

## Lung Cancer Diagnosis from Proteomic Analysis of Preinvasive Lesions

S.M. Jamshedur Rahman<sup>1</sup>, Adriana L. Gonzalez<sup>2</sup>, Ming Li<sup>5</sup>, Erin H. Seeley<sup>3,4</sup>, Lisa J. Zimmerman<sup>4</sup>, Xueqiong J. Zhang<sup>5</sup>, M. Lisa Manier<sup>1,3,4</sup>, Sandra J. Olson<sup>2</sup>, Ronak N. Shah<sup>1</sup>, Alison N. Miller<sup>1,10</sup>, Joe B. Putnam<sup>7</sup>, York E. Miller<sup>11</sup>, Wilbur A. Franklin<sup>12</sup>, William J. Blot<sup>8</sup>, David P. Carbone<sup>3,4,6,9</sup>, Yu Shyr<sup>5</sup>, Richard M. Caprioli<sup>3,4</sup>, and Pierre P. Massion<sup>1,9,10</sup>

### Abstract

Early detection may help improve survival from lung cancer. In this study, our goal was to derive and validate a signature from the proteomic analysis of bronchial lesions that could predict the diagnosis of lung cancer. Using previously published studies of bronchial tissues, we selected a signature of nine matrix-assisted laser desorption ionization mass spectrometry (MALDI MS) mass-to-charge ratio features to build a prediction model diagnostic of lung cancer. The model was based on MALDI MS signal intensity (MALDI score) from bronchial tissue specimens from our 2005 published cohort of 51 patients. The performance of the prediction model in identifying lung cancer was tested in an independent cohort of bronchial specimens from 60 patients. The probability of having lung cancer based on the proteomic analysis of the bronchial specimens was characterized by an area under the receiver operating characteristic curve of 0.77 (95% CI 0.66–0.88) in this validation cohort. Eight of the nine features were identified and validated by Western blotting and immunohistochemistry. These results show that proteomic analysis of endobronchial lesions may facilitate the diagnosis of lung cancer and the monitoring of high-risk individuals for lung cancer in surveillance and chemoprevention trials. *Cancer Res*; 71(8); 3009–17. ©2011 AACR.

### Introduction

Informative biomarkers for early detection would be of great utility in improving the outcome of patients with lung cancer, the leading cause of cancer deaths in the United States and worldwide (1). Recent attempts at early detection strategies using low-dose chest CT are showing promise in reducing lung cancer-related mortality (2). Yet, continued research is needed in assessing individuals for their risk in developing and for having lung cancer (3, 4).

**Authors' Affiliations:** <sup>1</sup>Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, <sup>2</sup>Department of Pathology, <sup>3</sup>Mass Spectrometry Research Center, <sup>4</sup>Department of Biochemistry, <sup>5</sup>Division of Cancer Biostatistics, Department of Biostatistics, <sup>6</sup>Division of Hematology and Oncology, Department of Medicine, <sup>7</sup>Department of Thoracic Surgery, <sup>8</sup>Epidemiology Division, Department of Medicine, <sup>9</sup>Department of Cancer Biology, Vanderbilt-Ingram Comprehensive Cancer Center, Vanderbilt University School of Medicine; <sup>10</sup>Veterans Affairs Medical Center, Nashville, Tennessee; <sup>11</sup>Pulmonary Division, Department of Medicine, Denver Veterans Affairs Medical Center; and <sup>12</sup>Department of Pathology, University of Colorado Denver, Denver, Colorado

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

**Corresponding Author:** Pierre P. Massion, Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Vanderbilt-Ingram Cancer Center, 640 PRB, Nashville, TN 37232-6838. Phone: 615-936-2256; Fax: 615-936-1790; E-mail: pierre.massion@vanderbilt.edu

doi: 10.1158/0008-5472.CAN-10-2510

©2011 American Association for Cancer Research.

Although only a minority of preinvasive lung lesions progress to cancer, they are associated with an increased risk of having lung cancer (5–10). These lesions are likely to harbor molecular alterations indicative of lung cancer risk in the entire affected epithelium, including tumors of different histological subtypes (11, 12). However, this histology-based assessment of these lesions is limited in many ways. First, grading is semiquantitative at best and the agreement between 2 observers for the reporting of the histopathology according to the World Health Organization (WHO) criteria has not been studied extensively (13, 14). Distinguishing mild from moderate dysplasia and severe dysplasia from carcinoma *in situ* is difficult, for example. Consequently, most studies evaluate bronchial preinvasive lesions that are either low or high grade. Second, the effect of the structural integrity of the biopsy on the interpretability of the histology features remains a significant obstacle to this area of research. Thus, the development of a quantitative measure of probability of having lung cancer based on the molecular analysis of the bronchial epithelium, and independent of the histological grade, may have important implications. Our study was therefore designed to measure the probability of having a lung cancer based on the proteomic analysis of endobronchial lesions.

We recently identified patterns of protein expression in fresh frozen bronchial lesions obtained by autofluorescence bronchoscopy and resected lung tissue samples at different stages of tumor progression using matrix-assisted laser desorption ionization mass spectrometry (MALDI MS; ref. 15).

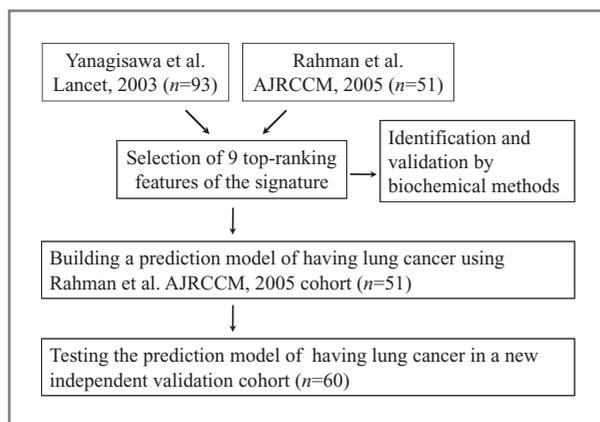
Our proteomic analysis generated signatures discriminating subtypes of lesions in a continuum from normal lung to invasive lung tumor.

