Proteomic Patterns of Preinvasive Bronchial Lesions

S. M. Jamshedur Rahman, Yu Shyr, Pinar B. Yildiz, Adriana L. Gonzalez, Huiming Li, Xueqiong Zhang, Pierre Chaurand, Kiyoshi Yanagisawa, Bonnie S. Slovis, Robert F. Miller, Mathew Ninan, York E. Miller, Wilbur A. Franklin, Richard M. Caprioli, David P. Carbone, and Pierre P. Massion

Division of Allergy, Pulmonary, and Critical Care Medicine, Department of Medicine; Departments of Biostatistics and Pathology; Mass Spectrometry Research Center; Division of Hematology and Oncology, Departments of Surgery and Cancer Biology, Vanderbilt-Ingram Comprehensive Cancer Center, Vanderbilt University School of Medicine; Veterans Affairs Medical Center, Nashville, Tennessee; and Department of Medicine and Pathology, University of Colorado Health Science Center, Denver, Colorado

Purpose: A proteomics approach is warranted to further elucidate the molecular steps involved in lung tumor development. We asked whether we could classify preinvasive lesions of airway epithelium according to their proteomic profile.

Experimental Design: We obtained matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiles from 10- μ m sections of fresh-frozen tissue samples: 25 normal lung, 29 normal bronchial epithelium, and 20 preinvasive and 36 invasive lung tumor tissue samples from 53 patients. Proteomic profiles were calibrated, binned, and normalized before analysis. We performed class comparison, class prediction, and supervised hierarchic cluster analysis. We tested a set of discriminatory features obtained in a previously published dataset to classify this independent set of normal, preinvasive, and invasive lung tissues.

Results: We found a specific proteomic profile that allows an overall predictive accuracy of over 90% of normal, preinvasive, and invasive lung tissues. The proteomic profiles of these tissues were distinct from each other within a disease continuum. We trained our prediction model in a previously published dataset and tested it in a new blinded test set to reach an overall 74% accuracy in classifying tumors from normal tissues.

Conclusions: We found specific patterns of protein expression of the airway epithelium that accurately classify bronchial and alveolar tissue with normal histology from preinvasive bronchial lesions and from invasive lung cancer. Although further study is needed to validate this approach and to identify biomarkers of tumor development, this is a first step toward a new proteomic characterization of the human model of lung cancer tumorigenesis.

Keywords: early detection; mass spectrometry; preneoplasia; profiling

The discovery of preinvasive lesions in the high-risk population places patients at increased risk of developing lung cancer (1–3). Although the natural history of these preinvasive lesions is poorly understood, increasing evidence suggests that, in the absence of treatment, high-grade preinvasive lesions, most commonly found in patients with a prior history or with concomitant cancer, will develop into invasive carcinoma in 30 to 50% of the cases, whereas the vast majority of low-grade preinvasive lesions remain stable or regress on follow-up (4–7). In addition, nonstep-

wise progression of preinvasive lesions makes prediction of lung cancer progression based on histologic grade unreliable. Therefore, the need for molecular biomarkers predictive of tumor progression is becoming more evident (8–10).

In a recent report, we used conventional matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI MS) to generate protein profiles that accurately classify and predict histologic subgroups of lung cancer (11). Using biostatistical methods to select differentially expressed peaks (MS signals) and after the development of a class prediction model (12), 82 discriminatory signals were found to classify normal lung from lung cancer tissue samples in a derivation and validation study design with excellent accuracy. Other recent studies indicated the importance of using protein expression profiles as a diagnostic or prognostic biomarker for patients with early-stage lung cancer using two-dimensional polyacrylamide gel electrophoresis analysis (13, 14), protein arrays (15), or surface-enhanced laser desorption/ionization MS (16). Although new strategies are needed for the early detection of lung cancer, clinical proteomics applied directly to the airway epithelium may provide a novel approach for the selection of biomarkers.

