gene in the list. The probability of obtaining at least 99 genes significant by chance (at the 0.01 level), if there were no real differences between the classes, was 0.024. We excluded four lung carcinoma biopsies subtyped as “non–small cell” from the histology training set cross-validation analysis. The designation of “non–small cell” encompasses multiple histologic subtypes and is not a World Health Organization category for histologic classification of resected tumors.

Among the lung histology classifier genes detected in the biopsy specimens, several have been identified in other studies that used the U95A microarray platform. These marker genes include *ERBB2*, *TTF-1*, *MUC1*, *BENE*, *SELENBP1*, *TGFBR2* (adenocarcinoma); *KIF5C*, *TMSNB*, *TUBB*, *FOXG1B*, *ESPL1*, *TRIM28* (small cell carcinoma); and *KRT17*, *KRT6E*, *BPAG1* (squamous cell carcinoma) (6, 7, 27). To further examine the association of the classifiers with lung cancer histology, we performed class prediction testing with a k -nearest neighbor (28) leave-one-out cross-validation. In this procedure, one sample is removed from the training set, a new gene set is generated from which a classifier is generated, and this classifier is applied to the sample left out. This procedure is repeated for all the samples. Three nearest neighbor classifiers generated in this manner correctly predicted the histologic class for 13 (68%) of 19 samples. A permutation analysis of the predictor was performed. On the basis of 1,000 random permutations, the classifier had a p value of 0.035, indicating that the misclassification rate of the predictor was significantly smaller than the misclassification rate of the permutations.

We tested the accuracy of the biopsy histology classifier model by using it to predict the histology of 29 independently obtained lung carcinoma resection specimens (histology validation set). The distribution of the histology validation set was adenocarcinoma ($n = 22$); small cell ($n = 2$); and squamous cell carcinoma ($n = 5$). The 99-gene histology classifier model was able to accurately predict histology in 25 (86%) of 29 tumors (Table 3). Four of the adenocarcinoma tumors were incorrectly classified as squamous cell carcinomas. Interestingly, histologic sections of these tumors showed areas of squamous differentiation within a predominantly glandular tumor, and in a previous study three of these adenocarcinomas segregated with squamous cell carcinomas in an unsupervised clustering procedure (6). Therefore, histologic heterogeneity may have accounted for misclassification by histology classifier genes in these tumors. The results of

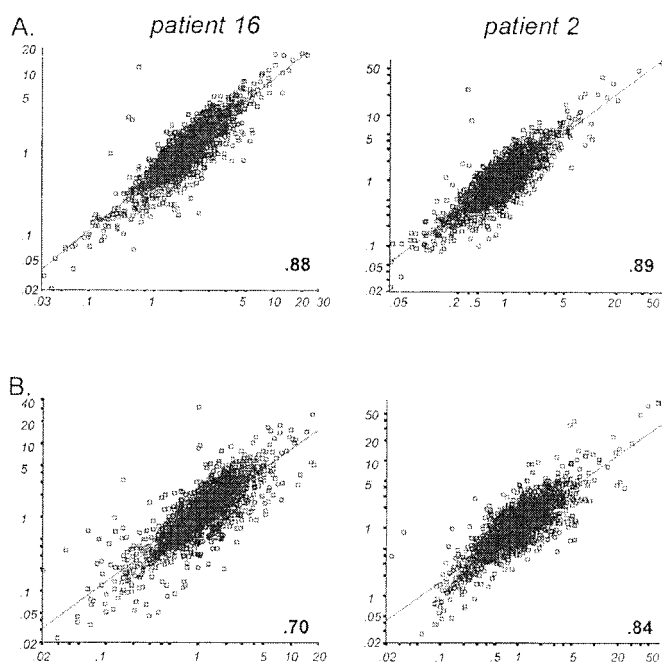


Figure 1. Scatter plots indicating log gene expression ratios, comparing amplification protocols and comparing biopsy with resected tumor. (A) Comparison of targets processed by standard protocol (*horizontal axis*) and by the modified Eberwine protocol (*vertical axis*). Total RNA was obtained from two microdissected resected tumors and was diluted 1:10 for processing by the modified Eberwine procedure. (B) Comparison of targets from microdissected resected tumor (*horizontal axis*) with paired biopsy specimen (*vertical axis*). The Pearson correlation coefficient for each experiment is indicated in **boldface**; $p < 0.05$ in each instance.