Here we took our original observation further and hypothesized that the bronchial epithelium expresses a subset of proteins that once increased confer an increased probability of having lung cancer. To assess the probability of having a lung cancer based on the analysis of bronchial lesions, we are moving away from a pathological scale based on the grade of the lesions to a quantitative scale based on the proteomic analysis of these bronchial specimens. We selected 9 MALDI MS features [mass-to-charge ratio ( $m/z$ )] as a signature from 2 previous studies (15, 16). The novel proteomic signature of 9 features gives rise to a score based on MALDI MS signal intensity that is used as a continuous variable to predict the diagnosis of lung cancer and circumvents many of the problems associated with a histology-based assessment of the lesions. We identified 8 of the 9 features and tested their expression by immunohistochemistry (IHC) in formalin fixed paraffin-embedded (FFPE) tissues and Western blotting.

## Materials and Methods

### Study design

The study design is summarized in Figure 1. To derive a bronchial epithelium-based signature predictive of lung cancer, results from 2 independent studies published from our laboratory were examined for top ranking discriminatory features ( $m/z$ ) identified by MALDI MS. The first study led to features distinguishing normal lung from invasive lung tumors ( $n = 93$ ; ref. 16), and the second study led to features distinguishing normal bronchial epithelium, preinvasive bron-



**Figure 1.** Overview of the study design. Proteomic data of 2 independent studies were examined for top ranking discriminatory features identified by MALDI MS. A signature of 9 features was selected from the intercept of best classifiers from these 2 previous studies based on their statistical significance with false discovery rate-adjusted  $P < 0.01$  and based on the expert visual confirmation of the characteristics of the peak. To assess the association of the signature with cancer status, we first build a model using Rahman and colleagues (15) cohort ( $n = 51$ ). The intensity of the signature (9  $m/z$  values) was transformed to MALDI MS score to predict the risk of having lung cancer. The model was then tested in the validation cohort ( $n = 60$ ).

chial lesions, and invasive lung tumors ( $n = 51$ ; ref. 15). Nine features ( $m/z$ ) were selected from these 144 patients as candidates of a novel signature from the intercept of best classifiers from these 2 previous studies based on their statistical significance with false discovery rate-adjusted  $P < 0.01$  and based on the expert visual confirmation of the characteristics of the peak. The  $m/z$  values of the selected features were 4,749, 4,965, 6,175, 8,451, 8,565, 9,955, 11,048, 12,275, and 12,345. The features ( $m/z$ ) were identified and their expression was verified by biochemical methods. To evaluate the proteomic signature of bronchial lesions as a biomarker diagnostic of lung cancer, we built a prediction model using the Rahman and colleagues cohort (51 patients, 41 with and 10 without lung cancer; ref. 15) and validated it in a new data set (60 patients, 40 with and 20 without lung cancer). The intensity of the signature (9  $m/z$  values) was transformed to a MALDI MS score to predict the diagnosis of having lung cancer (see Statistical Analysis).

The project was approved by the local Institutional Review Board and informed consent was obtained on all individuals at all institutions.

### Patients and tissue samples

A total of 69 bronchial biopsies and surgical lung tissue specimens were collected from the 60 patients in the validation cohort (Table 1 and Supplementary Table S1) who participated between 2003 and 2008 at Vanderbilt University Medical Center (VUMC), the Nashville Veteran Affairs Medical Center (VAMC), or the University of Colorado Health Science Center (UCHSC). Multiple bronchial specimens were obtained at the time of bronchoscopy or surgical resection. Patients undergoing bronchoscopy for clinical suspicion of lung cancer were consented for this research study and agreed to undergo autofluorescence bronchoscopy and to provide bronchial specimens at predetermined normal sites (with normal fluorescence ratio) and at suspicious sites (with increased fluorescence ratio; ref. 17) in addition to the samples taken for

**Table 1.** Characteristics of the patients in the validation cohort according to cancer status

Patients	No cancer, $n = 20$	Cancer, $n = 40$
Age, years (IQR) <sup>a</sup>	69 (62–74)	62 (54–71)
Gender		
Male, $n$ (%)	18 (90)	27 (68)
Female, $n$ (%)	2 (10)	13 (32)
Smoking status		
Never-smoker, $n$ (%)	2 (10)	2 (5)
Ex-smoker, $n$ (%)	8 (40)	24 (60)
Current, $n$ (%)	10 (50)	14 (35)
PKY (IQR)	50 (38–68)	50 (40–80)

Abbreviations: IQR, interquartile range; PKY, pack year (pack per day  $\times$  number of years smoked).

<sup>a</sup>IQR (1st–3rd) values are the median.

clinical diagnosis. Patients with lung cancer in both the validation cohort and the Rahman and colleagues (15) cohort included different subtypes and all stages of lung cancer (Supplementary Tables S1 and S2). As shown in these supplementary tables, some patients in both cohorts provided more than 1 tissue specimen. Specimens collected for research were snap frozen and later cut, stained, and graded by the pathologists (W.A.F. and A.L.G.) according to the WHO nomenclature (14). Only bronchial epithelium specimens were considered in the analysis so that no alveolar tissue sample was used in either the training or test set. Normal bronchial epithelium specimens were obtained from surgical specimens at least 2 cm away from the tumor. Two pathologists (A.L.G. and W.A.F.) reviewed the hematoxylin and eosin (H&E) stained sections of each lesion.

#### Acquisition of MALDI-time of flight MS data

Bronchial biopsy specimens were frozen in liquid nitrogen immediately after collection, embedded in Tissue-Tek Optimal Cutting Temperature compound (IMEB) and stored at  $-80^{\circ}\text{C}$ . For MALDI MS profiling, a 12  $\mu\text{m}$  section was cut and thaw-mounted directly onto a gold-coated MALDI plate followed by fixation in graded ethanol washes (70%, 90%, and 95% for 30 seconds each). An adjacent 7  $\mu\text{m}$  section was also cut and stained with H&E following standard protocols. Photomicrographs of H&E stained tissue sections were obtained with a Mirax Scan digital microscope slide scanner (Mirax) at a pixel resolution of 0.23  $\mu\text{m}$ . The serial 12  $\mu\text{m}$  histology images were annotated using Mirax viewer software by a pathologist (A.L.G.) using a small circular shape (200  $\mu\text{m}$ ) to mark areas of interest for matrix spotting (Supplementary Fig. S1). Color coding was used to designate areas of different histology grade within the tissue section. Annotated histology images were then overlaid with the MALDI plate images using Adobe Photoshop. Fiducial points were placed into at least 4 distinctive areas that are visible in the MALDI plate image and served as landmarks for registering the MALDI plate to the robotic spotter. Using these fiducial points, coordinates of the histology annotations are transformed into a system recognized by the spotter. An acoustic robotic microspotter (LabCyte) was used to place crystalline matrix (25 mg/mL sinapinic acid in 1:1 acetonitrile: 0.2% trifluoroacetic acid) spots ( $\sim 120$  pL) that were typically 180–220  $\mu\text{m}$  in diameter (18). Subsequent to spotting, spot placement was evaluated for accuracy compared to the location designated by the pathologist. Spots that deviated significantly from their desired locations were eliminated from further analysis. Tissue profile spectra were acquired from the spotted areas using an Autoflex II (Bruker Daltonics) MALDI-TOF MS equipped with a SmartBeam Laser (Nd:YAG, 355 nm) Bruker Autoflex II. An accelerating voltage 20 kV was used with an extraction voltage of 18.65 kV and an Einzel lens voltage of 6.00 kV. Delayed extraction (350 nanoseconds) was optimized for resolution at  $m/z$  12,000. Data were collected over the range of 2 to 40 kDa. A total of 400 laser shots were collected for each profile spectrum. Spectra were converted to ASCII files for preprocessing and statistical analysis.