Our primary goal is to establish an MALDI MS-based proteomics approach to obtain specific patterns of protein expression of the airway epithelium at different stages of tumor development. We hypothesize that airway epithelium (alveolar and bronchial) modify its proteomic expression profile during tumor development. In this first report, we demonstrated that proteomic profiles obtained by MALDI MS allow the classification of normal alveolar or normal bronchial epithelium, preinvasive lesions, and invasive lung tumors and may provide new insights to the biology of these preinvasive lesions of unclear clinical significance. Some of the results of these studies have been previously reported in abstract form (17).

METHODS

Patients and Tissue Samples

A total of 53 patients from Vanderbilt University Medical Center (VUMC) participated in this study between January 2002 and November 2003. A total of 110 samples were obtained from bronchoscopic or surgical procedures. Eighteen of 29 normal bronchial epithelium samples and 18 of 20 preinvasive tissues were obtained at the time of bronchoscopy, and all 25 normal alveolar and all 36 lung tumor tissues were obtained at the time of surgery. Normal alveolar specimens were obtained from at least 2 cm away from the tumor. Endobronchial biopsies were obtained by white-light bronchoscopy and laser-induced fluorescence endoscopy from VUMC and University of Colorado Health Science Center (UCHSC) between May 2002 and December 2003. Biopsies were obtained from patients with lung cancer or at risk of lung cancer from predetermined as well as suspicious sites. Of the 10 patients without concomitant lung cancer, five were recruited from the UCHSC lung cancer cohort (with sputum atypia), one patient was a donor of bronchial epithelium as a motor vehicle accident victim, and four patients were recruited from VUMC on the basis of clinical suspicion of lung cancer. Squamous metaplasia and mild dysplasia were

(Received in original form February 19, 2005; accepted in final form September 20, 2005)

Supported by the Damon Runyon Cancer Research Foundation (P.P.M. is a Damon Runyon-Lilly Clinical Investigator, CI 19-03). Also supported in part by the Lung SPOR P50 CA 90949 to D.P.C. and by the Flight Attendant Medical Research Institute to P.P.M.

Correspondence and requests for reprints should be addressed to Pierre P. Massion, M.D., Division of Allergy, Pulmonary and Critical Care Medicine, Vanderbilt-Ingram Comprehensive Cancer Center, 2220 Pierce Avenue, 640 Preston Research Building, Nashville, TN 37232-6838. E-mail: pierre.massion@vanderbilt.edu

Am J Respir Crit Care Med Vol 172, pp 1556–1562, 2005

Originally Published in Press as DOI: 10.1164/rccm.200502-2740C on September 22, 2005

Internet address: www.atsjournals.org

Figure 1. Representative raw matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI MS) spectra of normal, preinvasive, and invasive squamous carcinomas of the lung. Intensity ion current (*y* axis) plotted against increasing mass (*m/z*) ratios from 1,999 to 23,387 daltons (*x* axis).

grouped as low-grade preinvasive lesion, whereas moderate dysplasia, severe dysplasia, and carcinoma *in situ* were grouped as high-grade preinvasive lesion. Bronchial epithelial lesions were graded according to the World Health Organization nomenclature (18). The project was approved by the local institutional review board, and informed consent was obtained for all individuals.

Preparation of Tissue Samples

Tissue sections were cut into 10-μm sections and stored in a -140°C freezer until use. Only a minimum amount of optimal cutting temperature (OCT) medium was used to immobilize the tissues, avoiding em-

bedding the area to be sectioned for MALDI MS. Sections were directly transferred onto a gold-coated stainless steel sample plate (PE Biosystems, Foster City, CA) and a glass slide. Because of their small size, bronchial biopsies were embedded into OCT medium. To remove OCT medium from bronchial biopsy sections, water wash was performed at room temperature. The sections were dried for 45 min in a dessicator. An adjacent section was stained with hematoxylin and eosin as a guide, and the area to be analyzed by MALDI MS was precisely marked by one of the present authors (A.L.G.), the lung pathologist at Vanderbilt University, who was trained by another pathologist at UCHSC (W.A.F.). Samples from UCHSC were examined by W.A.F. Matrix

Figure 2. Supervised hierarchical cluster analysis of (A) low-grade (lg; *n* = 13) versus high-grade (hg; *n* = 7) preinvasive lesions and (B) high-grade preinvasive lesions versus invasive tumors (t; *n* = 36).

solution (sinapinic acid, 35 mg/dl in water:acetonitrile:trifluoroacetic acid, 500/500/1, vol/vol/vol) was deposited before MS analysis. Matrix deposition was performed with a fine capillary needle under microscopic guidance.

