

## A Panel of Sputum-Based Genomic Marker for Early Detection of Lung Cancer

Feng Jiang<sup>1</sup>, Nevins W. Todd<sup>2</sup>, Ruiyun Li<sup>3</sup>, Howard Zhang<sup>2</sup>, HongBin Fang<sup>4</sup>, and Sanford A. Stass<sup>1</sup>

### Abstract

Non-small cell lung cancer (NSCLC) is the leading cause of cancer death. Early detection of NSCLC will improve its outcome. We previously identified genetic signatures whose genomic copy number aberrations were associated with early stage NSCLC. Here, we aimed to develop a panel of genes that could be detected in sputum for NSCLC early detection. We first optimized a panel of genes by using an *in situ* minichip for measuring changes of the signatures in sputum of a case-control cohort of 49 NSCLC patients, 49 patients with chronic obstructive pulmonary disease (COPD), and 49 healthy smokers. We then validated the genes in an independent cohort of 69 NSCLC patients and 65 noncancer subjects. The results were compared with those of sputum cytology. Fifteen genes showed significant differences of their copy number changes in sputum between NSCLC and both COPD and healthy subjects. A logistic regression model with the best prediction was built on the basis of 6 genes, *ENO1*, *FHIT*, *HYAL2*, *SKP2*, *p16*, and *14-3-3zeta*. The composite of the 6 genes produced 86.7% sensitivity and 93.9% specificity in distinguishing stage I NSCLC patients from the noncancer individuals. Furthermore, the genes had higher sensitivity (86.9%) in identification of squamous cell carcinoma (SCC) than in adenocarcinoma of the lungs (80.8%;  $P < 0.05$ ). Validation of the genes in the independent cohort confirmed their diagnostic power that also showed higher accuracy for lung SCCs than for sputum cytology. The gene panel could provide sputum-based markers that have the potential to improve early detection of lung SCCs. *Cancer Prev Res*; 3(12); 1571–8. ©2010 AACR.

### Introduction

Non-small cell lung cancer (NSCLC) is the number one cancer killer in the United States and worldwide (1). NSCLC comprises 2 major histologic subtypes: squamous cell carcinoma (SCC) and adenocarcinoma (AC). The overall 5-year survival rate for stage I NSCLC patients who are typically treated with surgery remain up to 83%. In striking contrast, only 5% to 15% and less than 2% of patients with stage III and IV NSCLC are alive after 5 years (1). These statistics imply that identification of early NSCLC patients could potentially reduce the mortality (2). Chest X-ray has been used for its early detection, however, the sensitivity is low (3). Computed tomography (CT) provides excellent anatomic information and plays an increasing role in

noninvasive diagnosis of early NSCLC (4). However, its efficacy in reducing the mortality is not yet shown. Furthermore, if being proven to be effective, CT is likely to be more accurate for peripherally located ACs than for central lung cancers that mainly are SCCs (3). Therefore, the development of noninvasive approaches that can be used alone or complement CT in more precisely identifying NSCLC, particularly SCC is clinically important.

Because morphological changes of exfoliated bronchial epitheliums from sputum are associated with incident lung cancer, sputum cytology has been used for noninvasively detecting NSCLC, especially SCC tumors arising in central airway areas (3). However, the technique depends on the skills required for identifying subtle nuclear changes in cells. Therefore, the sensitivity of the sputum cytology is low (3). Instead of observing cellular morphological characterization, molecular study of sputum can identify the cells containing tumor-related genetic aberrations and therefore could be more sensitive than cytology in diagnosis of NSCLC (2). Various sputum biomarkers have been investigated for early detecting lung cancer or predicting the malignancy in high-risk populations. For instance, hypermethylation of *p16* was detected by PCR-based assay in sputum collected from patients with NSCLC 5 to 35 months before sputum cytologic diagnosis (4). Analysis of chromosomal aneusomy by fluorescence *in situ* hybridization could predict lung cancer incidence within 18 months

**Authors' Affiliations:** Departments of <sup>1</sup>Pathology, <sup>2</sup>Medicine, and <sup>3</sup>Surgery and <sup>4</sup>Division of Biostatistics of The University of Maryland Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, Maryland

**Note:** Supplementary data for this article are available at Cancer Prevention Research Online (<http://cancerprevres.aacrjournals.org/>).