#### Validation of the identified protein expression by Western immunoblotting

Validation of the identified proteins was performed by Western immunoblotting using surgically resected specimens of squamous cell carcinomas of the lung, as well as adjacent histologically normal tissue obtained from the Vanderbilt lung Specialized Programs of Research Excellence (SPORE) tissue repository. Samples were snap-frozen and stored at  $-80^{\circ}\text{C}$  until protein lysates were prepared. Protein concentration of the supernatant was determined by the bicinchoninic acid (BCA) protein assay (Thermo Scientific) and 40  $\mu\text{g}$  of total protein from each sample was separated on 15% SDS-polyacrylamide gels [ubiquitin, macrophage migration inhibitory factor (MIF), thymosin  $\beta_4$ , cytochrome *c*] or 4% to 12% Bis-Tris gels [Invitrogen; acyl-coA binding protein (ACBP) and cystatin A (CSTA)]. Proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane and treated with blocking solution before treatment with primary antibodies for thymosin  $\beta_4$  (1:1,000, Santa Cruz Biotechnology), ubiquitin (1:1,000, R&D Systems), ACBP (1:250, Abcam), CSTA (1:2,500, R&D Systems), cytochrome *c* (1:1,000, Cell Signaling), and MIF (1:500, R&D Systems). The blots were treated with horseradish peroxidase (HRP)-conjugated secondary antibodies and developed with the enhanced chemiluminescence system (Perkin-Elmer). Immunoblotting of  $\beta$ -actin (Sigma) was used as a loading control.

#### Validation of the identified protein expression by immunohistochemistry

Five micron tissue sections were cut from FFPE lung tissue blocks, deparaffinized by placing in 3 xylene baths, 10 minutes each, hydrated through a series of ethanol baths of decreasing concentration, and placed in a buffer bath of TBS. Endogenous peroxidase was quenched using Dako Peroxidase Blocking Reagent and nonspecific staining blocked using Dako Serum-Free Protein Block according to manufacturer's instructions. Optimal staining for each antibody was obtained by using a citrate antigen retrieval and the following primary antibody dilutions: thymosin  $\beta_4$  (1:500, BioDesign), ubiquitin (1:400, Dako), ACBP (1:80, Abcam), CSTA (1:8,000, Sigma), ProteinTech Group, cytochrome *c* (1:1,000, Cell Signaling), and MIF (1:200, Santa Cruz Biotechnology). Primary antibodies were applied at room temperature for 1 hour. All further steps and chromogen development were carried out according to manufacturer's instructions (Dako EnVision<sup>+</sup> System, HRP). Slides were counterstained using a Gills #1 hematoxylin, rinsed appropriately and dehydrated through a series of ethanol baths of increasing concentration into xylene. A nonaqueous coverslipping material was used (Cytoseal or Permount). Results of the IHC staining were analyzed by the pathologist (A.L.G.).

Tissue microarrays (TMA) of lung cancer subtypes from 59 patients were prepared from paraffin blocks. There were 21 squamous carcinomas, 20 adenocarcinomas, 4 bronchioalveolar carcinomas, 4 large cell carcinomas, 2 non-small cell lung cancer (NSCLC), 5 carcinoids, and 1 each of adenosquamous, sarcoma, and small cell lung cancer (SCLC) represented on the TMA. Archived tissue blocks from 1989 to 2002 were retrieved

from the files of VUMC and the Nashville VAMC pathology departments. For all tissue blocks, H&E-stained sections were reviewed by our pathologist (A.L.G.). Areas to be punched for array production were carefully marked. Tissues from 26 patients had spots in triplicates, 13 of which also had single spot of corresponding normal tissue obtained at least 2 cm away from the tumor. Tissues from 33 patients had spots in duplicates. Cores 0.6 mm in diameter were taken from the selected area of each specimen and inserted into a recipient paraffin block. Five-micrometer sections were cut from the arrays and mounted onto charged slides. Every 15th section was stained by H&E to confirm the presence of the histological feature of interest (tumor or normal histology). Immunohistochemical staining of the TMA for thymosin  $\beta$ 4, ACBP, CSTA, and MIF were was performed according the protocol described earlier. IHC staining was scored 0 to 3 based on intensity by ALG. Scoring for ubiquitin was not performed because of poor staining quality.

### Statistical analysis

Before statistical analysis, the MALDI MS data were pre-processed for calibration, baseline correction, denoising, normalization, peak detection, and quantification by a preprocessing package: Wavespec software (19; The raw data can be found at <http://www.vicc.org/biostatistics/supp.php>). We selected 9 features from the intercept of best classifiers to distinguish between normal- or low-grade preinvasive lesions and high-grade or invasive cancer from 2 previously described datasets (15, 16). The selection was based on the statistical significance of the discriminatory features ( $m/z$ ) with false discovery rate-adjusted  $P < 0.01$  and based on the expert visual confirmation of the characteristics of the peak. With these predetermined 9 candidate features, we carried out the following analysis to assess their ability to predict whether these bronchial samples were associated with a patient with a lung cancer. First, we generated a weighted compound score (MALDI score) based on the weighted flexible compound covariate method (WFCCM; refs. 16 and 20): the preselected 9 features were reduced to 1 "summarized" feature by building MALDI score for each patient (see Supplementary Materials). For easier interpretation and comparisons, the MALDI scores were rescaled to a range between 0 and 5. Second, a logistic regression model was applied to test the hypothesis that the summarized 9 features profile (MALDI score) and histology group are associated, thus, this MALDI score has predictive ability in differentiating cancer patients from normal controls. The model was derived and internally validated using the Rahman and colleagues cohort (15). A bias-corrected estimate of predictive accuracy was obtained by bootstrap method (21). The class prediction performance was assessed by a widely used measurement of diagnostic discrimination: the area under receiver operating characteristic (ROC) curve and it follows the standards for reporting of diagnosis accuracy (22). Third, we then applied this internally validated model to the new independent validation cohort described earlier by carrying the estimated coefficients of this model to the independent validation set. The prediction ability in terms of area under ROC curve and the predicted