TABLE 2. HISTOLOGY CLASSIFIERS OF BIOPSY SPECIMENS IDENTIFIED BY F TEST

Adenocarcinoma		Small Cell		Squamous Cell	
Affymetrix ID	Symbol	Affymetrix ID	Symbol	Affymetrix ID	Symbol
33325_at	RPS6KA2	36701_at	DKFZp564N1662	34301_r_at	KRT17
37760_at	BAIAP2	37580_at	SH3GL3	41641_at	C4.4A
33218_at	ERBB2	35778_at	KIF5C	39016_r_at	KRT6E
33754_at	TTF-1	38279_at	GNAZ	39015_f_at	KRT6E
927_s_at	MUC1	41388_at	MEIS2	40304_at	BPAG1
1368_at	IL1R1	39642_at	ELOVL2	35820_at	SAP-3
36528_at	ASL	36815_at	AF038185	38356_at	FST
634_at	PRSS8	37530_s_at	RELN	1898_at	TRIM29
38028_at	DAT1	36491_at	TMSNB	40518_at	PTPRC
37639_at	HPN	36029_at	C11orf8		
38342_at	PHF15	36941_at	AF1Q		
33331_at	BENE	38146_at	KIAA0535		
37405_at	SELENBP1	41356_at	BCL11A		
41177_at	FLJ12443	33637_g_at	NY-ESO-1		
38095_i_at	HLA-DPB1	39387_at	SEPHS1		
39698_at	HOP	39605_att	FOXGIB		
37754_at	LGALS3BP	1787_at	CDKNIC		
943_at	RUNX1	36200_at	BAT8		
38047_at	BPMS	38163_at	RIMS2		
33327_at	C11 orf9	40041_at	HEC		
32249_at	HFL1	34417_at	FLJ36166		
988_at	CEACAM1	39590_at	APBA2		
36076_g_at	RABL4	1373_at	TCF3		
37001_at	CAPN2	35226_at	EYA2		
35276_at	CLDN4	39332_at	TUBB		
40504_at	PON2	38634_at	RBP1		
38783_at	MUC1	1490_at	L-myc		
40848_g_at	MICAL2	1713_s_at	CDKN2A		
34235_at	GPR116	41199_s_at	SFPQ		
41176_at	FLJ12443	38933_at	KIFC1		
39345_at	NpC2	36761_at	ZNF339		
40928_at	WSB1	38158_at	ESPL1		
34876_at	CPD	33425_at	TRIM28		
33774_at	CASP8	543_g_at	CRABP1		
40297_at	STEAP	41342_at	RANBP1		
1815_g_at	TGFBR2	1782_s_at	STMN1		
1915_s_at	FOS	2054_g_at	NCAD		
35341_at	TRIM38	39324_at	FLJ12377		
37430_at	ALOX15B	37985_at	LMNB1		
		41084_at	MGC51082		
		37302_at	CENPF		
		35312_at	MCM2		
		33157_at	INSM1		
		39980_at	VRK1		
		36990_at	UCHL1		
		39710_at	P311		
		1544_at	BLM		
		41355_at	BCL11A		
		1909_at	BCL2		
		37210_at	INA		
		39677_at	KIAA0186		

histology training and validation set class prediction analyses indicate that gene expression profiles of lung biopsies were representative of histologically specific subtypes of lung carcinoma.

Prognosis

We examined whether biopsy gene expression signatures could predict another clinically relevant end point, prognosis. Among the 23 patients who underwent lung biopsy, 6 cancer deaths occurred within 12 months. These patients were classified as high risk for early cancer death. We identified genes associated with high-risk and low-risk outcome, using the Maximum Difference Subset (MDSS) algorithm. This tool combines standard statistical tests (pooled variance *t* test) and machine prediction learning to identify class predictors with higher specificity and

accuracy compared with other classification algorithms (19). In the biopsy data set, MDSS identified 42 genes associated with cancer death within 12 months (Table 4). We tested the accuracy of these predictors to classify risk for cancer death. The overall outcome training set class prediction accuracy rate was 87% (20 of 23 predicted correctly), with a *p* value of 0.008 based on 1,000 random permutations of the class labels.

