(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 October 2013 (17.10.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/154998 A1

(51) International Patent Classification:
G01N 33/53 (2006.01) *C12Q 1/00* (2006.01)

(21) International Application Number:
PCT/US2013/035632

(22) International Filing Date:
8 April 2013 (08.04.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/621,854 9 April 2012 (09.04.2012) US

(71) Applicant: **DUKE UNIVERSITY**; 2812 Erwin Road, Suite 306, Box 90083, Durham, NC 27705 (US).

(72) Inventors: **PATZ, Edward, F.**; 6721 Creek Wood Drive, Chapel Hill, NC 27514 (US). **CAMPA, Michael, Joseph**; 105 East Markam Avenue, Durham, NC 27701 (US).

(74) Agent: **TAYLOR, Arles, A., Jr.**; Jenkins, Wilson, Taylor & Hunt, P.A., Suite 1200, University Tower, 3100 Tower Boulevard, Durham, NC 27707 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: SERUM BIOMARKERS AND PULMONARY NODULE SIZE FOR THE EARLY DETECTION OF LUNG CANCER

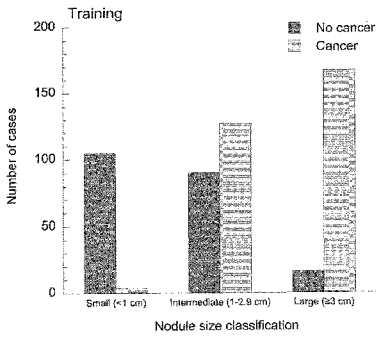


FIG. 1A

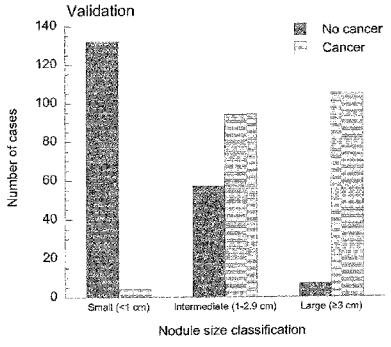


FIG. 1B

(57) Abstract: Methods are provided for determining the probability of lung cancer in patients with indeterminate pulmonary nodules. Also provided are methods for managing treatment of subjects with potential lung cancer. The methods can include determining an amount of each member of a panel of biomarkers in a sample from the subject, wherein, for example, the panel of biomarkers comprises at least two of the proteins alpha-1-antitrypsin, carcinoembryonic antigen, squamous cell carcinoma antigen, retinol binding protein, transferrin, and haptoglobin; and assigning the subject to a group having a higher or lower probability of lung cancer based on the determined amount of each biomarker the panel and based on the size of indeterminate pulmonary nodule.

WO 2013/154998 A1

DESCRIPTION

SERUM BIOMARKERS AND PULMONARY NODULE SIZE FOR THE EARLY DETECTION OF LUNG CANCER

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims benefit of U.S. Provisional Application Serial No. 61/621,854, titled "Serum Biomarkers and Pulmonary Nodule Size for the Early Detection of Lung Cancer," filed April 9, 2012, which is herein incorporated by reference in its entirety.

TECHNICAL FIELD

The presently disclosed subject matter pertains to the use of biomarkers in the detection of lung cancer in subjects with indeterminate pulmonary nodules and in the management of treatment in subjects with potential lung cancer.

BACKGROUND

Lung cancer accounts for more cancer deaths than any other malignancy. Despite advances in diagnostic capabilities and treatment, lung cancer mortality has not significantly changed over the past several decades. Most patients present with inoperable disease when therapeutic options including chemotherapy and radiotherapy are rarely curative. Screening studies for lung cancer with chest radiographs and sputum cytology have failed to show that this approach will decrease the number of patients that die from the disease. See Marcus et al. (2000).

More recent trials with low-dose spiral computed tomography (CT) have found it can detect small tumors, but no study has shown that this will reduce lung cancer mortality, the ultimate goal of a screening program. See Bach et al. (2007). These trials also illustrate a number of challenging issues with regard to imaging as a solitary early detection strategy for lung cancer.

In particular, indeterminate pulmonary nodules or focal pulmonary abnormalities are common radiographic findings and present a diagnostic

dilemma. See Brandman et al. (2011) and Weir et al. (2011). With improvements in spatial resolution on CT, and increased numbers of thoracic CT studies performed for a wide variety of indications, the number of patients with indeterminate pulmonary nodules continues to rise. In the recent National Lung Cancer Screening Trial (NLST) report, over 24% of CT-screened participants had a pulmonary abnormality necessitating further evaluation because of the concern for lung cancer. See Aberle et al. (2011). All of these indeterminate abnormalities create an undesirable burden on the health care system and the patient, since each lesion must be followed up, with the understanding that the vast majority of them are benign and the prevalence of lung cancer is low. While several strategies to evaluate indeterminate lesions have been suggested, to date there is no evidence-based consensus as to the appropriate manner for management of these patients.

Current guidelines recommend multiple sequential imaging studies to assess for growth, or an invasive procedure, depending on clinical risk for disease, clinical presentation, and the size and morphology of the lesion. From an imaging perspective, fluorodeoxyglucose(^{18}F)-positron emission tomography (FDG-PET) has been introduced as a metabolic marker of disease for the purpose of distinguishing a benign from malignant focal pulmonary abnormality. See Erasmus et al. (1999) and Patz et al. (1993). However, due to the predominately small size of nodules and the sheer number of lesions that need to be evaluated, PET can be cost prohibitive and impractical. Further, recent data examining ground glass nodules suggest that these lesions typically have a more indolent behavior, and the development of additional guidelines for these abnormalities is ongoing. See Goo et al. (2011); Kim et al. (2010), Kim et al. (2007); and Lee et al. (2011).

Accordingly, there is a need for additional methods of assessing the probability of lung cancer in subjects presenting indeterminate pulmonary nodules or other abnormalities. There is also a need for additional methods of managing treatment of subjects with potential lung cancer. In particular, there is an ongoing need for risk assessment and management strategies for

subjects with indeterminate pulmonary nodules and/or potential lung cancer. Such methods should avoid delays in diagnosis, decrease radiation exposure due to multiple imaging studies, and/or reduce the need for invasive procedures.

SUMMARY

In some embodiments, the presently disclosed subject matter provides a method for assigning a subject having an indeterminate pulmonary nodule or abnormality to a group having higher or lower probability of lung cancer, the method comprising: a) providing a subject having an indeterminate pulmonary nodule or abnormality, the indeterminate pulmonary nodule or abnormality having a size; b) determining an amount of each member of a panel of biomarkers in a sample from the subject, wherein the panel of biomarkers comprises at least two of the proteins alpha-1-antitrypsin (AAT), carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCC), retinol binding protein (RBP), transferrin, and haptoglobin; and c) assigning the subject to a group having a higher or lower probability of lung cancer based on the determined amount of each biomarker member in the panel and based on the size of the indeterminate pulmonary nodule or abnormality.

In some embodiments, the panel of biomarkers comprises at least AAT, CEA, and SCC. In some embodiments, the panel of biomarkers is AAT, CEA, and SCC.