Proteomic Analysis

Protein expression profiles were obtained as described previously (11, 19–21). Briefly, each spectrum was acquired over the surface of the matrix spot. In this analysis, signals in the range between 2,000 and 20,000 mass-to-charge ratios were considered. Each spectrum underwent smoothing, baseline correction, and internal calibration, with peaks from internal hemoglobin α and β chains using Data Explorer software (Applied Biosystems, Foster City, CA). Peak list obtained from the spectra was normalized and binned with a custom algorithm written by the group at VUMC, resulting in 1,261 bins in total, and then submitted for statistical analyses. Representative spectra from each histologic subtype are shown in Figure 1. Reproducibility and variability issues of spectra acquisition were addressed in previous reports (11, 22, 23) and confirmed in this dataset. The intrasample variability for the top 100 peaks was less than 30% of the overall variability (intra- and intersample variability).

Alignment, Binning, and Normalization of the MS Profiles

Alignment of MALDI MS peaks across multiple spectra was accomplished by use of a computing algorithm that simultaneously determines an optimal set of “bins” for categorizing each peak as a specific protein (24). Optimal bins are identified as those that maximize the number of single peaks within each bin across the spectra while minimizing the number of bins that have multiple peaks within a spectrum. We have demonstrated that this is an effective approach in a recent proteomics study of lung cancer (11). Spectra were normalized to each other according to the maximum sum of total ion current before binning. The maximum sum was divided by the sum of each spectrum to obtain the normalization factor for each spectrum. This factor was then used to multiply each data point for each spectrum.

Statistical Analysis

The primary objective of this study was to identify a set of proteins expressed differentially among study groups. The statistical analyses were focused on the following steps:

1. Selection of MS features that were differentially expressed between the study groups was performed. The selection was based on the Kruskal-Wallis test, Fisher's exact test (dichotomize the expression level as present or not), the permutation t test, significance analysis of microarrays (25), weighted gene analysis (26), and the modified information score method (27). The cutoff points for each method were $p < 0.0005$, $p < 0.0005$, $p < 0.0005$, 2, 2, and 0, respectively. Because there is no standard statistical method for analyzing the proteomic data, the selection of the “winner” markers was based on these six different statistical methods. A feature was included on the final list if it met at least two of these six criteria.
2. The “winner” features that have reasonable predictive power

were examined. The statistical class-prediction model based on the selected features was applied to determine whether the proteomic patterns could be used to classify normal from preinvasive and invasive tissue samples. The weighted flexible compound covariate method (WFCCM) (11, 12, 28, 29) was used in the class-prediction model based on the selected features to determine whether the proteomic patterns could be used to classify tissue samples among groups. We estimated the misclassification rate using the leave-one-out cross-validation class-prediction method based on the WFCCM.

3. The prediction model generated from the training set was applied to the blinded test cohort, and the samples were classified based on the closeness of the distance from the two classes.
4. The agglomerative hierarchic clustering algorithm (28) and multi-dimensional scaling (MDS) method were applied to investigate the pattern among the statistically significant discriminatory features as well as disease status using M. Eisen's software (30) and MATLAB (MathWorks Inc., Natick, MA), respectively. MDS can be considered to be an alternative to factor analysis. In general, the goal of the analysis is to detect meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities (distances) between the investigated objects. Multiple comparisons in this study were adjusted by using the false discovery rate (FDR). FDR is a new approach to the multiple comparisons problem. Instead of controlling the chance of any false positives (as in Bonferroni's adjustment), the FDR of a set of predictions is the expected percentage of false predictions in the set of markers.

