

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 October 2005 (20.10.2005)

PCT

(10) International Publication Number  
**WO 2005/098445 A2**

(51) International Patent Classification<sup>7</sup>: **G01N 33/574**

(21) International Application Number:  
PCT/US2005/010575

(22) International Filing Date: 30 March 2005 (30.03.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/557,404 30 March 2004 (30.03.2004) US

(71) Applicant (for all designated States except US): **EAST-ERN VIRGINIA MEDICAL SCHOOL** [US/US]; 825 Fairfax Avenue, P.O. Box 1980, Norfolk, VA 23501-6006 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SEMMES, Oliver, John** [US/US]; 704 Abilene Court, Chesapeake, VA 23322 (US). **CAZARES, Lisa, H.** [US/US]; 721 Fairfax Avenue, Norfolk, VA 23501-6006 (US).

(74) Agents: **WIXON, Henry, N.** et al.; Wilmer Cuter Pickering Hale and Dorr LLP, 1455 Pennsylvania Avenue, N.W., Washington, DC 2004 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

**WO 2005/098445 A2**

(54) Title: LUNG CANCER BIOMARKERS

(57) Abstract: Disclosed are protein biomarkers and their use in diagnosing lung cancer or to make a negative diagnosis in patients. Also disclosed are kits for the diagnosis of lung cancer that detect the protein biomarkers of the invention, as well as methods using a plurality of classifiers to make a probable diagnosis of lung cancer. In certain aspects of the invention, the methods include use of a decision tree analysis. Various computer readable media and their use according to the invention are also disclosed.

## LUNG CANCER BIOMARKERS

### STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

5 The present invention was made with Government support under grant number CA85067 awarded by the National Institutes of Health/National Cancer Institute. The Government may have certain rights in the invention.

### BACKGROUND OF THE INVENTION

10 Lung cancer is the most common form of cancer in the world. Typical diagnosis of lung cancer combines x-ray with sputum cytology. Unfortunately, by the time a patient seeks medical attention for their symptoms, the cancer is at such an advanced state it is usually incurable. Consequently, research has been focused on early detection of tumor markers before the cancer becomes clinically apparent and while the cancer is still localized and amenable to therapy.

15 Particular interest has been given to the identification of antigens associated with the lung cancer proteome. These antigens have been used in screening, diagnosis, clinical management, and potential treatment of lung cancer. For example, carcinoembryonic antigen (CEA) has been used as a tumor marker of several cancers, including lung cancer. (Nutini, et al. 1990. "Serum NSE, CEA, CT, CA 15-3 levels in human lung cancer," *Int J Biol Markers* 5:198-202).

20 Squamous cell carcinoma antigen (SCC) is another established serum marker. (Margolis, et al. 1994. "Serum tumor markers in non-small cell lung cancer," *Cancer* 73:605-609.). Other serum antigens for lung cancer include antigens recognized by monoclonal antibodies (MAb) 5E8, 5C7, and 1F10, the combination of which distinguishes between patients with lung cancer from those without. (Schepart, et al. 1988. "Monoclonal antibody-mediated detection of lung cancer

antigens in serum," *Am Rev Respir Dis* 138:1434-8). Serum CA 125, initially described as an ovarian cancer-associated antigen, has been investigated for its use as a prognostic factor in lung cancer. (Diez, et al. 1994. "Prognostic significance of serum CA 125 antigen assay in patients with non-small cell lung cancer," *Cancer* 73:136876). Other tumor markers studied for utilization in multiple biomarker assays for lung cancer include carbohydrate antigen CA19-9, neuron specific enolase (NSE), tissue polypeptide antigen (TPA), alpha fetoprotein (AFP), HCG beta subunit, and LDH. (Mizushima, et al. 1990. "Clinical significance of the number of positive tumor markers in assisting the diagnosis of lung cancer with multiple tumor marker assay," *Oncology* 47:43-48; Lombardi, et al. 1990. "Clinical significance of a multiple biomarker assay in patients with lung cancer," *Chest* 97:639-644; and Buccheri, et al. 1986. "Clinical value of a multiple biomarker assay in patients with bronchogenic carcinoma," *Cancer* 57:2389-2396).

Monoclonal antibodies to the antigens associated with lung cancer have been generated and examined as possible diagnostic and/or prognostic tools. For example, monoclonal antibodies for lung cancer were first developed to distinguish non-small cell lung carcinoma (NSCLC) which includes squamous, adenocarcinoma, and large cell carcinomas from small cell lung carcinomas (SCLC). (Mulshine, et al. 1983. "Monoclonal antibodies that distinguish non-small-cell from small-cell lung cancer," *J Immunol* 121:497-502). Other antibodies have also been developed as immunocytochemical stains for sputum samples to predict the progression of lung cancer. (Tockman, et al. 1988. "Sensitive and specific monoclonal antibody recognition of human lung cancer antigen on preserved sputum cells: a new approach to early lung cancer detection," *J Clin Oncol* 6:1685-1693). U.S. Pat. No. 4,816,402 discloses a murine hybridoma monoclonal antibody for determining bronchopulmonary carcinomas and possibly adenocarcinomas. Some monoclonal antibodies utilized in immunohistochemical studies of lung

carcinomas include MCA 44-3A6, L45, L20, SLC454, L6, and YH206. (Radosevich, et al. 1985. "Monoclonal antibody 44-3A6 as a probe for a novel antigen found on human lung carcinomas with glandular differentiation," *Cancer Res* 45:5808-5812).

In U.S. Pat. Nos. 5,589,579 and 5,773,579, a lung cancer marker antigen specific for non-  
5 small cell lung carcinoma was identified and designated LCGA (also known as HCAVIII and HCA XII). The antigen was found useful in methods for detection of non-small cell lung cancer and for potential production of antibodies and probes for treatment compositions.

Despite the numerous examples of isolated lung cancer antigens and subsequent production of MAb to these antigens, none has yet emerged that has changed clinical practice.  
10 (Mulshine, et al., "Applications of monoclonal antibodies in the treatment of solid tumors," In: *Biologic Therapy of Cancer*. Edited by V. T. Devita, S. Hellman, and S. A. Rosenberg. Philadelphia: J B Lippincott, 1991, pp. 563-588). Thus far, the immunoassays developed have failed to meet the need for early detection .

In addition, proteomic research similarly has not satisfied this need. Proteomic research  
15 traditionally involved two-dimensional gel electrophoresis to detect protein expression differences in tissue and body fluid specimens between healthy (control) groups and disease groups (Srinivas, P.R., et al., *Clin Chem.* 47:1901-1911 (2001); Adam, B.L., et al., *Proteomics* 1:1264-1270 (2001)). Although two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) has been the classical approach in exploring the proteome for separation and detection of  
20 differences in protein expression, it has its limitations in that it is cumbersome, labor intensive, suffers reproducibility problems, and is not easily applied in the clinical setting.

Overall, despite the identification and extensive study of several potential tumor markers, none has been found to have clinical utility as a diagnostic marker or screening tool for lung

cancer. It seems probable that given the complexity of the genetic and molecular alterations that occur in lung cancer cells, the expression pattern of these complex changes may hold more vital information in screening, diagnosis and prognosis than the individual molecular changes themselves.

5       Recent technological advances in proteomics have permitted the development of diagnostic tests for the detection of some cancers. For example, one such technology includes the ProteinChip® surface-enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS) (Kuwata, H., et al., *Biochem. Biophys. Res. Commun.* 245:764-773 (1998); Merchant, M. et al., *Electrophoresis* 21:1164-1177 (2000)). This system uses surface-enhanced 10 laser desorption/ionization time-of-flight (SELDI-TOF) mass spectrometry to detect proteins bound to a protein chip array. The SELDI system is an extremely sensitive and rapid method that analyzes complex mixtures of proteins and peptides. Applications of this technology show great potential for the early detection of prostate, breast, ovarian, bladder, and head and neck cancers (Li, J., et al., *Clin. Chem.* 48:1296-1304 (2002); Adam, B., et al., *Cancer Res.* 62:3609-15 3614 (2002); Cazares, L.H., et al., *Clin. Cancer Res.* 8:2541-2552 (2002); Petricoin, E.F., et al., *Lancet* 359:572-577 (2002); Petricoin, E.F. et al., *J. Natl. Cancer Inst.* 94:1576-1578 (2002); Vlahou, A., et al., *Amer. J. Pathology* 158:1491-1502 (2001); Wadsworth, J.T., et al., *Arch. Otolaryngol. Head Neck Surg.* 130:98-104 (2004)). For example, U.S. Provisional Application 20 15 Number 60/496,682 describes the use of SELDI ProteinChip® technology as a tool of interrogation for head and neck squamous cell carcinoma ("HNSCC") patients. This application describes how serum from HNSCC patients was compared to normal controls in order to develop HNSCC protein fingerprints for the diagnosis of HNSCC. However, to date, the use of SELDI had not been used to identify protein biomarkers for the detection of lung cancer.

Continued efforts to identify protein profiles or patterns that differentiate cancer from non-cancer could lead to earlier detection of lung cancer and the development of diagnostic tests for lung cancer. There is a need, then, for methods and compositions for the diagnosis of lung cancer that can be performed relatively fast and inexpensively, yet are clinically useful. The 5 present invention addresses this and other needs.

## SUMMARY OF THE INVENTION

The present invention provides, for the first time, novel protein markers that are differentially present in the samples of patients with lung cancer and in the samples of control 10 subjects. The present invention also provides sensitive and methods and kits that can be used as an aid for the diagnosis of lung cancer by detecting these novel markers. The measurement of these markers, alone or in combination, in patient samples, provides information that can be correlated with a probable diagnosis of lung cancer or a negative diagnosis (e.g., normal or disease-free). All the markers are characterized by molecular weight. The markers can be 15 resolved from other proteins in a sample by, e.g., chromatographic separation coupled with mass spectrometry, or by traditional immunoassays. In preferred embodiments, the method of resolution involves Surface-Enhanced Laser Desorption/Ionization ("SELDI") mass spectrometry, in which the surface of the mass spectrometry probe comprises absorbents that bind to the marker.

20 In one form of the invention, a method for aiding in, or otherwise making, a diagnosis includes detecting at least one protein biomarker in a test sample from a subject. The protein biomarkers have a molecular weight selected from the group consisting of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,

4301  $\pm$  21, 4645  $\pm$  23, 9495  $\pm$  47, 11625  $\pm$  60, 9288  $\pm$  46, 8631  $\pm$  43, 8933  $\pm$  45, 11728  $\pm$  59,  
14105  $\pm$  70, 11940  $\pm$  60, 8861  $\pm$  44, 9150  $\pm$  46, 10264  $\pm$  51, 17047  $\pm$  85, 10461  $\pm$  52, 13354  $\pm$   
67, 7471  $\pm$  37, 3821  $\pm$  19, 12135  $\pm$  60, 5968  $\pm$  30, 4614  $\pm$  23, 5182  $\pm$  25, 4069  $\pm$  20, 4634  $\pm$  23,  
11600  $\pm$  58, 30133  $\pm$  150, 11939  $\pm$  60, 17894  $\pm$  89, 11723  $\pm$  58, 11493  $\pm$  57, 4959  $\pm$  25, 2013  $\pm$   
5 10, 4370  $\pm$  22, 45862  $\pm$  226, 15105  $\pm$  75, 20898  $\pm$  104, 38099  $\pm$  190, 5873  $\pm$  27, 3668  $\pm$  18, 9091  
 $\pm$  45, 8491  $\pm$  42, 3391  $\pm$  16, 4130  $\pm$  20, 3136  $\pm$  15, 3441  $\pm$  17, 30952  $\pm$  154, 4029  $\pm$  20, 11253  $\pm$   
56, 3820  $\pm$  19, 3506  $\pm$  17, 4571  $\pm$  23, 6933  $\pm$  34, 3887  $\pm$  19, 8602  $\pm$  43, 4644  $\pm$  23, 8630  $\pm$  43,  
and 8674  $\pm$  43 Daltons. The method further includes correlating the detection with a probable  
diagnosis of lung cancer or a negative diagnosis.

10 In one embodiment, the correlation takes into account the amount of the marker or  
markers in the sample and/or the frequency of detection of the same marker or markers in a  
control.

In another embodiment, gas phase ion spectrometry is used for detecting the marker or  
markers. For example, laser desorption/ionization mass spectrometry can be used.

15 In another embodiment, laser desorption/ionization mass spectrometry used to detect  
markers comprises: (a) providing a substrate comprising an adsorbent attached thereto; (b)  
contacting the sample with the adsorbent; and (c) desorbing and ionizing the marker or markers  
with the mass spectrometer. Any suitable adsorbent can be used to bind one or more markers.  
For example, the adsorbent on the substrate can be a cationic adsorbent, an antibody adsorbent,  
20 etc.

In another embodiment, an immunoassay can be used for detecting the marker or  
markers.

In certain forms of the invention, the markers in the test sample from a subject may be detected in the following groups and may have the following molecular weights: about 3820, 3506, 4571, and 6933 Daltons or about 8603, 3887, 4644, 8630, 4301, and 8674 Daltons.

In another form of the invention, a method for monitoring the effectiveness of lung cancer treatment in a patient is provided. This method comprises obtaining a biological sample from a patient undergoing treatment for lung cancer, detecting the quantity of at least one protein biomarker in said sample, said protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ ,  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,  $4614 \pm 23$ ,  $5182 \pm 25$ ,  $4069 \pm 20$ ,  $4634 \pm 23$ ,  $11600 \pm 58$ ,  $30133 \pm 150$ ,  $11939 \pm 60$ ,  $17894 \pm 89$ ,  $11723 \pm 58$ ,  $11493 \pm 57$ ,  $4959 \pm 25$ ,  $2013 \pm 10$ ,  $4370 \pm 22$ ,  $45862 \pm 226$ ,  $15105 \pm 75$ ,  $20898 \pm 104$ ,  $38099 \pm 190$ ,  $5873 \pm 27$ ,  $3668 \pm 18$ ,  $9091 \pm 45$ ,  $8491 \pm 42$ ,  $3391 \pm 16$ ,  $4130 \pm 20$ ,  $3136 \pm 15$ ,  $3441 \pm 17$ ,  $30952 \pm 154$ ,  $4029 \pm 20$ ,  $11253 \pm 56$ ,  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ ,  $6933 \pm 34$ ,  $3887 \pm 19$ ,  $8602 \pm 43$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $8674 \pm 43$  Daltons, comparing the quantity of said at least one protein biomarker to a known standard, and determining the effectiveness of said lung cancer treatment. The known standard can be a biological sample from a healthy control or a biological sample obtained from the lung cancer patient prior to the lung cancer treatment.

In accordance with the present invention, at least one of the biomarkers described herein may be detected. It is to be understood, and is described herein, that one or more of the biomarkers may be detected and subsequently analyzed, including all of the biomarkers. Further,

it is to be understood that the failure to detect one or more of the biomarkers of the invention, or the detection thereof at levels or quantities that may correlate with the absence of clinical or pre-clinical lung cancer, may be useful and desirable as a means of diagnosing the absence of clinical or pre-clinical lung cancer, and that the same forms a contemplated aspect of the present

5 invention.

In yet another aspect of the invention, kits that may be utilized to detect the biomarkers described herein and may otherwise be used to diagnose, or otherwise aid in the diagnosis of lung cancer, are provided. In one form of the invention, a kit may include a substrate comprising an adsorbent attached thereto, wherein the adsorbent is capable of retaining at least one protein

10 biomarker having a molecular weight selected from the group consisting of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ ,  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,  $4614 \pm 23$ ,  $5182 \pm 25$ ,  $4069 \pm 20$ ,  $4634 \pm 23$ ,  $11600 \pm 58$ ,  $30133 \pm 150$ ,  $11939 \pm 60$ ,  $17894 \pm 89$ ,  $11723 \pm 58$ ,  $11493 \pm 57$ ,  $4959 \pm 25$ ,  $2013 \pm 10$ ,  $4370 \pm 22$ ,  $45862 \pm 226$ ,  $15105 \pm 75$ ,  $20898 \pm 104$ ,  $38099 \pm 190$ ,  $5873 \pm 27$ ,  $3668 \pm 18$ ,  $9091 \pm 45$ ,  $8491 \pm 42$ ,  $3391 \pm 16$ ,  $4130 \pm 20$ ,  $3136 \pm 15$ ,  $3441 \pm 17$ ,  $30952 \pm 154$ ,  $4029 \pm 20$ ,  $11253 \pm 56$ ,  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ ,  $6933 \pm 34$ ,  $3887 \pm 19$ ,  $8602 \pm 43$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $8674 \pm 43$  Daltons; and instructions to detect the protein biomarker by contacting

20 a test sample with the adsorbent and detecting the biomarker retained by the adsorbent.

In yet another embodiment of the invention, the kit may include a substrate comprising an adsorbent attached thereto, wherein the adsorbent is capable of retaining at least one protein biomarker having a molecular weight selected from the group consisting of about  $4748 \pm 25$ ,

8603  $\pm$  43, 8675  $\pm$  43, 7566  $\pm$  38, 7972  $\pm$  40, 8812  $\pm$  44, 7766  $\pm$  38, 7835  $\pm$  39, 7925  $\pm$  40, 3886  $\pm$  19, 4301  $\pm$  21, 4645  $\pm$  23, 9495  $\pm$  47, 11625  $\pm$  60, 9288  $\pm$  46, 8631  $\pm$  43, 8933  $\pm$  45, 11728  $\pm$  59, 14105  $\pm$  70, 11940  $\pm$  60, 8861  $\pm$  44, 9150  $\pm$  46, 10264  $\pm$  51, 17047  $\pm$  85, 10461  $\pm$  52, 13354  $\pm$  67, 7471  $\pm$  37, 3821  $\pm$  19, 12135  $\pm$  60, 5968  $\pm$  30, 4614  $\pm$  23, 5182  $\pm$  25, 4069  $\pm$  20, 4634  $\pm$  23, 11600  $\pm$  58, 30133  $\pm$  150, 11939  $\pm$  60, 17894  $\pm$  89, 11723  $\pm$  58, 11493  $\pm$  57, 4959  $\pm$  25, 2013  $\pm$  10, 4370  $\pm$  22, 45862  $\pm$  226, 15105  $\pm$  75, 20898  $\pm$  104, 38099  $\pm$  190, 5873  $\pm$  27, 3668  $\pm$  18, 9091  $\pm$  45, 8491  $\pm$  42, 3391  $\pm$  16, 4130  $\pm$  20, 3136  $\pm$  15, 3441  $\pm$  17, 30952  $\pm$  154, 4029  $\pm$  20, 11253  $\pm$  56, 3820  $\pm$  19, 3506  $\pm$  17, 4571  $\pm$  23, 6933  $\pm$  34, 3887  $\pm$  19, 8602  $\pm$  43, 4644  $\pm$  23, 8630  $\pm$  43, and 8674  $\pm$  43 Daltons; and instructions to detect the protein biomarker by contacting 10 a test sample with the adsorbent and detecting the biomarker retained by the adsorbent.

In yet another aspect of the invention, methods of using a plurality of classifiers to make a probable diagnosis of lung cancer or a negative diagnosis are provided. In one form of the invention, a method includes a) obtaining mass spectra from a plurality of samples from normal subjects and subjects diagnosed with lung cancer; b) applying a decision tree analysis to at least a 15 portion of the mass spectra to obtain a plurality of weighted base classifiers comprising a peak intensity value and an associated threshold value; and c) making a probable diagnosis of lung cancer or a negative diagnosis based on a linear combination of the plurality of weighted base classifiers. In certain forms of the invention, the method may include using the peak intensity value and the associated threshold value in linear combination to make a probable diagnosis of 20 lung cancer or to make a negative diagnosis.

It is a further object of the invention to provide computer program media storing computer instructions therein for instructing a computer to perform a computer-implemented process for developing and/or using a plurality of classifiers to make a probable diagnosis of

lung cancer or a negative diagnosis using at least one protein biomarker having a molecular weight selected from the group consisting of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ ,  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,  $4614 \pm 23$ ,  $5182 \pm 25$ ,  $4069 \pm 20$ ,  $4634 \pm 23$ ,  $11600 \pm 58$ ,  $30133 \pm 150$ ,  $11939 \pm 60$ ,  $17894 \pm 89$ ,  $11723 \pm 58$ ,  $11493 \pm 57$ ,  $4959 \pm 25$ ,  $2013 \pm 10$ ,  $4370 \pm 22$ ,  $45862 \pm 226$ ,  $15105 \pm 75$ ,  $20898 \pm 104$ ,  $38099 \pm 190$ ,  $5873 \pm 27$ ,  $3668 \pm 18$ ,  $9091 \pm 45$ ,  $8491 \pm 42$ ,  $3391 \pm 16$ ,  $4130 \pm 20$ ,  $3136 \pm 15$ ,  $3441 \pm 17$ ,  $30952 \pm 154$ ,  $4029 \pm 20$ ,  $11253 \pm 56$ ,  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ ,  $6933 \pm 34$ ,  $3887 \pm 19$ ,  $8602 \pm 43$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $8674 \pm 43$  Daltons.

Preferably, the protein biomarkers are selected from the group having a molecular weight of about  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ , and  $6933 \pm 34$  Daltons protein biomarkers or about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ ,  $4301 \pm 21$ , and  $8674 \pm 43$  Daltons protein biomarkers.

15

## BRIEF DESCRIPTION OF THE FIGURES

FIGS. 1A - 1C show a representative SELDI spectra from bronchial lavage fluid ("BALF") of lung cancer patients. FIG 1A exhibits the SELDI spectra for peaks between 2000 Da - 10000 Da; FIG 1B shows the peaks from 10000 Da - 20000 Da; and FIG. 1C exhibits the spectra for peaks from 20000 Da - 100000 Da.

20 FIG. 2A shows a representative SELDI gelview from bronchial lavage samples of lung cancer patients compared with lavage samples from normal controls. The two "boxes" identify peaks with average masses of about 3820 and about 4069 Daltons that are underexpressed in lung cancer samples compared to the control samples. FIG. 2B shows the expression levels of

these proteins in the bronchial lavage fluid of lung cancer patients compared with the lavage fluid from normal controls. "—" indicates the mean normalized intensity.

FIG. 3A shows a representative SELDI spectra from bronchial lavage samples of lung cancer patients compared with samples from normal controls ranging from 20,000 to 60,000 m/z.

5 The "box" identifies a peak with an average mass of about 30132 Da that is overexpressed in lung cancer samples compared to normal samples. FIG. 3B shows the expression level of the about 30132 Da protein in the bronchial lavage samples of lung cancer patients compared with samples from normal controls. "—" indicates the mean normalized intensity while "■" and "●" indicate values of individual control (normal or unininvolved) and lung cancer patients,

10 respectively.