**Corresponding Author:** Feng Jiang, Department of Pathology, The University of Maryland School of Medicine, 10 South Pine Street, MSTF 7th Floor, Baltimore, MD 21201. Phone: 410-706-4854; Fax: 410-706-8481. E-mail: [fjiang@som.umaryland.edu](mailto:fjiang@som.umaryland.edu)

**doi:** 10.1158/1940-6207.CAPR-10-0128

©2010 American Association for Cancer Research.

before clinical diagnosis (5). However, none of the techniques and tested markers, thus far, was sufficiently able to achieve the required characteristics for early detection of NSCLC. To improve reliability of molecular analysis of sputum for diagnosis of NSCLC, we previously used magnetic assisted cell sorting to enrich bronchial epithelial cells from sputum (6). By using this approach, we concentrated the bronchial epithelial cells to 40% purity from 1.1% of the starting population, yielding enough bronchial epithelium for cytologic and molecular analyses of sputum (6). Furthermore, taking advantage of the developments in microarray, and based on the principle of *in situ* hybridization, we developed an *in situ* minichip assay that examined genomic copy number changes of multiple genes at once (7). In addition, using high-resolution comparative genomic hybridization microarrays, we identified a set of genetic signatures whose copy number changes were associated with early stage NSCLC (8). We further demonstrated that assessing changes of some genes of the signatures could identify cancer cells not only in the cytologically positive sputum, but also in cytologically negative sputum of NSCLC patients (9). However, the sensitivity obtained by examining the individual genes is not, as yet, high enough for clinical applications.

Lung tumor is a heterogeneous disease and develops from field defect lesions of respiratory airways in complex and multistep processes (10). Therefore, we hypothesized that simultaneously assessing changes of multiple tumor-related genes in sputum could diagnose early NSCLC with high sensitivity and specificity. To verify the hypothesis, from our previously defined genetic signatures of NSCLC, we optimized a small panel of genes that could reliably be measurable in sputum for the early detection of NSCLC, particularly lung SCC, with higher accuracy than each individual gene. We then validated the gene panel in an independent cohort of lung cancer patients. The validation confirmed the efficiency of the genes that provided higher diagnostic power than cytology.

## Materials and Methods

### Patients and control subjects

This study consisted of 2 phases. The first phase was to optimize a panel of genes whose genomic copy number changes could be detected in sputum for diagnosis of NSCLC with high sensitivity and specificity. To this end, a training set of 49 patients with stage I lung NSCLC, 49 patients with chronic obstructive pulmonary disease (COPD), and 49 healthy smokers who served as "normal controls" were recruited from University of Maryland Medical Center. The cancer patients, COPD patients, and normal controls were matched in the ratio of 1:1 by age, gender, and smoking history as a case-control cohort (Table 1). The second phase of the study was to validate the diagnostic performance of the genes that were identified from the first phase in an independent cohort. A testing set of 69 NSCLC patients including 35 lung SCC patients and 34 AC patients, and 65 cancer-free individuals con-

**Table 1.** Characteristics of a case-control cohort of stage I NSCLC patients, healthy smokers, and COPD patients

	Primary NSCLC (n = 49)	Healthy smokers (n = 49)	COPD patients (n = 49)
Age <sup>a</sup>	69 (54–76)	67 (58–72)	69 (57–73)
Sex			
Female	15	15	15
Male	34	34	34
Smoking status, pack-years	37 ± 23	36 ± 28	38 ± 22
Stage	All are stage I		
Histologic types			
SCC	23		
AC	26		
Location of tumor			
Central	22		
Peripheral	27		

<sup>a</sup>Data are presented as median (range).

sisting of 26 healthy smokers and 39 COPD patients were enrolled from Baltimore VA Medical Center (Supplementary Table 1). Because the testing cohort was enrolled from a different medical center, the cohort was considered as an independent set. The control subjects were patients who visited the centers for minor illness other than respiratory disease and had no diagnosis of any cancer. The healthy smokers were recruited from the same medical centers as those were for cancer and COPD. Therefore, they were considered to be enrolled from the same larger population resources as did the clinical cases and COPD subjects. The controls in the independent cohort were also matched to cases by age, sex, race, and cigarette smoking status represented by pack-years. All study subjects were given written, informed consent for the interviews under our approved International Review Board (IRB) protocols from the 2 medical centers. The surgical-pathologic staging of NSCLC was determined according to the TNM classification of the International Union Against Cancer (UICC) with the American Joint Committee on Cancer (AJCC) and the International Staging System for Lung Cancer that was revised in 1997 (11–12). The location of tumors was recorded with regard to central or peripheral (within 2 cm of costal pleura) site (13). Diagnosis and classification of COPD severity was made according to Global Initiative for Chronic Obstructive Lung Disease (<http://www.goldcopd.org>).

### Collecting, processing, and cytologic studying sputum

All participants in the 2 different medical centers underwent the same sputum collection using an induced protocol over the same interval of time (6, 7, 9, 14–16). Sputum was collected from lung cancer patients before they had received

preoperative adjuvant chemotherapy or radiotherapy. Respiratory epithelial cells were enriched from sputum by using our developed protocol with magnetic nanobeads (6). Two cytocentrifuge slides were made from each enriched sputum sample. One slide was used for checking quality of the specimen. The main variables applied to select sputum for the study were high cellularity, good nuclear morphology, and lack of purulence, debris, and residual cytoplasm (9, 15). Another one was stained with Papanicolaou stain for cytologic diagnosis using the classification of Saccomanno (17). Positive cytology included both carcinomas *in situ* and invasive carcinoma (18). The rest of cells from each sputum sample were then stored in a  $-80^{\circ}\text{C}$  freezer until further processing for the following study.