RESULTS

Study Population

To obtain specific patterns of protein expression of the airway epithelium at different stages of tumor development, we obtained MALDI MS profiles on 110 tissue samples from a total of 53 individuals, 43 of whom were patients with concomitant lung cancer. The patients' characteristics are presented in Table 1. The nature of the samples obtained in patients with specific tumor histology is provided in Table 2. Because of the heterogeneous nature of lung cancer and because the nature of the cell type of origin of these tumors remains a subject of controversy, both alveolar epithelium and bronchial epithelium were included as controls for lung tumors examined. Some patients provided more than one type of tissue sample as shown in Table 3. These samples included 25 histologically normal lung (alveolar space), 29 histologically normal bronchial epithelium, 20 preinvasive tumor tissue samples (13 low-grade and 7 high-grade), and 36 invasive tumor tissue samples.

TABLE 1. PATIENTS' CHARACTERISTICS

	Total (<i>n</i> = 53)	Alveolar (<i>n</i> = 25/53)	Bronchial (<i>n</i> = 21/53)	Preinvasive (<i>n</i> = 14/53)	Invasive (<i>n</i> = 34/53)
Age	65 ± 11	68 ± 11	64 ± 11	62 ± 10	65 ± 12
Sex					
Male	35	19	13	8	24
Female	18	6	8	6	10
Smoking status					
Nonsmoker	4	1	3		1
Ex-smoker	30	18	9	4	25
Smoker	19	6	9	10	9
Smoking, pack/yr	56 ± 34	56 ± 35	55 ± 38	67 ± 22	52 ± 33
With cancer	43	25	15	7	34
Without cancer	10	0	6	7	0

TABLE 2. NATURE OF 110 SAMPLES AND ASSOCIATION WITH CANCER

	Alveolar (<i>n</i> = 25 from 25 patients)	Bronchial (<i>n</i> = 29 from 21 patients)	Preinvasive (<i>n</i> = 20 from 14 patients)	Invasive (<i>n</i> = 36 from 34 patients)
Samples associated with cancer, no.	25	20	11	36
Squamous cell carcinoma	8	8	1	17
Adenocarcinoma	17	6	5	17
Non-small cell lung cancer	0	3	4	2
Small cell lung cancer	0	3	1	0
Samples not associated with cancer, no.	—	9	9	—

Preinvasive Lesions Differ from Normal Airway Epithelium and from Invasive Lung Cancer

For each histologic subgroup, we obtained a proteomic profile that distinguishes from other lesions with predictive accuracy between 83 and 100%. Results are summarized in Table 4. Examples of such profiles are provided in Figure 1. Classification and misclassification rates were calculated by use of leave-one-out cross-validation class prediction method based on our covariate method of analysis. To avoid the possibility of overfitting, we reported only one set of “winner” features (based on the pre-termination cut-off of the *p* values), which was applied to the prediction model. Normal bronchial epithelium proteomic profiles clearly differ from those of invasive cancer and all other classes. Preinvasive lesions provided different proteomic patterns from normal bronchial epithelium or invasive cancer. Importantly, and despite our relatively small sample size, we found patterns characteristic of either low- or high-grade lesions based on the different expression of 14 features (Figure 2A). We performed the same comparisons using agglomerative hierarchic cluster analysis with very similar results. For example, hierarchic cluster analysis of high-grade preinvasive lesion versus invasive tumor is shown in Figure 2B. The potential issue of multiple comparisons in high-dimensional data analysis is adjusted by the FDR (*see* STATISTICAL ANALYSIS).