FIG. 4 depicts a schematic of the decision tree classification system based on bronchial lavage fluid samples, which is described in Example 1. The squares in bold are the primary nodes and the non-bolded squares indicate terminal nodes. The mass value in the root nodes are followed by  $\leq$  the intensity value.

15 FIGS. 5A - 5C shows a representative SELDI spectra from sera of lung cancer patients. FIG 5A shows peaks from 2000 - 10000 Da; FIG. 5B shows peaks from 10000 - 20000 Da; and FIG. 5C shows peaks from 20000 - 100000 Da.

FIG. 6 shows a representative SELDI spectra (A) and gelview (B) from sera of lung cancer patients ("LuCA") compared with sera from healthy smokers ("Norm smoker"), healthy 20 non-smokers ("Norm Nonsmoker"), and non-cancer patients with abnormal CT's ("nonCA AbCT") ranging from about 7,500 to 10,000 m/z. The "boxes" identify peaks with average masses of about 7766, 8603, and 8933 Daltons that are differentially expressed in lung cancer samples compared to normal samples. Specifically, it is shown that the about 7766 Dalton

biomarker is underexpressed in lung cancer serum while the about 8603 and 8933 Dalton biomarkers are overexpressed in lung cancer serum compared to non-cancer serum. FIG. 6C shows the expression levels of the about 7766, 8603, and 8933 Da proteins in the serum samples of lung cancer patients ("LuCA") compared with the serum samples from healthy smokers

.5 ("smoker"), healthy non-smokers ("nonsmoker"), and non-cancer patients with abnormal CT's ("AbCT"). "—" indicates the mean normalized intensity while "■", "●", "▲", and "◆" indicate values of individual LuCA, AbCT, normal nonsmoker, and normal smoker patients, respectively.

FIGS. 7A - 7D show the expression levels of the about 4748, 7566, 4301, and 4644

Dalton proteins, respectively, in the sera of lung cancer patients ("Lung CA") compared with

10 sera from healthy smokers ("Norm Smoker"), healthy non-smokers ("Norm Nonsmoker"), and non-cancer patients with abnormal CT's ("NoCA AbCT"). "—" indicates the mean normalized intensity while "■", "●", "▲", and "◆" indicate values of individual LuCA, AbCT, normal nonsmoker, and normal smoker patients, respectively.

FIGS. 8A and 8B depict the Receiver Operating Characteristic ("ROC") plots of one of

15 the peaks at about 8603 Da from lung cancer serum with the highest p-value in comparison with normal nonsmokers (A) and normal smokers (B). This peak is overexpressed in lung cancer patients. FIG. 8C depicts the ROC plot of the about 8674 Da peak from lung cancer sera compared to sera from normal nonsmokers while FIG. 8D depicts the ROC plot of the about 4301 Da peak from lung cancer sera in comparison with sera from normal smokers.

20 FIG. 9 depicts a schematic of the decision tree classification system based on serum utilized in Example 2. The squares in bold are the primary nodes and the non-bolded squares indicate terminal nodes. The mass value in the root nodes are followed by  $\leq$  the intensity value. "Dis" means diseased patient; "nondis" means a non-diseased patient.

FIG. 10 depicts various protein peaks that were differentially expressed in serum and bronchial lavage ("BAL") samples from lung cancer patients compared to normal controls.

FIG. 11 illustrates one example of a central processing unit for implementing a computer process in accordance with a computer implemented embodiment of the present invention.

5 FIG. 12 illustrates one example of a block diagram of internal hardware of the central processing unit of FIG. 11.

### **DESCRIPTION OF THE PREFERRED EMBODIMENTS**

For the purposes of promoting an understanding of the principles of the invention, 10 reference will now be made to preferred embodiments and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of the invention is thereby intended, such alteration and further modifications of the invention, and such further applications of the principles of the invention as illustrated herein, being contemplated as would normally occur to one skilled in the art to which the invention relates.

15 The present invention relates to methods for aiding in a diagnosis of, and methods for diagnosing lung cancer. Protein biomarkers have been identified that may be utilized to aid in the diagnosis of and/or to diagnose lung cancer or to make a negative diagnosis. Such protein biomarkers are also provided herein.

The methods of the present invention effectively differentiate between individuals with 20 lung cancer and normal individuals. As defined herein, normal individuals are individuals with a negative diagnosis with respect to lung cancer. That is, normal individuals do not have lung cancer. The method includes detecting a protein biomarker in a test sample from a subject. For example, the protein biomarkers having a molecular weight of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675$

$\pm$  43, 7566  $\pm$  38, 7972  $\pm$  40, 8812  $\pm$  44, 7766  $\pm$  38, 7835  $\pm$  39, 7925  $\pm$  40, 3886  $\pm$  19, 4301  $\pm$  21,

4645  $\pm$  23, 9495  $\pm$  47, 11625  $\pm$  60, 9288  $\pm$  46, 8631  $\pm$  43, 8933  $\pm$  45, 11728  $\pm$  59, 14105  $\pm$  70,

11940  $\pm$  60, 8861  $\pm$  44, 9150  $\pm$  46, 10264  $\pm$  51, 17047  $\pm$  85, 10461  $\pm$  52, 13354  $\pm$  67, 7471  $\pm$  37,

3821  $\pm$  19, 12135  $\pm$  60, 5968  $\pm$  30, 4614  $\pm$  23, 5182  $\pm$  25, 4069  $\pm$  20, 4634  $\pm$  23, 11600  $\pm$  58,

5 30133  $\pm$  150, 11939  $\pm$  60, 17894  $\pm$  89, 11723  $\pm$  58, 11493  $\pm$  57, 4959  $\pm$  25, 2013  $\pm$  10, 4370  $\pm$  22, 45862  $\pm$  226, 15105  $\pm$  75, 20898  $\pm$  104, 38099  $\pm$  190, 5873  $\pm$  27, 3668  $\pm$  18, 9091  $\pm$  45, 8491  $\pm$  42, 3391  $\pm$  16, 4130  $\pm$  20, 3136  $\pm$  15, 3441  $\pm$  17, 30952  $\pm$  154, 4029  $\pm$  20, 11253  $\pm$  56, 3820  $\pm$  19, 3506  $\pm$  17, 4571  $\pm$  23, 6933  $\pm$  34, 3887  $\pm$  19, 8602  $\pm$  43, 4644  $\pm$  23, 8630  $\pm$  43, and 8674  $\pm$

43 Daltons have been identified that aid in the probable diagnosis of lung cancer or aid in a

10 negative diagnosis. In accordance with the present invention, at least one of the protein biomarkers is detected. Preferably, two or more, three or more, four or more, five or more, ten or more, fifteen or more, twenty or more, thirty or more, or all sixty protein biomarkers are detected and the presence or absence of such biomarkers is correlated to a diagnosis of lung cancer. As used herein, the term "detecting" includes determining the presence, the absence, the quantity, or 15 a combination thereof, of the protein biomarkers. The quantity of the biomarkers may be represented by the peak intensity as identified by mass spectrometry, for example, or concentration of the biomarkers.

In certain forms of the invention, selected groups of protein biomarkers find utility in diagnosing lung cancer. For example, the following groups of markers find utility in making, or 20 otherwise aiding in making, a specific diagnosis: (1) the about 3820  $\pm$  19 Dalton biomarker; (2) the about 3820  $\pm$  19 and 3506  $\pm$  17 Dalton biomarkers; (3) the about 3820  $\pm$  19, 3506  $\pm$  17, and 4571  $\pm$  23 Dalton biomarkers; (4) the about 3820  $\pm$  19, 3506  $\pm$  17, 4571  $\pm$  23, and 6933  $\pm$  34 Dalton biomarkers; (5) the about 3820  $\pm$  19 and 6933  $\pm$  34 Dalton biomarkers; (6) the about

8603  $\pm$  43 Dalton biomarker, (7) the about 8603  $\pm$  43 and 3887  $\pm$  19 Dalton biomarkers, 8) the about 8603  $\pm$  43, 3887  $\pm$  19, and 4644  $\pm$  23 Dalton biomarkers, (9) the about 8603  $\pm$  43, 3887  $\pm$  19, 4644  $\pm$  23 and 8630  $\pm$  43 Dalton biomarkers, (10) the about 8603  $\pm$  43, 3887  $\pm$  19, 4644  $\pm$  23, 8630  $\pm$  43, and 4301  $\pm$  21 Dalton biomarkers, and (11) the about 8603  $\pm$  43, 3887  $\pm$  19, 4644  $\pm$  23, 8630  $\pm$  43, 4301  $\pm$  21, and 8674  $\pm$  43 Dalton biomarkers. Preferably, the about 3820  $\pm$  19 Dalton biomarker, the about 8603  $\pm$  43 Dalton biomarker, the combination of the about 3820  $\pm$  19, 3506  $\pm$  17, 4571  $\pm$  23, and 6933  $\pm$  34 Dalton biomarkers, or the combination of the about 8603  $\pm$  43, 3887  $\pm$  19, 4644  $\pm$  23, 8630  $\pm$  43, 4301  $\pm$  21, and 8674  $\pm$  43 Dalton biomarkers are used.

10 "Protein biomarker" as used herein is defined as any molecule, such as a peptide or protein fragment which is useful in differentiating lung cancer samples from normal samples. The biomarker is typically differentially present or expressed in lung cancer patients relative to normal patients. However, some biomarkers, while not being differentially expressed between two classes may, nevertheless, be classified as a biomarker according to the present invention to 15 the extent that they are significant in delineating subsets of groups in a classification tree.

The differential expression, such as the over- or under-expression, of selected biomarkers relative to normal individuals may be correlated to lung cancer. By differentially expressed, it is meant herein that the protein biomarkers may be found at a greater or smaller level in one disease state compared to another, or that it may be found at a higher frequency (e.g. intensity) in 20 one or more disease state. For example, the underexpression of the about 3820  $\pm$  19 and 4069  $\pm$  20 Dalton biomarkers by at least two-fold, three-fold, four-fold, preferably five-fold, relative to the normal patient may be correlated with the probable diagnosis of lung cancer. Furthermore, the underexpression of the about 7766  $\pm$  38, 4748  $\pm$  25, 7566  $\pm$  38, and 4644  $\pm$  23 Dalton

biomarkers relative to the normal patient may be correlated with a probable diagnosis of lung cancer. In addition, the overexpression of the about  $30132 \pm 150$ ,  $8603 \pm 43$ ,  $8933 \pm 45$ , and  $4301 \pm 21$  Dalton biomarkers relative to the normal patient may be correlated with a probable diagnosis of lung cancer. Moreover, the about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  
5  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ ,  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,  $4614 \pm 23$ ,  $5182 \pm 25$ ,  $4069 \pm 20$ ,  $4634 \pm 23$ ,  $11600 \pm 58$ ,  $30133 \pm 150$ ,  $11939 \pm 60$ ,  $17894 \pm 89$ ,  $11723 \pm 58$ ,  $11493 \pm 57$ ,  $4959 \pm 25$ ,  $2013 \pm 10$ ,  $4370 \pm 22$ ,  $45862 \pm 226$ ,  
10  $15105 \pm 75$ ,  $20898 \pm 104$ ,  $38099 \pm 190$ ,  $5873 \pm 27$ ,  $3668 \pm 18$ ,  $9091 \pm 45$ ,  $8491 \pm 42$ ,  $3391 \pm 16$ ,  $4130 \pm 20$ ,  $3136 \pm 15$ ,  $3441 \pm 17$ ,  $30952 \pm 154$ ,  $4029 \pm 20$ ,  $11253 \pm 56$ ,  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ ,  $6933 \pm 34$ ,  $3887 \pm 19$ ,  $8602 \pm 43$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $8674 \pm 43$  Daltons biomarkers have been found to be differentially expressed in lung cancer patients relative to normal patients. In particular, for example, the about  $30132$ ,  $8603$ ,  $8933$ , and  $4301$  Dalton biomarkers have been found to be overexpressed in lung cancer patients and the about  $3820$ ,  
15  $4069$ ,  $7766$ ,  $4748$ ,  $7566$ , and  $4644$  Dalton biomarkers have been found to be under-expressed in lung cancer patients.

Moreover, combinations of groupings of biomarkers in classification trees have been found to be useful to identify lung cancer-positive and lung cancer-negative patients. A classification tree may be produced using one or more of the protein biomarkers of this invention in connection with a threshold value. The threshold value may be based on the protein biomarker and its use in the classification tree. The threshold value represents the normalized peak intensity of the biomarkers. As more fully described in Examples 1 and 2, these threshold

values may represent the normalized peak intensity of a particular biomarker which is related to the concentration of the biomarker. The normalization process may involve using the total ion current as a normalization factor. The normalization process could alternatively involve reporting the peak intensity relative to the peak intensity of an internal or external control. For 5 example, a known protein may be added to the system. Additionally, a known product produced by the test subject, such as albumin, may act as an internal standard or control. It is understood that the threshold values identified in FIGS. 4 and 9 are relative to the control used in Examples 1 and 2, respectively. However, as one having ordinary skill in the art would appreciate, this threshold may be different based on the internal or external control.

10 For example, FIG. 4 depicts a suitable classification tree that may be used to distinguish lung cancer and normal patients. In one group, the presence of the about 3820 Dalton biomarker at a threshold value of less than or equal to 0.322 and the presence of the about 3506 Dalton biomarker at a threshold value of less than or equal to 0.162 may be correlated to a normal diagnosis. In another group, the presence of the about 3820 Dalton biomarker at a peak intensity 15 threshold value of less than or equal to 0.322, the presence of the about 3506 Dalton biomarker at a peak intensity value of greater than 0.162, the presence of the about 4571 Dalton Biomarker at a peak intensity value of less than or equal to 0.642, and the presence of the about 6933 Dalton biomarker at a threshold value of less than or equal to 0.066 may be correlated to a normal diagnosis. In another group, the presence of the about 3820 Dalton biomarker at a peak intensity 20 threshold value of less than or equal to 0.322, the presence of the about 3506 Dalton biomarker at a peak intensity value of greater than 0.162, the presence of the about 4571 Dalton Biomarker at a peak intensity value of less than or equal to 0.642, and the presence of the about 6933 Dalton biomarker at a threshold value of greater than 0.066 may be correlated to a probable diagnosis of

lung cancer. In another group, the presence of the about 3820 Dalton biomarker at a peak intensity threshold value of less than or equal to 0.322, the presence of the about 3506 Dalton biomarker at a peak intensity value of greater than 0.162, and the presence of the about 4571 Dalton Biomarker at a peak intensity value of greater than 0.642 may be correlated to a normal diagnosis. In another group, the presence of the about 3820 Dalton biomarker at a peak intensity threshold value of greater than 0.322 and the presence of the about 6933 Dalton biomarker at a peak intensity of less than or equal to about 1.618 may be correlated to a normal diagnosis.

Finally, the presence of the about 3820 Dalton biomarker at a peak intensity threshold value of greater than 0.322 and the presence of the about 6933 Dalton biomarker at a peak intensity greater than 1.618 may be correlated to either a normal or lung cancer diagnosis. Preferably, the combination of these groupings makes up a single classification tree for a diagnosis of lung cancer. However, the present invention contemplates the use of these individual groupings alone or in combination with other groupings to aid in the diagnosis or identification of lung cancer-positive and lung cancer-negative patients. Thus, one or more of such groupings, preferably two or more, or more preferably, all of these groupings aid in the diagnosis.

FIG. 9 depicts another suitable classification tree that may also be used to distinguish lung cancer and normal patients. In one group, the value of the about 3887 Dalton biomarker multiplied by 0.734, subtracted from the value of the about 8602 Dalton biomarker multiplied by 0.679, at a threshold value of less than or equal to 0.815, the value of the about 3887 Dalton biomarker multiplied by 0.667, subtracted by the value of the about 4644 Dalton biomarker multiplied by 0.335, added to the value of the about 8630 Dalton biomarker multiplied by 0.666, at a threshold value of less than or equal to 3.30 may be correlated to a probable diagnosis of lung cancer. In another group, the value of the about 3887 Dalton biomarker multiplied by

0.734, subtracted from the value of the 8602 Dalton biomarker multiplied 0.679, at a threshold value of less than or equal to 0.815 the value of the about 3887 Dalton biomarker multiplied by 0.667, subtracted from the value of the 4644 Dalton biomarker multiplied by 0.335, added to the value of the about 8630 Dalton biomarker multiplied by 0.666, at a threshold value of greater than or equal to 3.30 and the about 3887 Dalton biomarker, less than or equal to the value of 5 5.975 may be correlated with a normal diagnosis while a value greater 5.975 may be correlated to either a lung cancer or normal diagnosis. In another group, the value of the about 3887 Dalton biomarker multiplied by 0.734, subtracted from the value of the about 8602 Dalton biomarker multiplied by 0.679, at a threshold value of greater than 0.815, and the value of the about 4301 10 Dalton biomarker multiplied by -0.905, subtracted by the value of the about 8630 biomarker multiplied by 0.426 less than or equal to a threshold value of -1.119 may be correlated to a normal diagnosis. In another group, the value of the about 3887 Dalton biomarker multiplied by 0.734, subtracted from the value of the about 8602 Dalton biomarker multiplied by 0.679, at a threshold value of greater than 0.815, and the value of the about 4301 Dalton biomarker 15 multiplied by -0.905 subtracted by the value of the about 8630 biomarker multiplied by 0.426 greater than a threshold value of -1.119 and if the value of the biomarker at or about 8674 is less than or equal to 0.531 may be correlated to a normal diagnosis, while a value greater than 0.531 may be correlated to a probable diagnosis of lung cancer.

In another form of the invention, the drug responder status of a biological sample of a 20 lung cancer patient may be determined. A drug responder state is a state of a biological sample in response to the use of a drug. Biological statuses may also include beginning states, intermediate states, and terminal states. For example, different biological statuses may include the beginning state, the intermediate state, and the terminal state of a disease such as lung cancer.

In connection with this aspect of the invention, the different biological statuses may correspond to samples from treated lung cancer patients that are associated with respectively different drugs or drug types. In an illustrative example, mass spectra of samples from lung cancer patients who were treated with a drug of known effect are created. The mass spectra 5 associated with the drug of known effect may represent drugs of the same type as the drug of known effect. For instance, the mass spectra associated with drugs of known effect may represent drugs with the same or similar characteristics, structure, or the same basic effect as the drug of known effect. Many different analgesic compounds, for example, may all provide pain relief to a person. The drug of known effect and drugs of the same or similar type might all 10 regulate the same biochemical pathway in a person to produce the same effect on a person. Characteristics of the biological pathway (e.g., up- or down-regulated proteins) may be reflected in the mass spectra.

Data analysis can include the steps of determining signal strength (e.g., height of peaks, area of peaks) of a biomarker detected and removing “outliers” (data deviating from a 15 predetermined statistical distribution). For example, the observed peaks can be normalized, a process whereby the height of each peak relative to some reference is calculated. For example, a reference can be background noise generated by instrument and chemicals (e.g., energy absorbing molecule) which is set as zero in the scale. The signal strength can then be detected for each biomarker or other substances can be displayed in the form of relative intensities in the 20 scale desired (e.g., 100). Alternatively, a standard may be included with the sample so that a peak from the standard can be used as a reference to calculate relative intensities of the signals observed for each biomarker or other markers detected.

The method includes detecting at least one protein biomarker. However, any number of biomarkers may be detected. It is preferred that at least two protein biomarkers are detected in the analysis. However, it is realized that three, four, or more, including all, of the biomarkers described herein may be utilized in the diagnosis. Thus, not only can one or more markers be 5 detected, one to 60, preferably two to 60, two to 20, two to 10 biomarkers, two to 5 biomarkers, or some other combination, may be detected and analyzed as described herein. In addition, other protein biomarkers not herein described may be combined with any of the presently disclosed protein biomarkers to aid in the diagnosis of lung cancer. Moreover, any combination of the above protein biomarkers may be detected in accordance with the present invention.

10 The detection of the protein biomarkers described herein in a test sample may be performed in a variety of ways. In one form of the invention, a method for detecting the biomarker includes detecting the biomarker by gas phase ion spectrometry utilizing a gas phase ion spectrometer. The method may include contacting a test sample having a biomarker, such as the protein biomarkers described herein, with a substrate comprising an adsorbent thereon under 15 conditions to allow binding between the biomarker and adsorbent and detecting the biomarker bound to the adsorbent by gas phase ion spectrometry.

A wide variety of adsorbents may be used. The adsorbents may include a hydrophobic group, a hydrophilic group, a cationic group, an anionic group, a metal ion chelating group, or antibodies that specifically bind to an antigenic biomarker, or some combination thereof (such as 20 a “mixed mode” adsorbent). Exemplary adsorbents that include a hydrophobic group include matrices having aliphatic hydrocarbons, such as C<sub>1</sub>-C<sub>18</sub> aliphatic hydrocarbons and matrices having aromatic hydrocarbon functional groups, including phenyl groups. Exemplary adsorbents that include a hydrophilic group include silicon oxide, or hydrophilic polymers such as

polyalkylene glycol, polyethylene glycol, dextran, agarose or cellulose. Exemplary adsorbents that include a cationic group include matrices of secondary, tertiary or quaternary amines. Exemplary adsorbents that have an anionic group include matrices of sulfate anions and matrices of carboxylate anions or phosphate anions. Exemplary adsorbents that have metal chelating 5 groups include organic molecules that have one or more electron donor groups which may form coordinate covalent bonds with metal ions, such as copper, nickel, cobalt, zinc, iron, aluminum and calcium. Exemplary adsorbents that include an antibody include antibodies that are specific for any of the biomarkers provided herein and may be readily made by methods known to the skilled artisan.

10        Alternatively, the substrate can be in the form of a probe, which may be removably insertable into a gas phase ion spectrometer. For example, a substrate may be in the form of a strip with adsorbents on its surface. In yet other forms of the invention, the substrate can be positioned onto a second substrate to form a probe which may be removably insertable into a gas phase ion spectrometer. For example, the substrate can be in the form of a solid phase, such as a 15 polymeric or glass bead with a functional group for binding the marker, which can be positioned on a second substrate to form a probe. The second substrate may be in the form of a strip, or a plate having a series of wells at predetermined locations. In this manner, the biomarker can be adsorbed to the first substrate and transferred to the second substrate which can then be submitted for analysis by gas phase ion spectrometry.