#### Analyzing genomic copy number changes of the genes on sputum by using *in situ* minichip assay

The 15 genes that were identified from our previous work (8, 19), whose copy number changes were associated with stage I NSCLC are listed in Table 2. We made probes that specifically covered genomic sequences of the target genes as previously described (7). The centromeric probes (CEP) for the chromosomes on which the genes were located were purchased from either Vysis or Genetix USA Inc. The CEPs were used as internal controls for assessing changes of the target genes. *In situ* minichips consisting of the probes were made and then hybridized on the specimens as previously described (7, 15). After washing, the slides were examined under microscopes equipped with appropriate filter sets (Leica Microsystems). Cells (1,000) were counted on each slide. Signals for each probe on the *in situ* minichip were analyzed blindly by investigators who did not have knowledge of case-control status. The criteria for defining a gene with abnormal copy numbers and a positive cell with abnormal genetic changes were used as described in our previous reports (7, 15). Briefly, more or less intense signals from the specific gene probe than from the corresponding CEP indicated a gain or loss of the gene. A cell that had 3 or more copies or exhibited hemizygous or homozygous loss of a gene was considered as a positive cell. Furthermore, percentage of the positive cells (PPC) with genetic changes of a gene in a given sample was used to represent level of the abnormalities of the gene in the sample.

#### Statistical analysis

In the first phase, to define an optimal panel of genes for distinguishing cancer patients from either healthy smokers

or COPD patients, we used receiver-operator characteristics (ROC) curve and the area under ROC curve (AUC) analyses to determine diagnostic sensitivity and specificity of each gene for lung cancer by selecting optimal threshold of its PPC level. Furthermore, on the basis of the results generated from the ROC curves and their AUC values, logistic regression was applied to create prediction model building. The genes were fitted into logistic regression models, and the stepwise backward model selection was further performed to determine the best discriminating combinations of genes for the diagnosis of NSCLC in sputum. In addition, we used MedCalc-User-friendly statistical software (<http://www.medcalc.be/index.php>) to compare the AUCs of different ROC curves. In the second phase, we utilized contingency table and logistic regression analysis to determine the associations between the PPCs of the genes and both clinicopathologic and demographic characteristics of the cases and controls. All tests were 2-sided and a *P* value of  $<0.05$  was considered statistically significant.

## Results

#### Performance of the specific genomic probes and *in situ* minichip assay on sputum

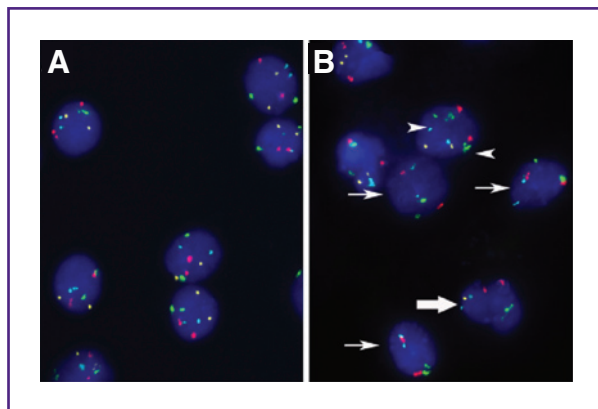
To determine whether all the genes could be reliably detected in sputum by the minichip, we first randomly selected sputum samples from 10 healthy smokers. The probes were then tested on the specimens. As shown in Figure 1A, the specific signals of each probe on interphases were very bright and sharp. Hybridization efficiency of the probes on the normal interphases was 100% as was that of the commercial control probes (Fig. 1A). Furthermore, when an *in situ* minichip containing only a 1-color fluorescence-labeled probe in each element was tested in the specimens, the normal diploid cell population always comprised 2 aqua, 2 green, 2 red, and 2 yellow signals for each probe (Supplementary Fig. 1). Altogether, the *in situ* minichip worked well on sputum, and the specific probes for the genes can be easily visualized and reliably counted on sputum specimens.

Whenever displaying aberrations, 9 genes (*ENO1*, *SKP2*, *NME2*, *S100A1*, *POLR*, *14-3-3zeta*, *EEF1A2*, *FGF4*, and *hTERT*) had gains of their genomic copy numbers, whereas 6 genes (*FHIT*, *TGFBR*, *HYAL2*, *CD74*, *P16*, and *SP-A*) exhibited losses of the copy numbers, as compared with their corresponding CEPs (Fig. 1B). When PPC value was

**Table 2.** The 15 genes whose genomic changes are associated with stage I NSCLC

Gene	Loc.	Gene	Loc.	Gene	Loc.	Gene	Loc.	Gene	Loc.
<i>ENO1</i>	1p36.23	<i>S100A1</i>	1q21	<i>FHIT</i>	3p14	<i>HYAL2</i>	3p21	<i>TGFBR</i>	3p22
<i>POLR</i>	3q28	<i>SKP2</i>	5p13	<i>hTERT</i>	5p15	<i>CD74</i>	5q32	<i>P16</i>	9p21
<i>14-3-3-zeta</i>	8q23	<i>SP-A</i>	10q22	<i>FGF4</i>	11q13	<i>NME2</i>	17q21	<i>EEF1A2</i>	20q13

NOTE: Loc., cytogenetic location of a gene in the chromosomes.