Distinguishing Normal from Preinvasive Lesions from Cancer within a Continuum

In an effort to distinguish proteomic profiles of normal, preinvasive, and invasive tumors in a continuum, we analyzed the profiles obtained from normal bronchial epithelium, preinvasive bronchial epithelium, and squamous carcinomas of the lung. Results of supervised clustering analysis (Figure 3A) and of multidimensional scaling analysis of these three groups (Figure 3B) show distinct clustering of the groups in a continuum. Bio-

markers found in preinvasive and in invasive lesions but not in normal airway epithelium of subjects with or at risk for lung cancer would be of particular interest in the development of markers for early detection. We searched for such features in our dataset and, although some specific features are completely absent from normal bronchial tissues, they were found in 0 to 15% of low-grade preinvasive lesions, in 14 to 43% of high-grade lesions, and in 6 to 24% of the squamous invasive tumor tissues; none of these candidate biomarkers reached statistical significance set at *p* < 0.0005.

Proteomic Profiles of Preinvasive Lesions Resemble Those of Normal Epithelium

After training our prediction model to detect features that discriminate between normal bronchial epithelium and invasive tumor, we asked how the model would classify low- and high-grade lesions. Proteomic profiles obtained from 12 of 13 low-grade and all high-grade preinvasive lesions were classified by 54 discriminatory features as normal airway epithelium.

Validation of Prediction Model in Normal and Invasive Lung Cancer

Finally, we trained our prediction model in our previously published dataset of 14 normal lung tissue samples and 66 tumors (11) and tested the discriminatory profile in our new and independent test set of 36 tumors, 25 alveolar lung tissues, and 29 bronchial specimens. The two datasets were binned and normalized together. We reached an overall 74% rate of accuracy in classifying tumors from normal tissues. Specifically, we correctly classified 18 of 25 of the alveolar lung tissues, 23 of 29 of the normal bronchial specimens, and 26 of 36 tumors.

DISCUSSION

We report for the first time proteomic expression profiles specific to different stages of lung tumor development. We detected MALDI MS signals that were discriminatory and predictive of alveolar and bronchial epithelium, low- and high-grade preinvasive lesions, and invasive lung tumors with overall 90% accuracy. When proteomic profiles of preinvasive lesions were tested using patterns that distinguish between normal airway and invasive tumors obtained in a training set, the majority of preinvasive lesions were classified as normal-appearing epithelium. We also demonstrated in a training and test paradigm that a set of discriminatory features obtained in an independent, previously published set of normal and tumor samples correctly classified 74% of our samples as tumor or normal epithelium.

Although the risk of developing lung cancer increases with the presence of preinvasive lesions, the molecular determinants predicting the irreversible progression to lung cancer have not been identified. Previous studies have addressed the genetic basis for classification of preinvasive lesions in a model of tumor development (1). Yet, none of these biomarkers are predictive

TABLE 3. NUMBER OF PATIENTS PROVIDING DIFFERENT TISSUES

Patients without cancer providing tissues, no.	10
Bronchial	3
Bronchial and preinvasive	3
Preinvasive	4
Patients with cancer providing tissues, no.	43
Alveolar	2
Alveolar and bronchial	1
Alveolar, bronchial and invasive	6
Alveolar, bronchial, preinvasive and invasive	1
Alveolar and invasive	15
Bronchial	1
Bronchial and preinvasive	4
Bronchial and invasive	1
Bronchial, preinvasive and invasive	1
Preinvasive	1
Invasive	10

TABLE 4. CLASSIFICATION OF TISSUE SAMPLES ON THE BASIS OF DISCRIMINATORY FEATURE

Class Comparisons	No. Misclassified	Correctly Classified (%)	No. Discriminatory Features
Normal alveolus and normal bronchus (54) versus low-grade preinvasive (13)	2/54, 0/13	97	115
Normal alveolus and normal bronchus (54) versus high-grade preinvasive (7)	3/54, 1/7	93	36
Normal alveolus and normal bronchus (54) versus invasive tumor (36)	5/54, 2/36	92	54
Normal alveolus and normal bronchus (54) versus low- and high-grade preinvasive and invasive tumor (56)	12/54, 7/56	83	59
Normal bronchus (29) versus low- and high-grade preinvasive and squamous tumor (37)	0/29, 5/37	92	116
Low-grade preinvasive (13) versus high-grade (7) preinvasive lesion	0/13, 0/7	100	14
Low-grade preinvasive (13) versus invasive tumor (36)	1/13, 3/36	92	19
High-grade preinvasive (7) versus invasive tumor (36)	1/7, 1/36	95	55
Normal alveolus (25) versus normal bronchus (29)	0/25, 2/29	96	68
Normal alveolus and normal bronchus and low-grade preinvasive (67) versus high-grade preinvasive and invasive tumor (43)	9/67, 8/43	85	44