20        The probe can be in the form of a wide variety of desired shapes, including circular, elliptical, square, rectangular, or other polygonal or other desired shape, as long as it is removably insertable into a gas phase ion spectrometer. The probe is also preferably adapted or otherwise configured for use with inlet systems and detectors of a gas phase ion spectrometer.

For example, the probe can be adapted for mounting in a horizontally and/or vertically translatable carriage that horizontally and/or vertically moves the probe to a successive position without requiring, for example, manual repositioning of the probe.

The substrate that forms the probe can be made from a wide variety of materials that can 5 support various adsorbents. Exemplary materials include insulating materials, such as glass and ceramic; semi-insulating materials, such as silicon wafers; electrically-conducting materials (including metals such as nickel, brass, steel, aluminum, gold or electrically-conductive polymers); organic polymers; biopolymers, or combinations thereof.

In other embodiments of the invention, depending on the nature of the substrate, the 10 substrate surface may form the adsorbent. In other cases, the substrate surface may be modified to incorporate thereon a desired adsorbent. The surface of the substrate forming the probe can be treated or otherwise conditioned to bind adsorbents that may bind markers if the substrate cannot bind biomarkers by itself. Alternatively, the surface of the substrate can also be treated or otherwise conditioned to increase its natural ability to bind desired biomarkers. Other probes 15 suitable for use in the invention may be found, for example, in PCT international publication numbers WO 01/25791 (Tai-Tung et al.) and WO 01/71360 (Wright et al.).

The adsorbents may be placed on the probe substrate in a wide variety of patterns, including a continuous or discontinuous pattern. A single type of adsorbent, or more than one type of adsorbent, may be placed on the substrate surface. The patterns may be in the form of 20 lines, curves, such as circles, or any such other shape or pattern as desired.

The method of production of the probes will depend on the selection of substrate materials and/or adsorbents as known in the art. For example, if the substrate is a metal, the surface may be prepared depending on the adsorbent to be applied thereon. For example, the

substrate surface may be coated with a material, such as silicon oxide, titanium oxide or gold, that allows derivatization of the metal surface to form the adsorbent. The substrate surface may then be derivatized with a bifunctional linker, one of which binds, such as covalently binds, with a functional group on the surface and the opposing end of the linker may be further derivatized 5 with groups that function as an adsorbent. As a further example, a substrate that includes a porous silicon surface generated from crystalline silicon can be chemically modified to include adsorbents for binding markers. Additionally, adsorbents with a hydrogel backbone can be formed directly on the substrate surface by *in situ* polymerization of a monomer solution which includes, for example, substituted acrylamide or acrylate monomers, or derivatives thereof that 10 include a functional group of choice as adsorbent.

In preferred forms of the invention, the probe may be a chip, such as those available from Ciphergen Biosystems, Inc. (Palo Alto, CA). The chip may be a hydrophilic, hydrophobic, anion-exchange, cation-exchange, immobilized metal affinity or preactivated protein chip array. The hydrophobic chip may be a ProteinChip® H4, which includes a long-chain aliphatic surface 15 that binds proteins by reverse phase interaction. The hydrophilic chip may be ProteinChips® NP1 and NP2 which include a silicon dioxide substrate surface. The cation exchange ProteinChip® array may be ProteinChip® WCX2, a weak cation exchange array with a carboxylate surface to bind cationic proteins. Alternatively, the chip may be an anion exchange protein chip array, such as SAX1 (strong anion exchange) ProteinChip® which is made from 20 silicon-dioxide-coated aluminum substrates, or ProteinChip® SAX2 with a higher capacity quaternary ammonium surface to bind anionic proteins. A further useful chip may be the immobilized metal affinity capture chip (IMAC3) having nitrilotriacetic acid on the surface. Further alternatively, ProteinChip® PS1 is available which includes a carbonyldiimidazole

surface which covalently reacts with amino groups or may be ProteinChip® PS2 which includes an epoxy surface which covalently reacts with amine and thiol groups.

In accordance with the present invention, the probe contacts a test sample. The test sample may be obtained from a wide variety of sources. The sample is typically obtained from 5 biological fluid from a subject or patient who is being tested for lung cancer or from a normal individual who may be thought to be of risk for the disease. A preferred biological fluid is blood, blood sera, or bronchial lavage ("BAL") fluid. Other biological fluids in which the biomarkers may be found include, for example, seminal fluid, seminal plasma, lymph fluid, mucus, nipple secretions, sputum, tears, saliva, urine, or other similar fluid. Moreover, the 10 biological sample may include tissue, including bronchial/lung tissue, or any other similar tissue.

If necessary, the sample can be solubilized in or mixed with an eluant prior to being contacted with the probe. The probe may contact the test sample solution by a wide variety of techniques, including bathing, soaking, dipping, spraying, washing, pipetting or other desirable methods. The method is performed so that the adsorbent of the probe preferably contacts the test 15 sample solution. Although the concentration of the biomarker or biomarkers in the sample may vary, it is generally desirable to contact a volume of test sample that includes about 1 attomole to about 100 picomoles of marker in about 1 µl to about 500 µl solution for binding to the adsorbent.

The sample and probe contact each other for a period of time sufficient to allow the 20 biomarker to bind to the adsorbent. Although this time may vary depending on the nature of the sample, the nature of the biomarker, the nature of the adsorbent and the nature of the solution the biomarker is dissolved in, the sample and adsorbent are typically contacted for a period of about 30 seconds to about 12 hours, preferably about 30 seconds to about 15 minutes.

The temperature at which the probe contacts the sample will depend on the nature of the sample, the nature of the biomarker, the nature of the adsorbent and the nature of the solution the biomarker is dissolved in. Generally, the sample may be contacted with the probe under ambient temperature and pressure conditions. However, the temperature and pressure may vary as 5 desired. In presently preferred embodiments of the invention, for example, the temperature may vary from about 4°C to about 37°C.

After the sample has contacted the probe for a period of time sufficient for the marker to bind to the adsorbent or substrate surface should no adsorbent be used, unbound material may be washed from the substrate or adsorbent surface so that only bound materials remain on the 10 respective surface. The washing can be accomplished by, for example, bathing, soaking, dipping, rinsing, spraying or otherwise washing the respective surface with an eluant or other washing solution. A microfluidics process is preferably used when a washing solution such as an eluant is introduced to small spots of adsorbents on the probe. The temperature of the washing solution may vary, but is typically about 0°C to about 100°C, and preferably about 4°C and about 15 37°C.

A wide variety of washing solutions may be utilized to wash the probe substrate surface. The washing solutions may be organic solutions or aqueous solutions. Exemplary aqueous solutions may be buffered solutions, including HEPES buffer, a Tris buffer, phosphate buffered saline or other similar buffers known to the art. The selection of a particular washing solution 20 will depend on the nature of the biomarkers and the nature of the adsorbent utilized. For example, if the probe includes a hydrophobic group and a sulfonate group as adsorbents, such as the SCXI ProteinChip® array, then an aqueous solution, such as a HEPES buffer, may be used. As a further example, if a probe includes a metal binding group as an adsorbent, such as with the

Ni(II) ProteinChip® array, than an aqueous solution, such as a phosphate buffered saline may be preferred. As yet a further example, if a probe includes a hydrophobic group as an adsorbent, such as with the HF ProteinChip® array, water may be a preferred washing solution.

An energy absorbing molecule, such as one in solution, may be applied to the markers or 5 other substances bound on the substrate surface of the probe. As used herein, an “energy absorbing molecule” refers to a molecule that absorbs energy from an energy source in a gas phase ion spectrometer, which may assist the desorption of markers or other substances from the surface of the probe. Exemplary energy absorbing molecules include cinnamic acid derivatives, sinapinic acid, dihydroxybenzoic acid and other similar molecules known to the art. The energy 10 absorbing molecule may be applied by a wide variety of techniques previously discussed herein for contacting the sample and probe substrate, including, for example, spraying, pipetting or dipping, preferably after the unbound materials are washed off the probe substrate surface.

After the biomarker is appropriately bound to the probe, the biomarker may be detected, quantified and/or its characteristics may be otherwise determined using an appropriate detection 15 instrument, preferably a gas phase ion spectrometer. As known in the art, gas phase ion spectrometers include, for example, mass spectrometers, ion mobility spectrometers, and total ion current measuring devices.

In a preferred embodiment, a mass spectrometer is utilized to detect the biomarkers bound to the substrate surface of the probe. The probe, with the bound marker on its surface, 20 may be introduced into an inlet system of the mass spectrometer. The marker may then be ionized by an ionization source, such as a laser, fast atom bombardment, plasma or other suitable ionization sources known to the art. The generated ions are typically collected by an ion optic assembly and a mass analyzer then disperses and analyzes the passing ions. The ions exiting the

mass analyzer are detected by a detector. The detector translates information of the detected ions into mass-to-charge ratios. Detection and/or quantitation of the marker will typically involve detection of signal intensity.

In further preferred forms of the invention, the mass spectrometer is a laser desorption 5 time-of-flight mass spectrometer, and further preferably surface enhanced laser desorption time-of-flight mass spectrometry (SELDI) is utilized. SELDI is an improved method of gas phase ion spectrometry for biomolecules. In SELDI, the surface on which the analyte is applied plays an active role in the analyte capture and/or desorption.

As known in the art, in laser desorption mass spectrometry, a probe with a bound marker 10 is introduced into an inlet system. The marker is desorbed and ionized into the gas phase by a laser ionization source. The ions generated are collected by an ion optic assembly. Ions are accelerated in a time-of-flight mass analyzer through a short high voltage field and allowed to drift into a high vacuum chamber. The accelerated ions strike a sensitive detector surface at a far end of the high vacuum chamber at a different time. As the time-of-flight is a function of the 15 mass of the ions, the elapsed time between ionization and impact can be used to identify the presence or absence of molecules of specific mass. Quantitation of the biomarkers, either in relative or absolute amounts, may be accomplished by comparison of the intensity of the displayed signal of the biomarker to a control amount of a biomarker or other standard as known in the art. The components of the laser desorption time-of-flight mass spectrometer may be 20 combined with other components described herein and/or known to the skilled artisan that employ various means of desorption, acceleration, detection, or measurement of time.

In further embodiments, detection and/or quantitation of the biomarkers may be accomplished by matrix-assisted laser desorption ionization (MALDI). MALDI also provides

for vaporization and ionization of biological samples from a solid-state phase directly into the gas phase. As known in the art, the sample, including the desired analyte, is dissolved or otherwise suspended in, a matrix that co-crystallizes with the analyte, preferably to prevent the degradation of the analyte during the process.

5 An ion mobility spectrometer can be used to detect and characterize the biomarkers described herein. The principle of ion mobility spectrometry is based on different mobility of ions. Specifically, ions of a sample produced by ionization move at different rates, due to their difference in, for example, mass, charge, or shape, through a tube under the influence of an electric field. The ions (typically in the form of a current) are registered at the detector which  
10 can then be used to identify a marker or other substances in the sample. One advantage of ion mobility spectrometry is that it can operate at atmospheric pressure.

A total ion current measuring device can be used to detect and characterize the biomarkers described herein. This device can be used, for example, when the probe has a surface chemistry that allows only a single type of marker to be bound. When a single type of  
15 marker is bound on the probe, the total current generated from the ionized biomarker reflects the nature of the marker. The total ion current produced by the biomarker can then be compared to stored total ion current of known compounds. Characteristics of the biomarker can then be determined.

Data generated by desorption and detection of the biomarkers can be analyzed with the  
20 use of a programmable digital computer. The computer program generally contains a readable medium that stores codes. Certain code can be devoted to memory that includes the location of each feature on a probe, the identity of the adsorbent at that feature and the elution conditions used to wash the adsorbent. Using this information, the program can then identify the set of

features on the probe defining certain selectivity characteristics, such as types of adsorbent and eluants used. The computer also contains code that receives data on the strength of the signal at various molecular masses received from a particular addressable location on the probe as input. This data can indicate the number of biomarkers detected, optionally including the strength of 5 the signal and the determined molecular mass for each biomarker detected. As described above, the data may be normalized according to known methods, such as by determining the signal strength (e.g., height of peaks or area of peaks) of a biomarker detected and removing any strength (e.g., height of peaks or area of peaks) of a biomarker detected and removing any “outliers.”

The computer can transform the resulting data into various formats for displaying. In one 10 format, referred to as “spectrum view or retentate map,” a standard spectral view can be displayed, wherein the view depicts the quantity of biomarker reaching the detector at each particular molecular weight. In another format, referred to as “peak map,” only the peak height and mass information are retained from the spectrum view, yielding a cleaner image and enabling markers with nearly identical molecular weights to be more easily seen. In yet another 15 format, referred to as “gel view,” each mass from the peak view can be convened into a grayscale image based on the height of each peak, resulting in an appearance similar to bands on electrophoretic gels. In a further format, referred to as “3-D overlays,” several spectra can be overlaid to study subtle changes in relative peak heights. In yet a further format, referred to as “difference map view,” two or more spectra can be compared, conveniently highlighting unique 20 biomarkers and biomarkers which are up- or down-regulated between samples. Biomarker profiles (spectra) from any two samples may be compared visually.

Using any of the above display formats, it can be readily determined from the signal display whether a biomarker having a particular molecular weight is detected from a sample.

Moreover, from the strength of signals, the amount of markers bound on the probe surface can be determined.

In preferred forms of the invention, a single decision tree classification algorithm is utilized to analyze the data generated from SELDI. Algorithms used to generate such classifications are known in the art. For example, algorithms used to generate classification trees, such as from Classification Logic, based on cumulative probability, PeakMiner (Internet address: [www.evms.edu/vpc/seldi](http://www.evms.edu/vpc/seldi)), or Classification And Regression Tree (CART) (Breiman, L., Friedman, J., Olshen, R., and Stone, C. J. (1984) *Classification and Regression Trees* Chapman and Hall, New York), and those developed by known methods that are suitable for the generation of such classification trees; for example, genetic cluster, logistical regression, surface vector machine, and neural nets can be used. (Jain et al. "Statistical Pattern Recognition: A Review," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, Jan. 2000). For example, one such algorithm is more specifically described in Examples 1 and 2 herein.

The test samples may be pre-treated prior to being subject to gas phase ion spectrometry. For example, the samples can be purified or otherwise pre-fractionated to provide a less complex sample for analysis. The optional purification procedure for the biomolecules present in the test sample may be based on the properties of the biomolecules, such as size, charge and function. Methods of purification include centrifugation, electrophoresis, chromatography, dialysis or a combination thereof. As known in the art, electrophoresis may be utilized to separate the biomolecules in the sample based on size and charge. Electrophoretic procedures are well-known to the skilled artisan, and include isoelectric focusing, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), agarose gel electrophoresis, and other known methods of electrophoresis.

The purification step may be accomplished by a chromatographic fractionation technique, including size fractionation, fractionation by charge and fractionation by other properties of the biomolecules being separated. As known in the art, chromatographic systems include a stationary phase and a mobile phase, and the separation is based upon the interaction of the 5 biomolecules to be separated with the different phases. In preferred forms of the invention, column chromatographic procedures may be utilized. Such procedures include partition chromatography, adsorption chromatography, size-exclusion chromatography, ion-exchange chromatography and affinity chromatography. Such methods are well-known to the skilled artisan. In size-exclusion chromatography, it is preferred that the size fractionation columns 10 exclude molecules whose molecular mass is greater than about 10,000 Da.

In a preferred form of the invention, the sample is purified or otherwise fractionated on a bio-chromatographic chip by retentate chromatography before gas phase ion spectrometry. A preferred chip is the Protein Chip™ available from Ciphergen Biosystems, Inc. (Palo Alto, CA). As described above, the chip or probe is adapted for use in a mass spectrometer. The chip 15 comprises an adsorbent attached to its surface. This adsorbent can function, in certain applications, as an *in situ* chromatography resin. In operation, the sample is applied to the adsorbent in an eluant solution. Molecules for which the adsorbent has affinity under the wash condition bind to the adsorbent. Molecules that do not bind to the adsorbent are removed with the wash. The adsorbent can be further washed under various levels of stringency so that 20 analytes are retained or eluted to an appropriate level for analysis. An energy absorbing molecule can then be added to the adsorbent spot to further facilitate desorption and ionization. The analyte is detected by desorption from the adsorbent, ionization and direct detection by a detector. Thus, retentate chromatography differs from traditional chromatography in that the

analyte retained by the affinity material is detected, whereas in traditional chromatography, material that is eluted from the affinity material is detected.

The biomarkers of the present invention may also be detected, qualitatively or quantitatively, by an immunoassay procedure. The immunoassay typically includes contacting a 5 test sample with an antibody that specifically binds to or otherwise recognizes a biomarker, and detecting the presence of a complex of the antibody bound to the biomarker in the sample. The immunoassay procedure may be selected from a wide variety of immunoassay procedures known to the art involving recognition of antibody/antigen complexes, including enzyme immunoassays, competitive or non-competitive, and including enzyme-linked immunosorbent 10 assays (ELISA), radioimmunoassay (RIA), and Western blots, and use of multiplex assays, including use of antibody arrays, wherein several desired antibodies are placed on a support, such as a glass bead or plate, and reacted or otherwise contacted with the test sample. Such assays are well-known to the skilled artisan and are described, for example, more thoroughly in *Antibodies: A Laboratory Manual* (1988) by Harlow & Lane; *Immunoassays: A Practical* 15 *Approach*, Oxford University Press, Gosling, J.P. (ed.) (2001) and/or *Current Protocols in Molecular Biology* (Ausubel et al.) which is regularly and periodically updated.

The antibodies to be used in the immunoassays described herein may be polyclonal antibodies and may be obtained by procedures which are well-known to the skilled artisan, including injecting purified biomarkers into various animals and isolating the antibodies 20 produced in the blood serum. The antibodies may alternatively be monoclonal antibodies whose method of production is well-known to the art, including injecting purified biomarkers into a mouse, for example, isolating the spleen cells producing the anti-serum, fusing the cells with tumor cells to form hybridomas and screening the hybridomas. The biomarkers may first be

purified by techniques similarly well-known to the skilled artisan, including the chromatographic, electrophoretic and centrifugation techniques described previously herein. Such procedures may take advantage of the protein biomarker's size, charge, solubility, affinity for binding to selected components, combinations thereof, or other characteristics or properties 5 of the protein. Such methods are known to the art and can be found, for example, in *Current Protocols in Protein Science*, J. Wiley and Sons, New York, NY, Coligan et al. (Eds.) (2002); Harris, E.L.V., and S. Angal in *Protein purification applications: a practical approach*, Oxford University Press, New York, NY (1990). Once the antibody is provided, a biomarker can be detected and/or quantitated by the immunoassays previously described herein.

10        Although specific procedures for immunoassays are well-known to the skilled artisan, generally, an immunoassay may be performed by initially obtaining a sample as previously described herein from a test subject. The antibody may be fixed to a solid support prior to contacting the antibody with a test sample to facilitate washing and subsequent isolation of the antibody/protein biomarker complex. Examples of solid supports are well-known to the skilled 15 artisan and include, for example, glass or plastic in the form of, for example, a microtiter plate. Antibodies can also be attached to the probe substrate, such as the ProteinChip® arrays described herein.

After incubating the test sample with the antibody, the mixture is washed and the antibody-marker complex may be detected. The detection can be accomplished by incubating 20 the washed mixture with a detection reagent, and observing, for example, development of a color or other indicator. Any detectable label may be used. The detection reagent may be, for example, a second antibody which is labeled with a detectable label. Exemplary detectable labels include magnetic beads (e.g., DYNABEADS™), fluorescent dyes, radiolabels, enzymes

(e.g., horseradish peroxide, alkaline phosphatase and others commonly used in enzyme immunoassay procedures), and colorimetric labels such as colloidal gold, colored glass or plastic beads. Alternatively, the marker in the sample can be detected using an indirect assay, wherein, for example, a labeled antibody is used to detect the bound marker-specific antibody complex 5 and/or in a competition or inhibition assay wherein, for example, a monoclonal antibody which binds to a distinct epitope of the biomarker is incubated simultaneously with the mixture. The amount of an antibody-marker complex can be determined by comparing to a standard.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, 10 preferably from about 5 minutes to about 24 hours. However, the incubation time will depend upon the particular immunoassay, biomarker, and assay conditions. Usually the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as about 0°C to about 40°C.

Kits are provided that may, for example, be utilized to detect the biomarkers described 15 herein. The kits can, for example, be used to detect any one or more of the biomarkers described herein, which may advantageously be utilized for diagnosing or aiding in the diagnosis of lung cancer or in a negative diagnosis.

In one embodiment, a kit may include a substrate that includes an adsorbent thereon, wherein the adsorbent is preferably suitable for binding one or more protein biomarkers 20 described herein, and instructions to detect the biomarker by contacting a test sample as described herein with the adsorbent and detecting the biomarker retained by the adsorbent. In certain embodiments, the kits may include an eluant, or instructions for making an eluant, wherein the combination of the eluant and the adsorbent allows detection of the protein

biomarkers by, for example, use of gas phase ion spectrometry. Such kits can be prepared from the materials described herein.

In yet another embodiment, the kit may include a first substrate that includes an adsorbent thereon (e.g., a particle functionalized with an adsorbent) and a second substrate onto 5 which the first substrate can be positioned to form a probe which is removably insertable into a gas phase ion spectrometer. In other embodiments, the kit may include a single substrate which is in the form of a removably insertable probe with adsorbents on the substrate. In yet another embodiment, the kit may further include a pre-fractionation spin column (e.g, K-30 size exclusion column).

10 The kit may further include instructions for suitable operating parameters in the form of a label or a separate insert. For example, the kit may have standard instructions informing a consumer or other individual how to wash the probe after a particular form of sample is contacted with the probe. As a further example, the kit may include instructions for pre-fractionating a sample to reduce the complexity of proteins in the sample.

15 In a further embodiment, a kit may include an antibody that specifically binds to the marker and a detection reagent. Such kits can be prepared from the materials described herein. The kit may further include pre-fractionation spin columns as described above, as well as instructions for suitable operating parameters in the form of a label or a separate insert.

In yet another aspect of the invention, methods of using a plurality of classifiers to make 20 a probable diagnosis of lung cancer are provided. In one form of the invention, a method includes a) obtaining mass spectra from a plurality of samples from normal subjects and subjects diagnosed with lung cancer; b) applying a decision tree analysis to at least a portion of the mass spectra to obtain a plurality of weighted base classifiers, wherein the classifiers include a peak

intensity value and an associated threshold value; and c) making a probable diagnosis of lung cancer or a negative diagnosis based on a linear combination of the plurality of weighted base classifiers. In certain forms of the invention, the method includes using the peak intensity value and the associated threshold value in linear combination to make a probable diagnosis of lung 5 cancer or a negative diagnosis. The preferred algorithm and data treatment is more fully described in Examples 1 and 2.