probability of having lung cancer with the change of MALDI score were provided for validation set. To complete our analysis, we applied a permutation analysis for an alternative interpretation on the same hypothesis we were testing, and the permutation test reconfirmed such association. From patients who provided more than 1 bronchial biopsy, including biopsies of histological normal and invasive cancer tissues, we used the average of these signal intensities to represent the patient in the analysis (see Supplementary Materials).

## Results

### Patient characteristics

Characteristics of the 60 patients in the validation cohort are summarized in Table 1. Characteristics of individual patients of the validation cohort and the Rahman and colleagues cohort (15) are given in Supplementary Tables S1 and S2. Representative images of H&E stained tissue sections of each histological type and images with spots for automatic targeted MALDI matrix deposition are shown in Supplementary Figure S1.

### Performance of the signature and assessment of the probability of having lung cancer

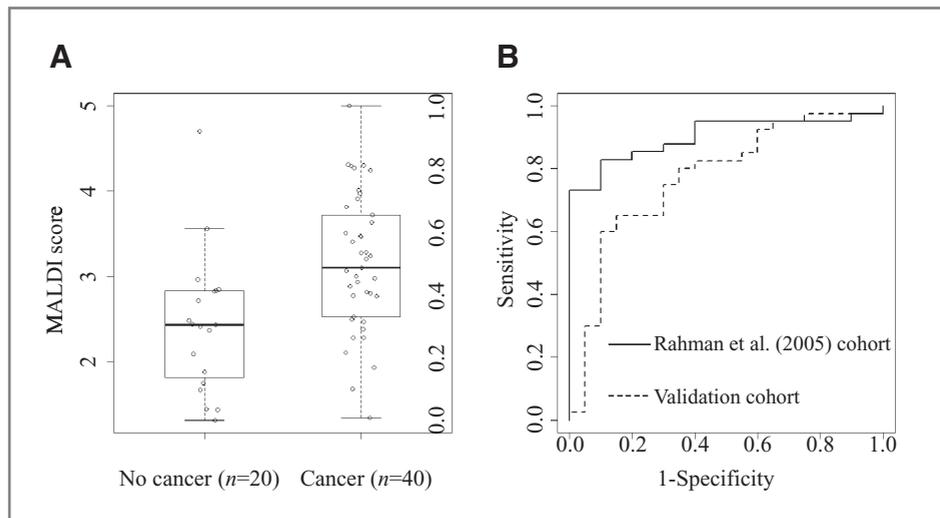
The proteomic signature of 9 features ( $m/z$ ) was first used to build a prediction model using Rahman and colleagues (15) cohort to assess the probability of having lung cancer. The resultant MALDI score obtained as a weighted sum (see Statistical Section and Supplementary Materials) was used as a continuous variable in the validation cohort. The model showed that the MALDI score in this independent data set is statistically significantly associated with the cancer status. Based on the model, the estimated probability of having lung cancer can be calculated for any patient in the cohort with a MALDI score value. The estimated probabilities for some values of MALDI score are summarized in Table 2. To understand the characteristics of MALDI score better, 2 graphical summaries are provided in Figure 2. The distribution of MALDI scores across patients with and without cancer in the validation cohort ( $n = 60$ ) is presented in Figure 2A and

**Table 2.** Probability of predicting lung cancer based on MALDI MS score

Score	<i>P</i>	95% CI
0	0.10	0.00–0.3
1	0.25	0.00–0.5
2	0.48	0.29–0.67
3	0.72	0.60–0.85
4	0.88	0.75–1
5	0.95	0.87–1

Abbreviation: MALDI MS, matrix-assisted laser desorption ionization mass spectrometry.

**Figure 2.** Performance of the signature. A, boxplots representing the distribution of MALDI scores from patients of the validation cohort ( $n = 60$ ). Patients with ( $n = 40$ ) and without ( $n = 20$ ) lung cancer were analyzed. Weighted compound score was generated based on the weighted flexible compound covariate method (WFCCM). The 9 candidate features ( $m/z$ ) were reduced to 1 summarized feature to build MALDI score for each patient. B, receiver-operator characteristic curves demonstrating performance of the signature to classify patients with and without cancer in the published cohort (solid line) and in the validation cohort (dotted line). AUC, area under the curve.



MALDI score prediction performance through the prediction model, is presented by an area under the ROC curve in Figure 2B. By bootstrap method, the bias-corrected area under the ROC curve is 0.90 [with 95% CI (0.82, 0.98)] in the Rahman and colleagues (15) cohort and 0.77 [95% CI (0.66, 0.88)] in the validation cohort.

#### Identification of the MALDI MS features of the signature

Out of 9 features we have identified 8 corresponding to 6 different proteins listed in Table 3. Two of the 6 proteins were newly identified in this study. One of those features at  $m/z$  9,955 was identified ACBP following in-gel digestion and LC-MS/MS analysis (see Supplementary Methods and Supplementary Fig. S2). Based on the features identified, the  $m/z$  value was consistent with the loss of the initial methionine residue and an acetylation modification to the N-terminal

serine residue. Examination of the 42 Da mass shift in the b-ion series in the MS/MS spectra for the acetylated peptide, SQAEFEK and its miscleavage product SQAEFEKAAEEVR, supports such a modification as shown in Supplementary Figure S3A. Another protein identified was CSTA. Analysis of the gel band corresponding to  $m/z$  11,048 was found to contain an acetylated lysine residue in the peptide sequence NKDDELDTGF. The modification was confirmed by examination of the MS/MS spectra containing the 42 Da mass shift in the b-ion series as shown in Supplementary Figure S3B. The unmodified NKDDELDTGF peptide was also identified, consistent with the  $m/z$  11,006 observed in the MALDI MS spectrum. All peptides identified for both ACBP and CSTA are listed in Supplementary Table S3. The other 4 proteins identified by our group are thymosin  $\beta$ 4 (23), ubiquitin (16), cytochrome *c* (24), and MIF (25).