In some embodiments, the method comprises assigning the nodule or abnormality to a size category selected from the group comprising a small size category, an intermediate size category, and a large size category. In some embodiments, the nodule or abnormality is assigned to the small size category when the largest of the length, width and height of the nodule or abnormality is less than one centimeter; assigned to the intermediate size category when the largest of the length, width and height of the nodule or abnormality is at least one centimeter and less than three centimeters; or assigned to the large size category when the largest of the length, width and height of the nodule or abnormality is three centimeters or more.

In some embodiments, the method comprises applying a predetermined algorithm to the determined amount of each biomarker in the panel and the size of the indeterminate pulmonary nodule or abnormality; and using the results of the algorithm to assign the subject to a group having a higher or lower probability of lung cancer. In some embodiments, the algorithm is a logistic regression analysis or a classification and regression tree (CART) analysis. In some embodiments, the predetermined algorithm is selected based on the size of the indeterminate pulmonary nodule or abnormality.

In some embodiments, a CART analysis algorithm is selected when the indeterminate pulmonary nodule or abnormality has a largest dimension that is 3 centimeters or more. In some embodiments, a logistic regression analysis algorithm is selected when the indeterminate pulmonary nodule or abnormality has a largest dimension that is less than 3 centimeters. In some embodiments, the logistic regression analysis algorithm is a weighted logistic regression analysis algorithm when the indeterminate pulmonary nodule or abnormality has a largest dimension that is at least one centimeter and less than 3 centimeters.

In some embodiments, the sample is a serum sample. In some embodiments, the subject is a human subject.

In some embodiments, the presently disclosed subject matter provides a method for managing treatment of a subject with potential lung cancer, the method comprising: a) providing a subject having an indeterminate pulmonary nodule or abnormality, the indeterminate pulmonary nodule or abnormality having a size; b) determining an amount of each member of a panel of biomarkers in a sample from a subject having an indeterminate pulmonary nodule or abnormality, wherein the panel of biomarkers comprises at least two of the proteins AAT, CEA, SCC, RBP, transferrin and haptoglobin; c) assigning the subject to a group having a higher or lower probability of lung cancer based on the determined amount of each biomarker member in the panel and based on the size of the indeterminate pulmonary nodule or abnormality; and d) managing the treatment of the subject with potential lung cancer based on the group to

which the subject is assigned and based on the size of the indeterminate pulmonary nodule or abnormality. In some embodiments, the managing comprises one or more of scheduling a CT scan, scheduling a biopsy, and scheduling a PET scan.

In some embodiments, the panel of biomarkers comprises at least AAT, CEA, and SCC. In some embodiments, the panel of biomarkers is AAT, CEA, and SCC.

Accordingly, it is an object of the presently disclosed subject matter to provide a novel method for detecting lung cancer. This and other objects are achieved in whole or in part by the presently disclosed subject matter.

An object of the presently disclosed subject matter having been stated above, other objects and advantages will become apparent upon a review of the following descriptions, figures and examples.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a bar graph showing the number of malignant ("Cancer", light grey bars) and benign ("No cancer", dark grey bars) pulmonary nodules in the subjects of a training set based on nodule size, i.e., small, less than 1 centimeters (cm); intermediate, 1-2.9 cm; or large, ≥ 3 cm.

Figure 1B is a bar graph showing the number of malignant ("Cancer", light grey bars) and benign ("No cancer", dark grey bars) pulmonary nodules in the subjects of a validation set based on nodule size, i.e., small, less than 1 centimeters (cm); intermediate, 1-2.9 cm; or large, ≥ 3 cm.

Figure 2 is a graphical depiction of the results of Classification and Regression Tree (CART) analysis of a training set of 509 serum specimens from individuals with (Cancer) or without (No Cancer) lung cancer. The analysis had 9 terminal nodes and was based on pulmonary nodule size (SIZE) and serum concentration of two protein biomarkers, carcinoembryonic antigen (CEA) and alpha-1 antitrypsin (AAT). The bar at the bottom of each node box indicates the percentage of specimens from individuals without cancer (grey shading) and with cancer (black shading) in the node.

DETAILED DESCRIPTION

In accordance with the presently disclosed subject matter, methods are provided for the detection of lung cancer and for managing the treatment of subjects with potential lung cancer. Lung cancer can be detected based on analyzing, in combination, (i) the levels of biomarkers in a panel of biomarkers and (ii) the size of an indeterminate pulmonary nodule or abnormality present in the subject being assessed. The presently provided methods can suggest which patients with indeterminate pulmonary nodules or abnormalities have lung cancer (for example, can suggest which patients have a high risk of having lung cancer).

The presently disclosed subject matter will now be described more fully hereinbelow. The presently disclosed subject matter can, however, be embodied in different forms and should not be construed as limited to the embodiments set forth hereinbelow and in the accompanying Examples. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the embodiments to those skilled in the art.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this presently described subject matter belongs. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.

I. DEFINITIONS

While the following terms are believed to be well understood by one of ordinary skill in the art, the following definitions are set forth to facilitate explanation of the presently disclosed subject matter.

Following long-standing patent law convention, the terms “a” and “an” mean “one or more” when used in this application, including the claims.

Unless otherwise indicated, all numbers expressing quantities of size, biomarker concentration, probability, percentage and so forth used in the specification and claims are to be understood as being modified in all instances by the term “about”. For example, the amounts can vary by about

10%, 5%, 1%, or 0.5%. Accordingly, unless indicated to the contrary, the numerical parameters set forth in this specification and attached claims are approximations that can vary depending upon the desired properties sought to be obtained by the presently disclosed subject matter.

The term "and/or" when used in describing two or more items or conditions refers to situations where all named items or conditions are present or applicable, or to situations wherein only one (or less than all) of the items or conditions is present or applicable.

The term "comprising", which is synonymous with "including," "containing," or "characterized by" is inclusive or open-ended and does not exclude additional, unrecited elements or method steps. "Comprising" is a term of art used in claim language which means that the named elements are essential, but other elements can be added and still form a construct within the scope of the claim.

As used herein, the phrase "consisting of" excludes any element, step, or ingredient not specified in the claim. When the phrase "consists of" appears in a clause of the body of a claim, rather than immediately following the preamble, it limits only the element set forth in that clause; other elements are not excluded from the claim as a whole.

As used herein, the phrase "consisting essentially of" limits the scope of a claim to the specified materials or steps, plus those that do not materially affect the basic and novel characteristic(s) of the claimed subject matter.

With respect to the terms "comprising," "consisting essentially of," and "consisting of", where one of these three terms is used herein, the presently disclosed and claimed subject matter can include the use of the other two terms.

"Amino acid sequence" and terms such as "peptide", "polypeptide" and "protein" are used interchangeably herein, and are not meant to limit the amino acid sequence to the complete, native amino acid sequence (i.e. a sequence containing only those amino acids found in the protein as it occurs in nature) associated with the recited protein molecule. The proteins and protein fragments of the presently disclosed subject matter can be produced

by recombinant approaches or can be isolated from a naturally occurring source. The protein fragments can be any size, and for example can range in size from four amino acid residues to the entire amino acid sequence minus one amino acid.