The methods of the present invention have other applications as well. For example, the biomarkers can be used to screen for compounds that modulate the expression of the biomarkers *in vitro* or *in vivo*, which compounds in turn may be useful in treating or preventing lung cancer 10 in patients. In another example, the biomarkers can be used to monitor the response to treatments for lung cancer. In yet another example, the biomarkers can be used in heredity studies to determine if the subject is at risk for developing lung cancer.

Thus, for example, the kits of this invention could include a solid substrate having an 15 cation exchange function, such as a protein biochip (e.g., a Ciphergen WCX2 ProteinChip array, e.g., ProteinChip array) and a sodium acetate buffer for washing the substrate, as well as instructions providing a protocol to measure the biomarkers of this invention on the chip and to use these measurements to diagnose lung cancer.

Compounds suitable for therapeutic testing may be screened initially by identifying 20 compounds which interact with one or more biomarkers listed herein. By way of example, screening might include recombinantly expressing a biomarker listed herein, purifying the biomarker, and affixing the biomarker to a substrate. Test compounds would then be contacted with the substrate, typically in aqueous conditions, and interactions between the test compound and the biomarker are measured, for example, by measuring elution rates as a function of salt

concentration. Certain proteins may recognize and cleave one or more biomarkers listed herein, in which case the proteins may be detected by monitoring the digestion of one or more biomarkers in a standard assay, e.g., by gel electrophoresis of the proteins.

- In a related embodiment, the ability of a test compound to inhibit the activity of one or 5 more of the biomarkers listed herein may be measured. One of skill in the art will recognize that the techniques used to measure the activity of a particular biomarker will vary depending on the function and properties of the biomarker. For example, an enzymatic activity of a biomarker may be assayed provided that an appropriate substrate is available and provided that the concentration of the substrate or the appearance of the reaction product is readily measurable.
- 10 The ability of potentially therapeutic test compounds to inhibit or enhance the activity of a given biomarker may be determined by measuring the rates of catalysis in the presence or absence of the test compounds. The ability of a test compound to interfere with a non-enzymatic (e.g., structural) function or activity of one of the biomarkers listed herein may also be measured. For example, the self-assembly of a multi-protein complex which includes one of the biomarkers 15 listed herein may be monitored by spectroscopy in the presence or absence of a test compound. Alternatively, if the biomarker is a non-enzymatic enhancer of transcription, test compounds which interfere with the ability of the biomarker to enhance transcription may be identified by measuring the levels of biomarker-dependent transcription *in vivo* or *in vitro* in the presence and absence of the test compound.
- 20 Test compounds capable of modulating the activity of any of the biomarkers listed herein may be administered to patients who are suffering from or are at risk of developing lung cancer or other cancer. For example, the administration of a test compound which increases the activity of a particular biomarker may decrease the risk of lung cancer in a patient if the activity of the

particular biomarker *in vivo* prevents the accumulation of proteins for lung cancer. Conversely, the administration of a test compound which decreases the activity of a particular biomarker may decrease the risk of lung cancer in a patient if the increased activity of the biomarker is responsible, at least in part, for the onset of lung cancer.

5 At the clinical level, screening a test compound includes obtaining samples from test subjects before and after the subjects have been exposed to a test compound. The levels in the samples of one or more of the biomarkers listed herein may be measured and analyzed to determine whether the levels of the biomarkers change after exposure to a test compound. The samples may be analyzed by mass spectrometry, as described herein, or the samples may be

10 analyzed by any appropriate means known to one of skill in the art. For example, the levels of one or more of the biomarkers listed herein may be measured directly by Western blot using radio- or fluorescently-labeled antibodies which specifically bind to the biomarkers.

Alternatively, changes in the levels of mRNA encoding the one or more biomarkers may be measured and correlated with the administration of a given test compound to a subject. In a

15 further embodiment, the changes in the level of expression of one or more of the biomarkers may be measured using in vitro methods and materials. For example, human tissue cultured cells which express, or are capable of expressing, one or more of the biomarkers listed herein may be contacted with test compounds. Subjects who have been treated with test compounds will be routinely examined for any physiological effects which may result from the treatment. In

20 particular, the test compounds will be evaluated for their ability to decrease disease likelihood in a subject. Alternatively, if the test compounds are administered to subjects who have previously been diagnosed with lung cancer, test compounds will be screened for their ability to slow or stop the progression of the disease.

### Computer Implementation

The techniques of the present invention may be implemented on a computing system 104 such as that depicted in FIG. 11. In this regard, FIG. 11 is an illustration of a computer system 104 which is also capable of implementing some or all of the computer processing in accordance 5 with at least one computer implemented embodiment of the present invention.

Viewed externally, in FIG. 11, a computer system designated by reference numeral 104 has a computer portion 112 having drives 502 and 504, which are merely symbolic of a number of disk drives which might be accommodated by the computer system. Typically, these could include a floppy disk drive 502, a hard disk drive (not shown externally) and a CD ROM 504.

10 The number and type of drives vary, typically with different computer configurations. Disk drives 502 and 504 are in fact optional, and for space considerations, can be omitted from the computer system.

The computer system 104 also has an optional display monitor 110 upon which visual information pertaining to cells being normal or abnormal, suspected normal, suspected abnormal, 15 etc. can be displayed. In some situations, a keyboard 116 and a mouse 114 are provided as input devices through which input may be provided, thus allowing input to interface with the central processing unit (CPU) 604 (FIG. 12). Then again, for enhanced portability, the keyboard 116 can be either a limited function keyboard or omitted in its entirety. In addition, the mouse 114 optionally is a touch pad control device, or a track ball device, or even omitted in its entirety as 20 well, and similarly may be used as an input device. In addition, the computer system 104 may also optionally include at least one infrared (or radio) transmitter and/or infrared (or radio) receiver for either transmitting and/or receiving infrared signals.

Although computer system 104 is illustrated having a single processor, a single hard disk drive 614 and a single local memory, the system 104 is optionally suitably equipped with any multitude or combination of processors or storage devices. Computer system 104 is, in point of fact, able to be replaced by, or combined with, any suitable processing system operative in accordance with the principles of the present invention, including hand-held, laptop/notebook, mini, mainframe and super computers, as well as processing system network combinations of the same.

FIG. 12 illustrates a block diagram of the internal hardware of the computer system 104 of FIG. 11. A bus 602 serves as the main information highway interconnecting the other components of the computer system 104. CPU 604 is the central processing unit of the system, performing calculations and logic operations required to execute a program. Read only memory (ROM) 606 and random access memory (RAM) 608 constitute the main memory of the computer system 104. Disk controller 610 interfaces one or more disk drives to the system bus 602. These disk drives are, for example, floppy disk drives such as 502, CD ROM or DVD (digital video disks) drive 504, or internal or external hard drives 614. As indicated previously, these various disk drives and disk controllers are optional devices.

A display interface 618 interfaces display 110 and permits information from the bus 602 to be displayed on the display 110. Again as indicated, display 110 is also an optional accessory. For example, display 110 could be substituted or omitted. Communications with external devices, for example, the other components of the system described herein, occur utilizing communication port 616. For example, optical fibers and/or electrical cables and/or conductors and/or optical communication (e.g., infrared, and the like) and/or wireless communication (e.g., radio frequency (RF), and the like) can be used as the transport medium between the external

devices and communication port 616. Peripheral interface 620 interfaces the keyboard 116 and the mouse 114, permitting input data to be transmitted to the bus 602.

In alternate embodiments, the above-identified CPU 604, may be replaced by or combined with any other suitable processing circuits, including programmable logic devices, 5 such as PALs (programmable array logic) and PLAs (programmable logic arrays). DSPs (digital signal processors), FPGAs (field programmable gate arrays), ASICs (application specific integrated circuits), VLSIs (very large scale integrated circuits) and the like.

Any presently available or future developed computer software language and/or hardware components can be employed in such embodiments of the present invention. For example, at 10 least some of the functionality mentioned above could be implemented using Extensible Markup Language (XML), HTML, Visual Basic, C, C++, or any assembly language appropriate in view of the processor(s) being used. It could also be written in an interpretive environment such as Java and transported to multiple destinations to various users.

One of the implementations of the invention is as sets of instructions resident in the 15 random access memory 608 of one or more computer systems 104 configured generally as described above. Until required by the computer system 104, the set of instructions may be stored in another computer readable memory, for example, in the hard disk drive 614, or in a removable memory such as an optical disk for eventual use in the CD-ROM 504 or in a floppy disk (e.g., floppy disk 702) for eventual use in a floppy disk drive 502. Further, the set of 20 instructions (such as those written in Java, HTML, XML, Standard Generalized Markup Language (SGML), and/or Structured Query Language (SQL)) can be stored in the memory of another computer and transmitted via a transmission medium such as a local area network or a wide area network such as the Internet when desired by the user. One skilled in the art knows

that storage or transmission of the computer program medium changes the medium electrically, magnetically, or chemically so that the medium carries computer readable information.

Reference will now be made to specific examples illustrating the biomarkers, kits, computer program media and methods above. It is to be understood that the examples are 5 provided to illustrate preferred embodiments and that no limitation of the scope of the invention is intended thereby.

#### **EXAMPLE 1**

##### **Bronchial Lavage Samples**

Bronchial lavage samples were obtained from Dr. William Rom at New York University. 10 After informed consent, bronchial lavage samples were obtained from lung cancer patients and from controls. The bronchial lavage samples were separated out, aliquotted, and frozen at -80° C until thawed specifically for SELDI analysis.

##### **Patient and Donor Cohorts**

15 Specimens from two groups of patients were used in this study: 13 samples from patients diagnosed with lung cancer and 61 samples from normal, control patients (including samples from the non-cancerous lung from 10 of the 13 lung cancer patients).

##### **SELDI Protein Profiling**

20 Bronchial lavage samples were processed for SELDI analysis as previously described using the IMAC3 ProteinChip® pre-treated with CuSO<sub>4</sub> (Merchant, M., et al., *Electrophoresis* 21:1164-1177 (2000)). Briefly, 200 µl of undiluted bronchial lavage fluid was added to the ProteinChips® with the aid of a bio-processor. Each bronchial lavage sample was assayed in

duplicate, with duplicate samples randomly placed on different ProteinChips®. ProteinChips® were then incubated at room temperature followed by washes of PBS and water. Arrays were allowed to air dry and a saturated solution of sinapinic acid (Ciphergen Biosystems, Fremont, CA.) in 50 % (v/v) acetonitrile, 0.5% (v/v) trifluoroacetic acid was added to each spot. The 5 ProteinChips® were analyzed using the SELDI ProteinChip® System (PBS-II, Ciphergen Biosystems, Inc.). Spectra were collected by the accumulation of 192 shots in the positive mode using a laser intensity of 220. The protein masses were calibrated externally using purified peptide standards (Ciphergen Biosystems, Inc.) Instrument settings were optimized using a pooled serum standard.

10

### Data Analysis

The data consisted of a learning set consisting of 61 normal samples and 13 lung cancer samples. This learning set was then subjected to five-fold cross validation to determine whether the classification rate was retained.

15

### Peak Detection

Spectra were analyzed in the mass range of 2-100kDa with the Ciphergen ProteinChip® software (version 3.2) and normalized using total ion current. Peak detection and clustering were performed using Ciphergen's Biomarker Wizard tool, using values of 3 for signal to noise 20 threshold, 10% peak threshold and a mass window of 0.2%. All the labeled peaks were exported from SELDI to an Excel spreadsheet.

**Classification and Regression Tree (CART) Analysis**

Construction of the decision tree classification algorithm was performed as described by Breiman, L., et al., *Classification and Regression Trees*, (1984), using a learning data set consisting of 74 samples (61 normal and 13 lung cancer). Details regarding the Classification and Regression Tree (CART) and the artificial intelligence bioinformatics algorithm incorporated within the BioMarker Patterns software program have also been described in Bertone, P., et al., *Nucleic Acids Res.* 29: 2884-2898 (2001); Kosuda, S., et al., *Ann. Nucl. Med.* 16: 263-271 (2002). Classification trees split the data into two bins or nodes, using one rule at a time in the form of a question. The splitting decision was based on the presence or absence and the intensity levels of one peak. Therefore, each peak or cluster identified from the SELDI profile was a variable in the classification process. For example, the answer to "does mass A have an intensity less than or equal to X" splits the data into two nodes, a left node for yes and a right node for no. This "splitting" process continues until terminal nodes are reached and further splitting has no gain in data classification. Classification of terminal nodes was determined by the group ("class") of samples (i.e., Lung Cancer, Normal) representing the majority of samples in that node. Classification trees were constructed using the learning set and then subjected to five-fold cross validation. Multiple classification trees were generated using this process, and the best performing tree was chosen for further testing.

**20 Statistical Analysis**

Specificity was calculated as the ratio of the number of negative samples correctly classified to the total number of true negative samples. Sensitivity was calculated as the ratio of the number of correctly classified diseased samples to the total number of diseased samples.

Comparison of relative peak intensity levels between groups was calculated using the Student's t-test.

### **Data Analysis**

5 Data from the BioMarker Wizard analysis was exported into a spreadsheet, and the intensity values for each peak were averaged for duplicate samples. This analysis identified a large number of peaks per spectrum. Of these, 102 common peaks or clusters were obtained from the IMAC3 protein profiling. As shown in FIG. 10, 31 of these peaks were found to have significant differential expression levels between lung cancer and control bronchial lavage fluid.

10

### **CART Analysis**

Using all 102 peaks, classification trees were created using the learning set with V-fold cross validation. This type of cross validation uses random numbers to split up the data in the learning set for testing each tree. Based on CART analysis, the underexpression of a protein  
15 peak at 3820 was found and used in the best performing classification tree as the first primary splitter. Figure 2 is a representative gel-view showing the underexpression of this peak, as well as the 4069 Da peak, in the lung cancer BAL samples when compared to control BAL samples. Figure 2 also shows the plotted averaged normalized intensity values for the 3820 and 4069 Dalton peaks and shows that the average expression for these peaks is five-fold lower in lung  
20 cancer BAL samples compared to the average expression in the control BAL samples. Furthermore, Figure 3 shows a representative spectra and the plotted averaged normalized intensity values for the 30132 Dalton peak which is found to be overexpressed in lung cancer BAL samples as compared to control samples. As seen in Figure 3, there appears to be a pattern

shift in the diseased spectra from a peak below 30 kDa to the higher molecular weight peak of 30132 Da. This may be indicative of post-translational modifications.

All 102 peaks were used to construct the decision tree classification algorithm. One sample classification algorithm used 4 masses between 3-7 kDa to generate 6 terminal nodes (Figure 4). The classification algorithm used 10 additional peaks (from FIG. 10) as surrogates or competitors. Once the algorithm identified the most discriminatory peaks, the classification rule was quite simple.

The classification tree analysis yielded a total of 4 trees with classification rates above 85% correct. The most accurate tree correctly classified 96.7% of the normal and 100% of the lung cancer BAL samples in the learning set (see Table 1). This classification tree algorithm was subjected to five-fold cross validation and the correct classification rate was retained. 86.9% of the controls and 84.6% of lung cancer samples were correctly identified in the cross validation set (see Table 1). The topology of the classification tree consisted of 4 primary peaks (3820, 3506, 4571, and 6933 Da) and 6 terminal nodes (see Figure 4).

A summation of the classification results from the 6 terminal nodes is presented for the learning and cross validation sets in Table 1 seen below.

**Table 1. Decision Tree Classification of the Lung Cancer Learning and Cross Validation Sets Based on Bronchial Lavage Fluid**

**20 A. Learning Set**

Class	Total Cases	Percent Correct	Normal	Cancer
Normal	61	96.72	59	2
Cancer	13	100	0	13

## B. Cross Validation

Class	Total Cases	Percent Correct	Normal	Cancer
Normal	61	86.89	53	8
Cancer	13	84.62	2	11

## Reproducibility

- 5 A key aspect of any clinical approach for reliable disease diagnostics and early detection is reproducibility. The reproducibility of SELDI data has been demonstrated previously using a pooled normal serum sample (Adam, B.L., et al., *Cancer Res.* 62:3609-3614 (2002)). The intra-assay and inter-assay coefficient of variance (CV) for peak masses is routinely 0.05% with normalized intensity CV values of 15-20%.

10

## EXAMPLE 2

### Serum Samples

- Serum samples were obtained from Dr. William Rom at New York University. After informed consent, whole blood was drawn from lung cancer patients, non-cancerous patients  
15 with abnormal lung CT scans, healthy smokers, and healthy non-smokers. The serum was separated out, aliquotted, and frozen at -80° C until thawed specifically for SELDI analysis.

### **Patient and Donor Cohorts**

Specimens from four groups of patients were used in this study: 21 samples from patients diagnosed with lung cancer, 16 samples from healthy smokers, 10 samples from healthy non-smokers, and 4 samples from non-cancer patients with abnormal lung CT scans.

5

### **SELDI Protein Profiling**

Serum samples were processed for SELDI analysis as previously described using the IMAC3 ProteinChip® pre-treated with CuSO<sub>4</sub> (Merchant, M., et al., *Electrophoresis* 21:1164-1177 (2000)). Briefly, 20μl of serum was pre-treated with 8M urea, 1% CHAPS and vortexed 10 for 10 minutes at 4° C. A further dilution was made in 1M urea, 0.125% CHAPS and PBS.

Diluted serum was then added to the ProteinChips® with the aid of a bio-processor. Each serum sample was then assayed in duplicate. The ProteinChips® were analyzed using the SELDI ProteinChip® System (PBS-II, Ciphergen Biosystems, Inc.). Spectra were collected by the accumulation of 192 shots in the positive mode. The protein masses were calibrated externally 15 using purified peptide standards (Ciphergen Biosystems, Inc.) Instrument settings were optimized using a pooled serum standard.

### **Data Analysis**

The data consisted of a learning set consisting of 30 "normal" samples (including samples from 16 healthy smokers, 10 healthy non-smokers, and 4 non-cancer patients with abnormal lung 20 CT scans), and 21 lung cancer samples. This learning set was then subjected to five-fold cross validation to determine whether the same classification rate was retained.

**Peak Detection**

Spectra were analyzed in the mass range of 2-100 kDa with the Ciphergen ProteinChip® software (version 3.2) and normalized using total ion current. Peak detection and clustering were performed using Ciphergen's Biomarker Wizard tool, using values of 3 for signal to noise 5 threshold, 10% peak threshold and a mass window of 0.2%. All the labeled peaks were exported from SELDI to an Excel spreadsheet.

**Classification and Regression Tree (CART) and Statistical Analysis**

Construction of the decision tree classification algorithm was performed as described in 10 Example 1, using a learning data set consisting of 51 samples (30 normal and 21 lung cancer). Multiple classification trees were generated using this process, and the best performing tree was chosen for further testing. Specificity and sensitivity were also calculated as described in Example 1.

**15 Data Analysis**

Data from the BioMarker Wizard analysis was exported into a spreadsheet, and the intensity values for each peak were averaged for duplicate samples. This analysis identified a large number of peaks per spectrum. Of these, 60 common peaks or clusters were obtained from the IMAC3 protein profiling. 27 of these peaks were found to have significant differential 20 expression levels between lung cancer and control serum (See Figure 10 which lists 20 of these peaks).

### CART Analysis

Using all 60 peaks, classification trees were created using the learning set with V-fold cross validation. This type of cross validation uses random numbers to split up the data in the learning set for testing each tree. Based on CART analysis, the overexpression of a protein peak 5 at 8603 was found and used in the best performing classification tree as the first primary splitter. Figure 6 is a representative gel-view showing the overexpression of this peak in the lung cancer serum when compared to "normal" serum (including serum from healthy smokers, healthy non-smokers, and non-cancerous patients with abnormal lung CT scans). Figure 6 also shows the plotted averaged normalized intensity values for the 8603 Dalton peak and shows that the 10 average expression is higher in lung cancer serum samples compared to the average expression in the "normal" serum samples. (The ROC plots for this 8603 Da biomarker are shown in FIGS. 8A and B compared to normal nonsmokers and normal smokers.) FIG. 6 further shows that the 8933 Dalton peak is also overexpressed in lung cancer serum when compared to "normal" serum while the 7766 Dalton peak is underexpressed in lung cancer serum as compared to normal 15 serum. As seen in FIG. 6, peak expression of the group with the abnormal CT scan most closely matched the lung cancer group in most cases, while the healthy smokers and healthy non-smokers had similar patterns. FIGS. 7A, B, and D also show that the 4748, 7566, and 4644 Da peaks are underexpressed in lung cancer serum as compared to "normal" controls while FIG. 7C shows that the 4301 biomarker is overexpressed in lung cancer serum as compared to "normal" 20 controls. In addition, FIGS. 8C and D show ROC plots of other peaks with high p-values in comparison with healthy smokers and healthy nonsmokers, including the 8674 and 4301 Da peaks, which were both used in the best performing classification tree.

All 60 peaks were used to construct the decision tree classification algorithm. One sample classification algorithm used 6 masses between 3-9 kDa to generate 7 terminal nodes (Figure 9). Once the algorithm identified the most discriminatory peaks, the classification rule was quite simple.

5 The most accurate tree correctly classified 100% of the normal and 100% of the lung cancer serum samples in the learning set (see FIG. 9). This classification tree algorithm was subjected to five-fold cross validation and the correct classification rate was retained. 83.3% of the "normal" samples and 81.0% of lung cancer samples were correctly identified in the cross validation set (see Table 2). The topology of the classification tree consisted of 6 primary peaks  
10 (8602, 3887, 4644, 8630, 4301, and 8674 Da) and 7 terminal nodes (see Figure 9).

A summation of the classification results from the 7 terminal nodes is presented for the learning and cross validation sets in Table 2 seen below.

15 **Table 2. Decision Tree Classification of the Lung Cancer Learning and Cross Validation Sets Based on Serum**

**A. Learning Set**

Class	Total Cases	Percent Correct	Normal	Cancer
Normal	30	100	30	0
Cancer	21	100	0	21

### B. Cross Validation

Class	Total Cases	Percent Correct	Normal	Cancer
Normal	30	83.3	25	5
Cancer	21	81.0	4	17

Samples from head and neck squamous cell carcinoma ("HNSCC") patients and healthy smokers were also tested using the above described classification tree in Figure 9. A summation  
 5 of the classification results from the 7 terminal nodes is presented in Table 3 seen below.