To determine whether the outcome classifiers identified in expression profiling of lung cancer biopsies were applicable to other lung cancer gene expression data sets, we examined whether our genes were associated with cancer-free survival in an independent set of homogenized tumors resected from a large cohort of Massachusetts-based patients with lung adenocarcinoma (outcome validation set) (7). We determined that 9 of the

TABLE 3. PREDICTION OF RESECTED TUMOR HISTOLOGY

Specimen	Histology	Prediction
AD20009	AD	SQ
AD20014	AD	AD
AD20033	AD	AD
AD21001	AD	AD
AD21002	AD	AD
AD21006	AD	SQ
AD21011	AD	AD
AD21012	AD	AD
AD21013	AD	AD
AD21014	AD	AD
AD22003	AD	AD
AD22005	AD	SQ
AD22009	AD	SQ
AD22010	AD	AD
AD22037	AD	AD
AD22048	AD	AD
AD22051	AD	AD
AD23005	AD	AD
AD99015	AD	AD
AD99034	AD	AD
AD99035	AD	AD
AD99043	AD	AD
SM21015	SM	SM
SM22060	SM	SM
SQ22002	SQ	SQ
SQ22004	SQ	SQ
SQ22016	SQ	SQ
SQ99011	SQ	SQ
SQ99014	SQ	SQ

Definition of abbreviations: AD = adenocarcinoma; SM = small cell carcinoma; SQ = Squamous cell carcinoma.

42 genes associated with risk for one-year cancer death in our outcome training set were associated with cancer-free survival in the Massachusetts-based outcome validation data set, using the log rank test ($p < 0.05$; Figure 2). These genes were as follows: *CCNB1*, *FHL2*, *HLA-DPBI*, *LOXL2*, *IRS-1*, *PLOD2*, *MTHFD2*, *TGFB1*, and *TRIPBR2*. This result suggests that despite differences in histologic subtypes, specimen types, and amplification protocols, selected outcome genes may be applicable to the prediction of lung carcinoma outcome in other patients.

Immunohistochemistry

Because tumor behavior may be modulated by signals from the tumor and its surrounding microenvironment, we examined immunolocalization of representative outcome marker proteins to determine whether expression was detectable in tumor cells. Antibodies were selected on the basis of commercial availability. Immunoreactivity for both FHL2 (nuclear) and cyclin B1 (cytoplasmic) was detectable in tumor cells, suggesting that biopsy gene expression signatures are derived from tumor cells (Figure 3).

DISCUSSION

Lung cancer biopsy gene expression profiles identify unique tumoral signatures that provide information about tissue morphology and clinical outcome. Using validated methods of gene identification that account for the statistical problems associated with multiple comparisons, the present study identified 42 genes associated with high risk for cancer death within one year. The use of specimens acquired by lung biopsy procedures to identify genes associated with clinical outcome suggests several applications as biomarkers of prognosis or treatment response.

The relevance of the outcome marker genes identified in the biopsy specimens is supported by other studies indicating that

several genes are associated with prognosis in patients with lung carcinoma or other carcinomas. Examples include *MYC*, encoding the nuclear transcription factor c-Myc, which functions in cell growth and proliferation and is frequently amplified in lung carcinoma (29). Increased expression of c-Myc is associated with adverse prognosis in lymphoma and node-negative breast carcinoma (30, 31). *CCNB1* encodes the cell cycle-regulatory protein cyclin B1, which regulates the G₂-M transition. Increased expression of cyclin B1 is associated with poor survival in esophageal carcinoma and in non-small cell lung carcinoma (32, 33). *FHL2* encodes four and a half of LIM-only protein, which is a β -catenin-binding protein with *trans*-activation activity (34). FHL2 expression is increased in hepatoblastoma and is associated with cyclin D1 promoter activation in a β -catenin-dependent fashion. Whereas FHL2 is not directly associated with cancer outcome, cyclin D1 expression is associated with decreased survival in resected lung carcinomas (35). *HLA-DPBI*, which encodes a human MHC Class II lymphocyte antigen β chain, was associated with improved survival in our data set. A similar association was reported in a gene profiling study of diffuse large B cell lymphoma specimens. Lower expression of *HLA-DPBI* and other MHC Class II genes was associated with poor patient survival and decreased tumor immunosurveillance (36).