**Fig. 1.** Genetic analysis of exfoliated respiratory epitheliums from sputum by the *in situ* minichip for copy number changes of a panel of 6 genes. A, bronchial epithelial cells enriched from sputum of a healthy control shows 2 yellow signals of the *HYAL2* probe, 2 red signals of the CEP 3 that is used as internal control probe for assessing changes of *HYAL2*, 2 green signals from the *14-3-3-zeta* probe, and blue signals of the CEP8 that is used as internal control probe for assessing changes of *14-3-3-zeta*. B, bronchial epitheliums obtained from a lung cancer patient shows homozygous deletions of *HYAL2* denoted by thin arrow (no yellow signal, 2 red signals), hemizygous deletion of *HYAL2* denoted by wide arrow (a single yellow signal, 2 red signals), and gains of amplified *14-3-3-zeta* denoted by arrowhead (multiple green signals, 2 blue signals). The minichip can simultaneously detect changes of 12 probes consisting of 6 probes for target genes and 6 probes for their corresponding CEPs. The figure only shows 4 probes. Original magnification, 400 $\times$ .

used to determine level of copy number changes of a gene, all 15 genes had higher sputum PPC values in cancer patients than in cancer-free individuals (all  $P < 0.05$ ; Supplementary Table 2). Furthermore, 8 of the 15 genes also displayed higher PPCs in COPD patients than in healthy smokers (all  $P < 0.05$ ). The data suggest that aberrant changes of the 8 genes could also occur in sputum of the COPD patients. However, when the PPCs of the genes in sputum of NSCLC patients were compared with those in healthy smokers and COPD patients, the genes

exhibited considerably higher PPC levels in the cancer patients than in the healthy individuals and COPD subjects (all  $P < 0.05$ ; Supplementary Table 2). Collectively, the 15 genes would provide potential sputum-based biomarkers in distinguishing NSCLC patients from both healthy smokers and COPD patients.

**Optimizing a panel of sputum genetic markers for NSCLC**

ROC analyses were first applied to evaluate the capability of using the genes to discriminate cancer patients from healthy smokers and COPD patients. The 15 genes exhibited 0.623 to 0.678 and 0.614 to 0.67 AUC values, respectively (Supplementary Table 3). When optimum cutoffs were selected, they yielded moderate sensitivities (57.2%–79.8%) and specificities (59.1%–79.4%) in distinguishing NSCLC patients from the healthy smokers. Furthermore, the genes produced 58.2% to 78.3% sensitivities and 58.6% to 77.9% specificities in discriminating NSCLC patients from the COPD patients (Supplementary Table 3).

Logistic regression of the 15 genes using a backward elimination approach was performed to optimize a small panel of markers for diagnosis of NSCLC with high sensitivity and specificity. One of the logistic regression models was built on the basis of 6 genes, *ENO1*, *FHIT*, *HYAL2*, *SKP2*, *p16*, and *14-3-3zeta*, which in combination provided the best prediction in distinguishing NSCLC patients from healthy smokers. The composite panel of the genes created 0.846 AUC that were considerably higher than 0.623 to 0.678 AUC values of individual genes (all  $P < 0.05$ ). Accordingly, given a specificity of 93.9% in distinguishing NSCLC patients from healthy smokers, the panel of the genes revealed a sensitivity of 86.7% (Table 3), which was significantly higher than 57.2% to 79.8% sensitivities generated by individual ones ( $P < 0.05$ ). Similarly, the gene panel also provided the best prediction of NSCLC patients in discriminating cancer patients from COPD subjects with 0.825 AUC, producing 81.6% sensitivity and 93.9% specificity (Table 3). Interestingly, the genes yielded higher

**Table 3.** Diagnostic performance of the composite of 6 genes in a cohort of 49 stage I NSCLC patients, 49 healthy smokers, and 49 COPD patients

	Distinguishing NSCLC patients from healthy smokers		Distinguishing NSCLC patients from COPD patients	
	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %
All cases	83.67 (41/49)	93.88 (46/49)	81.63 (40/49)	93.88 (46/49)
Cases with different histologic types				
SCC	86.96 (20/23) <sup>a</sup>	93.88 (46/49)	86.96 (20/23) <sup>a</sup>	93.88 (46/49)
AC	80.77 (21/26) <sup>a</sup>	93.88 (46/49)	76.92 (20/26) <sup>a</sup>	93.88 (46/49)
Cases with different tumor locations				
Central	86.36 (19/22) <sup>a</sup>		86.36 (19/22) <sup>a</sup>	
Peripheral	81.48 (22/27) <sup>a</sup>		77.78 (21/27) <sup>a</sup>	

<sup>a</sup> $P \leq 0.05$ .