**Table 3. Decision Tree Classification of the HNSCC and Healthy Smoker Samples**

Class	Total Cases	Percent Correct	Normal	Cancer
HNSCC	24	37.5	9	15
Smokers	76	89.5	8	68

10 **Discussion**

Using SELDI/TOF-MS techniques, the present inventors have surprisingly achieved 86.89% specificity and 84.62% sensitivity for the detection of lung cancer from bronchial lavage fluid samples and 83.3% specificity and 81.0% sensitivity from serum samples, in a rapid and reproducible manner. While lung cancer is most often related to smoking, many of the control  
 15 bronchial lavage and serum samples used in the preceding examples were obtained from normal individuals lacking this risk factor. As seen in Figures 6-8, the protein expression patterns for healthy smokers were more similar to the patterns of lung cancer patients than were the patterns of non-smokers. Significantly, the differences between healthy smokers and lung cancer patients

were expected to be less than those between normal healthy controls and lung cancer patients, since progression from normal to cancer is multifocal and heterogeneous. This suggests that some “healthy” smokers may well be on the way to developing lung cancer without overt clinical signs.

5        Many protein peaks were found to be differentially expressed with high statistical significance in lung cancer compared to control samples (Figure 10). It is notable that while not all of these significant peaks were used in the classification tree algorithms, the present invention contemplates the use of the differentially expressed markers. Unlike statistical tools that look only for single variables that can act as a predictor, CART analysis examines combinations of 10 variables. A significant p-value may be achieved when testing for a group mean difference for a single protein peak. The classification algorithm is able to examine a number of different variables at once, looking for a combination of peak expression that gives the best classification. Furthermore, a peak without a significant p-value when tested between groups, may in fact be relevant to the classification algorithm. For instance, two of the peaks used in the best 15 performing classification tree for bronchial lavage fluid shown in Figure 4 (3506 and 4571 Da) were individually not expressed differentially between the two groups of lavage fluid. However, they were significant to the classification tree to delineate subsets of groups that had been stratified by the significant peak at 3820 Da. The combinations that resulted in maximum sensitivity/specificity for differentiating lung cancer from the non-cancer groups used the 20 patterns of several different masses. One of these masses, the 3820 Da peak, is under-expressed in lung cancer and is one example of how SELDI technology may aid both the discovery of new biologic markers for lung cancer as well as provide analysis of differences in protein expression patterns.

The use of the presently most preferred lung cancer classification systems described herein relies on the protein "fingerprint" pattern of two different groupings of masses. The first grouping, for bronchial lavage samples, has four masses: about 3820, about 3506, about 4571, and about 6933 Daltons. The second grouping, for serum samples, has six masses: about 8603, 5 about 3887, about 4644, about 8630, about 4301, and about 8674 Daltons. These masses have been found to be reproducibly and reliably detected. The mass values and the expression levels (i.e., the values of each peak) for these biomarkers enabled a correct classification or diagnosis. Importantly, knowing the identities of these biomarkers for the purpose of differential diagnosis is not required.

10 In addition to being an important diagnostic tool, SELDI protein profiles can also be utilized before, during, and after treatment of lung cancer in order to determine whether or not a particular cancer treatment is successful and to enable the monitoring of patients for persistent or recurrent disease.

SELDI protein fingerprinting represents a paradigm shift from traditional cancer 15 diagnostic approaches. The discovery of potentially new protein biomarkers is facilitated by SELDI/TOF-MS. While not intending to be bound by a particular theory, it appears that the protein pattern, rather than individual protein alteration, may be more important for differentiating normal healthy individuals from those who have, or are likely to develop, lung cancer. The high sensitivity and specificity achieved in this study using SELDI/TOF-MS 20 techniques, coupled with a robust artificial intelligence classification algorithm, identified protein patterns in serum that distinguished healthy controls from lung cancer patients. This technique provides data that are easy to accumulate and should lend itself readily to clinical use.

While the invention has been illustrated and described in detail in the drawings and foregoing description, the same is to be considered as illustrative and not restrictive in character, it being understood that only the preferred embodiments have been shown and described and that all changes and modifications that come within the spirit of the invention are desired to be 5 protected. In addition, all references and patents cited herein are indicative of the level of skill in the art and are hereby incorporated by reference in their entirety.

## CLAIMS

What is claimed is:

1. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a biological sample from a patient suspected of suffering from lung cancer; detecting at least one protein biomarker in said sample, said protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ ,  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,  $4614 \pm 23$ ,  $5182 \pm 25$ ,  $4069 \pm 20$ ,  $4634 \pm 23$ ,  $11600 \pm 58$ ,  $30133 \pm 150$ ,  $11939 \pm 60$ ,  $17894 \pm 89$ ,  $11723 \pm 58$ ,  $11493 \pm 57$ ,  $4959 \pm 25$ ,  $2013 \pm 10$ ,  $4370 \pm 22$ ,  $45862 \pm 226$ ,  $15105 \pm 75$ ,  $20898 \pm 104$ ,  $38099 \pm 190$ ,  $5873 \pm 27$ ,  $3668 \pm 18$ ,  $9091 \pm 45$ ,  $8491 \pm 42$ ,  $3391 \pm 16$ ,  $4130 \pm 20$ ,  $3136 \pm 15$ ,  $3441 \pm 17$ ,  $30952 \pm 154$ ,  $4029 \pm 20$ ,  $11253 \pm 56$ ,  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ ,  $6933 \pm 34$ ,  $3887 \pm 19$ ,  $8602 \pm 43$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $8674 \pm 43$  Daltons; wherein said detecting of said at least one protein biomarker is correlated with a diagnosis of lung cancer in said patient.
2. The method of claim 1, wherein said detection step further comprises identifying the differential expression of said at least one protein biomarker.

3. The method of claim 1, wherein the correlation takes into account the presence or absence of the said at least one protein biomarker in the sample and the frequency of detection of the same said at least one protein biomarker in a control.

5 4. The method of claim 3, wherein the correlation further takes into account the quantity of said at least one protein biomarker in the sample compared to a control quantity of the said at least one protein biomarker.

5. The method of claim 1, wherein at least one protein biomarker is selected from the group 10 consisting of protein biomarkers having a molecular weight of about  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ , and  $6933 \pm 34$  Dalton biomarkers.

6. The method of claim 5, wherein said method comprises determining the quantity of the protein biomarkers having a molecular weight of about  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ , and 15  $6933 \pm 34$  Dalton biomarkers.

7. The method of claim 1, wherein at least one protein biomarker is selected from the group consisting of protein biomarkers having a molecular weight of about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ ,  $4301 \pm 21$ , and  $8674 \pm 43$  Dalton biomarkers.

20

8. The method of claim 7, wherein said method comprises determining the quantity of the protein biomarkers having a molecular weight of about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ ,  $4301 \pm 21$ , and  $8674 \pm 43$  Dalton biomarkers.

9. The method of claim 1, wherein said detecting at least one protein biomarker is performed by mass spectrometry.

5 10. The method of claim 9, wherein said mass spectroscopy is laser desorption mass spectroscopy.

11. The method of claim of claim 10, wherein said mass spectroscopy is surface enhanced laser desorption/ionization mass spectroscopy.

10

12. The method of claim 11, wherein the laser desorption/ionization mass spectroscopy includes providing a substrate comprising an adsorbent attached thereto, contacting the biological sample with the adsorbent, desorbing and ionizing the biomarkers from the substrate, and detecting the desorbed/ionized biomarkers with a mass spectrometer.

15

13. The method of claim 12, further comprising purifying the biological sample prior to contacting the sample with the adsorbent.

14. The method of claim 1, wherein said detecting at least one protein biomarker in a  
20 biological sample from a subject is performed by immunoassay.

15. The method of claim 14, wherein said immunoassay is an enzyme immunoassay.

16. The method of claim 1, wherein the biological sample is selected from the group consisting of body fluid and tissue.

17. The method of claim 1, wherein the biological sample is blood serum.

5

18. The method of claim 1, wherein the biological sample is bronchial lavage fluid.

19. The method of claim 1, wherein the biological sample is selected from the group consisting of seminal fluid, seminal plasma, saliva, blood, lymph fluid, lung/bronchial washes, mucus, feces, nipple secretions, sputum, tears, or urine.

10 20. The method of claim 1, wherein two to sixty biomarkers are detected.

21. The method of claim 1, wherein said method comprises detecting the presence or absence 15 of protein biomarkers having a molecular weight selected from the group consisting of about 3820  $\pm$  19, 3506  $\pm$  17, 4571  $\pm$  23, and 6933  $\pm$  34 Daltons, and correlating the detection with a probable diagnosis of lung cancer.

22. The method of claim 21, wherein the presence or absence of the protein biomarker 20 having a molecular weight of about 3820  $\pm$  19 Daltons is detected.

23. The method of claim 21, wherein the presence or absence of the protein biomarker having a molecular weight of about 3820  $\pm$  19 and about 3506  $\pm$  17 Daltons is detected.

24. The method of claim 21, wherein the presence or absence of the protein biomarker having a molecular weight of about  $3820 \pm 19$ , about  $3506 \pm 17$ , and about  $4571 \pm 23$  Daltons is detected.

5

25. The method of claim 21, wherein the presence or absence of the protein biomarker having a molecular weight of about  $3820 \pm 19$ , about  $3506 \pm 17$ , about  $4571 \pm 23$ , and about  $6933 \pm 34$  Daltons is detected.

10 26. The method of claim 21, wherein the presence or absence of the protein biomarker having a molecular weight of about  $3820 \pm 19$  and about  $6933 \pm 34$  Daltons is detected.

27. The method of claim 21, wherein said detecting is performed by mass spectroscopy.

15 28. The method of claim 27, wherein said mass spectroscopy is laser desorption mass spectroscopy.

29. The method of claim 28, wherein said mass spectroscopy is surface enhanced laser desorption/ionization mass spectroscopy.

20

30. The method of claim 29, wherein the laser desorption/ionization mass spectroscopy includes providing a substrate comprising an adsorbent attached thereto, contacting the biological sample with the adsorbent, desorbing and ionizing the biomarkers from the substrate, and detecting the desorbed/ionized biomarkers with a mass spectrometer.

25

31. The method of claim 30, further comprising purifying the biological sample prior to contacting the test sample with the adsorbent.

32. The method of claim 21, wherein said detecting is performed by an immunoassay.

33. The method of claim 32, wherein said immunoassay is an enzyme immunoassay.

5

34. The method of claim 1, wherein said method comprises detecting the presence or absence of protein biomarkers having a molecular weight selected from the group consisting of about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ ,  $4301 \pm 21$ , and  $8674 \pm 43$  Daltons, and correlating the detection with a probable diagnosis of lung cancer.

10

35. The method of claim 34, wherein the presence or absence of the protein biomarker having a molecular weight of about  $8603 \pm 43$  Daltons is detected.

36. The method of claim 34, wherein the presence or absence of the protein biomarker having a molecular weight of about  $8603 \pm 43$  and about  $3887 \pm 19$  Daltons is detected.

15

37. The method of claim 34, wherein the presence or absence of the protein biomarker having a molecular weight of about  $8603 \pm 43$ , about  $3887 \pm 19$ , and about  $4644 \pm 23$  Daltons is detected.

20

38. The method of claim 34, wherein the presence or absence of the protein biomarker having a molecular weight of about  $8603 \pm 43$ , about  $3887 \pm 19$ , about  $4644 \pm 23$ , and about  $8630 \pm 43$  Daltons is detected.

39. The method of claim 34, wherein the presence or absence of the protein biomarker having a molecular weight of about  $8603 \pm 43$ , about  $3887 \pm 19$ , about  $4644 \pm 23$ , about  $8630 \pm 43$ , and about  $4301 \pm 21$  Daltons is detected.

5 40. The method of claim 34, wherein the presence or absence of the protein biomarker having a molecular weight of about  $8603 \pm 43$ , about  $3887 \pm 19$ , about  $4644 \pm 23$ , about  $8630 \pm 43$ , about  $4301 \pm 21$ , and about  $8674 \pm 43$  Daltons is detected.

41. The method of claim 34, wherein said detecting is performed by mass spectroscopy.

10 42. The method of claim 41, wherein said mass spectroscopy is laser desorption mass spectroscopy.

15 43. The method of claim 42, wherein said mass spectroscopy is surface enhanced laser desorption/ionization mass spectroscopy.

20 44. The method of claim 43, wherein the laser desorption/ionization mass spectroscopy includes providing a substrate comprising an adsorbent attached thereto, contacting the biological sample with the adsorbent, desorbing and ionizing the biomarkers from the substrate, and detecting the desorbed/ionized biomarkers with a mass spectrometer.

45. The method of claim 44, further comprising purifying the biological sample prior to contacting the test sample with the adsorbent.

25

46. A kit comprising a substrate comprising:

(a) an adsorbent attached thereto, wherein the adsorbent is capable of retaining at least one protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ , and  $6933 \pm 34$  Daltons, and

5 (b) instructions to detect the protein biomarker by contacting a test sample with the adsorbent and detecting the biomarker retained by the adsorbent.

47. The kit of claim 46, wherein the substrate is a probe adapted for use with a gas phase ion spectrometer, said probe having a surface onto which the adsorbent is attached.

10

48. The kit of claim 46, wherein the adsorbent is a metal chelate adsorbent.

49. The kit of claim 46, wherein the adsorbent comprises a cationic group.

15 50. The kit of claim 46, wherein the substrate comprises a plurality of different types of adsorbent.

51. The kit of claim 46, wherein the adsorbent is an antibody that specifically binds to the biomarker.

20

52. The kit of claim 46, wherein the kit further comprises an eluant wherein the biomarker is retained on the adsorbent when washed with the eluant.

53. A kit comprising a substrate comprising:

(a) an adsorbent attached thereto, wherein the adsorbent is capable of retaining at least one protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ ,  $4301 \pm 21$ , and  $8674 \pm 43$  Daltons, and

(b) instructions to detect the protein biomarker by contacting a test sample with the adsorbent and detecting the biomarker retained by the adsorbent.

54. The kit of claim 53, wherein the substrate is a probe adapted for use with a gas phase ion

10 spectrometer, said probe having a surface onto which the adsorbent is attached.

55. The kit of claim 53, wherein the adsorbent is a metal chelate adsorbent.

56. The kit of claim 53, wherein the adsorbent comprises a cationic group.

15

57. The kit of claim 53, wherein the substrate comprises a plurality of different types of adsorbent.

58. The kit of claim 53, wherein the adsorbent is an antibody that specifically binds to the biomarker.

20

59. The kit of claim 53, wherein the kit further comprises an eluant wherein the biomarker is retained on the adsorbent when washed with the eluant.

60. A kit, comprising:

(a) a substrate comprising an adsorbent attached thereto, wherein the adsorbent is capable of retaining at least one protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  
5  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ ,  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,  $4614 \pm 23$ ,  $5182 \pm 25$ ,  $4069 \pm 20$ ,  $4634 \pm 23$ ,  $11600 \pm 58$ ,  $30133 \pm 150$ ,  $11939 \pm 60$ ,  $17894 \pm 89$ ,  $11723 \pm 58$ ,  $11493 \pm 57$ ,  $4959 \pm 25$ ,  $2013 \pm 10$ ,  $4370 \pm 22$ ,  $45862 \pm 226$ ,  
10  $15105 \pm 75$ ,  $20898 \pm 104$ ,  $38099 \pm 190$ ,  $5873 \pm 27$ ,  $3668 \pm 18$ ,  $9091 \pm 45$ ,  $8491 \pm 42$ ,  $3391 \pm 16$ ,  
 $4130 \pm 20$ ,  $3136 \pm 15$ ,  $3441 \pm 17$ ,  $30952 \pm 154$ ,  $4029 \pm 20$ ,  $11253 \pm 56$ ,  $3820 \pm 19$ ,  $3506 \pm 17$ ,  
 $4571 \pm 23$ ,  $6933 \pm 34$ ,  $3887 \pm 19$ ,  $8602 \pm 43$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $8674 \pm 43$  Daltons; and  
(b) instructions to detect the protein biomarker by contacting a test sample with the adsorbent and detecting the biomarker obtained by the adsorbent.

15

61. The kit of claim 60, wherein the substrate is a probe adapted for use with a gas phase ion spectrometer, said probe having a surface onto which the adsorbent is attached.

62. The kit of claim 60, wherein the adsorbent is a metal chelate adsorbent.

20

63. The kit of claim 60, wherein the adsorbent comprises a cationic group.

64. The kit of claim 60, wherein the substrate comprises a plurality of different types of adsorbent.

65. The kit of claim 60, wherein the adsorbent is an antibody that specifically binds to the 5 biomarker.

66. The kit of claim 60, wherein the kit further comprises an eluant wherein the biomarker is retained on the adsorbent when washed with the eluant.

10 67. A method of using a plurality of classifiers to make a probable diagnosis of lung cancer or a negative diagnosis, comprising the steps of obtaining mass spectra from a plurality of samples from normal subjects and subjects diagnosed with lung cancer; and, applying a decision tree analysis to at least a portion of the mass spectra to obtain a plurality of weighted base classifiers comprising a peak intensity value and an associated threshold value, said values used 15 in linear combination to make a probable diagnosis of at least one of lung cancer and a negative diagnosis.

20 68. A computer program medium storing computer instructions therein for instructing a computer to perform a computer-implemented process of aiding in a diagnosis of lung cancer, comprising:

(a) first computer program code means for detecting at least one protein biomarkers in a test sample from a subject, said protein biomarkers having a molecular weight selected from the group consisting of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ , 25  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,

4614  $\pm$  23, 5182  $\pm$  25, 4069  $\pm$  20, 4634  $\pm$  23, 11600  $\pm$  58, 30133  $\pm$  150, 11939  $\pm$  60, 17894  $\pm$  89, 11723  $\pm$  58, 11493  $\pm$  57, 4959  $\pm$  25, 2013  $\pm$  10, 4370  $\pm$  22, 45862  $\pm$  226, 15105  $\pm$  75, 20898  $\pm$  104, 38099  $\pm$  190, 5873  $\pm$  27, 3668  $\pm$  18, 9091  $\pm$  45, 8491  $\pm$  42, 3391  $\pm$  16, 4130  $\pm$  20, 3136  $\pm$  15, 3441  $\pm$  17, 30952  $\pm$  154, 4029  $\pm$  20, 11253  $\pm$  56, 3820  $\pm$  19, 3506  $\pm$  17, 4571  $\pm$  23, 6933  $\pm$

5 34, 3887  $\pm$  19, 8602  $\pm$  43, 4644  $\pm$  23, 8630  $\pm$  43, and 8674  $\pm$  43 Daltons; and

(b) second computer program code means for correlating the detection with a probable diagnosis of lung cancer or a negative diagnosis.

69. The medium of claim 68, wherein the at least one protein biomarker has a molecular weight of about 3820  $\pm$  19 Dalton protein biomarkers.

70. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about 3820  $\pm$  19 and 3506  $\pm$  17 Dalton biomarkers.

15 71. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about 3820  $\pm$  19, 3506  $\pm$  17, and 4571  $\pm$  23 Dalton biomarkers.

72. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about 3820  $\pm$  19, 3506  $\pm$  17, 4571  $\pm$  23, and 6933  $\pm$  34 Dalton biomarkers.

20 73. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about 3820  $\pm$  19 and 6933  $\pm$  34 Dalton biomarkers.

74. The medium of claim 68, wherein the at least one protein biomarker has a molecular weight of about 8603  $\pm$  43 Dalton protein biomarkers.

75. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about 8603  $\pm$  43 and 3887  $\pm$  19 Dalton biomarkers.

30 76. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about 8603  $\pm$  43, 3887  $\pm$  19, and 4644  $\pm$  23 Dalton biomarkers.

77. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ , and  $8630 \pm 43$  Dalton biomarkers.

5 78. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $4301 \pm 21$  Dalton biomarkers.

10 79. The medium of claim 68, wherein the protein biomarkers have a molecular weight of about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ ,  $4301 \pm 21$ , and  $8674 \pm 43$  Dalton biomarkers.

80. The method of claim 1, wherein the protein biomarker has a molecular weight of about  $3820 \pm 19$  Daltons.

15 81. The method of claim 80, wherein said method comprises determining the quantity of the protein biomarker having a molecular weight of about  $3820 \pm 19$  Daltons.

82. The method of claim 1, wherein the protein biomarker has a molecular weight of about  $8603 \pm 43$  Daltons.

20 83. The method of claim 82, wherein said method comprises determining the quantity of the protein biomarker having a molecular weight of about  $8603 \pm 43$  Daltons.

25 84. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a biological sample from a patient suspected of suffering from lung cancer, detecting, by surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS), at least one protein biomarker in said sample, said protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ , and  $6933 \pm 34$  Daltons, wherein said detecting of said at least one protein biomarker is correlated 30 with a diagnosis of lung cancer in said patient.

85. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a biological sample from a patient suspected of suffering from lung cancer, detecting, by surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS), at least one protein biomarker in said sample, said protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $8603 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ ,  $4301 \pm 21$ , and  $8674 \pm 43$  Daltons, wherein said detecting of said at least one protein biomarker is correlated with a diagnosis of lung cancer in said patient.
86. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a body fluid sample from a patient suspected of suffering from lung cancer, detecting, by surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS), a protein biomarker in said sample having a molecular weight of about  $3820 \pm 19$  Daltons, wherein said detecting of said protein biomarker is correlated with a diagnosis of lung cancer in said patient.
87. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a body fluid sample from a patient suspected of suffering from lung cancer, detecting, by surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS), the quantity of a protein biomarker in said sample having a molecular weight of about  $3820 \pm 19$  Daltons, wherein underexpression of said protein biomarker is correlated with a diagnosis of lung cancer in said patient.
88. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a body fluid sample from a patient suspected of suffering from lung cancer, detecting, by surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS), a protein biomarker in said sample having a molecular weight of about  $8603 \pm 43$  Daltons, wherein said detecting of said protein biomarker is correlated with a diagnosis of lung cancer in said patient.

89. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a body fluid sample from a patient suspected of suffering from lung cancer, detecting, by surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS), the quantity of a protein biomarker in said sample having a molecular weight of about  $8603 \pm 43$  Daltons, wherein overexpression of said protein biomarker is correlated with a diagnosis of lung cancer in said patient.
90. A method for monitoring the effectiveness of lung cancer treatment in a patient comprising obtaining a biological sample from a patient undergoing treatment for lung cancer; detecting the quantity of at least one protein biomarker in said sample, said protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ ,  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,  $4614 \pm 23$ ,  $5182 \pm 25$ ,  $4069 \pm 20$ ,  $4634 \pm 23$ ,  $11600 \pm 58$ ,  $30133 \pm 150$ ,  $11939 \pm 60$ ,  $17894 \pm 89$ ,  $11723 \pm 58$ ,  $11493 \pm 57$ ,  $4959 \pm 25$ ,  $2013 \pm 10$ ,  $4370 \pm 22$ ,  $45862 \pm 226$ ,  $15105 \pm 75$ ,  $20898 \pm 104$ ,  $38099 \pm 190$ ,  $5873 \pm 27$ ,  $3668 \pm 18$ ,  $9091 \pm 45$ ,  $8491 \pm 42$ ,  $3391 \pm 16$ ,  $4130 \pm 20$ ,  $3136 \pm 15$ ,  $3441 \pm 17$ ,  $30952 \pm 154$ ,  $4029 \pm 20$ ,  $11253 \pm 56$ ,  $3820 \pm 19$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ ,  $6933 \pm 34$ ,  $3887 \pm 19$ ,  $8602 \pm 43$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $8674 \pm 43$  Daltons; comparing the quantity of said at least one protein biomarker to a known standard; and determining the effectiveness of said lung cancer treatment.
91. The method of claim 90, wherein the known standard is a biological sample from a healthy control.
92. The method of claim 90, wherein the known standard is a biological sample obtained from said lung cancer patient prior to said lung cancer treatment.
93. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a bronchial lavage fluid sample from a patient suspected of suffering from lung cancer; detecting at least one protein biomarker in said sample, said protein biomarker selected from the group

consisting of protein biomarkers having a molecular weight of about  $3821 \pm 19$ ,  $12135 \pm 60$ ,  $5968 \pm 30$ ,  $4614 \pm 23$ ,  $5182 \pm 25$ ,  $4069 \pm 20$ ,  $4634 \pm 23$ ,  $11600 \pm 58$ ,  $30133 \pm 150$ ,  $11939 \pm 60$ ,  $17894 \pm 89$ ,  $11723 \pm 58$ ,  $11493 \pm 57$ ,  $4959 \pm 25$ ,  $2013 \pm 10$ ,  $4370 \pm 22$ ,  $45862 \pm 226$ ,  $15105 \pm 75$ ,  $20898 \pm 104$ ,  $38099 \pm 190$ ,  $5873 \pm 27$ ,  $3668 \pm 18$ ,  $9091 \pm 45$ ,  $8491 \pm 42$ ,  $3391 \pm 16$ ,  $4130 \pm 20$ ,  $3136 \pm 15$ ,  $3441 \pm 17$ ,  $30952 \pm 154$ ,  $4029 \pm 20$ ,  $3506 \pm 17$ ,  $4571 \pm 23$ ,  $6933 \pm 34$ ,  $3820 \pm 19$ , and  $11253 \pm 56$  Daltons; wherein said detecting of said at least one protein biomarker is correlated with a diagnosis of lung cancer in said patient.

94. A method for aiding in a diagnosis of lung cancer in a patient comprising obtaining a serum sample from a patient suspected of suffering from lung cancer; detecting at least one protein biomarker in said sample, said protein biomarker selected from the group consisting of protein biomarkers having a molecular weight of about  $4748 \pm 25$ ,  $8603 \pm 43$ ,  $8675 \pm 43$ ,  $7566 \pm 38$ ,  $7972 \pm 40$ ,  $8812 \pm 44$ ,  $7766 \pm 38$ ,  $7835 \pm 39$ ,  $7925 \pm 40$ ,  $3886 \pm 19$ ,  $4301 \pm 21$ ,  $4645 \pm 23$ ,  $9495 \pm 47$ ,  $11625 \pm 60$ ,  $9288 \pm 46$ ,  $8631 \pm 43$ ,  $8933 \pm 45$ ,  $11728 \pm 59$ ,  $14105 \pm 70$ ,  $11940 \pm 60$ ,  $8861 \pm 44$ ,  $9150 \pm 46$ ,  $10264 \pm 51$ ,  $17047 \pm 85$ ,  $10461 \pm 52$ ,  $13354 \pm 67$ ,  $7471 \pm 37$ ,  $8602 \pm 43$ ,  $3887 \pm 19$ ,  $4644 \pm 23$ ,  $8630 \pm 43$ , and  $8674 \pm 43$  Daltons; wherein said detecting of said at least one protein biomarker is correlated with a diagnosis of lung cancer in said patient.

1/15

Spectra Example BALF

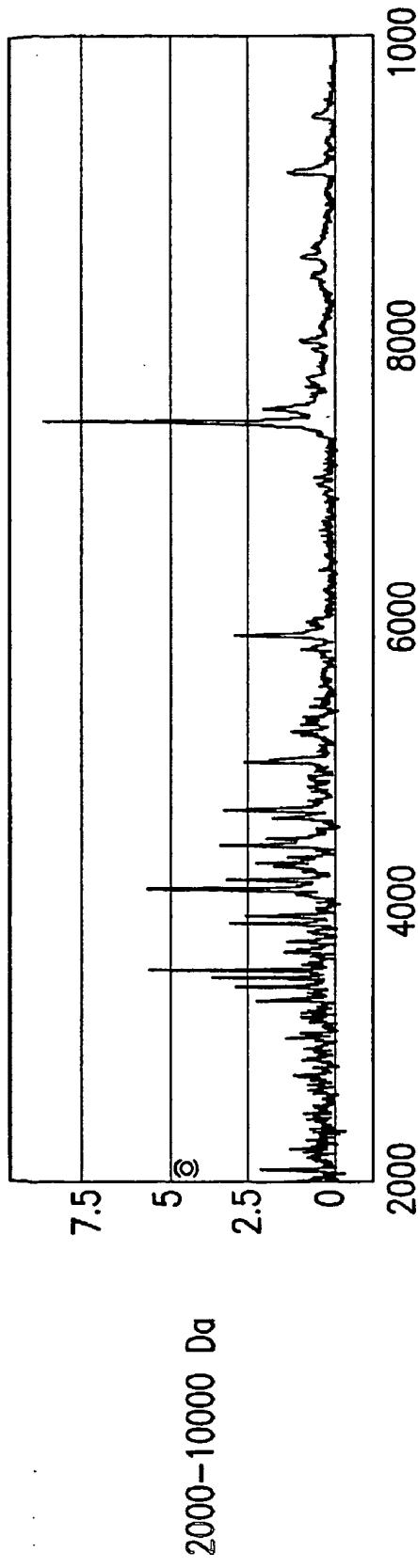


FIG. 1 A

Spectra Example BALF

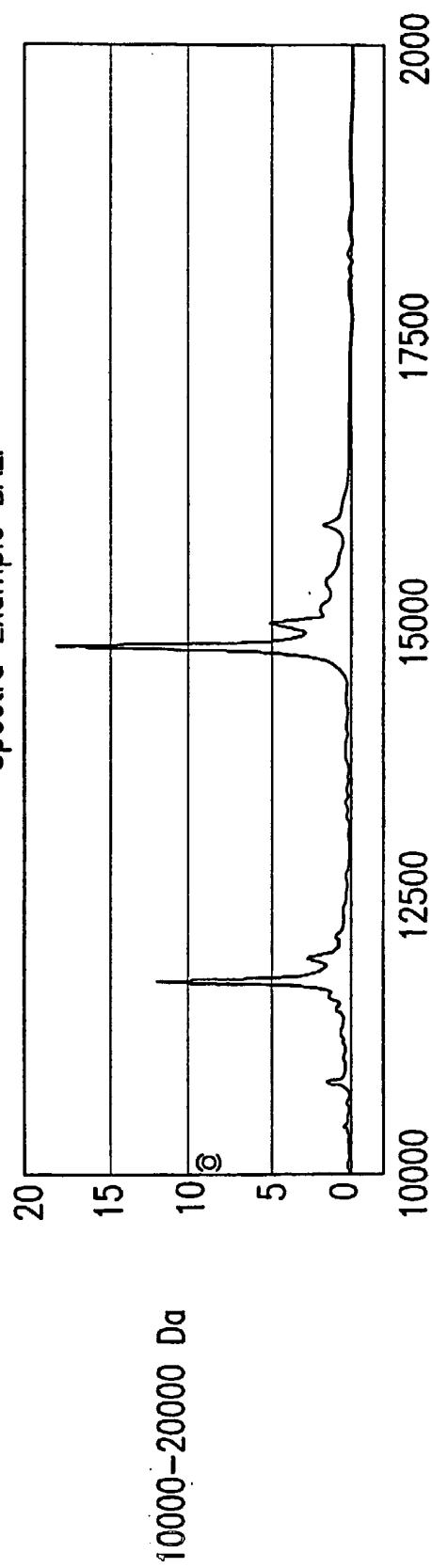


FIG. 1 B

Spectra Example BALF

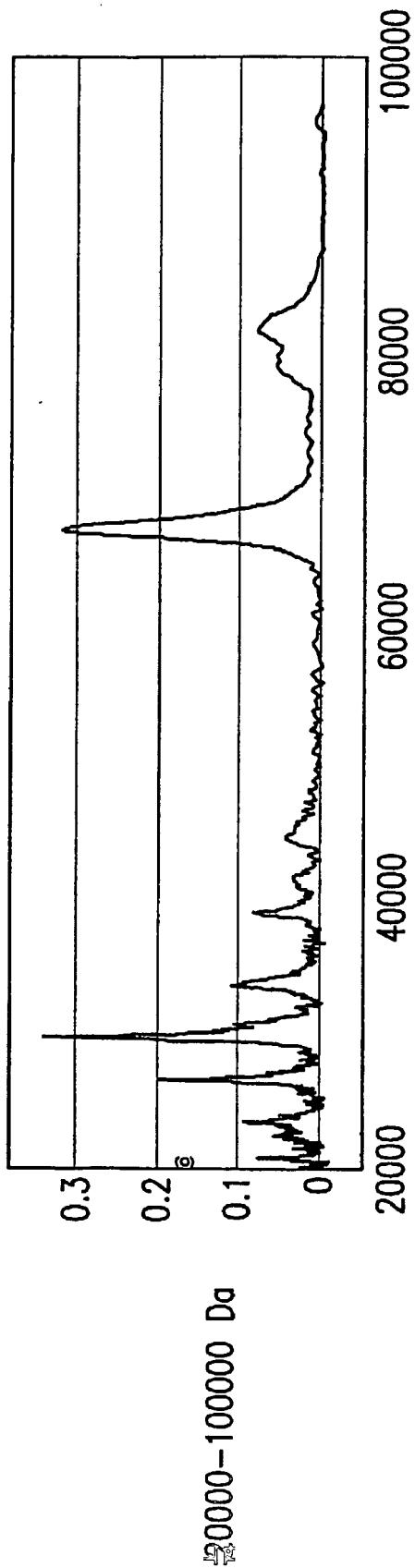


FIG. 1 C

3/15

## Low Molecular Peaks Found Underexpressed in Lung Cancer Bronchial Lavage Fluid

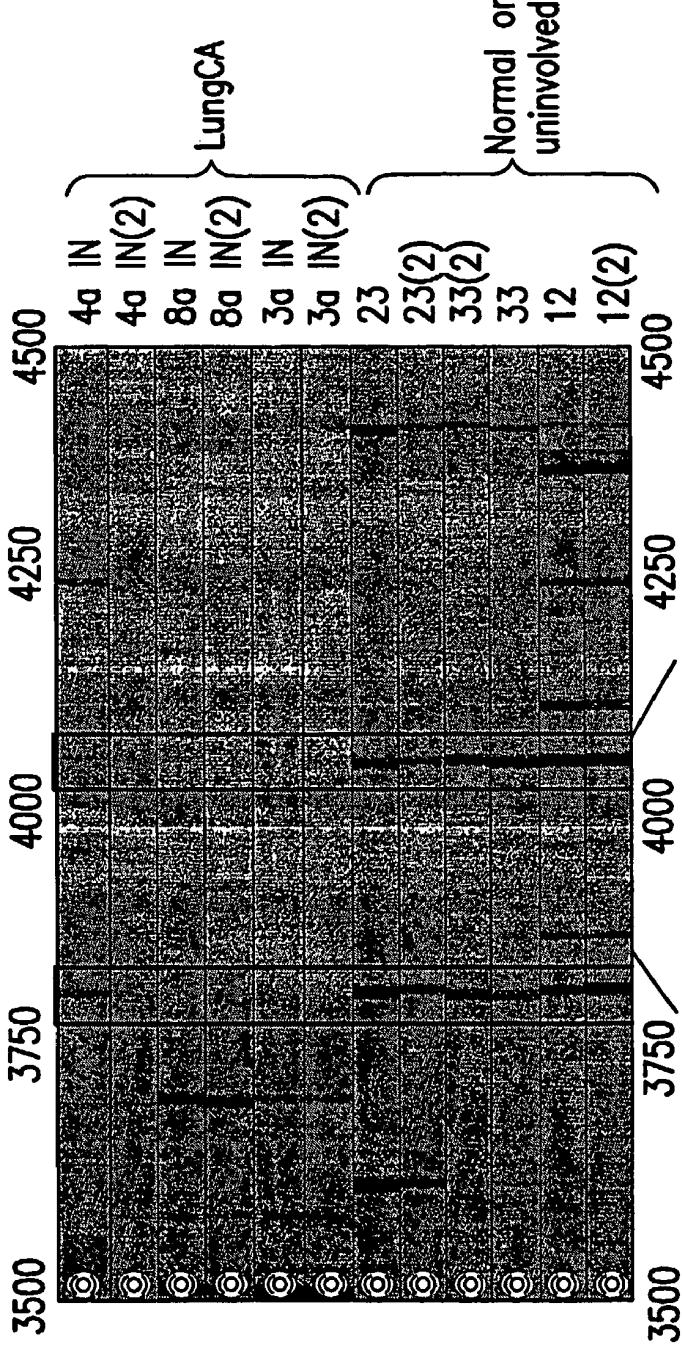


FIG. 2A → 4069

m/z 3820

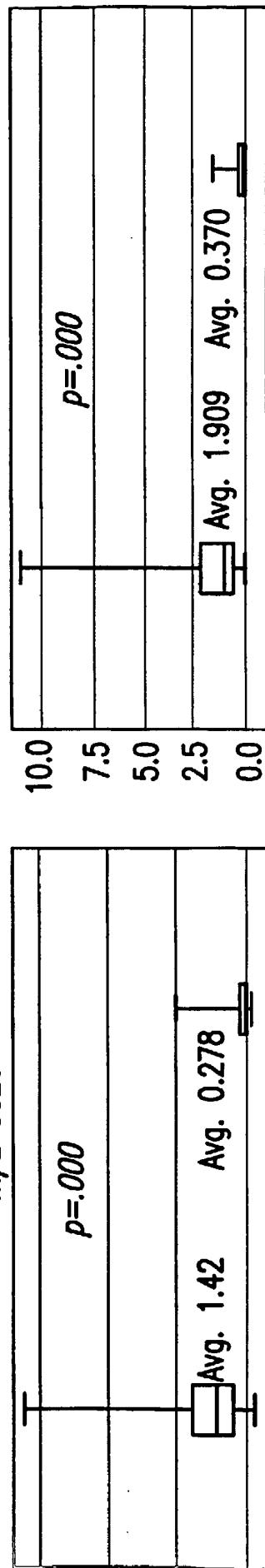


FIG. 2B

Groups	LungCa
Control	100
Groups	115

4/15

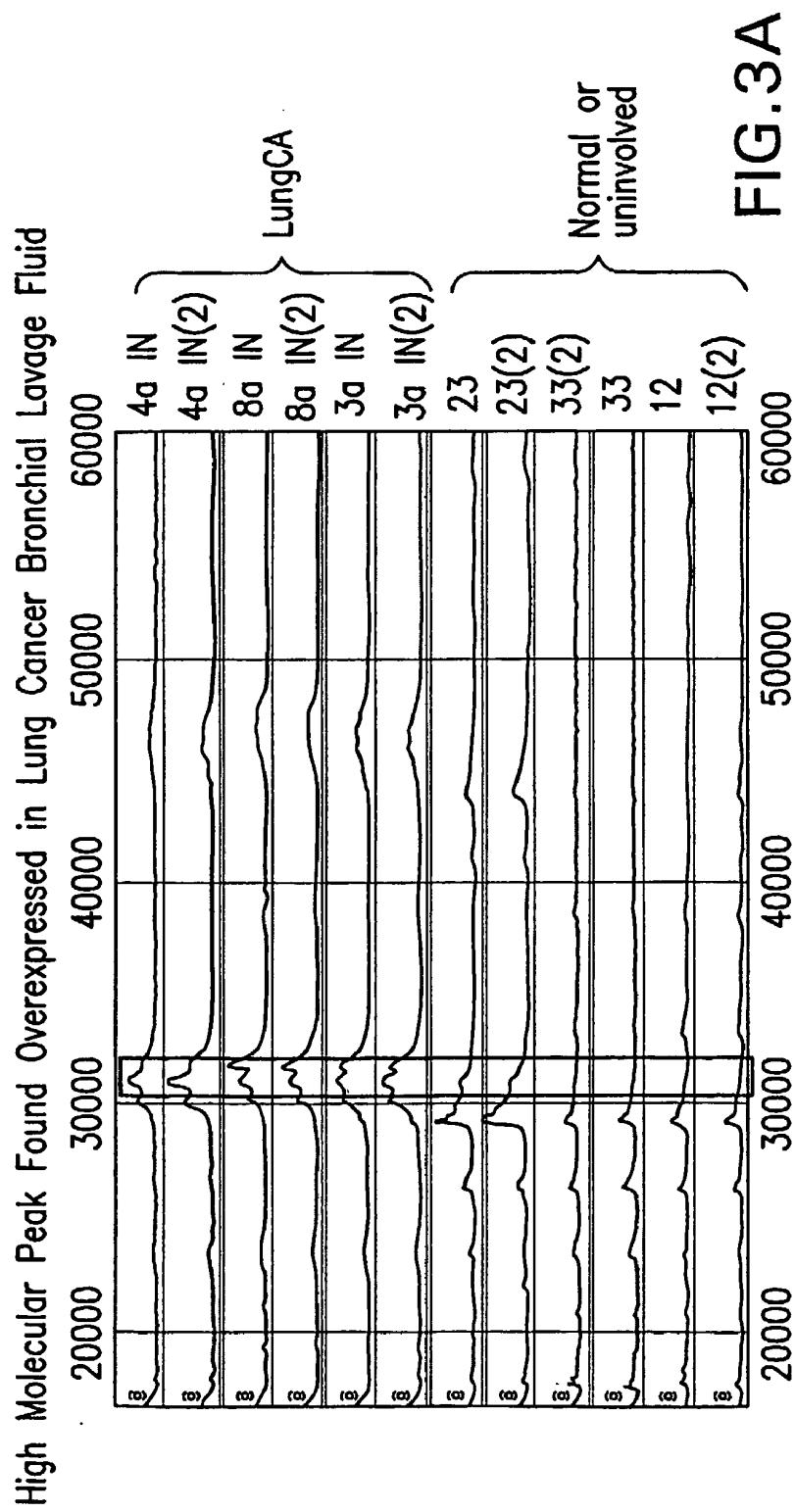


FIG. 3A

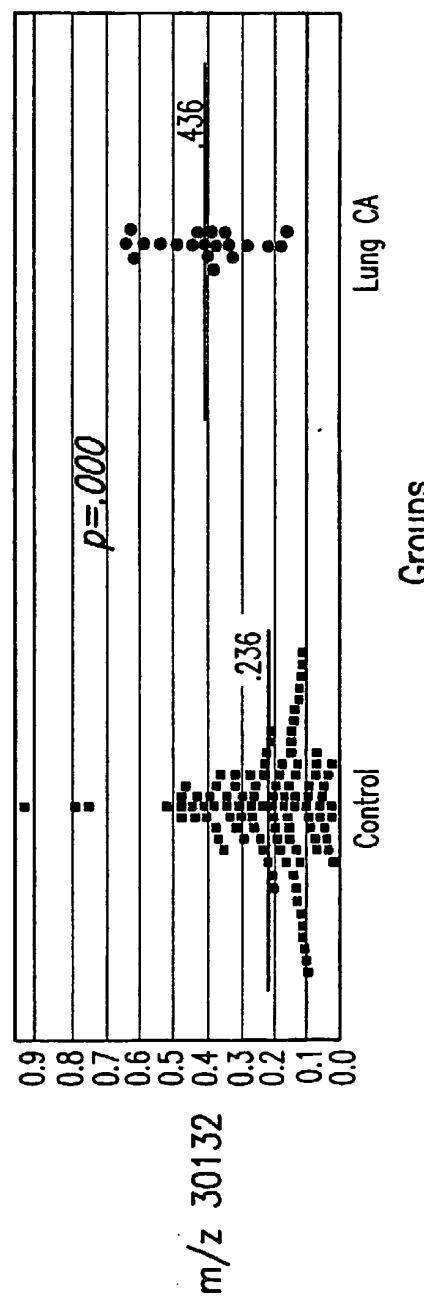
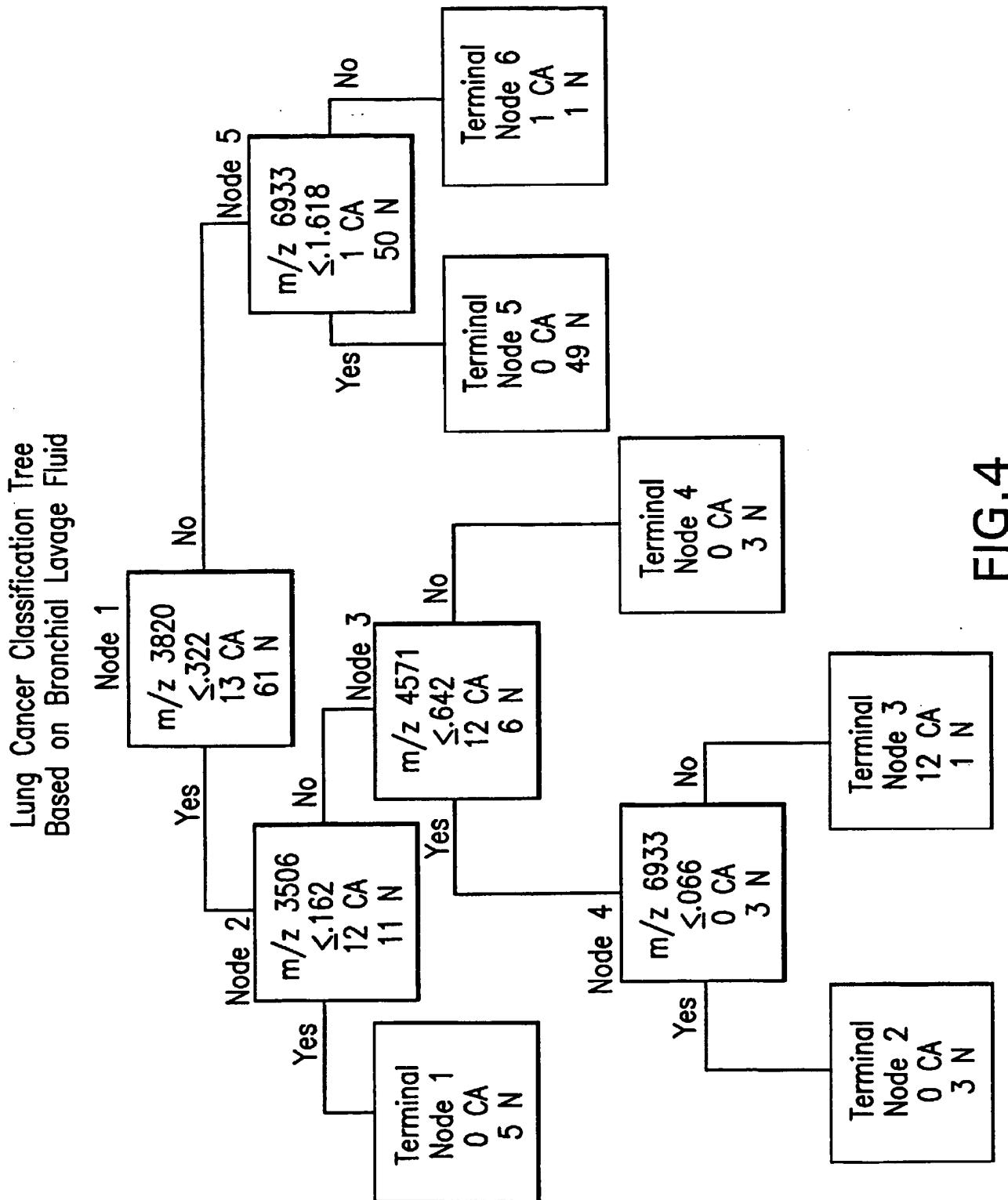