The 5-year survival rate for lung cancer is about 15%, which is markedly lower than the rates for other common cancers of the breast, colon, and prostate (37). This discrepancy may be due to biological differences such as histologic heterogeneity or to the absence of proven screening programs that effectively detect cancers at an early, curable stage. However, even for surgically resected early Stage I non-small cell lung carcinomas, the recurrence rate is 3–5% annually and the 5-year survival rate is about 70%. Studies suggest that gene expression profiles of early-stage lung adenocarcinomas may predict risk for death (7, 8) and therefore may be useful to identify individuals who would be most likely to benefit from systemic therapy delivered before or after resection. Data from early-stage lung cancer systemic therapy trials indicate that neoadjuvant chemotherapy combined with radiation therapy (38) and adjuvant chemotherapy (39) may provide a survival benefit for a small proportion of patients. The potential role of lung biopsy gene expression profiling in the management of early-stage non-small cell carcinoma would be to identify patients with high-risk tumors who would be most likely to benefit from neoadjuvant systemic therapy. The potential utility of this approach has been demonstrated in breast carcinoma. Gene profiles obtained from breast tumors have been shown to predict a short-term clinical response to neoadjuvant docetaxel (40).

Another potential role for gene profiling of lung cancer biopsies that might be applicable to the large proportion of patients with lung cancer with unresectable tumors is selection of chemotherapy agents. Advanced stage non-small cell carcinomas and small cell carcinomas are treated by systemic chemotherapy. For non-small cell lung carcinomas, the average response rate in previously untreated patients ranges widely, from 13 to 42% (2); yet there are no reliable biomarkers to guide the selection of particular regimens to patients who are most likely to benefit. *In vitro* studies show that the response of lung cancer cells and other cancer cells to single chemotherapy agents can be predicted by distinct gene expression profiles (41, 42). These results suggest that gene profiling may complement decisions regarding the selection of systemic chemotherapeutic agents. This hypothesis is supported by B cell lymphoma clinical trials that identified tumor gene expression predictors of patient survival after chemotherapy treatment (43, 44). Interestingly, adverse prognosis genes were associated with a proliferation functional class

TABLE 4. SURVIVAL CLASSIFIERS

Rank	Accession No.	Gene	Molecular Function
High risk	1. 37724_at	MYC	Regulation of gene transcription
	2. 1495_at	TGFB1	Growth factor binding
	3. 33439_at	SNF1LK	Protein tyrosine kinase
	4. 35977_at	DKK1	Signal transduction
	5. 32065_at	CREM	Signal transduction
	6. 33127_at	LOXL2	Scavenger receptor activity
	7. 39277_at	OSMR	DNA binding
	8. 41049_at	IRS1	Signal transduction
	9. 34795_at	PLOD2	Protein modification
	10. 38422_s_at	FHL2	Oncogenesis
	11. 35291_at	BAG2	Chaperone activity
	12. 36497_at	C14orf78	
	13. 37312_at	TRIP-Br2	
	14. 40074_at	MTHFD2	Oxidoreductase activity
	15. 32066_g_at	CREM	Signal transduction
	16. 32186_at	SLC7A5	Amino acid transport
	17. 34563_at	KIF14	ATP binding
	18. 37474_at	OIPS	Protein binding
	19. 34777_at	ADM	Hormone activity
	20. 31863_at	KIAA0179	
	21. 36873_at	VLDLR	Signal transduction
	22. 547_s_at	NR4A2	Transcription factor activity
	23. 1973_s_at	MYC	Regulation of gene transcription
	24. 41419_at	CED-6	Signal transducer activity
	25. 32067_at	CREM	Signal transduction
	26. 41449_at	SGCE	Cell-matrix adhesion
	27. 1945_at	CCNB1	G ₂ /M transition of mitotic cell cycle
	28. 37623_at	NR4A2	Transcription factor activity
	29. 34721_at	FKBP5	FK506 binding
	30. 33534_at	ESM1	Insulin-like growth factor binding
Low risk	1. 35207_at	SCNN1A	Ion channel activity
	2. 39514_at	GADD45G	DNA repair
	3. 37405_at	SELENBP1	Selenium binding
	4. 33754_at	TTF-1	Transcription factor activity
	5. 1664_at	HG3543-HT3739	
	6. 38095_i_at	HLA-DPB1	Class II major histocompatibility complex
	7. 38754_at	P8	Induction of apoptosis
	8. 33052_at	PLA2G10	Phospholipase A ₂ activity
	9. 39698_at	HOP	Transcription factor activity
	10. 38028_at	DATI	
	11. 41779_at	RGS16	Signal transduction
	12. 37021_at	CTSH	Cathepsin H activity