The term "antibody" includes any antibody fragments that bind with sufficient specificity to a protein of interest.

The phrase "detection molecule" is used herein in its broadest sense to include any molecule that can bind with sufficient specificity to one of the members of the biomarker panel to allow for detection of the particular biomarker member in the presence or absence of the other members of the panel. To allow for detection can mean to determine the presence or absence of the particular biomarker member and, in some embodiments, can mean to determine the amount of the particular biomarker. Detection molecules can include antibodies and antibody fragments.

As used herein, the term "sample" is used in its broadest sense. In one sense, it is meant to include a specimen from a biological source. Biological samples can be obtained from animals (including humans) and encompass fluids (e.g., blood, mucus, urine, saliva), solids, tissues, cells, and gases. Biological samples include blood products, such as plasma, serum and the like.

The term "size" when used herein with regard to an indeterminate pulmonary nodule or abnormality can refer to the measure of the largest dimension (i.e., length, width, height) of the nodule or abnormality. When the nodule or abnormality is approximately spherical, the largest dimension can also be the diameter. Thus, "size", "diameter," and "largest dimension" can be used interchangeably herein. The size can be determined from an imaging scan (e.g., a CT scan) or by physical examination if the nodule is removed or exposed during surgery. In some embodiments, "size" can refer to a size category encompassing a range of particular measured sizes.

The phrase "a specific binding partner for each of the detection molecules" is used herein to include any molecule that binds with sufficient specificity to one of the detection molecules to allow for detection of the particular detection molecule in the presence or absence of the detection

molecules for the other members of the biomarker panel. For example, in some embodiments the specific binding partner can be a secondary antibody that recognizes the detection molecule that is a primary antibody. In some embodiments the specific binding partner can be a molecule that specifically binds to a group on the detection molecule such as, for example, a biotin group on the detection molecule.

As used herein, the term "subject" refers to any animal (e.g., a mammal), including, but not limited to, humans, non-human primates, rodents, and the like, which is to be the recipient of a particular treatment. The terms "subject" and "patient" are used interchangeably herein, such as but not limited to in reference to a human subject.

As used herein, the terms "subject suspected of having lung cancer" and "subject with potential lung cancer" include a subject that presents one or more symptoms indicative of lung cancer or is being screened for lung cancer (e.g., during a routine physical). A subject suspected of having lung cancer can also have one or more risk factors for developing lung cancer (e.g., a history of smoking, a history of exposure to carcinogens (including second hand smoke), a family history of cancer, genetic markers for lung cancer, a prior cancer, etc.). A subject suspected of having cancer has generally not been tested for cancer. In some embodiments, a "subject suspected of having lung cancer" is an individual who has had a CT scan or other imaging study showing an indeterminate pulmonary nodule or other abnormality. The term further includes people who once had lung cancer (e.g., an individual in remission).

II. REPRESENTATIVE EMBODIMENTS

In some embodiments, the presently disclosed subject matter provides methods related in part to a panel of serum biomarkers useful for the detection, desirably early detection, of lung cancer and the management of patients with potential lung cancer. In some aspects, the panel of serum biomarkers provided herein, in combination with indeterminate pulmonary nodule size, addresses certain limitations of early detection of tumors by CT screening alone or by biomarker screening alone.

Indeterminate pulmonary nodules are a common radiographic finding and present a diagnostic dilemma. See Brandman et al. (2011) and Weir et al. (2011). These anatomic abnormalities typically require further evaluation because of the concern for lung cancer. Patient work-up can depend on the clinical scenario and the size and morphology of the abnormality. The majority of lesions are small (< 1 cm), and most patients are followed with sequential CT studies, as previously recommended for the management of small pulmonary nodules. See MacMahon et al. (2005). Other individuals are subjected to additional imaging with PET or undergo a biopsy in order to establish a diagnosis. A more efficient and cost-effective risk-assessment strategy that avoids delays in diagnosis, decreases radiation exposure due to multiple imaging studies, and reduces the need for invasive procedures could have a significant impact on patient management. The integration of biomarkers into the diagnostic strategy could meet these challenges. However, the discovery and validation of optimal tumor specific markers has been more difficult than initially envisioned.

The concept of biomarkers is founded on the biological properties of cancer as a systemic disease. As a malignancy develops, it secretes proteins required for growth and metastasis and sheds cells into the circulation. The host responds by inducing changes in tissue architecture and vasculature in the microenvironment of the incipient tumor, as well as systemically mounting an immunological defense. This can include innate and adaptive responses with migration of inflammatory cells including macrophages, histiocytes and lymphocytes into the tumor, and the production of autoantibodies. See Petersen et al. (2006); Zhong et al. (2006); Welsh et al. (2005); and Condeelis et al. (2006). Thus, a combination of tumor-expressed and host response proteins, if identified, can be useful to develop a profile of cancer for clinical screening.

Alpha-1-antitrypsin (AAT) has been associated with both the detection and etiology of lung cancer. See Ljujic et al. (2006) and Zelvyte et al. (2004). Transferrin, retinol binding protein (RBP), haptoglobin and AAT were previously identified as being differentially expressed serum proteins in patients with and without lung cancer. See PCT International Publication

No. WO 2008/144034. Carcinoembryonic antigen (CEA) and squamous cell carcinoma antigen (SCC) are two additional proteins whose serum levels are known to vary depending on lung cancer status. While each of these markers has been associated with cancer, they are independently insufficient to direct patient care. Since lung cancer is a heterogeneous disease, a panel of markers that covers the broad clinical phenotype appears to be needed. Classification and Regression Tree (CART) analysis has indicated that some panels of biomarkers can correctly classify lung cancer sera and non-cancer control sera with 77.8% sensitivity and 75.4% specificity. See Patz et al. (2007) and PCT International Publication No. WO 2008/144034.

According to the presently disclosed subject matter, it has been found that combining nodule size data with serum biomarker data can be useful in increasing the overall sensitivity and specificity of methods for assessing lung cancer probability in patients presenting indeterminate pulmonary nodules. Thus, in some embodiments, in the methods of the presently disclosed subject matter, data of serum biomarker levels and nodule size are assessed in combination using derived algorithms to determine the risk for lung cancer across a spectrum of nodule sizes. Stratification of patients into high and low-risk groups by use of these algorithms can guide patient treatment. More particularly, in some embodiments, a Lung Cancer Biomarker Assay (LCBA) can be used to measure the serum levels of a panel of biomarkers (e.g., CEA, AAT, and SCC). LCBA data and nodule size data is statistically analyzed (e.g., by CART analysis or logistic regression analysis) and the results used to assign patients to a high or low-risk group. The risk designation can more efficiently guide further evaluation of the patients, rather than treating all patients with indeterminate pulmonary nodules or all patients with similar size pulmonary nodules with the same management plan. In some embodiments, serum CEA levels and nodule size are assessed in combination to determine the risk for lung cancer.