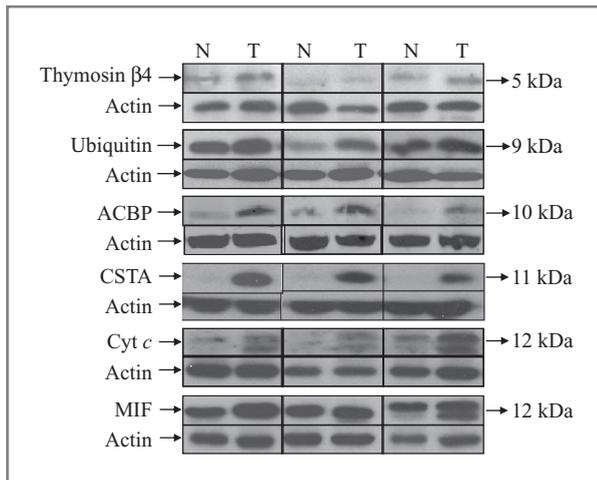
#### Proteomic validation of the MALDI MS features components of the signature

Overexpression of each of the 6 identified proteins was then confirmed by Western blotting in lung cancers as shown in Figure 3. Representative data of 3 surgical specimens of squamous cell carcinomas of the lung and adjacent histologically normal lung tissue are presented. Protein expression was further validated in FFPE squamous cell carcinoma of the lung and normal tissues by IHC as shown in Figure 4. Squamous cell carcinomas showed moderate to strong cytoplasmic and nuclear expression of all these proteins. Thymosin  $\beta$ 4 and cytochrome *c* were predominantly expressed in the cytoplasm. In normal bronchial epithelium weak to moderate staining of MIF was observed, while staining was negative to weak for thymosin  $\beta$ 4, ubiquitin, ACBP, CSTA, and cytochrome *c*. We also validated the expression of thymosin  $\beta$ 4, ACBP, CSTA, and MIF in a TMA representing lung cancer tissues from 59 patients by IHC. The immunostaining signals of these proteins were localizing to the cytoplasm. The summary of the results (intensity score 0–3) are presented in Supplementary Table S4. Of interest, CSTA staining intensity

**Table 3.** Identity of the candidate features in the signature

Selected candidate features, $m/z$	Identification
4,749	Not determined
4,965	Thymosin $\beta$ 4
6,175	Macrophage migration inhibitory factor <sup>++</sup>
8,451	Des-ubiquitin (Ubiquitin-Gly-Gly)
8,565	Ubiquitin
9,955	Acyl-coA binding protein
11,048	Cystatin A
12,275	Cytochrome <i>c</i>
12,345	Macrophage migration inhibitory factor

Abbreviation:  $m/z$ , mass-to-charge ratio.



**Figure 3.** Validation of the expression of candidate biomarker proteins by Western immunoblotting. Tissue lysates were prepared from matched pairs of fresh-frozen normal (N) and squamous lung tumor (T) tissues. Three representative pairs of tissues were used for each immunoblotting. Amount of protein loaded in each lane was 40  $\mu$ g. One of the lanes was added with molecular weight marker in each gel. Immunoblotting of  $\beta$ -actin was performed as a loading control.

was associated with prolonged survival ( $P = 0.03$ ). We found no significant association between staining intensity scores and gender, smoking status, histology, or pathological stage.

## Discussion

In this study we report a proteomic signature derived from the bronchial mucosa that confers a quantifiable probability of having lung cancer. We identified all but 1 of the 9 features of this signature (Table 3). Overexpression of the identified proteins was confirmed by Western blotting (Fig. 3) and immunohistochemical staining (Fig. 4 and Supplementary Table S4) using normal lung and NSCLC. The bronchial biopsy proteomic signature provides a quantitative measure of the probability of having lung cancer that goes beyond the histological evaluation of preinvasive lesions.

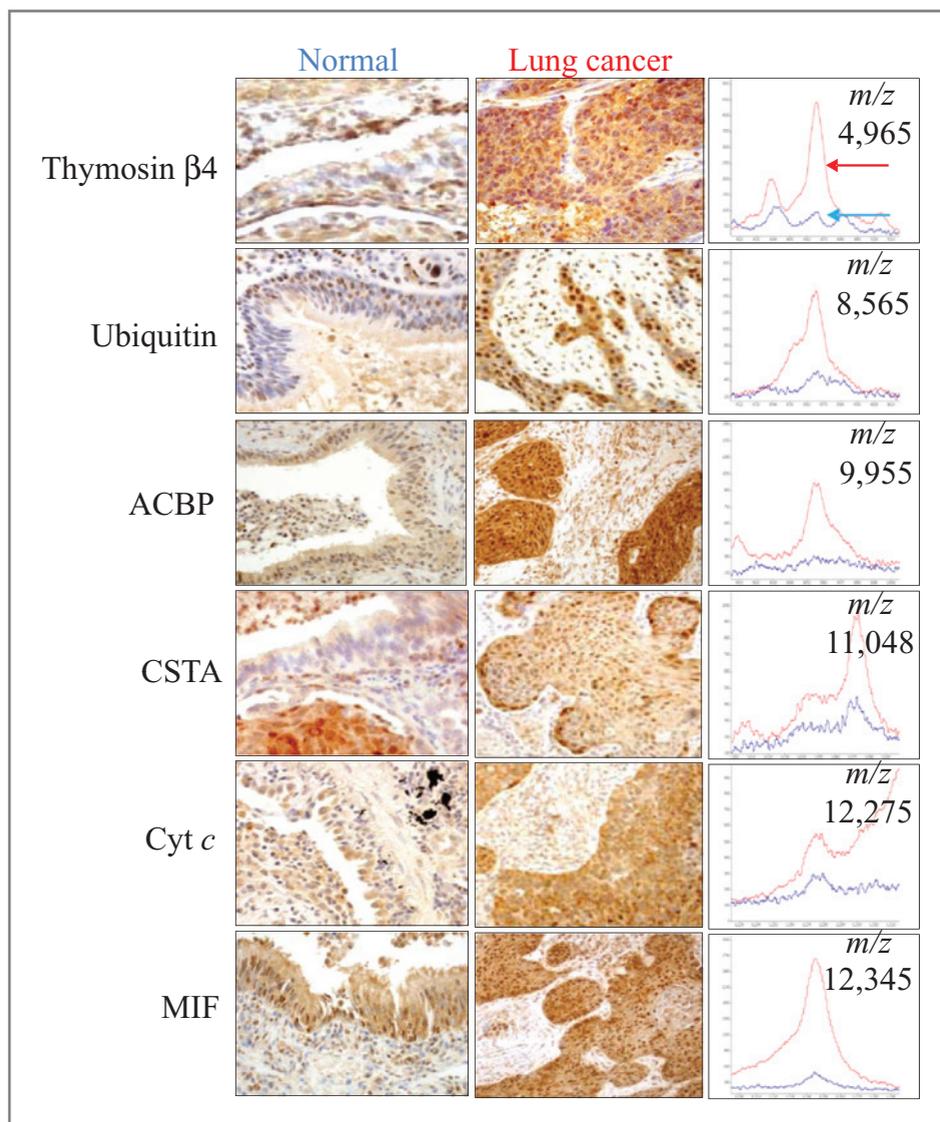
Preinvasive lesions are prevalent in the airways of patients at increased risk for and with lung cancer (8–10, 26, 27). Our ultimate goal is to predict the probability of developing lung cancer and to identify the molecular determinants of lung cancer development. In this study, we made an important step toward this goal by providing a quantitative measure of the probability of having lung cancer in individuals presenting with endobronchial lesions (Fig. 2A). Importantly, our diagnostic biomarker signature goes beyond histologic interpretation. It is quantifiable and used as a continuous variable. It was formally validated in an independent data set of specimens from 3 institutions (Supplementary Table S1). The differential expression of these candidate biomarker proteins revealed by proteomic technologies also allows to detect changes in protein expression during lung tumor progression which may otherwise remain undetected by histology alone. Moreover, the development of a proteomic diagnostic signature for lung cancer provides an opportunity to further our under-