whereas favorable outcome was associated with MHC Class II function (43). In our lung biopsy data set, proliferation genes (*CCNB1*, *MYC*, *FHL2*, and *NR4A2*) and MHC Class II genes (*HLA-DPB1*) were similarly associated with adverse and favorable outcomes, respectively. Further characterization of the function of these genes in lung carcinogenesis may lead to the development of novel targeted therapies.

Some methodologic limitations apply to our approach. First, our use of residual biopsy specimens did not consistently provide enough cellular material for gene expression analysis according to standard amplification protocols. Rather, we used a modified protocol that incorporated a second round of amplification and therefore increased the opportunity for variability and inconsistency in the data. However, our validation experiments and

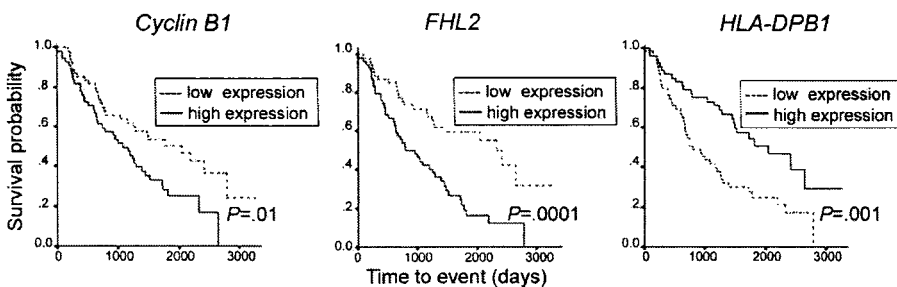


Figure 2. Kaplan–Meier survival plots of representative genes identified in patients with lung adenocarcinoma undergoing lung biopsy as predictors of cancer death within 12 months. Gene expression data for patients with adenocarcinoma were accessed from a data set that was acquired from 109 patients with early-stage resected tumors. For log rank analysis of survival for selected genes, specimens were classified as high expression (n = 55) or low expression (n = 54) on the basis of gene expression relative to the median across all specimens; p < 0.05 in each instance.