Any reduction in the number of imaging examinations or invasive procedures needed to establish or exclude the diagnosis of lung cancer by

using a less expensive, non-invasive test is desirable, if it can be done without compromising patient outcomes. According to the presently disclosed methods, 97% of all lung cancer patients in a population can be placed in a high-risk group, while 95% of all patients without cancer can be placed in a low-risk group.

Accordingly, in some embodiments methods are provided for assigning a subject with an indeterminate pulmonary nodule or abnormality to a group having a higher or lower probability of lung cancer. In some embodiments, the methods comprise determining the level of CEA in a serum sample from the subject and assigning the subject to a group having a higher or lower probability of lung cancer based on the determined amount of CEA and based on the size of the indeterminate pulmonary nodule or abnormality.

In some embodiments, the methods comprise: determining the level of each of a panel of biomarkers in a serum sample from the subject with an indeterminate pulmonary nodule or abnormality, wherein the panel of biomarkers comprises at least two proteins that are differentially expressed in lung cancer patients; and assigning the subject to the group having a higher or lower probability of lung cancer based on the determined amount of each biomarker in the panel and based on the size of the indeterminate pulmonary nodule or abnormality. In some embodiments, the panel of biomarkers includes at least two of the proteins AAT, CEA, and SCC. In some embodiments, the panel of biomarkers includes CEA and AAT. In some embodiments, the panel of biomarkers includes at least the three proteins AAT, CEA, and SCC. In some embodiments, the panel of biomarkers is AAT, CEA and SCC. In some embodiments, biomarkers other than or in addition to AAT, CEA, and SCC (e.g., including, but not limited to, RBP, transferrin, and haptoglobin) can be included in the biomarker panel.

Thus, the presently disclosed subject matter is not limited to the panels of biomarkers described above. Any marker that correlates with lung cancer or the progression of lung cancer can be included in the biomarker panel provided herein, and is within the scope of the presently disclosed subject matter. Any suitable method can be utilized to identify additional

lung cancer biomarkers suitable for use in the presently disclosed methods, including but not limited to, the methods described in PCT International Publication No. 2008/144034. For example, biomarkers that are known or identified as being up or down-regulated in lung cancer using methods known to those of ordinary skill in the art can be employed. Additional biomarkers can include one or more of polypeptides, small molecule metabolites, lipids, and nucleotide sequences. Markers for inclusion on a panel can be selected by screening for their predictive value using any suitable method, including but not limited to, those methods described in PCT International Publication No. 2008/144034, which is incorporated herein by reference in its entirety.

The level of each of the presently disclosed panel of biomarkers can be determined in a variety of animal tissues. In some embodiments, the biomarkers can be detected in animal tissue or bodily fluids. In some embodiments, the biomarkers can be detected in bodily fluids including plasma, serum, whole blood, mucus, and/or urine. In a particular embodiment, the biomarkers can be detected in serum.

Any suitable method can be employed for determining the level of each of the panel of biomarkers, as would be apparent to one skilled in the art upon a review of the present disclosure. For example, methods for detecting biomarkers can include, but are not limited to, gas chromatography (GC), liquid chromatography/mass spectroscopy (LC-MS), gas chromatography/mass spectroscopy (GC-MS), nuclear magnetic resonance (NMR), magnetic resonance imaging (MRI), Fourier Transform InfraRed (FT-IR), and inductively coupled plasma mass spectrometry (ICP-MS). It is further understood that mass spectrometry techniques include, but are not limited to, the use of magnetic-sector and double focusing instruments, transmission quadrapole instruments, quadrupole ion-trap instruments, time-of-flight instruments (TOF), Fourier transform ion cyclotron resonance instruments (FT-MS), and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS).

In some embodiments, protein biomarkers can be detected using technologies well known to those of skill in the art such as gel

electrophoresis, immunohistochemistry, and antibody binding. Methods for generating antibodies to a polypeptide of interest are well known to those of ordinary skill in the art. An antibody against a protein biomarker of the presently disclosed subject matter can be any monoclonal or polyclonal antibody, so long as it suitably recognizes the protein biomarker. In some embodiments, antibodies are produced using the protein biomarker as the immunogen according to any conventional antibody or antiserum preparation process. The presently disclosed subject matter provides for the use of both monoclonal and polyclonal antibodies. In addition, a protein used herein as the immunogen is not limited to any particular type of immunogen. For example, fragments of the protein biomarkers of the presently disclosed subject matter can be used as immunogens. The fragments can be obtained by any method including, but not limited to, expressing a fragment of the gene encoding the protein, enzymatic processing of the protein, chemical synthesis, and the like.

The antibodies of the presently disclosed subject matter can be useful for detecting the protein biomarkers. For example, antibody binding is detected by techniques known in the art (e.g., radioimmunoassay, ELISA (enzyme-linked immunosorbant assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, *in situ* immunoassays (e.g., using colloidal gold, enzyme or radioisotope labels, for example), Western blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. One example of an immunoassay is described in U.S. Pat. Nos. 5,599,677 and 5,672,480, each of which is herein incorporated by reference. Upon review of the present disclosure, those skilled in the art will be familiar with numerous specific immunoassay formats and variations thereof that can be useful for carrying out the methods of the presently disclosed subject matter.

The subject of, or provided in, the presently disclosed methods is typically a subject that has been determined to have an indeterminate

pulmonary nodule or abnormality based upon an imaging study (e.g., a prior imaging study, such as a prior CT study). The size of the nodule or abnormality can be determined, or can have been determined, for example, from a prior imaging study, in accordance with any routinely available technique. The imaging study can have been performed for any reason. For example, the subject can be a subject considered at risk for developing lung cancer based upon smoking history or other risk factors and who has undergone a screening CT study that found an indeterminate pulmonary nodule. However, the imaging can have been performed for reasons unrelated to lung cancer screening (e.g., pulmonary embolism, pneumonia, interstitial lung disease, chest pain, etc.). In some embodiments, the subject is a human subject with an indeterminate pulmonary nodule or abnormality.

In some embodiments, the method comprises assigning the nodule or abnormality to a size category (e.g., small, intermediate, or large) based upon the size (e.g., the diameter) of the nodule or abnormality. The diameter can be the largest of the length, height, or width of the nodule or abnormality. The diameter can be determined based upon data extracted from the CT image used to originally determine the presence of the nodule or abnormality. In some embodiments, nodules or abnormalities that have a diameter that is less than one centimeter are assigned to a small size category. In some embodiments, nodules or abnormalities that have a diameter that is at least one centimeter, but less than three centimeters, are assigned to an intermediate size category. In some embodiments, nodules or abnormalities that have a diameter that is three centimeters or more are assigned to a large size category.

In some embodiments, the presently disclosed methods can comprise statistically analyzing the amounts of each biomarker and the nodule or abnormality size. In some embodiments, the statistical analysis can comprise applying a predetermined algorithm to the determined amount of each biomarker in the panel and the size of the indeterminate pulmonary nodule or abnormality. The results of the algorithm can be employed to assign a subject to a group having a higher or lower probability of lung cancer.