standing about the role of these proteins in various cellular interactions and in pathways of lung tumorigenesis. In Figure 2A, we showed the difference in the MALDI score between patients with and without lung cancer giving the existing overlap. Although such overlap between boxplots is common in the biomarker field, logistical regression analysis indicated that the difference is statistically significant. Yet, these results do not necessarily imply a clinically useful classifier. Further prospective validation of the biomarker is therefore needed to address clinical utility. Finally, this signature may lead to an improved understanding of the progression of preinvasive lesions to an invasive phenotype, potentially enabling earlier therapeutic intervention, and allowing us to further select patients to enroll in surveillance studies.

Scarcity and the small size of preinvasive lesions of the airways remain the major challenges for studies of bronchial mucosa in lung tumor development. To accomplish our goal, we obtained specimens from 3 institutions (VUMC and VAMC in Nashville, TN, and UCHSC in Denver, CO). We also took advantage of significant improvements in proteomic technologies, including automated spotting and related software, MS instrumentation, database resources, and statistical analysis tools (19). Histology directed spotting of matrix using Mirax viewer software, for example, allowed us to limit data acquisition specifically from areas containing greater than 70% of the epithelial cells (18, 28).

Technological advances have also allowed us to confidently identify the proteins represented by our signature. Out of the 6 identified proteins contained in the signature, CSTA and ACBP deserve special attention. Although the role of ACBP in lung tumorigenesis is unknown, ACBP plays a crucial role in lipid metabolism and acts as acyl-CoA transporter and intracellular pool former. In spite of the fact that ACBP is ubiquitously expressed (29), its differential expression in normal epithelium and invasive tissues was clearly demonstrated by MALDI MS. To our knowledge, this is the first report of overexpression of ACBP in lung tumorigenesis.

Cystatins are members of a protein family with endogenous inhibitors of cysteine proteases such as cathepsins B, H, and L. Studies demonstrated an inverse relationship between increased cathepsin B expression and decreased CSTA in a variety of human tumors (30). The importance of CSTA in the development of malignant tumors (31) and its overexpression in lung cancer have been reported (32). Lung tumor cells are suggested to produce both cysteine proteases and cystatins *in vitro* which are regulated differently in histologic subtypes of lung cancers; however, their diagnostic and prognostic role in large pathological materials has not been studied extensively. One study involving preinvasive bronchial lesions and various lung carcinomas suggests that CSTA expression is associated with malignant transformation and predicts a risk of tumor recurrence as well as poor survival (32). Our studies suggest that CSTA is one of the promising candidate biomarkers for lung cancer and deserves further investigation in lung tumorigenesis. Of interest also is that cathepsin B which is a cysteine protease cleaves 2 terminal glycine residues of ubiquitin generating des-ubiquitin (33). Both monomeric ubiquitin and des-ubiquitin corresponding to  $m/z$  values 8,565 and



**Figure 4.** Validation of the expression of identified proteins by IHC. Photomicrographs (40 $\times$ ) of IHC staining of normal bronchial epithelium (left) and invasive squamous carcinoma of the lung (middle) are shown for 6 identified proteins. Scale bar, 50  $\mu$ m. Peaks (*m/z*) corresponding to 6 identified proteins from the average spectra of normal bronchial epithelium (blue arrow) and invasive tumor (red arrow) of the normalized MALDI MS data are represented in the right panel.

8,451 are among the 9 candidates of the signature described in this study, and in our previous study in established tumors (16). Overexpression of cathepsin B inhibitor CSTA, and des-ubiquitin, a product of cathepsin B proteolytic activity on ubiquitin, suggests a delicate balance between cathepsin B and CSTA may somehow perturb cellular homeostasis contributing to the overexpression of des-ubiquitin.

MIF is expressed and secreted in response to mitogens and integrin-dependent cell adhesion. Extracellular MIF is required in the steady-state activation of Rho GTPase family members, leading to cell growth and migratory phenotypes. Previous reports demonstrated that MIF expression is increased in premalignant, malignant, and metastatic tumors and a potential therapeutic target in NSCLC (34, 35). By use of a siRNA knockdown of MIF, the migratory and invasive potential of human lung adenocarcinoma cells was abolished (36).