FIG. 3B

5/15



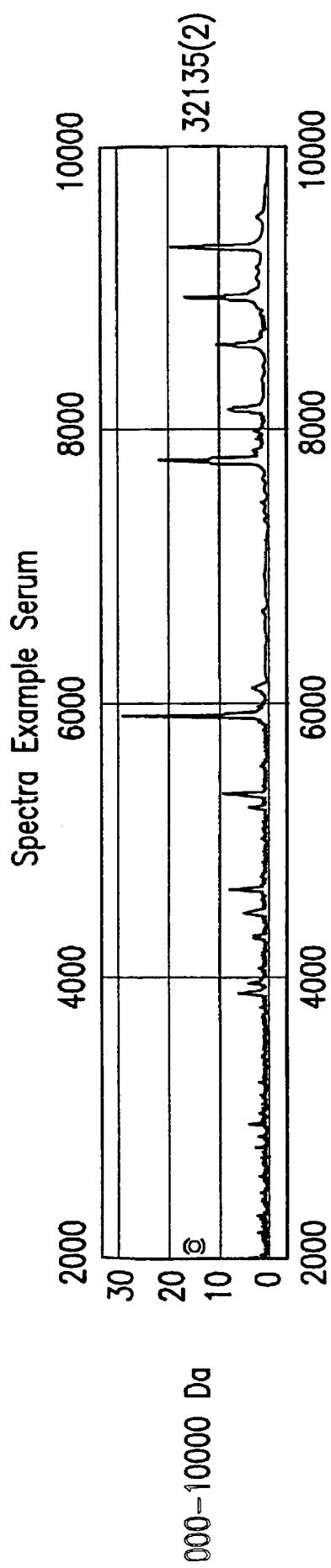


FIG. 5A

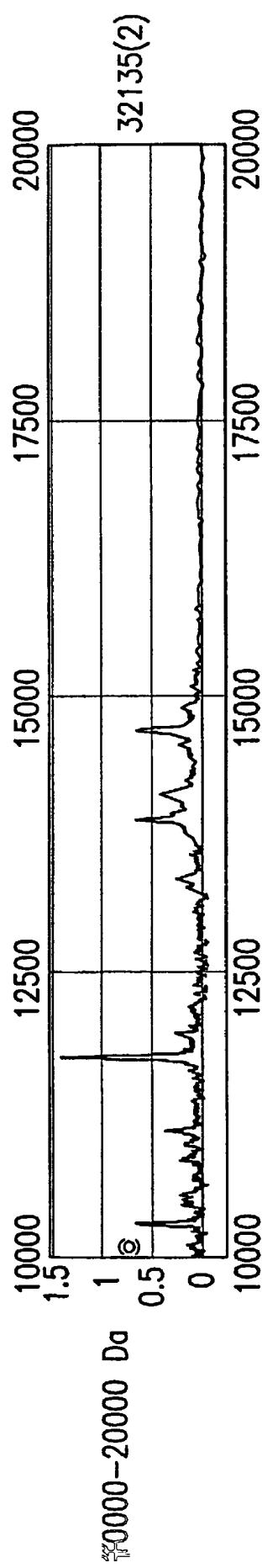


FIG. 5B

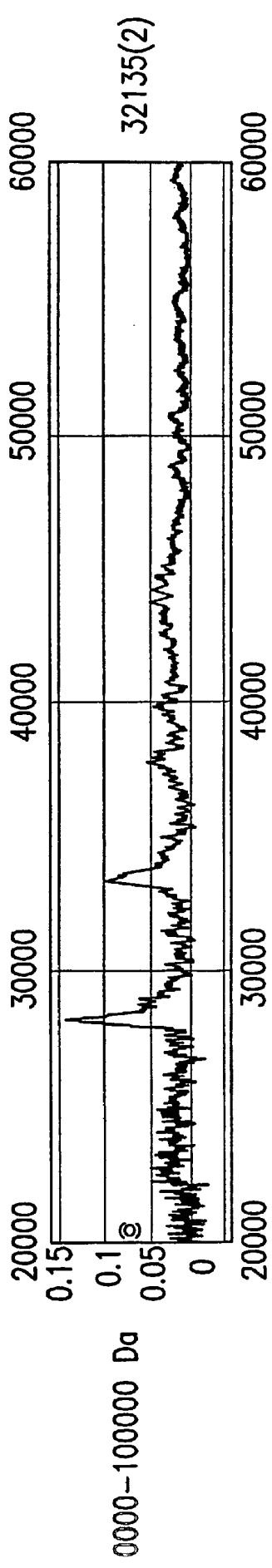
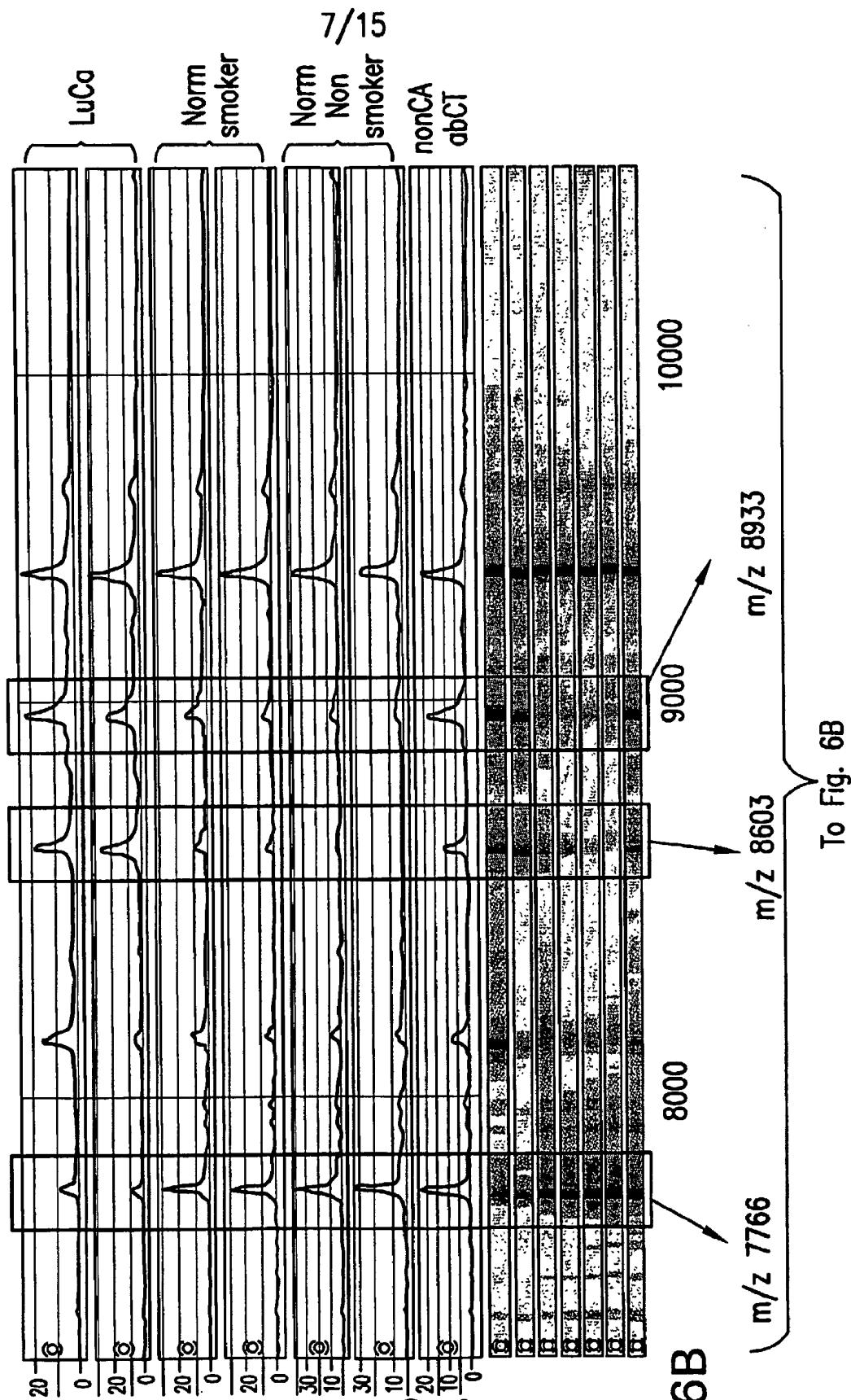


FIG. 5C

FIG. 6A

Example: Peaks Found Differentially Expressed in Serum



8/15

From Fig. 6B

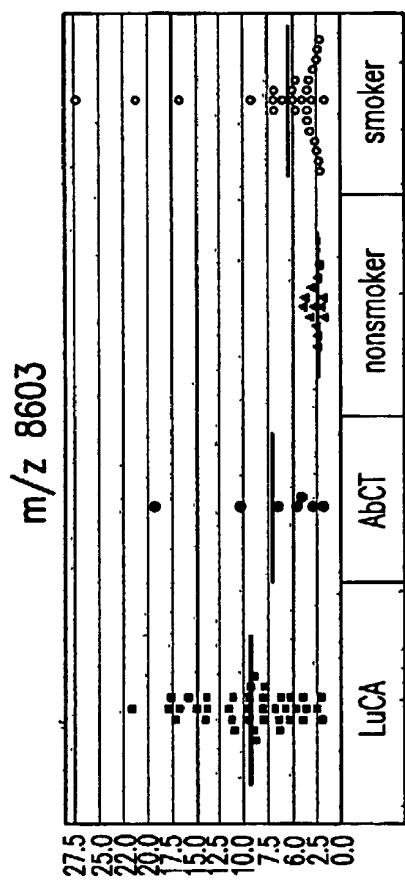
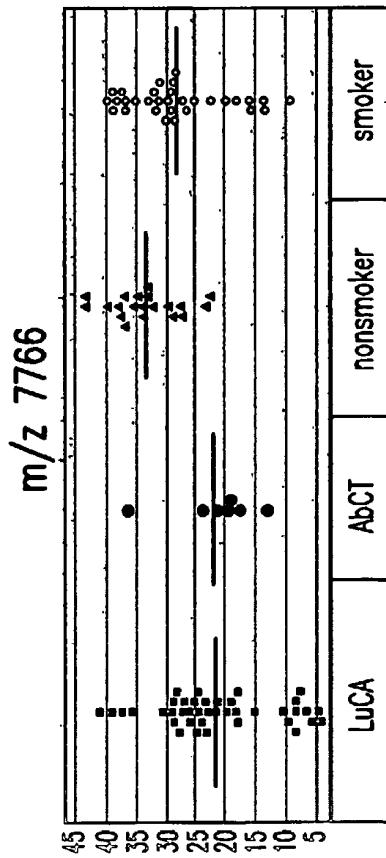
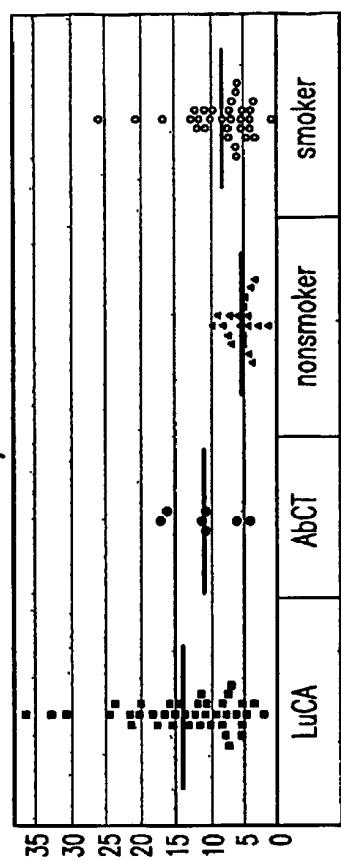
 $m/z$  8933