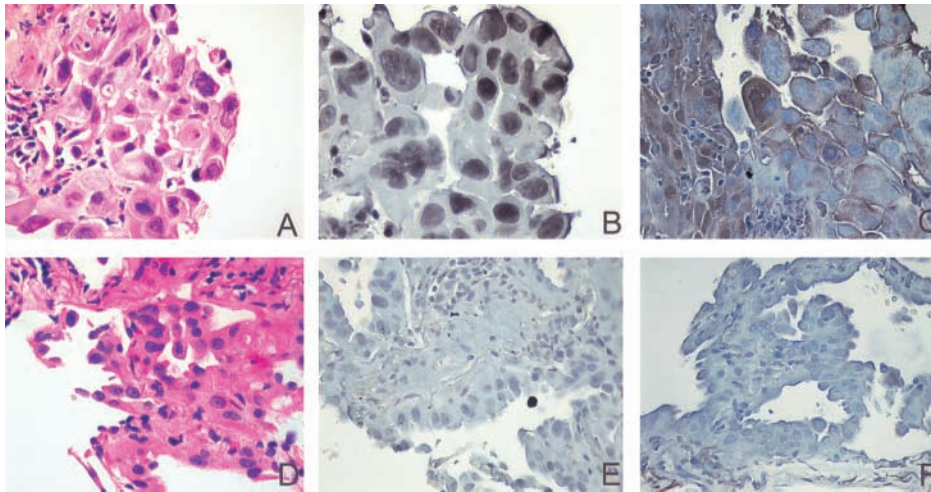


Figure 3. FHL2 and cyclin B1 immunostaining. Two representative biopsy specimens from Patient 13 (A–C) and Patient 6 (D–F) were immunostained with antibody to FHL2 (B and E) and cyclin B1 (C and F). Staining is detectable in tumor cells of Specimen 13 but is absent in Specimen 6; this correlated with gene signal intensity in these specimens. (A and D) Hematoxylin and eosin stain. Original magnification: (A–F) $\times 150$.

those performed by others indicate that experimental variability attributable to amplification procedures is small and that data produced from small specimens are reliable. Our technical adequacy rate was higher than those reported by other studies that examined gene expression profiles of lung and breast biopsies (25, 45). Second, the sample size was relatively small, which may introduce bias and reduce the ability to generalize our results to other lung cancer populations. To address this issue, we examined the ability of the outcome classifier model to predict cancer-free survival in a large independent gene expression data set of lung adenocarcinoma tumors. Despite differences in tumor specimen composition and in experimental protocols, several of our cancer outcome classifier genes were similarly associated with cancer-free survival in Massachusetts-based lung adenocarcinoma cases. Future prospective validation of the gene classifier model in an independent cohort of patients undergoing biopsy will reduce confounding by technical and clinical factors and will confirm the generalizability of the results. Third, because our data set was composed entirely of lung carcinoma biopsies, we could not examine the utility of biopsy gene profiles to distinguish malignant tumors from benign nodules. Experience with screening chest CT indicates a high prevalence of nodules (25–66%) of which only a small fraction (1–3%) are malignant (46). Although nodule size and interval change in size are useful tools to distinguish malignant from benign lesions, it is possible that gene expression profiles of CT-detected nodules may enhance diagnostic algorithms and the clinical utility of the procedure.

Other reports support the potential utility of biopsy gene profiles in the clinical management of breast carcinoma. Compared with breast biopsies, lung biopsy is associated with a higher risk of complications such as bleeding and pneumothorax. We addressed this risk in our study procedures by utilizing residual specimens from clinically indicated diagnostic lung biopsies; thus no medical risk was attributable to procedures utilized for gene expression analysis of lung biopsies. The gene expression signatures generated by the lung biopsies are robust, clinically relevant, and have the potential to improve lung cancer treatment and outcome. The procedures are safe and feasible; we suggest that the efficacy and utility of this strategy are now appropriate for assessment by prospective clinical trials.

Conflict of Interest Statement: A.C.B. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; L.S. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; G.D.N.P. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; K.L.W. does not have a financial relationship with a commercial entity that has

an interest in the subject of this manuscript; L.W. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; J.H.M.A. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; R.A.F. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; C.A.P. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: Analyses were performed with BRB ArrayTools, developed by Dr. Richard Simon and Amy Peng, and with MDSS, developed by Dr. James Lyons-Weiler and Satish Patel. We also thank Vladan Milkovic and Diane Alexis for technical assistance.