of progression to invasive cancer. The proteomic analysis of tumor development might provide a new understanding of the pathologic states a given tumor may undergo before acquiring its invasive phenotype. Our previous study, which defined profiles that allow classification of lung tumors from normal lung and from tumors associated with lymph node involvement (11), clearly showed the strength of tissue-based proteomics as an analytic tool.

This study provides evidence for a specific phenotype of the airway epithelium as it progresses from normal to preinvasive and to invasive cancer. The findings of the supervised cluster and MDS analyses, although they do not prove true progression, suggest a continuum between proteomic patterns of lesions as they progress toward invasive lesions. Yet, we did not find significant MALDI MS features robust enough to pass our conservative statistical approach that would be present in preinvasive and invasive lesions but not in normal bronchial epithelium. Several considerations may explain this lack of specific markers of tumor development. The study design (cross-sectional as opposed to longitudinal), the sample size, and the analysis of only a fraction of the proteome offered by the MALDI MS approach need to be taken into consideration. Biological variability may also prevent us from finding markers of tumor development because only a small percentage of preinvasive lesions are known to transform into cancer (7). The samples examined may not have harbored these features. A prospective study following the

natural history of a large number of these lesions may answer this question.

This report is our first step toward the selection of proteomic profiles predictive of lung cancer development. We found that this tissue-based proteomics approach allows us not only to recapitulate the classification of preinvasive lesions on the basis of histologic grade but to provide biological information about these lesions. Indeed, when testing proteomic predictors of lung cancer versus normal tissue to classify profiles from preinvasive lesions (as being closer to normal or cancer), both low- and high-grade lesions clustered with normal epithelium. These findings suggest that only a subset of preinvasive lesions (yet to be identified in a prospective study) may in fact progress to an invasive phenotype, and histology by itself may not be the ideal surrogate marker of tumor progression. This observation is in agreement with higher regression and low progression of preinvasive lesions shown in earlier studies (7, 31, 32).

With the current and independent set of tissues, we were able to reproduce similar results to those obtained in our previous report (11). Despite the inclusion of normal bronchial samples in the test set that were not represented in the training set (i.e., those of a separate operator [S.M.J.R.]) and of differences in the data processing (*see METHODS*), we were able to correctly classify 74% of the samples as normal epithelium or as invasive tumor. These results further validate our proteomic strategy.

Figure 3. Distinction of normal from preinvasive lesions from cancer within a continuum based on MALDI MS profiles. Supervised hierarchic cluster analysis (A) and multidimensional scaling (B) of normal bronchial epithelia ($n = 29$), preinvasive lesions ($n = 20$), and invasive squamous tumors ($n = 17$). Multidimensional scaling plot of normal lung (blue circles), preinvasive lesion of the lung (green squares), and invasive lung cancer (red triangles) tissues in 38 dimensional space according to the expression of 38 discriminatory features among the groups.

In our previous report (12), only alveolar tissues were representative of the control group. In the current study, because 18 of 20 preinvasive tissues were obtained from bronchial biopsies, we included a series of bronchial tissue as controls in addition to alveolar tissue. The difference between profiles of normal alveolar epithelium and bronchial epithelium is interesting and important for future research related to cellular origin of squamous carcinoma and adenocarcinoma.