Thymosin  $\beta$ 4 sequesters cytoplasmic monomeric actin (37) and is a potent regulator of actin polymerization. Overexpression of thymosin  $\beta$ 4 in tumors has been suggested to stimulate lung tumor metastasis by activating cell migration and angiogenesis (38). The gene expression of thymosin  $\beta$ 4 was also significantly associated with metastasis in NSCLC patients and it was suggested as a prognostic parameter for NSCLC (39). A combined proteomic indicator including thymosin  $\beta$ 4, thymosin  $\beta$ 10, and calmodulin was shown to be able to classify NSCLC patients according to good or poor prognosis (40).

Our study has limitations. The patient sample size, especially for noncancer cases, in the training and validation sets was not large. Although the demographics of our study population reflect well individuals who are at high risk for lung cancer, the results may not generalize to those with lower epidemiological risk. Our proteomic signature

represents a direct measure of alterations of the airway epithelium in the field of carcinogenesis, but the signature has not yet been prospectively validated as a predictor of lung cancer development. Also, we could not evaluate the possibility that the presence of such a signature represents a true susceptibility factor rather than a tumor marker. Our study also does not address the diagnostic value of the proteomic analysis of alveolar atypical hyperplasia thought to be possible precursors of adenocarcinoma *in situ* (formerly called bronchioalveolar carcinoma or BAC). Among other study limitations, although the methodology used to assess the probability of having lung cancer may be found useful as an MS-based analysis, the signature may require further development in the form of an immune-based multiplex assay to reach broad clinical utility.

Looking ahead we see this novel signature as a significant advance in the search for early detection strategies. As opposed to histologic evaluation which provides a few discrete subtypes with potential significant interobserver variability (41, 42), the quantitative nature of our signature, after further validation, may be able to better assess the diagnosis of lung cancer from endobronchial biopsies. We will need to further validate the diagnostic potential of the signature prospectively in a cohort of high-risk individuals. Determining the probability of having lung cancer based on a novel proteomic signature of bronchial lesions may have important clinical

implications. This signature may facilitate the diagnosis of lung cancer and the monitoring of high-risk individuals for lung cancer in surveillance and chemoprevention trials.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

### Acknowledgments

We thank all patients who volunteered to participate in this study and provided tissue samples. We thank Candace Murphy and Anna Ostrander for assistance with sample collection, Harriet Davis and Gabe Garcia for clinical data management, Jamie Allen for assistance with sample processing for MALDI MS data acquisition and organization, and Salisha Hill and Kristen Cheek for their assistance with LC-MS/MS.

### Grant Support

This work was supported by the National Institute of Health (CA102353 to P.P. Massion, the Lung SPORE P50 CA 90949 to D.P. Carbone, and the Lung SPORE P50 CA58187 to Y.E. Miller and W.A. Franklin), Vanderbilt-Ingram Cancer Center Core Support Grant (P30 CA68485 to P.P. Massion and R.M. Caprioli), Department of Defense (W81XWH-05-1-0179 to R.M. Caprioli) and NIH/NIGMS (5R01-GM58008 to R.M. Caprioli).

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 July 13, 2010; revised February 17, 2011; accepted February 21, 2011; published online April 12, 2011.

### References

- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. Epub 2011 Feb 4.
- Marshall E. Cancer screening. The promise and pitfalls of a cancer breakthrough. *Science* 2010;330:900-1.
- Franklin WA, Gazdar AF, Haney J, Wistuba II, La Rosa FG, Kennedy T, et al. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *J Clin Invest* 1997; 100:2133-7.
- Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007;13:361-6.
- Kennedy TC, McWilliams A, Edell E, Sutedja T, Downie G, Yung R, et al. Bronchial intraepithelial neoplasia/early central airways lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* 2007;132:221S-33S.
- Bota S, Auliac JB, Paris C, Metayer J, Sesboue R, Nouvet G, et al. Follow-up of bronchial precancerous lesions and carcinoma *in situ* using fluorescence endoscopy. *Am J Respir Crit Care Med* 2001; 164:1688-93.
- Miller YE, Blatchford P, Hyun DS, Keith RL, Kennedy TC, Wolf H, et al. Bronchial epithelial Ki-67 index is related to histology, smoking, and gender, but not lung cancer or chronic obstructive pulmonary disease. *Cancer Epidemiol Biomarkers Prev* 2007;16:2425-31.
- Paris C, Benichou J, Bota S, Sagnier S, Metayer J, Eloy S, et al. Occupational and nonoccupational factors associated with high grade bronchial pre-invasive lesions. *Eur Respir J* 2003;21:332-41.
- Keith RL, Miller YE, Gemmill RM, Drabkin HA, Dempsey EC, Kennedy TC, et al. Angiogenic squamous dysplasia in bronchi of individuals at high risk for lung cancer. *Clin Cancer Res* 2000;6:1616-25.
- Banerjee AK. Preinvasive lesions of the bronchus. *J Thorac Oncol* 2009;4:545-51.
- Loewen G, Natarajan N, Tan D, Nava E, Klippenstein D, Mahoney M, et al. Autofluorescence bronchoscopy for lung cancer surveillance based on risk assessment. *Thorax* 2007;62:335-40.
- McWilliams A, Mayo J, MacDonald S, IeRiche JC, Palcic B, Szabo E, et al. Lung cancer screening: a different paradigm. *Am J Respir Crit Care Med* 2003;168:1167-73.
- Franklin WA. Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia. *Chest* 2000;117:80S-9S.
- Brambilla E, Travis WD, Colby TV, Corrin B, Shimosato Y. The new World Health Organization classification of lung tumours. *Eur Respir J* 2001;18:1059-68.
- Rahman SM, Shyr Y, Yildiz PB, Gonzalez AL, Li H, Zhang X, et al. Proteomic patterns of preinvasive bronchial lesions. *Am J Respir Crit Care Med* 2005;172:1556-62.
- Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 2003;362:433-9.
- Hirsch FR, Prindiville SA, Miller YE, Franklin WA, Dempsey EC, Murphy JR, et al. Fluorescence versus white-light bronchoscopy for detection of preneoplastic lesions: a randomized study. *J Natl Cancer Inst* 2001;93:1385-91.
- Aerni HR, Cornett DS, Caprioli RM. Automated acoustic matrix deposition for MALDI sample preparation. *Anal Chem* 2006;78: 827-34.
- Chen S, Li M, Hong D, Billheimer D, Li H, Xu BJ, et al. A novel comprehensive wave-form MS data processing method. *Bioinformatics* 2009;25:808-14.
- Shyr Y, Kim K. Weighted flexible compound covariate method for classifying microarray data. In: Berrar D, editor. *A Practical Approach to Microarray Data Analysis*. New York: Kluwer Academic; 2003:186-200.
- Efron BaT R. *An introduction to the Bootstrap*. New York: Chapman and Hall; 1993.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49:7-18.

23. Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat Med* 2001;7:493–6.
24. Hardesty WM. Proteomic analysis and classification of metastatic melanoma by MALDI imaging mass spectrometry. Nashville, TN: Vanderbilt University; 2010.
25. Chaurand P, Fouchecourt S, DaGue BB, Xu BJ, Reyzer ML, Orgebin-Crist MC, et al. Profiling and imaging proteins in the mouse epididymis by imaging mass spectrometry. *Proteomics* 2003;3:2221–39.
26. Haussinger K, Becker H, Stanzel F, Kreuzer A, Schmidt B, Strausz J, et al. Autofluorescence bronchoscopy with white light bronchoscopy compared with white light bronchoscopy alone for the detection of precancerous lesions: a European randomised controlled multicentre trial. *Thorax* 2005;60:496–503.
27. Auerbach O, Forman JB, Gere JB, Kassouny DY, Muehsam GE, Petrick TG, Smolin HJ, et al. Changes in the bronchial epithelium in relation to smoking and cancer of the lung: report of progress. *N Engl J Med* 1957;256:97–104.
28. Cornett DS, Reyzer ML, Chaurand P, Caprioli RM. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat Methods* 2007;4:828–33.
29. Neess D, Kailerich P, Sandberg MB, Helledie T, Nielsen R, Mandrup S. ACBP—a PPAR and SREBP modulated housekeeping gene. *Mol Cell Biochem* 2006;284:149–57.
30. Li W, Ding F, Zhang L, Liu Z, Wu Y, Luo A, et al. Overexpression of stefin A in human esophageal squamous cell carcinoma cells inhibits tumor cell growth, angiogenesis, invasion, and metastasis. *Clin Cancer Res* 2005;11:8753–62.
31. Strojjan P, Budihna M, Smid L, Svetic B, Vrhovec I, Kos J, et al. Prognostic significance of cysteine proteinases cathepsins B and L and their endogenous inhibitors stefins A and B in patients with squamous cell carcinoma of the head and neck. *Clin Cancer Res* 2000;6:1052–62.
32. Leinonen T, Pirinen R, Bohm J, Johansson R, Rinne A, Weber E, et al. Biological and prognostic role of acid cysteine proteinase inhibitor (ACPI, cystatin A) in non-small-cell lung cancer. *J Clin Pathol* 2007;60:515–9.
33. Herring K. Identification of protein markers of drug-induced nephrotoxicity by MALDI MS: in vivo discovery of ubiquitin-T. Nashville: Vanderbilt University; 2009.
34. McClelland M, Zhao L, Carskadon S, Arenberg D. Expression of CD74, the receptor for macrophage migration inhibitory factor, in non-small cell lung cancer. *Am J Pathol* 2009;174:638–46.
35. Campa MJ, Wang MZ, Howard B, Fitzgerald MC, Patz EF Jr. Protein expression profiling identifies macrophage migration inhibitory factor and cyclophilin A as potential molecular targets in non-small cell lung cancer. *Cancer Res* 2003;63:1652–6.
36. Rendon BE, Roger T, Teneng I, Zhao M, Al-Abed Y, Calandra T, et al. Regulation of human lung adenocarcinoma cell migration and invasion by macrophage migration inhibitory factor. *J Biol Chem* 2007;282:29910–8.
37. Kobayashi T, Okada F, Fujii N, Tomita N, Ito S, Tazawa H, et al. Thymosin-beta4 regulates motility and metastasis of malignant mouse fibrosarcoma cells. *Am J Pathol* 2002;160:869–82.
38. Cha HJ, Jeong MJ, Kleinman HK. Role of thymosin beta4 in tumor metastasis and angiogenesis. *J Natl Cancer Inst* 2003;95:1674–80.
39. Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 2003;22:8031–41.
40. Xu BJ, Gonzalez AL, Kikuchi T, Yanagisawa K, Massion PP, Wu H, et al. MALDI-MS derived prognostic protein markers for resected non-small cell lung cancer. *Proteomics Clin Appl* 2008;2:1508–17.
41. Nicholson AG, Perry LJ, Cury PM, Jackson P, McCormick CM, Corrin B, et al. Reproducibility of the WHO/IASLC grading system for pre-invasive squamous lesions of the bronchus: a study of inter-observer and intra-observer variation. *Histopathology* 2001;38:202–8.
42. Kujan O, Khattab A, Oliver RJ, Roberts SA, Thakker N, Sloan P. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: an attempt to understand the sources of variation. *Oral Oncol* 2007;43:224–31.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Lung Cancer Diagnosis from Proteomic Analysis of Preinvasive Lesions

S.M. Jamshedur Rahman, Adriana L. Gonzalez, Ming Li, et al.

*Cancer Res* 2011;71:3009-3017. Published OnlineFirst April 12, 2011.

<b>Updated version</b>	Access the most recent version of this article at: doi: <a href="https://doi.org/10.1158/0008-5472.CAN-10-2510">10.1158/0008-5472.CAN-10-2510</a>
<b>Supplementary Material</b>	Access the most recent supplemental material at: <a href="http://cancerres.aacrjournals.org/content/suppl/2011/04/08/0008-5472.CAN-10-2510.DC1">http://cancerres.aacrjournals.org/content/suppl/2011/04/08/0008-5472.CAN-10-2510.DC1</a>

<b>Cited articles</b>	This article cites 36 articles, 12 of which you can access for free at: <a href="http://cancerres.aacrjournals.org/content/71/8/3009.full#ref-list-1">http://cancerres.aacrjournals.org/content/71/8/3009.full#ref-list-1</a>
<b>Citing articles</b>	This article has been cited by 3 HighWire-hosted articles. Access the articles at: <a href="http://cancerres.aacrjournals.org/content/71/8/3009.full#related-urls">http://cancerres.aacrjournals.org/content/71/8/3009.full#related-urls</a>

<b>E-mail alerts</b>	<a href="#">Sign up to receive free email-alerts</a> related to this article or journal.
<b>Reprints and Subscriptions</b>	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a> .
<b>Permissions</b>	To request permission to re-use all or part of this article, use this link <a href="http://cancerres.aacrjournals.org/content/71/8/3009">http://cancerres.aacrjournals.org/content/71/8/3009</a> . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.