References

- Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ, Thun MJ. Cancer statistics, 2004. *CA Cancer J Clin* 2004;54:8–29.
- Waters JS, O'Brien ME. The case for the introduction of new chemotherapy agents in the treatment of advanced non small cell lung cancer in the wake of the findings of the National Institute of Clinical Excellence (NICE). *Br J Cancer* 2002;87:481–490.
- Spiro SG, Porter JC. Lung cancer: where are we today? Current advances in staging and nonsurgical treatment. *Am J Respir Crit Care Med* 2002; 166:1166–1196.
- Powell CA, Spira A, Derti A, DeLisi C, Liu G, Borczuk A, Busch S, Sahasrabudhe S, Chen Y, Sugarbaker D, et al. Gene expression in lung adenocarcinomas of smokers and nonsmokers. *Am J Respir Cell Mol Biol* 2003;29:157–162.
- Sugita M, Geraci M, Gao B, Powell RL, Hirsch FR, Johnson G, Lapadat R, Gabrielson E, Bremnes R, Bunn PA, et al. Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma. *Cancer Res* 2002;62:3971–3979.
- Borcuk AC, Gorenstein L, Walter KL, Assaad AA, Wang L, Powell CA. Non-small-cell lung cancer molecular signatures recapitulate lung developmental pathways. *Am J Pathol* 2003;163:1949–1960.
- Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 2001;98:13790–13795.
- Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002;8:816–824.
- Datta D, Lahiri B. Preoperative evaluation of patients undergoing lung resection surgery. *Chest* 2003;123:2096–2103.
- British Thoracic Society Bronchoscopy Guidelines Committee, a Subcommittee of Standards of Care Committee of British Thoracic Society. British Thoracic Society guidelines on diagnostic flexible bronchoscopy. *Thorax* 2001;56:i1–i21.
- Ernst A, Silvestri GA, Johnstone D. Interventional pulmonary procedures: guidelines from the American College of Chest Physicians. *Chest* 2003; 123:1693–1717.