A variety of algorithms can be employed in the presently disclosed methods. The algorithms employed are not limited to those described in the Examples herein, but rather include algorithms as would be apparent to those of ordinary skill in the art upon a review of the instant disclosure. In some embodiments, the algorithm employed can be a decision tree analysis. In some embodiments, the algorithm employed can be a Classification and Regression Tree analysis (CART). In some embodiments, the algorithm employed can be a logistic regression analysis or a weighted logistic regression analysis.

The nodule or abnormality size or size category can be used to select the particular statistical treatment performed on data from a particular subject. For example, in some embodiments, a CART analysis algorithm is selected when the indeterminate nodule or abnormality has a diameter of three centimeters of more or is assigned to the large size category. In some embodiments, a logistic regression analysis algorithm is selected when the nodule or abnormality has a diameter than is less than three centimeters or has been assigned to the intermediate or small size category. In some embodiments, a weighted logistic regression analysis algorithm is selected when the nodule or abnormality has been assigned to the intermediate size category or has a diameter that is at least one centimeter but less than three centimeters.

In some embodiments, the subject is assigned to a high risk group for lung cancer. Subjects assigned to the high risk group are subjects who have greater than 60, 65, 70, 75, 80, 85, 90, or 95% chance of having lung cancer (e.g., of receiving a diagnosis of lung cancer in the future, based on, for example, histopathological confirmation). In some embodiments, the subject is assigned to a low-risk group for lung cancer. Subjects assigned to the low risk group for lung cancer are those subjects who are believed to be unlikely (e.g., have a less than 40%, 35, 30, 25, 20, 15, 10, or less than 5% chance) of having cancer.

In some embodiments, a method is provided for managing treatment of a subject with potential lung cancer. In some embodiments, the method comprises: determining the level of CEA in a serum sample from the

subject; assigning the subject to a group having a higher or lower probability of lung cancer based on the determined amount of CEA and based on the size of the indeterminate pulmonary nodule or abnormality; and managing the treatment of the subject with potential lung cancer based on the group to which the subject is assigned and based on the size of the indeterminate pulmonary nodule or abnormality.

In some embodiments, the method can comprise: determining the level of each of a panel of biomarkers in a serum sample from a subject with an indeterminate pulmonary nodule or abnormality; assigning the subject to a group having a higher or lower probability of lung cancer based on the determined amount of each biomarker in the panel and based on the size of the indeterminate pulmonary nodule or abnormality; and managing the treatment of the subject with potential lung cancer based on the group to which the subject is assigned and based on the size of the indeterminate pulmonary nodule or abnormality. In some embodiments, the panel of biomarkers comprises at least two of the proteins AAT, CEA, and SCC. In some embodiments, the panel of biomarkers includes CEA and AAT. In some embodiments, the panel comprises at least AAT, CEA and SCC. In some embodiments, the panel is AAT, CEA, and SCC. In some embodiments, biomarkers other than or in addition to AAT, CEA, and SCC (e.g., including, but not limited to, RBP, transferrin, and haptoglobin) can be included in the biomarker panel. In some embodiments, the panel comprises at least two proteins that are differentially expressed in lung cancer patients.

Managing treatment can comprise selecting appropriate time frames in which to schedule additional scans or studies, such as, but not limited to CT scans, biopsies, surgery, and PET scans, for the subject. Managing treatment can further comprise selecting an appropriate time frame in which to schedule a repeat assessment of biomarker levels or the assessment of additional biomarkers. For example, for subjects assigned to a low risk category, a future scan can be scheduled several months (e.g., 6, 7, 8, 9, 10, 11, 12, 14, 16, 18 months) or one or more years in the future. Subjects assigned to a high risk category can be scheduled for an additional imaging

scan or a more invasive screening procedure (e.g., a biopsy) within a shorter time frame (3 months, 2 months, 1 months, 4, 3, or 2 weeks or less), particularly if they have a large size nodule or abnormality. Assignment to a high risk category in conjugation with a large nodule or abnormality size can indicate a benefit to scheduling an immediate imaging scan or other procedure.

In some embodiments, a subject with an intermediate size nodule or abnormality assigned to a low risk category can be scheduled for a follow-up CT scan in 3 to 6 months. In some embodiments, a subject with an intermediate size nodule or abnormality assigned to either a high or low risk category can be scheduled for a follow-up CT scan in 3 months.

As is apparent from the foregoing embodiments, the presently disclosed methods are useful for screening patients with indeterminate pulmonary nodules or abnormalities for lung cancer, e.g., for the early detection of lung cancer, and for managing the treatment of patients with indeterminate pulmonary nodules or abnormalities. For example, in some embodiments, the combination of the panel of biomarkers and nodule or abnormality size can be useful for screening patients prior to additional imaging or other known methods for detecting lung tumors, to define patients at high risk or higher risk for lung cancer. In this manner, only those patients with both a high risk clinical profile and a test result according to the presently disclosed methods indicating a high or higher probability of lung cancer can be sent on to have a CT (e.g., a further CT) or other scan performed. Patients whose test results suggest a low probability of cancer can be reevaluated using the panel of serum biomarkers during a routine follow-up.

In some embodiments of the presently disclosed subject matter, a kit is provided for measuring an amount of each member of a panel of biomarkers in a sample of the subject. The kit can comprise: i) detection molecules specific for each of the biomarkers in the panel (e.g., at least two of CEA, SCC, and AAT or at least any two proteins that are differentially expressed in lung cancer patients), and ii) directions for measuring the amount of each member of the panel of biomarkers. The kit can also include

directions for using the determined biomarker levels in combination with nodule or abnormality size to determine the probability of lung cancer. The phrase "detection molecule" is used herein in its broadest sense to include any molecule that can bind with sufficient specificity to one of the members of the biomarker panel to allow for detection of the particular biomarker member in the presence or absence of the other members of the panel. To allow for detection can mean to determine the presence or absence of the particular biomarker member and, in some embodiments, can mean to determine the amount of the particular biomarker. Detection molecules can include antibodies and antibody fragments. In some embodiments, the detection molecules comprise a conjugated detectable group. In some embodiments, the detection molecules comprise antibodies specific for each of the protein biomarkers in the panel.

Approaches for producing a detectable signal include the use of radioactive labels (e.g., ^{35}S, ^{125}I, ^{131}I), fluorescent labels, enzyme labels (e.g., horseradish peroxidase, alkaline phosphatase), fluorescent labels (e.g., fluorescein) and so forth, in accordance with known techniques, as will be apparent to one skilled in the art upon review of the present disclosure. Many methods are known in the art for detecting binding in an immunoassay and are within the scope of the presently disclosed subject matter.