FIG. 6C

9/15

Two More Peaks Differentially Expressed in Serum

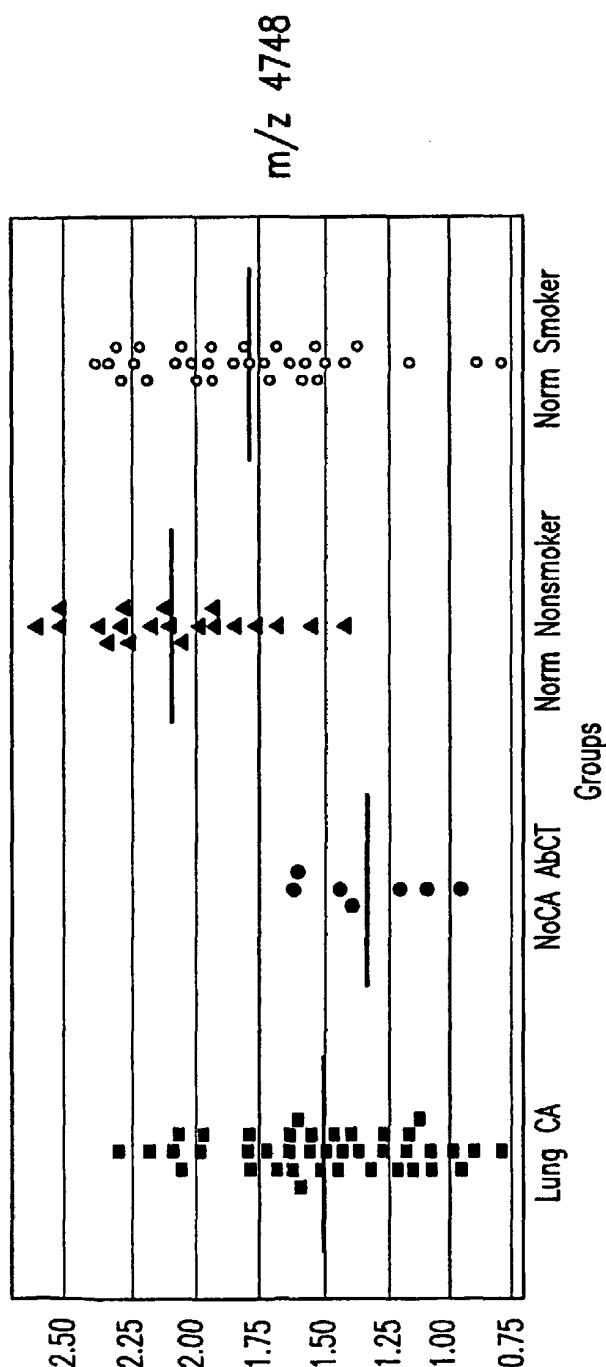


FIG. 7A

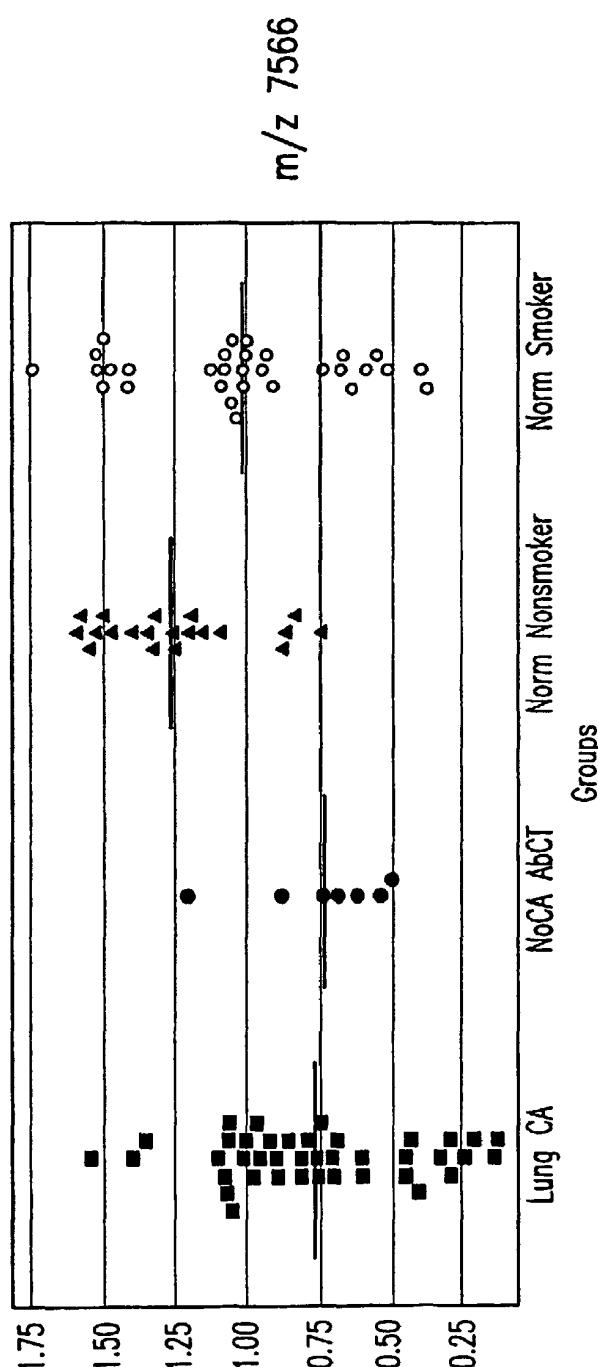


FIG. 7B

10/15

m/z 4301

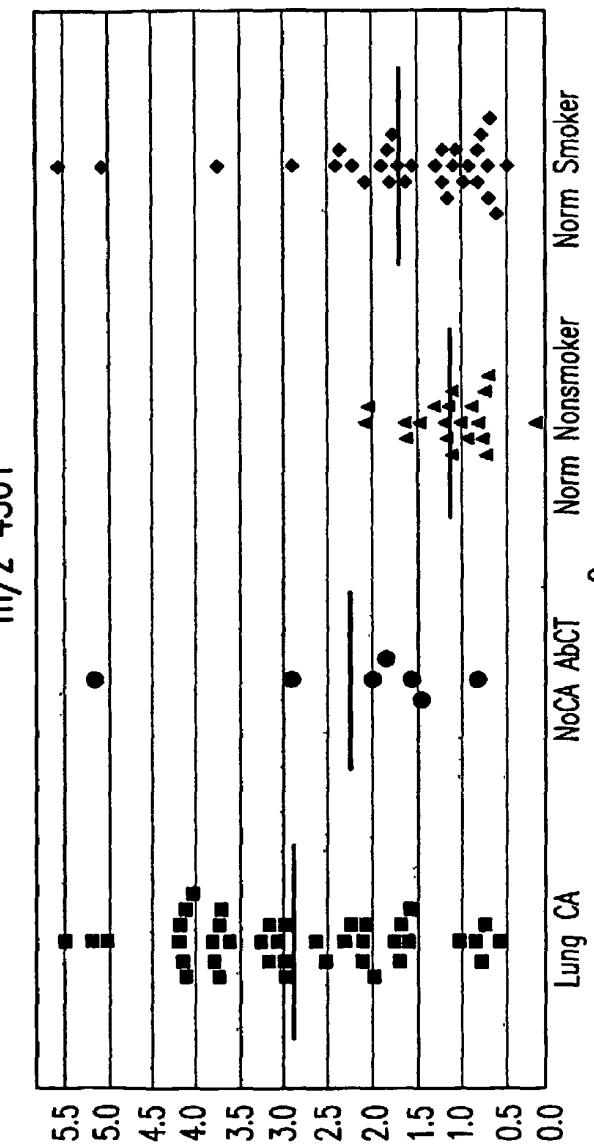


FIG. 7C

m/z 4644

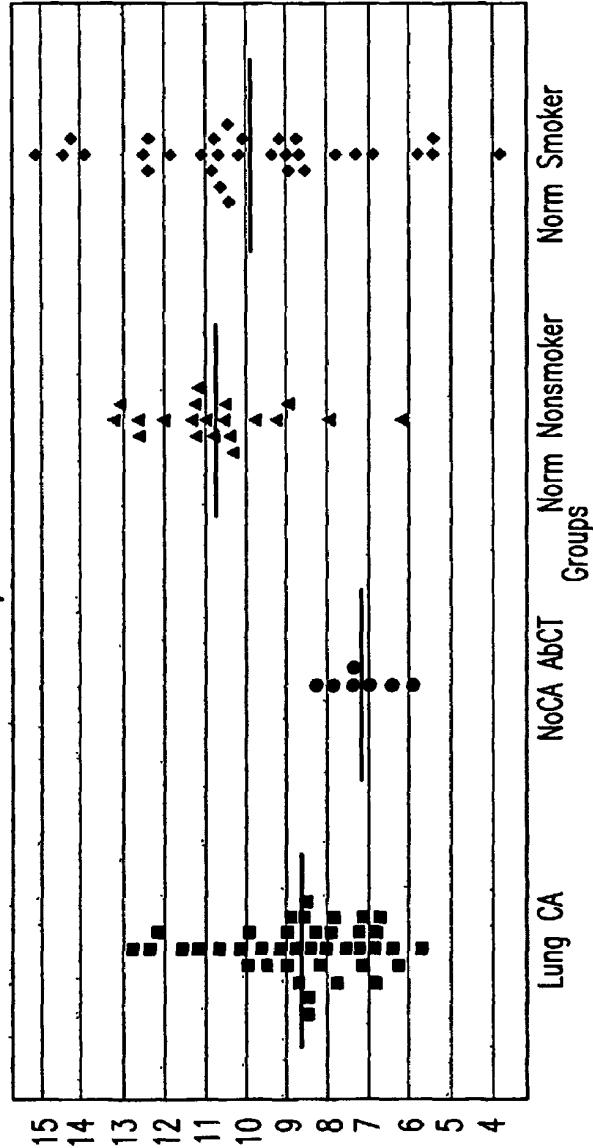


FIG. 7D

11/15

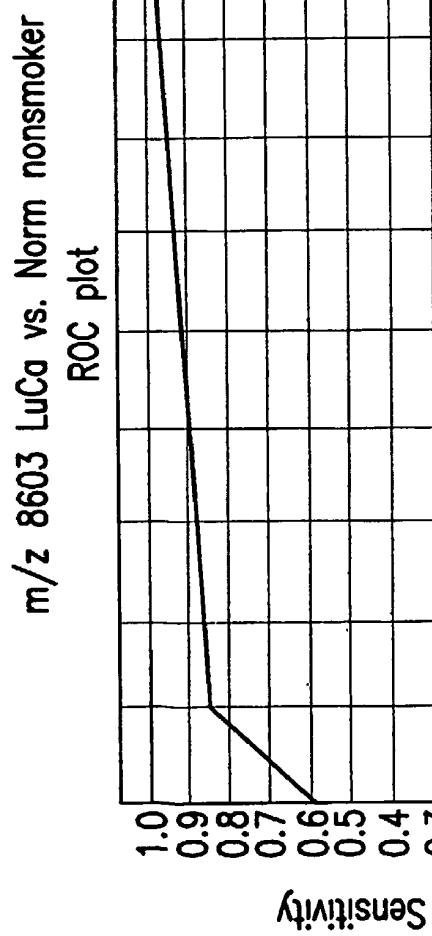


FIG. 8A

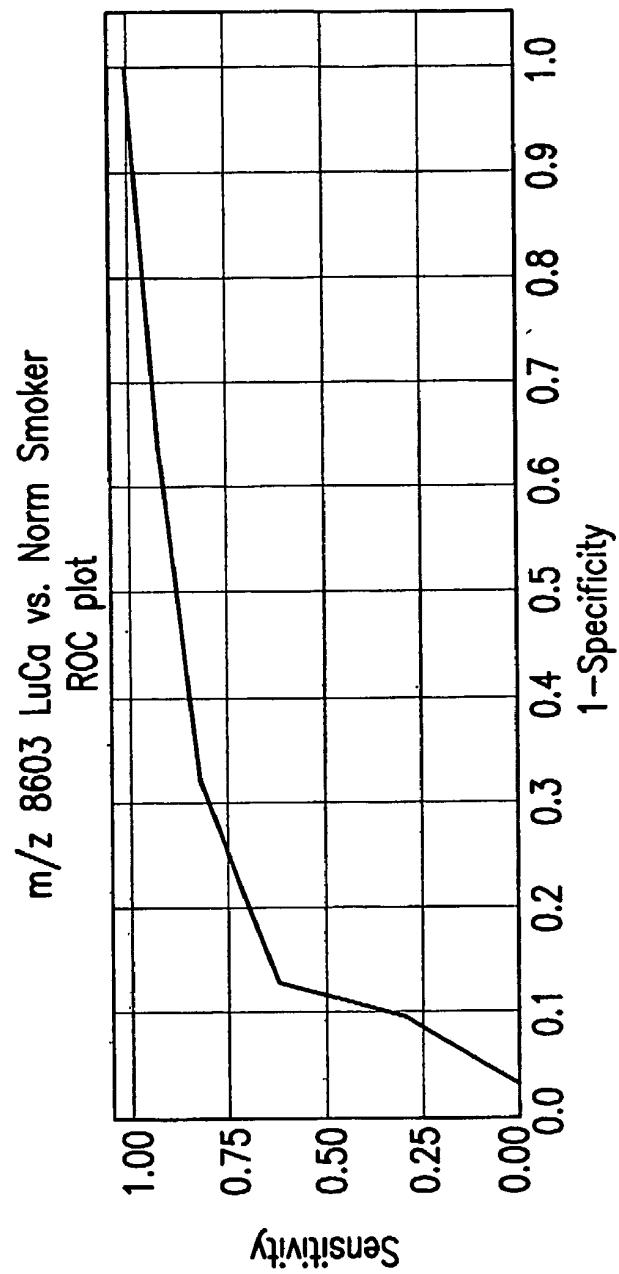


FIG. 8B

12/15

m/z 8674 LuCA vs. Norm nonsmoker  
ROC plot

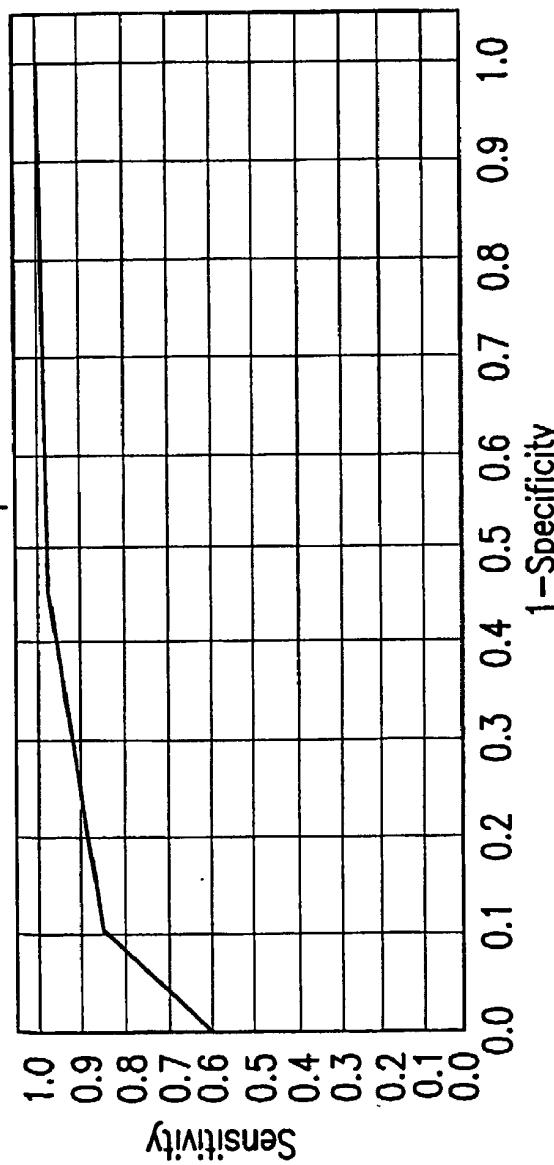


FIG. 8C

m/z 4301 LuCA vs. Norm Smoker  
ROC plot

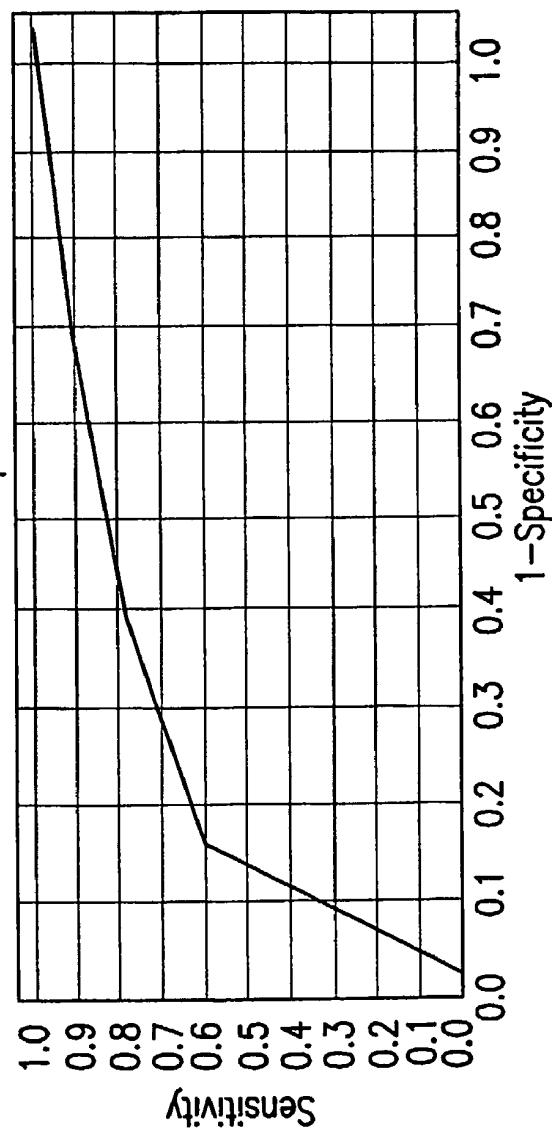


FIG. 8D

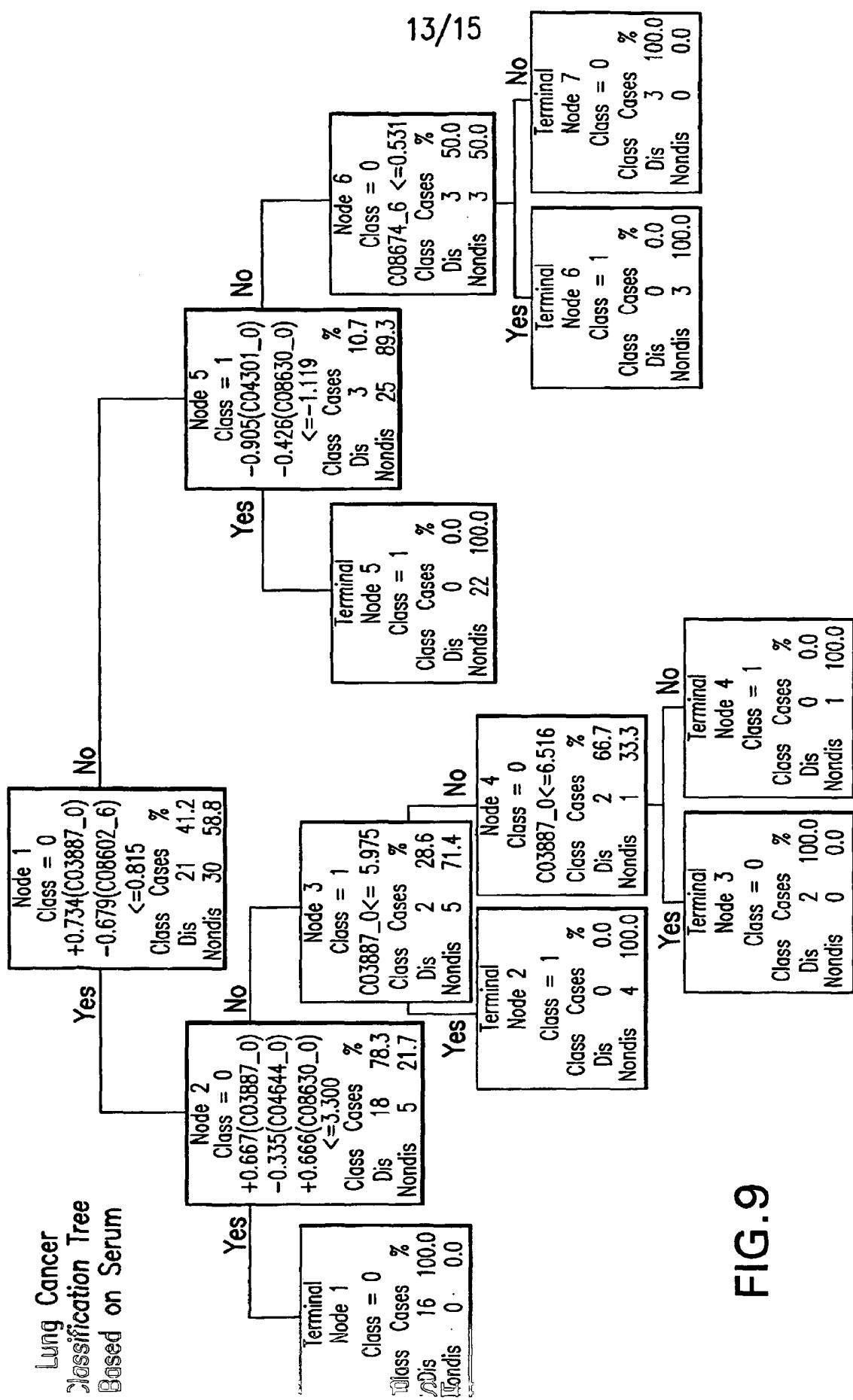


FIG. 9

14/15

SERUM

BAL

Cluster	Mass	p value	Cluster	Mass	p value
13	4748.476	0.0001475822	16	3820.5	0.0000209777
33	8603.114	0.0002715499	68	12135.2	0.0000325584
35	8674.844	0.0003035947	37	5967.6	0.0000600161
25	7566.195	0.0007761133	25	4613.8	0.0000762375
29	7972.365	0.0009560414	31	5182.4	0.0000762375
36	8812.247	0.0011378593	18	4069.2	0.0000808975
26	7766.334	0.0013586059	26	4634.1	0.0000910370
27	7835.187	0.0017213140	64	11599.8	0.0000965460
28	7924.831	0.0024354675	92	30132.7	0.0001023687
5	3885.948	0.0031362032	67	11939.2	0.0001150226
10	4301.109	0.0032292397	82	17893.8	0.0001150226
12	4644.565	0.0032347159	65	11723.3	0.0001291415
43	9494.965	0.0033479649	63	11493.1	0.0001534141
48	11624.92	0.0038867723	29	4959.3	0.0002690260
42	9288.409	0.0043245418	1	2012.6	0.0002842698
34	8631.375	0.0050149764	21	4370.2	0.0003172158
38	8933.498	0.0074213814	101	45862.4	0.0003172158
50	11728.34	0.0077272383	77	15104.5	0.0003537102
54	14104.64	0.0090369824	84	20898.2	0.0004628405
51	11939.74	0.0105434816	97	38098.9	0.0005147143
37	8860.63	0.0168997492	35	5873.3	0.0005426374
41	9149.755	0.0197091208	13	3668.3	0.0006690501
44	10264.19	0.0250278355	55	9090.7	0.0008224105
58	17047.25	0.0317797179	52	8491.1	0.0014281952
45	10460.91	0.0386595177	7	3390.8	0.0019113968
52	13353.53	0.0462630289	19	4130.0	0.0022056104
24	7470.722	0.0468173910	3	3136.1	0.0038450880

■ Used in tree as primary splitter

FIG 10

SUBSTITUTE SHEET (RULE 26)

15/15

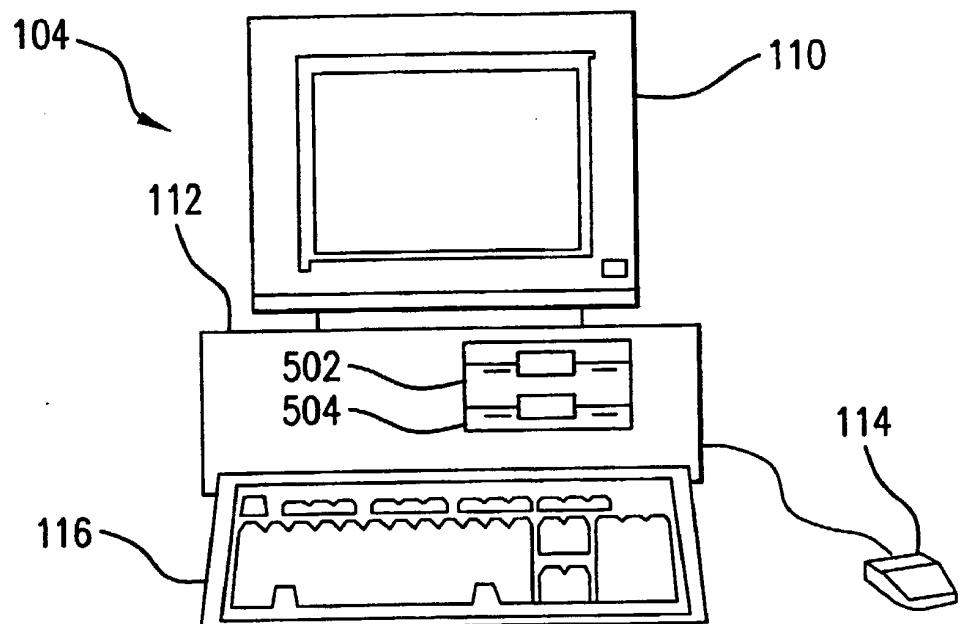


FIG. 11

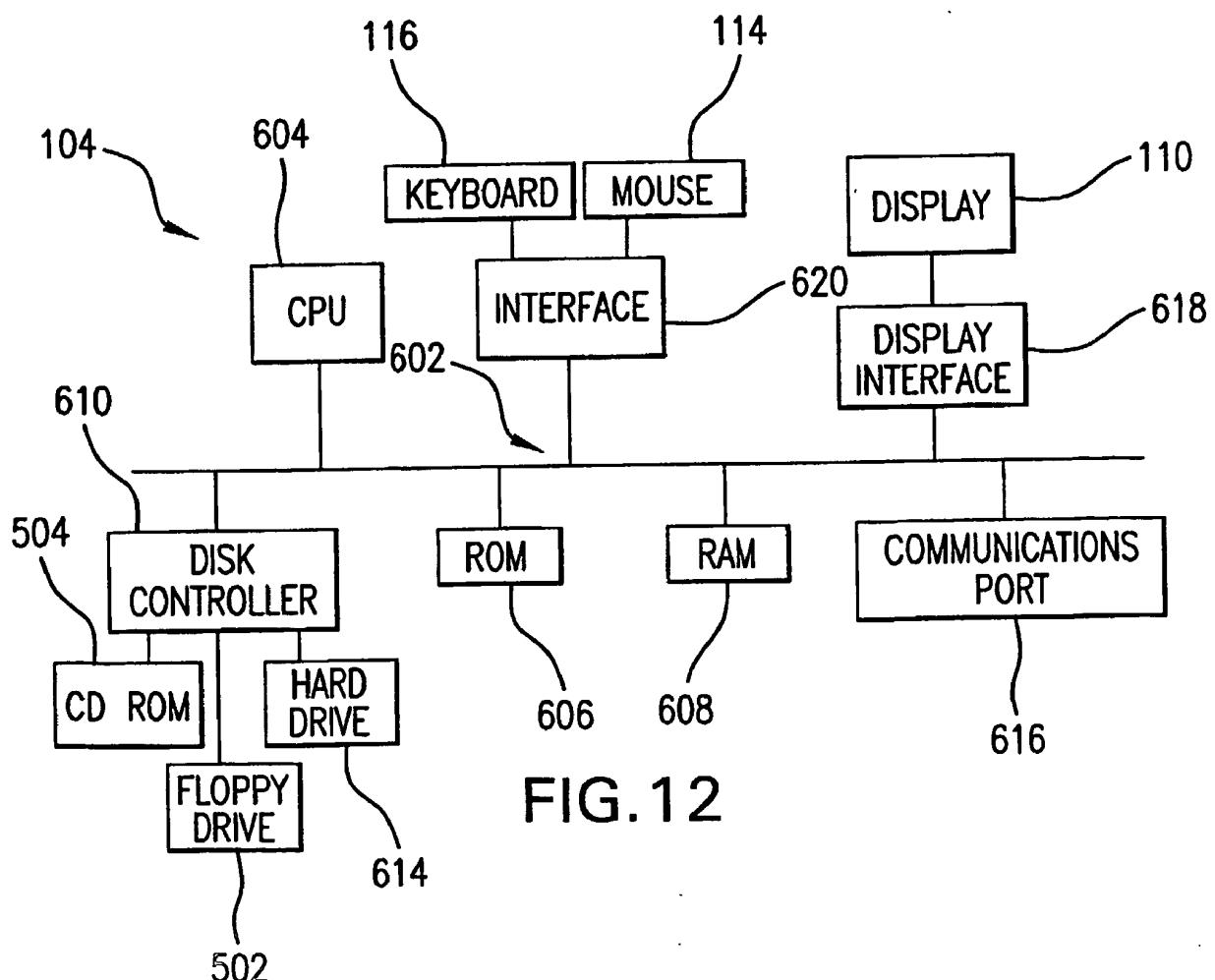


FIG. 12

专利名称(译)	肺癌生物标志物		
公开(公告)号	<a href="#">EP1735620A4</a>	公开(公告)日	2008-04-09
申请号	EP2005731474	申请日	2005-03-30
[标]申请(专利权)人(译)	东弗吉尼亚医学院		
申请(专利权)人(译)	东弗吉尼亚医学院		
当前申请(专利权)人(译)	东弗吉尼亚医学院		
[标]发明人	SEMMES OLIVER JOHN CAZARES LISA H		
发明人	SEMMES, OLIVER, JOHN CAZARES, LISA, H.		
IPC分类号	G01N33/544 G01N33/53 G01N33/574 G01N33/68		
CPC分类号	G01N33/6851 G01N33/57423 G01N2800/52		
代理机构(译)	LEE , NICHOLAS JOHN		
优先权	60/557404 2004-03-30 US		
其他公开文献	<a href="#">EP1735620A2</a>		
外部链接	<a href="#">Espacenet</a>		

## 摘要(译)

公开了蛋白质生物标志物及其在诊断肺癌或在患者中进行阴性诊断中的用途。还公开了用于诊断肺癌的试剂盒，其检测本发明的蛋白质生物标志物，以及使用多种分类器进行肺癌的可能诊断的方法。在本发明的某些方面，该方法包括使用决策树分析。还公开了根据本发明的各种计算机可读介质及其用途。

Class	Total Cases	Percent Correct	Normal	Cancer
Normal	61	96.72	59	2
Cancer	13	100	0	13