Limitations to this study include considerations about patient selection, methodology, sample size, and protein identification. First, most of our preinvasive lesions and our normal-appearing epithelium were obtained from patients with concomitant lung cancer. The natural history of these preinvasive lesions may be different from those of patients without concomitant lung cancer. Also, we lack a true biological "normal" epithelium for reference because the majority of our patients were known to have lung cancer and were current or ex-smokers. Second, technical considerations may have affected the outcome of the analysis. The tissues analyzed were obtained from different sources (surgically resected specimen vs. endobronchial biopsies) and processed slightly differently (without or with OCT embedding). Although OCT medium prevents optimal ionization of proteins (33), the necessary removal of water-soluble OCT could introduce a difference in methodology, yet across all histologic subtypes and therefore unlikely to affect the analysis. In addition, the resolution of the MS, daily variation of the vacuum, and laser intensity of the instrument itself and its overall performance are variables that have been addressed before (11). Third, a relatively small sample size of preinvasive lesions requires confirmation in a larger dataset. Nevertheless, our data strongly illustrate the strength of proteomic profiles to discriminate pathologic states along the various steps of tumor development. And fourth, we report the predictive value of protein profiles and not proteins or peptides. The identification of the proteins or peptides making these specific signatures, though critical, is beyond the scope of this study because we first need to validate them in an independent test set. Ultimately, selection and identification of proteins predictive of what preinvasive lesion is likely to develop to lung cancer is a critical goal of future investigation.

In conclusion, the original selection of these discriminatory profiles by MALDI MS is a critical first step in the identification of proteomic expression patterns along lung cancer development. Ultimately, this MALDI MS-based approach in tissue may provide characterization of the molecular determinants leading preinvasive lesions to an invasive phenotype.

Conflict of Interest Statement: None of the authors have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: The authors thank the individuals who provided their informed consent and participated in the study. They thank Lynne Fenner, Blake Mann, Darienne Adkins, Heather Templeton, Candace Murphy, and Anthony Frazier for their assistance in consenting individuals, in obtaining clinical data elements and biological samples. They also thank Hans Rudolf Aerni and Lisa Manier for their expertise in proteomic analysis.