In some embodiments, direct detection methods are provided, such as, for example, wherein the detection molecule is a primary antibody specific for a member of the biomarker panel and detection is by using a label on the primary antibody. In some embodiments, the detection molecule can be detected using an indirect method such as by detecting binding of a specific binding partner to the detection molecule. The specific binding partner can be any molecule that binds with sufficient specificity to the detection molecule to allow for detection of the particular detection molecule in the presence or absence of the detection molecules for the other members of the biomarker panel. In some embodiments, the detection molecule is a primary antibody and the primary antibody can be detected by detecting binding of a secondary antibody or a reagent or other specific binding partner to the primary antibody. For example, in some embodiments

the specific binding partner can be a secondary antibody that recognizes the detection molecule that is a primary antibody. In some embodiments the specific binding partner can be a molecule that specifically binds to a group on the detection molecule such as, for example, a biotin group on the detection molecule. In some embodiments, the binding partner can be labeled. In some embodiments, the binding partner is a secondary antibody that can be labeled.

For example, in some embodiments, indirect detection methods can involve a detection molecule that is an unlabeled primary antibody and a binding partner that is a labeled secondary antibody. This method can be more sensitive than direct detection methods due to signal amplification through more than one secondary antibody reaction with different antigenic sites on the primary antibody. In some embodiments, the indirect detection method is an immunofluorescence method, wherein the secondary antibody can be labeled with a fluorescent dye such as FITC, rhodamine or Texas red. In some embodiments, the indirect detection method is an immunoenzyme method, wherein the secondary antibody can be labeled with an enzyme such as peroxidase, alkaline phosphatase or glucose oxidase.

In some embodiments, an immunoassay can comprise antibodies specific for each of the members of the panel of protein biomarkers and an approach for producing a detectable signal. In some embodiments, the antibodies can be immobilized on a support (such as a bead, plate or slide) in accordance with known techniques and contacted with a test sample in liquid phase. The support can then be separated from the liquid phase and either the support phase or the liquid phase can be examined for the detectable signal that is related to the presence of the protein biomarker.

The presently disclosed subject matter includes kits for detecting each of the members of the panel of biomarkers. In some embodiments, the kit can comprise detection molecules, such as antibodies, specific for the protein biomarkers in the panel, the reagents necessary for producing a detectable signal as described above and buffers. In some embodiments, the kit can contain all of the components necessary to perform a detection

assay, including all controls, directions for performing assays, and any necessary software for analysis of the data of the presently disclosed methods and presentation of results. Indeed, the presently disclosed methods and/or kits can be employed using a suitably programmed computer, in some aspects.

Detection kits for carrying out the methods of the presently disclosed subject matter can be produced in a number of ways. In some embodiments, the detection kit can comprise a detection molecule that is an antibody or antibody fragment that specifically binds to a protein biomarker in the panel immobilized on a solid support, and a second antibody or antibody fragment specific for the first antibody or antibody fragment conjugated to a detectable group. In some embodiments, the kit can also include ancillary reagents such as buffering agents and protein stabilizing agents, and can include (where necessary) other members of the detectable signal-producing system of which the detectable group is a part (e.g., enzyme substrates); agents for reducing background interference in a test; control reagents; apparatus for conducting a test, and the like, as will be apparent to those skilled in the art upon a review of the instant disclosure.

In some embodiments, the detection kit can comprise antibodies or antibody fragments specific for each of the protein biomarkers in the panel, and a specific binding partner for each of the antibodies that is conjugated to a detectable group. Ancillary agents as described above can likewise be included. The test kit can be packaged in any suitable manner, typically with all groups in a single container along with a sheet or printed instructions for carrying out the test.

In some embodiments, the detection assay for the biomarker panels can be automated. Methods for the automation of immunoassays include those described in U.S. Pat. Nos. 5,885,530, 4,981,785, 6,159,750, and 5,358,691, each of which is herein incorporated by reference. In some embodiments, analysis of the biomarker data in combination with nodule size and presentation of results can also be automated. In this manner, a clinician can access the test results using any suitable approach or device. Thus, in some embodiments, a clinician need not understand the raw data,

as the data can be presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information to optimize care of the subject. The presently disclosed subject matter provides any method, system, and/or apparatus capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information providers, medical personnel, and subjects.

III. EXAMPLES

The presently disclosed subject matter will now be described more fully hereinafter with reference to the accompanying Examples, in which representative embodiments are shown. The presently disclosed subject matter can, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the embodiments to those skilled in the art.

EXAMPLE 1

GENERAL METHODS

A Lung Cancer Biomarker Assay (LCBA) was used to measure the serum levels of carcinoembryonic antigen (CEA), alpha-1-antitrypsin (AAT) and squamous cell carcinoma antigen (SCC). The make-up of the LCBA was as follows: CEA, Roche COBAS® Modular Analytics E170 (Roche Diagnostics, GmbH, Mannheim, Germany); AAT, Roche COBAS INTEGRA® Tina Quant Assay (Roche Diagnostics, GmbH, Mannheim, Germany); and SCC, CanAg® SCC EIA Assay (Fujirebio Diagnostics AB, Göteborg, Sweden). Assays were performed according to manufacturer instructions. All samples were run in duplicate for the training set. For the validation study, the SCC assay was run in duplicate, and the CEA and AAT assays were run on automated analyzers in singlicate.

All patients enrolled in the validation study (or patients who provided serum samples for the training study) had a CT study of the thorax, performed for a variety of indications including the following: possible respiratory neoplasm, pulmonary nodule, interstitial lung disease, pulmonary embolism, pneumonia, coronary artery scoring, or chest pain. CT studies were performed using multiple contiguous 5 mm sequential axial images through the thorax. Typically, the size of the nodule or focal abnormality was extracted from the official documented clinical CT examination report. In the validation study, the tumor size from 18 patients (4%) was determined by a pathology report. All patients with lung cancer had histological confirmation and pathological staging. All patients with a benign abnormality had histological diagnosis, CT resolution or 2 year stability of the lesion, or clinical observation without evidence of lung cancer.

EXAMPLE 2

TRAINING STUDY

Patient Samples: Serum samples were selected from an Institutional Review Board (IRB)-approved repository. The samples were all collected, processed, and stored in a similar fashion. Serum samples were selected from 298 patients with a confirmed diagnosis of lung cancer. In addition, 211 serum samples were selected from age, gender, and smoking history-matched individuals without cancer who met the following eligibility criteria: 1) no prior history of lung cancer; 2) no currently known extra-thoracic malignancy, and 3) a CT of the thorax with a pulmonary nodule. The distribution of benign and malignant nodules reflected the expected proportion of cases. See Figure 1A. Table 1 below shows the patient demographics for the training study.

Table 1. Patient Demographics and Clinical Profiles for the Training Set

Demographic	Cancer (n = 298)	Control (n = 211)
Age (years)	63.8 ± 10.6	62.5 ± 11.9
Range (years)	28-84	23-84
Gender		

Male	108	66
Female	113	69
Stage		
I	122	
II	40	
III	77	
IV	47	
unknown	12	
Histology		
Adenocarcinoma	145	
Adenosquamous cell carcinoma	3	
Bronchioloalveolar carcinoma	2	
Squamous cell carcinoma	74	
Large cell carcinoma	13	
Small cell carcinoma	2	
Non-small cell carcinoma, unspecified	40	
Adeno/BAC	11	
Large-cell/ adenocarcinoma	1	
Large-cell/ neuroendocrine carcinoma	1	
Large-cell/ small-cell carcinoma	1	
Small-cell/ squamous cell carcinoma	1	
Not specified	4	

Analysis of the Training Set: Training set data were analyzed using both classification and regression tree (CART) methodology and logistic regression. The CART model (see Figure 2) was generated using the Gini index that favors even splits with a minimum size of 5 for terminal nodes, and 10 for parent nodes. Ten-fold cross-validation was used in generating

this model. The sensitivity, specificity, and positive and negative predictive values for the training set were 90%, 82%, 88%, and 86%, respectively.