Downloaded from <http://aacrjournals.org/cancerpreventionresearch/article-pdf/31/2/1571/2248269/1571.pdf> by guest on 19 June 2024

sensitivities (86.9%, 86.9%) in discriminating SCC than in AC (80.8, 76.9%) from either healthy or COPD subjects (all  $P < 0.05$ ), while having the same specificity (93.9% vs. 93.9%; Table 3). Likewise, the panel of genes also resulted in higher sensitivity in identification of central tumors than in peripheral tumors (86.4% vs. 81.5%, 86.4% vs. 77.8%, all  $P < 0.05$ ; Table 3). The findings suggest that the panel of the 6 genes might provide more accurate diagnostic markers for SCCs and central tumors of the lungs than for AC and peripherally located lung tumors.

#### Validating the identified genes in an independent set of NSCLC patients

To further evaluate the diagnostic performance of the small gene panel, the 6 genes were assessed on sputum of the testing cohort. Overall, the genes displayed higher PPC levels in the lung cancer patients than in all noncancer individuals (all  $P < 0.05$ ). We used the optimal thresholds established in the training set to determine diagnostic value of the genes for lung cancer in the testing set. As shown in Table 4, the panel of the 6 genes produced 82.6% and 81.2% sensitivities, and 96.2% and 94.9% specificities in differentiating NSCLC patients from healthy smokers and COPD individuals, respectively. Furthermore, the genes used together generated 85.7% sensitivity and 96.2% specificity in discriminating lung SCC patients from healthy individuals, 85.7% sensitivity and 94.9% specificity in differentiating lung SCC patients from COPD subjects. However, the sensitivities of the gene panel were only 79.4% and 76.5% in distinguishing AC patients from healthy smokers and COPD patients, although maintain-

ing specificities at 96.15% and 94.9%, respectively (Table 4). Moreover, the gene panel resulted in 87.9% and 84.9% sensitivities in discriminating patients with central lung cancer from COPD patients and healthy smokers, whereas only 77.8% and 77.8% sensitivities for the patients with peripheral cancer ( $P < 0.05$ ). Altogether, the data confirmed that the composite panel of the 6 genes had higher diagnostic sensitivity for SCC and central tumors than for AC and peripheral tumors of the lungs (all  $P < 0.05$ ). In addition, the 6 genes combined had higher sensitivity in advanced stage lung SCCs (III-IV) as than in stages I and II SCCs ( $P < 0.05$ ), while maintaining specificity at more than 94.9% in identification of the SCC patients with all stages from both healthy smokers and COPD patients (Table 4). Finally, there was an association of the changes of the genes with smoking pack-year of cancer patients (Supplementary Table 4; all  $P < 0.05$ ). However, the changes were not related to the age, gender, and ethnicity of the participants (all  $P > 0.05$ ). Taken together, the validation data confirmed that the identified genes had the potential to be used as markers for SCCs and central tumors in the airways of the lungs.

Because sputum cytology had higher diagnostic accuracy for early lung SCCs than for ACs, we further examined the panel of the 6 genes in sputum and compared the results with that of cytology in distinguishing early stage NSCLC patients (stages I-II) from healthy smokers. Eighteen samples (41.9%) of sputum from the 43 early stage NSCLC patients were diagnosed to be positive by cytologic examination. All cytologically positive sputum specimens also exhibited copy number aberrations of the genes, thus were

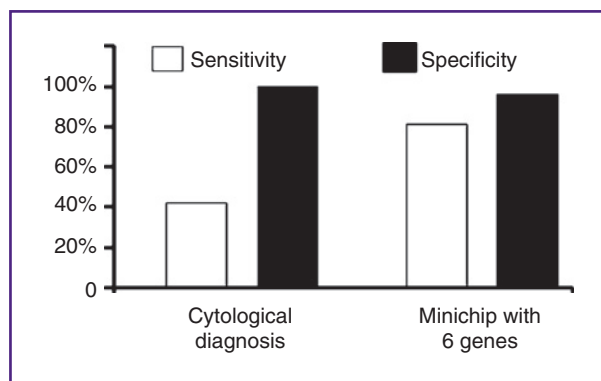
**Table 4.** Diagnostic performance of the composite panel of 6 genes in a cohort of 69 NSCLC patients with different stages, 26 healthy smokers, and 39 COPD patients