12. Geraghty PR, Kee ST, McFarlane G, Razavi MK, Sze DY, Dake MD. CT-guided transthoracic needle aspiration biopsy of pulmonary nodules: needle size and pneumothorax rate. *Radiology* 2003;229:475-481.
13. Kazerooni EA, Lim FT, Mikhail A, Martinez FJ. Risk of pneumothorax in CT-guided transthoracic needle aspiration biopsy of the lung. *Radiology* 1996;198:371-375.
14. Walter KL, Borczuk AC, Wang L, Assaad AM, Austin JHM, Pearson GDN, Shiau MC, Powell CA. Class prediction of lung nodule gene expression profiles [abstract]. *Chest* 2004;125:104S.
15. Kacharmina JE, Crino PB, Eberwine J. Preparation of cDNA from single cells and subcellular regions. *Methods Enzymol* 1999;303:3-18.
16. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003;31:e15.
17. Simon R, Radmacher R, Bittner M. BRB tools, version 3.0. Bethesda, MD: National Cancer Institute; 2003.
18. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 2003;95:14-18.
19. Lyons-Weiler J, Patel S, Bhattacharya S. A classification-based machine learning approach for the analysis of genome-wide expression data. *Genome Res* 2003;13:503-512.
20. Travis WD, Colby TV, Corrin B, Shimosato Y, Brambilla E. World Health Organization international histological classification of tumours: histological typing of lung and pleural tumors, 3rd ed. New York: Springer-Verlag; 1999.
21. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 1966;50:163-170.
22. Meier P, Kaplan E. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;158:457-481.
23. Sotiriou C, Powles TJ, Dowsett M, Jazaeri AA, Feldman AL, Assersohn L, Gadisetti C, Libutti SK, Liu ET. Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res* 2002;4:R3.
24. Luzzi V, Mahadevappa M, Raja R, Warrington JA, Watson MA. Accurate and reproducible gene expression profiles from laser capture microdissection, transcript amplification, and high density oligonucleotide microarray analysis. *J Mol Diagn* 2003;5:9-14.
25. Symmans WF, Ayers M, Clark EA, Stec J, Hess KR, Sneige N, Buchholz TA, Krishnamurthy S, Ibrahim NK, Buzdar AU, et al. Total RNA yield and microarray gene expression profiles from fine-needle aspiration biopsy and core-needle biopsy samples of breast carcinoma. *Cancer* 2003;97:2960-2971.
26. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* 2003;19:2448-2455.
27. Pedersen N, Mortensen S, Sorensen SB, Pedersen MW, Rieneck K, Bovin LF, Poulsen HS. Transcriptional gene expression profiling of small cell lung cancer cells. *Cancer Res* 2003;63:1943-1953.
28. Duda RO, Hart PE, Stork DG. Pattern classification, 2nd ed. New York: John Wiley & Sons; 2001.
29. Saksela K, Bergh J, Lehto VP, Nilsson K, Alitalo K. Amplification of the *c-myc* oncogene in a subpopulation of human small cell lung cancer. *Cancer Res* 1985;45:1823-1827.
30. Schlotter CM, Vogt U, Bosse U, Mersch B, Wassmann K. C-myc, not HER-2/neu, can predict recurrence and mortality of patients with node-negative breast cancer. *Breast Cancer Res* 2003;5:R30-R36.
31. Nagy B, Lundan T, Larramendy ML, Aalto Y, Zhu Y, Niini T, Edgren H, Ferrer A, Vilpo J, Elonen E, et al. Abnormal expression of apoptosis-related genes in haematological malignancies: overexpression of MYC is poor prognostic sign in mantle cell lymphoma. *Br J Haematol* 2003;120:434-441.
32. Takeno S, Noguchi T, Kikuchi R, Uchida Y, Yokoyama S, Muller W. Prognostic value of cyclin B1 in patients with esophageal squamous cell carcinoma. *Cancer* 2002;94:2874-2881.
33. Soria JC, Jang SJ, Khuri FR, Hassan K, Liu D, Hong WK, Mao L. Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. *Cancer Res* 2000;60:4000-4004.
34. Wei Y, Renard C-A, Labalette C, Wu Y, Levy L, Neuveut C, Prieur X, Flajolet M, Prigent S, Buendia M-A. Identification of the LIM protein FHL2 as a coactivator of β -catenin. *J Biol Chem* 2003;278:5188-5194.
35. Keum JS, Kong G, Yang SC, Shin DH, Park SS, Lee JH, Lee JD. Cyclin D1 overexpression is an indicator of poor prognosis in resectable non-small cell lung cancer. *Br J Cancer* 1999;81:127-132.
36. Rimsza LM, Roberts RA, Miller TP, Unger JM, LeBlanc M, Brazier RM, Weisenburger DD, Chan WC, Greiner TC, Muller-Hermelink HK, et al. Loss of MHC Class II gene and protein expression in diffuse large B cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival irrespective of other prognostic factors: a follow-up study from the leukemia and lymphoma molecular profiling project. *Blood* 2004;103:4251-4258.
37. Jemal A, Murray T, Samuels A, Ghafoor A, Ward E, Thun MJ. Cancer statistics, 2003. *CA Cancer J Clin* 2003;53:5-26.
38. Pisters KM, Ginsberg RJ, Giroux DJ, Putnam JB Jr, Kris MG, Johnson DH, Roberts JR, Mault J, Crowley JJ, Bunn PA Jr, Bimodality Lung Oncology Team. Induction chemotherapy before surgery for early-stage lung cancer: a novel approach. *J Thorac Cardiovasc Surg* 2000;119:429-439.
39. International Adjuvant Lung Cancer Trial Collaborative Group. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. *N Engl J Med* 2004;350:351-360.
40. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, Mohsin S, Osborne CK, Chamness GC, Allred DC, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003;362:362-369.
41. Staunton JE, Slonim DK, Collier HA, Tamayo P, Angelo MJ, Park J, Scherf U, Lee JK, Reinhold WO, Weinstein JN, et al. Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci USA* 2001;98:10787-10792.
42. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24:236-244.
43. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltman JM, et al. Lymphoma/Leukemia Molecular Profiling Project. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;346:1937-1947.
44. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 2002;8:68-74.
45. Lim EH, Aggarwal A, Agasthian T, Wong PS, Tan C, Sim E, Tan L, Goh PS, Wang SC, Khoo KL, et al. Feasibility of using low-volume tissue samples for gene expression profiling of advanced non-small cell lung cancers. *Clin Cancer Res* 2003;9:5980-5987.
46. Swensen SJ, Jett JR, Sloan JA, Midthun DE, Hartman TE, Sykes AM, Aughenbaugh GL, Zink FE, Hillman SL, Noetzel GR, et al. Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 2002;165:508-513.