Two separate models were developed using logistic regression analysis, including one that was weighted for sensitivity so that the number of cancers assigned to a low-risk group (i.e., false negative diagnoses) would be minimized. The standard model involved four data elements (nodule diameter, CEA, SCC and AAT results). The logistic regression model, along with odds ratio estimates for 1 unit change in a predictor are shown in Tables 2 and 3.

Table 2. Logistic Regression Model

Parameter	Estimate	SE	P-value
Intercept	-3.7966	0.5572	<0.0001
Nodule Size	0.1058	0.012	<0.0001
AAT	0.00475	0.00315	0.1316
CEA	0.2323	0.0499	<0.0001
SCC	0.4255	0.2348	0.0699

Table 3. Odds Ratio Estimates for Logistic Regression Model

Effect	Point Estimate	Confidence Limits
Nodule Size	1.112	1.086 - 1.138
ATT	1.005	0.999 - 1.011
CEA	1.261	1.144 - 1.391
SCC	1.530	0.966 - 2.425

This algorithm for the standard model uses the standard 0.50 as a decision point. A calculated probability of 0.50 or higher is considered to be cancer and a probability <0.50 is considered to be no cancer. The sensitivity, specificity, and positive and negative predictive values for the standard model training set were 81%, 84%, 88%, and 76%, respectively.

The sensitivity weighted model also involved the same four data elements (nodule diameter, CEA, SCC and AAT results). This algorithm was designed to increase sensitivity and uses a decision point of 0.30. A calculated probability of 0.30 or higher is considered to be cancer and a probability <0.30 is considered to be no cancer. The sensitivity, specificity, and positive and negative predictive values for the weighted model training

set were 94%, 63%, 78%, and 89%, respectively. Table 4 shows the sensitivity, specificity, and positive and negative predictive value for each of the nodule size categories in the training set.

Table 4. Logistic regression (LR) analysis of the training set according to size categories

Nodule Size Range	Cancer (N)	No Cancer (N)	Measure	LR Model with 0.5 Threshold	LR Model with 0.3 Threshold
Large	167	16	Sensitivity	100%	100%
			Specificity	0%	0%
			PPV	91%	91%
			NPV	NE	NE
Intermediate	126	90	Sensitivity	60%	87%
			Specificity	81%	47%
			PPV	82%	70%
			NPV	59%	72%
Small	4	105	Sensitivity	0%	75%
			Specificity	99%	87%
			PPV	0%	17%
			NPV	96%	99%

NE: Not Estimable

EXAMPLE 3

VALIDATION STUDY

Patient Selection: Patients with an indeterminate pulmonary nodule or focal pulmonary opacity detected on thoracic CT (performed for any indication) were enrolled in a validation study. All patients met the same eligibility criteria as for the training set. There were 203 patients with lung cancer and 196 patients without cancer. Patient demographics for the validation study are shown in Table 5. Serum was collected from each subject and stored at -80°C. After enrollment was complete, the samples were thawed for the first time and 500 µl aliquots were placed in de-identified tubes with randomly assigned unique study identification numbers.

Table 5. Patient Demographics and Clinical Profiles for the Validation Set

Demographic	Cancer (n = 203)	Control (n = 196)
Age (years)	68.6 ± 10.1	58.9 ± 12.8
Range (years)	39-100	26-88
Gender		
Male	121	83
Female	82	113
Stage		
I	94	
II	36	
III	42	
IV	27	
unknown	4	
Histology		
Adenocarcinoma	94	
Adenosquamous cell carcinoma	1	
Bronchioloalveolar carcinoma	5	
Squamous cell carcinoma	70	
Large cell carcinoma	6	
Small cell carcinoma	1	
Non-small cell carcinoma, unspecified	20	
Basaloid carcinoma	1	
Large-cell neuroendocrine carcinoma	3	
Pleomorphic carinoma	1	
Squamous cell and adenocarcinoma	1	

Statistical analysis for the independent blinded validation study:

The independent validation study was designed to test the accuracy of the CART and logistic regression models developed from the training set. Each model assigned each subject a "cancer" (high-risk group) or "no cancer" (low-risk group) designation. Following classification of all patients, the

diagnosis was revealed and the sensitivity, specificity, and positive and negative predictive values for each model were determined.

To power the validation study, a statistical model was developed that predicts the presence/absence of cancer among patients with a pulmonary nodule. The goal of this validation study was to assess the classification properties of the model within an independent blinded dataset. More specifically, this validation study determined whether the sensitivity and specificity of the derived classification algorithm is adequate for assessing the presence/absence of cancer within a patient group representative of all patients with a pulmonary nodule or focal pulmonary abnormality.

The study was powered to include at least 384 patients with an indeterminate pulmonary nodule. Power calculations assume the following: 1) approximately 50% of all patients will have cancerous nodules, 2) approximately 33% of nodules will be small (<1 cm), medium (1-<3 cm), or large (≥3 cm), and 3) among those patients with medium nodules, approximately 50% will be cancerous. However, among small and large sized nodules, that percentage will be approximately 5% and 83%, respectively.

The probability that a cancerous nodule is diagnosed as cancerous without the benefit of a pathology assessment is estimated to be 50%. If the serum based classification algorithm increased that probability to 75%, then it would be considered potentially clinically useful in the diagnosis of cancer in patients with a pulmonary nodule. If the classification algorithm increased the probability of detecting a cancerous tumor to 60%, it would not be worthy of clinical use. Therefore, the statistical hypothesis that would be tested for sensitivity is: H_0: $p<0.6$ *vs.* H_1: $p>0.75$ where p is the proportion of patients with tumors known to have cancerous nodules who are classified as having cancer according to the classification algorithm. A similar argument can be made for the determination of benign lesions when a nodule is truly benign. In that case, the statistical hypothesis that would be tested for specificity is: H_0: $q<0.6$ *vs.* H_1: $q>0.75$ where q is the proportion of patients known to have a benign pulmonary nodule who are classified as having benign disease.

Assuming at least 384 patients are included in this validation study, then approximately 192 nodules will be cancerous and 192 nodules will be benign. With these sample sizes, the hypotheses for sensitivity and specificity as described above can be conducted at the 0.1 level of significance with >99% power.

EXAMPLE 4

RESULTS OF THE INDEPENDENT VALIDATION STUDY

Characteristics of study participants in the independent validation study: As described in Example 3, 399 patients fulfilled the eligibility criteria and were enrolled in the independent validation study. The diagnosis of primary lung cancer was established in 203 patients (50.8%) and a benign abnormality in 196 (49.2%). The distribution of benign and malignant nodules according to lesion size (small, < 1 cm; intermediate, 1-<3 cm; and large, ≥ 3 cm) approximately reflected the expected proportion of cases. See Figure 1B.