	Distinguishing NSCLC patients from healthy smokers		Distinguishing NSCLC patients from COPD patients	
	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %
All cases	82.61 (57/69)	96.15 (25/26)	81.16 (56/69)	94.87 (37/39)
Different histologic types				
SCC	85.71 (30/35) <sup>a</sup>	96.15 (25/26)	85.71 (30/35) <sup>a</sup>	94.87 (37/39)
AC	79.41 (27/34) <sup>a</sup>	96.15 (25/26)	76.47 (26/34) <sup>a</sup>	94.87 (37/39)
Different tumor locations				
Central	87.88 (29/33) <sup>a</sup>		84.85 (28/33) <sup>a</sup>	
Peripheral	77.78 (28/36) <sup>a</sup>		77.78 (28/36) <sup>a</sup>	
Cases with different stages				
I	80.95 (17/21)	96.15 (25/26)	80.95 (17/21)	94.87 (37/39)
II	81.82 (18/22) <sup>b</sup>	96.15 (25/26)	77.27 (17/22) <sup>b</sup>	94.87 (37/39)
III-IV	84.62 (22/26) <sup>c</sup>	96.15 (25/26)	84.62 (22/26) <sup>c</sup>	94.87 (37/39)

<sup>a</sup> $P < 0.05$ ; 26 healthy smokers and 39 COPD patients.

<sup>b</sup>Comparison of the genetic test for diagnosis of stage I and stage II NSCLC with  $P > 0.05$ ; 26 healthy smokers and 39 COPD patients.

<sup>c</sup>Comparison of the genetic test for diagnosis of stages I-II and stages III-IV NSCLC with  $P < 0.05$ ; 26 healthy smokers and 39 COPD patients.



**Fig. 2.** Sensitivity and specificity of the composite of the genetic probes and cytologic analysis for detecting abnormal cells in sputum. Sensitivity of a test refers to the early stage NSCLC patients (21 stage I and 22 stage II cases) who have a positive result by the test. Specificity refers to the 26 healthy smokers who have a negative result by the test.

positive for the genetic test. Interestingly, of the 25 cytologically negative sputum samples from the NSCLC patients, 17 were positive for the minichip test with the panel of 6 genes. Overall, the sensitivity and specificity of sputum cytology in detection of stages I-II NSCLC were 41.9% and 100.0%, whereas the 6 genes in combination resulted in 81.4% sensitivity and 96.2% specificity. Therefore, the composite of use of the 6 genes had higher diagnostic sensitivity than for sputum cytology for early stage NSCLC ( $P < 0.01$ ; Fig. 2). Collectively, the results provide further evidence that the optimal set of the genes might be used as biomarkers for early lung SCC.

## Discussion

Of the 6 genes, 3 genes (*FHIT*, *HYAL2*, and *P16*) were previously investigated in sputum of lung cancer patients for their genomic copy number alterations by our group (6, 7, 9, 14–16). The present observations of the genes are consistent with our earlier findings, confirming their potential role as biomarkers for early NSCLC. Of the 3 newly characterized genes, *ENO1* is located at chromosome 1p36.23 that is one of the most frequent targets of genomic amplification in lung cancer (8, 19, 20). Rácz et al. (21) found that *ENO1* was amplified in tumors tissues of early stage of NSCLC, especially SCC. Our current finding is consistent with the previous reports and further suggests that *ENO1* might be a promising sputum-based biomarker for early SCC. *14-3-3zeta* located at 8q23.1, another important chromosomal region with genomic amplicons in lung cancer (8, 19). Using a functional genomic approach, we identified *14-3-3zeta* as a putative oncogene whose activation was common and driven by its genomic amplification in NSCLC (8, 19). The present data support our previous discoveries (8, 19, 22–23) and also demonstrate that assessing increased *14-3-3zeta* copy numbers in sputum might be useful in early diagnosis of SCCs. Similarly,

*SKP2* was located at a genomic amplified region, 5p13 (8, 19). We previously showed that *SKP2* played an oncogenic role in lung cancer, because inhibition of *SKP2* expression significantly increased p27 expression, reduced cell proliferation, and apoptosis in lung cancer cells (23). In good agreement with the findings, the current data demonstrate that increased *SKP2* copy numbers are common in respiratory epitheliums of lung SCC patients and more importantly, the assessment of the genetic aberrations in sputum holds diagnostic potential for early lung SCCs.

It is not surprising to find that changes of the genes are associated with smoking packer-year, because lung cancer is characterized by an accumulation of molecular genetic abnormalities resulting from repeated exposure of the respiratory tracts of chronic smokers to tobacco-related carcinogens (10). In particular, the relationship of SCCs to smoking is stronger compared with ACs (2). Furthermore, we found that these lung cancer-associated genetic changes were also related to COPD. The observation could be explained by possible pathologic relations between COPD and lung cancer (24, 25). Recent genetic linkage studies have identified common, biologically potential genetic markers in families who have susceptibility for both COPD and lung cancer (25). Therefore, discrimination of NSCLC from the benign disease has presented a difficult problem in past studies, in which a relatively high percentage of false-positive classification of NSCLC was observed. However, our present study shows that the copy number changes of the 6 genes display significant differences in sputum of NSCLC patients, COPD patients, and healthy smokers. More importantly, the genes used in combination could significantly differentiate lung cancer patients from both COPD subjects and healthy smokers with high accuracy. Nonetheless, long-term follow-up of the healthy smokers and COPD patients, whose sputum had the genetic aberrations, is needed to determine whether the result represents an early indicator of lung malignancy developed from tobacco or COPD-related epithelial damages before it is detectable by other means. Moreover, although higher levels of copy number changes of the genes were observed in more advanced stage NSCLC, the 6 genes produced 81.4% sensitivity for early stage SCCs that was dramatically higher than did cytologic diagnosis (41.9%), while keeping a similar specificity as did the cytology. Therefore, the panel of the genes could provide biomarkers that have higher diagnostic efficiency for early SCCs than for sputum cytology.