References

- Hirsch FR, Franklin WA, Gazdar AF, Bunn PA Jr. Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology. *Clin Cancer Res* 2001;7:5-22.
- Franklin WA. Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia. *Chest* 2000;117(4, Suppl 1):80S-89S.
- Prindiville SA, Byers T, Hirsch FR, Franklin WA, Miller YE, Vu KO, Wolf HJ, Baron AE, Shroyer KR, Zeng C, et al. Sputum cytological atypia as a predictor of incident lung cancer in a cohort of heavy smokers with airflow obstruction. *Cancer Epidemiol Biomarkers Prev* 2003;12:987-993.
- Risse EK, Vooijs GP, van't Hof MA. Diagnostic significance of "severe dysplasia" in sputum cytology. *Acta Cytol* 1988;32(5):629-634.
- Saccomanno G, Archer VE, Auerbach O, Saunders RP, Brennan LM. Development of carcinoma of the lung as reflected in exfoliated cells. *Cancer* 1974;33:256-270.
- Mulshine JL, Cuttitta F, Tockman MS, De Luca LM. Lung cancer evolution to preinvasive management. *Clin Chest Med* 2002;23:37-48.
- Breuer RH, Pasic A, Smit EF, van Vliet E, Vonk Noordegraaf A, Risse EJ, Postmus PE, Sutedja TG. The natural course of preneoplastic lesions in bronchial epithelium. *Clin Cancer Res* 2005;11:537-543.
- Wistuba II. Histologic evaluation of bronchial squamous lesions: any role in lung cancer risk assessment? *Clin Cancer Res* 2005;11:1358-1360.
- Chanin TD, Merrick DT, Franklin WA, Hirsch FR. Recent developments in biomarkers for the early detection of lung cancer: perspectives based on publications 2003 to present. *Curr Opin Pulm Med* 2004;10:242-247.
- Massion PP, Taflan PM, Shyr Y, Rahman SM, Yildiz P, Shakhthour B, Edgerton ME, Ninan M, Andersen JJ, Gonzalez AL. Early involvement of the phosphatidylinositol 3-kinase/Akt pathway in lung cancer progression. *Am J Respir Crit Care Med* 2004;170:1088-1094.
- Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, Roberts JR, Edgerton M, Gonzalez A, Nadaf S, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 2003;362:433-439.
- 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. pp. 186-200.
- Chen G, Gharib TG, Wang H, Huang CC, Kuick R, Thomas DG, Shedden KA, Misek DE, Taylor JM, Giordano TJ, et al. Protein profiles associated with survival in lung adenocarcinoma. *Proc Natl Acad Sci USA* 2003;100:13537-13542.
- Li C, Chen Z, Xiao Z, Wu X, Zhan X, Zhang X, Li M, Li J, Feng X, Liang S, et al. Comparative proteomics analysis of human lung squamous carcinoma. *Biochem Biophys Res Commun* 2003;309:253-260.
- Valle RP, Chavany C, Zhukov TA, Jendoubi M. New approaches for biomarker discovery in lung cancer. *Expert Rev Mol Diagn* 2003;3:55-67.
- Zhukov TA, Johanson RA, Cantor AB, Clark RA, Tockman MS. Discovery of distinct protein profiles specific for lung tumors and pre-malignant lung lesions by SELDI mass spectrometry. *Lung Cancer* 2003;40:267-279.
- Rahman S, Shyr Y, Yildiz P, Gonzalez A, Li H, Zhang X, Chaurand P, Yanagisawa K, Slovis B, Miller RD, et al. Proteomic patterns of preinvasive bronchial lesions in patients with lung cancer [abstract]. *Am J Respir Crit Care Med* 2004;169:A57.
- Travis W, Colby TV, Corrin B, Shimosato Y, Travis WD, Brambilla E. In: Sabin LH, editor. Histological typing of lung and pleural tumours (International Histological Classification of Tumours), 3rd ed. Berlin: Springer-Verlag; 1999.
- Chaurand P, Schwartz SA, Caprioli RM. Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections. *Curr Opin Chem Biol* 2002;6:676-681.
- 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;7:493-496.
- Xu BJ, Caprioli RM, Sanders ME, Jensen RA. Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. *J Am Soc Mass Spectrom* 2002;13:1292-1297.
- Schwartz SA, Weil RJ, Johnson MD, Toms SA, Caprioli RM. Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression. *Clin Cancer Res* 2004;10:981-987.
- Chaurand P, Sanders ME, Jensen RA, Caprioli RM. Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections. *Am J Pathol* 2004;165:1057-1068.
- Moore J, Parker J. Methods of microarray data analysis. Boston: Kluwer Academic; 2001.
- Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001;98:5116-5121.
- Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001;344:539-548.
- Ben'Dor AFN, Yakhini Z. Scoring genes for relevance. Tech Report AGL-2000-13, Agilent Labs, Agilent Technologies, 2000. Available from: <http://www.labs.agilent.com/resources/techreports.html> (accessed 2001).
- Yamagata N, Shyr Y, Yanagisawa K, Edgerton M, Dang TP, Gonzalez A, Nadaf S, Larsen P, Roberts JR, Nesbitt JC, et al. A training-testing

- approach to the molecular classification of resected non-small cell lung cancer. *Clin Cancer Res* 2003;9:4695-4704.
29. Tukey JW. Tightening the clinical trial. *Control Clin Trials* 1993;14:266-285.
30. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;95:14863-14868.
31. Bota S, Auliac JB, Paris C, Metayer J, Sesboue R, Nouvet G, Thiberville L. Follow-up of bronchial precancerous lesions and carcinoma in situ using fluorescence endoscopy. *Am J Respir Crit Care Med* 2001;164:1688-1693.
32. Banerjee AK, Rabbitts PH, George PJ. Are all high-grade preinvasive lesions premalignant, and should they all be treated? *Am J Respir Crit Care Med* 2002;165(10):1452-1453. [Author reply, 1453.]
33. Schwartz SA, Reyzer ML, Caprioli RM. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. *J Mass Spectrom* 2003;38:699-708.