Changes of the 6 genes in sputum are more closely associated with SCCs that are often centrally located as compared with ACs that tends to arise peripherally. Furthermore, the genetic changes are found to be more strongly related to advanced lung tumors as compared with early stage ones. As a result, the sputum biomarkers display high sensitivity in identification of advanced stage tumors and SCCs as compared with early stage biomarkers and ACs of the lungs. The observation would suggest that there are 2 major sources of the detected genetically altered cells in

sputum. First, the bronchial epitheliums with the genetic aberrations in sputum of the patients with centrally located advanced SCC are very likely derived from the tumor itself, thus producing high sensitivity and specificity. Second, the respiratory epitheliums with such genetic changes in sputum of early NSCLC patients might not be the cancer cells that directly shed from a primary tumor rather than the pulmonary epithelial cells that share clonally altered genetic lesions with the small peripheral lung tumors. This possibility is supported by that cigarette smoking produces a 'field defect' in airway epithelial cells such that genetic damage occurs throughout the lungs. Copy number changes in the pulmonary epitheliums that do not exfoliate from tumors may represent similar clonal genetic aberrations existed within the small tumors arising in peripheral lung, and hence could function as biomarkers for the malignancy.

CT provides excellent anatomic imaging, but has limitations in uncertain rate for central tumors that are mainly SCCs (3). This justifies the requirement for a new diagnostic mean of such lesions with suspicious imaging features. If genetic markers have high specificity and reasonable sensitivity in detection of centrally located lung SCCs, such markers can complement CT in early detection of NSCLC. This concept is supported by our recent study (15) in which, we showed that combining detection of genetic changes of bronchial epitheliums in sputum with CT scans could increase the diagnostic efficiency for stage I central lung tumors as compared with CT used alone. The finding from the present research is fairly encouraging, because the sensitivity and specificity of the composite use of the genes for identifying central lung SCC is much higher than in individual genes. Therefore, given the high accuracy, future combination of the panel of sputum-based markers with CT not only complements CT by overcoming the weakness of the imaging analysis that has low sensitivity in diagnosis of early central airway SCC, but also surmounts major obstacle of sputum-based biomarkers, by which, it is difficult to localize tumor mass location. Furthermore, although the solitary pulmonary nodule (SPN) is a common presentation of lung cancer, most SPNs are benign. The challenge in evaluating SPNs is to avoid invasive procedures in patients who have benign nodules, without allowing potentially curable bronchogenic carcinomas the time to progress to more advanced or even unresectable disease (26). The sputum-based bio-

markers could complement CT screening for precisely identifying early NSCLC from the subjects with indeterminate pulmonary nodules and/or radiographic emphysema. Therefore, the sputum biomarkers may potentially be an aid to decision making in the management of lung SPN (5).

Most recently, Varella-Garcia et al. (5) analyzed sputum by detecting chromosomal aneusomy with commercial chromosomal probes and showed that the test could predict lung cancer incidence with a sensitivity of 76% and a specificity of 94% within 18 months before lung cancer diagnosis. Differing from the report, here we aim to optimize a biomarker panel by targeting the lung cancer-associated genes for detection of lung cancer patients who already were diagnosed with the disease. The genetic markers may have higher sensitivity than chromosomal probes in diagnosis of NSCLC. However, future comparison of the chromosomal probes with the genetic probes in the same set of specimens for either early detecting lung cancer or screening for the disease in high-risk populations is required.

In conclusion, we have developed a panel of 6 genes whose copy number changes can reliably be measured in sputum. The assessment of the genes could potentially be used as a noninvasive diagnostic tool for early squamous cell lung cancer. Nevertheless, a large multicenter clinical project to further validate the full utility in large prospective cohorts is warranted before it could be adopted in routine clinical settings.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Grant Support

This work was supported in part by National Cancer Institute (NCI) grants CA-135382, CA-137742, and CA-133956, American Cancer Society Research Scholar Grant, a clinical innovator award from Flight Attendant Medical Research Institute, a scholar career development award from NIH K12RR023250-University of Maryland Multidisciplinary Research Career Development Program, and an exploratory research grant from Maryland Stem Cell Fund (F. J.).

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

Received 06/04/2010; revised 07/12/2010; accepted 07/23/2010; published OnlineFirst 09/23/2010

#### References

- Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* 2010;102:464-74.
- Minna JD, Roth JA, Gazdar AF. Focus on lung cancer. *Cancer Cell* 2002;1:49-52.
- Frost JK, Ball WC Jr, Levin ML, Tockman MS, Baker RR, Carter D, et al. Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins study. *Am Rev Respir Dis* 1984;130:549-54.
- Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K, et al. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res* 2006;66:3338-44.
- Varella-Garcia M, Schulte AP, Wolf HJ, Feser WJ, Zeng C, Braudrick S, et al. The detection of chromosomal aneusomy by fluorescence *in situ* hybridization in sputum predicts lung cancer incidence. *Cancer Prev Res* 2010;3:447-53.
- Qiu Q, Todd NW, Li R, Peng H, Liu Z, Yfantis HG, et al. Magnetic enrichment of bronchial epithelial cells from sputum for lung cancer diagnosis. *Cancer* 2008;114:275-83.
- Li R, Liu Z, Fan T, Jiang F. A novel multiple FISH array for the detection of genetic aberrations in cancer. *Lab Invest* 2006;86:619-27.

8. Jiang F, Yin Z, Caraway NP, Li R, Katz RL. Genomic profiles in stage I primary non small cell lung cancer using comparative genomic hybridization analysis of cDNA microarrays. *Neoplasia* 2004;6:623–35.
9. Li R, Todd NW, Qiu Q, Fan T, Zhao RY, Rodgers WH, et al. Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin Cancer Res* 2007;13:482–7.
10. Colby TV, Wistuba II, Gazdar A. Precursors to pulmonary neoplasia. *Adv Anat Pathol* 1998;5:205–9.
11. Hammar SP, Brambilla C, Pugatch B. Squamous cell carcinoma. In: Travis WD, Brambilla E, Muller-Hermelink HK, editors. *Pathology and Genetics, Tumours of the Lung, Pleura, Thymus and Heart*. Lyon, France: IARC Press;2005. p. 26–30.
12. Colby TV, Noguchi M, Henschke C. Adenocarcinoma. In: Travis WD, Brambilla E, Muller-Hermelink HK, editors. *Pathology and Genetics, Tumours of the Lung, Pleura, Thymus and Heart*. Lyon, France: IARC Press;2005. p. 35–44.
13. Diederich S, Wormanns D, Semik M, Thomas M, Lenzen H, Roos N, et al. Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. *Radiology* 2002;222:773–81.
14. Xie Y, Todd NW, Liu Z, Zhan M, Fang H, Peng H, et al. Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. *Lung Cancer* 2010;67:170–6.
15. Jiang F, Todd NW, Qiu Q, Liu Z, Katz RL, Stass SA. Combined genetic analysis of sputum and computed tomography for noninvasive diagnosis of non-small-cell lung cancer. *Lung Cancer* 2009;66:58–63.
16. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, et al. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer* 2010; Mar 2. [Epub ahead of print]
17. Saccomanno G, Archer VE, Auerbach O, Saunders RP, Brennan LM. Development of carcinoma of the lung as reflected in exfoliated cells. *Cancer* 1974;33:256–70.
18. Saccomanno G. *Diagnostic Pulmonary Cytology*. 2nd ed. Chicago (IL): American Society of Clinical Pathologists Press;1986.p.1–211.
19. Li R, Wang H, Bekele BN, Yin Z, Caraway NP, Katz RL, et al. Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach. *Oncogene*. 2006;18:2628–35.
20. Chang GC, Liu KJ, Hsieh CL, Hu TS, Charoenfuprasert S, Liu HK, et al. Identification of alpha-enolase as an autoantigen in lung cancer: its overexpression is associated with clinical outcomes. *Clin Cancer Res* 2006;12:5746–54.
21. Rác A, Brass N, Höfer M, Sybrecht GW, Remberger K, Meese EU. Gene amplification at chromosome 1pter-p33 including the genes PAX7 and ENO1 in squamous cell lung carcinoma. *Int J Oncol* 2000;17:67–73.
22. Fan T, Li R, Todd NW, Qiu Q, Fang HB, Wang H, et al. Up-regulation of 14-3-3zeta in lung cancer and its implication as prognostic and therapeutic target. *Cancer Res* 2007;16:7901–6.
23. Jiang F, Caraway NP, Li R, Katz RL. RNA silencing of S-phase kinase-interacting protein 2 inhibits proliferation and centrosome amplification in lung cancer cells. *Oncogene* 2005;12:3409–18.
24. Vineis P, Airoldi L, Veglia F, Olgiati L, Pastorelli R, Autrup H, et al. Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ* 2005;7486:277.
25. Young RP, Hopkins RJ, Hay BA, Epton MJ, Black PN, Gamble GD. Lung cancer gene associated with COPD: triple whammy or possible confounding effect? *Eur Respir J* 2008;5:1158–64.
26. Quint LE. Work-up of the solitary pulmonary nodule. *Cancer Imaging* 2003;3:119–21.