Results of CART analysis in the validation study: LCBA data and nodule size were entered into the CART algorithm derived from the training set, and a "cancer" (high-risk) or "no cancer" (low-risk) diagnosis was assigned for each patient. The sensitivity, specificity, and positive and negative predictive values for determining which patient had lung cancer in the validation study were 88%, 82%, 84%, and 87%, respectively. Of note, in the large nodule category, all patients with CEA > 3.22 ng/ml proved to have lung cancer (specificity =100%, sensitivity = 68%).

Results of the logistic regression analysis in the validation study: LCBA data and nodule size were used to assign a "cancer" (high-risk) or "no cancer" (low-risk) diagnosis to each patient using the two logistic regression models derived from the training set. The sensitivity, specificity, and positive and negative predictive values for determining which patient had lung cancer in the validation study were 80%, 89%, 89%, and 81%, respectively for the logistic regression standard model (LR-N, 0.5 Threshold), and 92%, 74%, 79%, and 90%, respectively, for the logistic regression weighted model (LR-W, 0.3 Threshold). The sensitivity and

specificity according to small, intermediate, and large nodule size categories for each model is shown in Table 6.

Table 6. Logistic regression analysis with the validation set according to size categories

Nodule Size Range	Cancer (N)	No Cancer (N)	Measure	Logistic with 0.5 Threshold	Logistic with 0.3 Threshold
Large	105	7	Sensitivity	100%	100%
			Specificity	0%	0%
			PPV	94%	94%
			NPV	NE	NE
Intermediate	94	57	Sensitivity	60%	86%
			Specificity	81%	42%
			PPV	84%	71%
			NPV	55%	65%
Small	4	132	Sensitivity	25%	25%
			Specificity	98%	92%
			PPV	25%	8%
			NPV	98%	98%

Of note, there were only a small number of lung cancers in the small size category (n=4). Both logistic regression models only had a sensitivity of 25%, although the specificity for the standard and weighted models was 98% and 92%, respectively. Those patients with lung cancers assigned to the low-risk group proved to have indolent slow growing tumors. None of the cases considered low-risk developed recurrent disease (up to at least 2 years following the diagnosis).

EXAMPLE 5

MANAGEMENT OF PATIENTS WITH INDETERMINATE PULMONARY NODULES

Because the probability of lung cancer typically increases with nodule size, in some embodiments, management of patients can begin with stratification according to nodule size categories (*i.e.*, small, intermediate, and large), and further evaluation can be determined by integrating biomarker results and the assignment to a risk group. Table 7 provides

recommendations for the follow-up and management of patients with indeterminate pulmonary nodules.

Table 7: Recommendations for Follow-up and Management of Nodules by Size and Risk Category

Nodule Size	Low-Risk Patient	High-Risk Patient
Small <1 cm	Follow-up CT in 9-12 months* (< 2% will develop NSCLC)	Follow-up CT in 6-9 months*
Intermediate 1- <3 cm	Follow-up CT in 3 months, PET and/or biopsy*	PET and/or biopsy
Large ≥3 cm	n/a	If CEA <3.22: immediate PET and/or consider biopsy If CEA >3.22: immediate biopsy and/or consider PET

*if no growth on CT, then repeat LCBA to determine current risk category and additional follow-up

n/a, not applicable: patients with nodules ≥ 3 cm are considered high risk

In the small nodule category (<1 cm diameter lesions), the prevalence of lung cancer is low (< 5% of all nodules). In this category, the standard model would place approximately 3% of individuals in a high-risk group, and would estimate that approximately 25% will eventually be diagnosed with lung cancer. A follow-up CT within 6 months is appropriate. The majority of patients in the small nodule size category, however, will be assigned to the low-risk group. The standard model estimates that only 2% of individuals in the low-risk group will ever be diagnosed with lung cancer. These patients do not need as close surveillance as those in the high-risk group, and a follow-up CT in 9-12 months can be adequate. In those few individuals in the low-risk group that eventually demonstrate nodule growth and are diagnosed with lung cancer, the risk of cancer progression to more advanced stage disease during the observation period appears low.

In the intermediate category (1- <3 cm diameter lesions), the use of a weighted logistic regression model appears appropriate, as it minimizes the

number of lung cancer patients assigned to the low-risk category. Eighty-six percent (86%) of all lung cancers in this size category were placed in the high-risk group. 71% of patients in this high risk category were diagnosed with lung cancer. Individuals assigned to this category could have an immediate or close (within 3 months) follow-up FDG-PET or biopsy. Individuals assigned to the low-risk group can have a follow-up in 3-6 months to assess for growth, with the understanding that approximately one third of individuals in this group will eventually be diagnosed with lung cancer. Alternatively, an immediate FDG-PET study could be performed.

In the large nodule size category, the prevalence of lung cancer is high. Some studies estimate that >90% of patients with >3 cm diameter lesions will have lung cancer. See Midthun et al. (1992). None of the algorithms were sufficient to exclude lung cancer, although all patients in this study with a lesion >3 cm and elevated CEA (>3.22 ng/ml) had lung cancer (100% specificity, 68% sensitivity). Patients in the large size category can have a FDG-PET scan, or an immediate biopsy (or resection), particularly if CEA is elevated. This again will depend on the clinical context.

Using a combination of a CART derived feature (CEA > 3.22 ng/ml) for the largest nodules and logistic regression algorithms developed in this study, the overall sensitivity, specificity, and positive and negative predictive values for the diagnosis of lung cancer are 92%, 79%, 82%, and 91%, respectively.

REFERENCES

All references listed herein, including patents, patent applications, data base citations and scientific literature, are incorporated herein by reference in their entireties to the extent that they supplement, explain, provide a background for, or teach methodology, techniques, and/or compositions employed herein.

Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.*;365(5): 395-409, 2011.

Bach, P. B., Jett, J. R., Pastorino, U., Tockman, M. S., Swensen, S. J., and Begg, C. B. Computed tomography screening and lung cancer outcomes. *Jama*, 297*:* 953-961, 2007.

Brandman S, Ko JP. Pulmonary nodule detection, characterization, and management with multidetector computed tomography. *J Thorac Imaging*; 26(2)**:** 90-105, 2011.

Condeelis, J. and Pollard, J. W. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell*, 124*:* 263-266, 2006.

Erasmus JJ, Patz EF, Jr. Positron emission tomography imaging in the thorax. *Clin Chest Med*; 20(4)**:** 715-24,1999.

Goo JM, Park CM, Lee HJ. Ground-glass nodules on chest CT as imaging biomarkers in the management of lung adenocarcinoma. *AJR Am J Roentgenol*;196(3)**:** 533-43, 2011.

Kim HK, Choi YS, Kim J, Shim YM, Lee KS, Kim K. Management of multiple pure ground-glass opacity lesions in patients with bronchioloalveolar carcinoma. *J Thorac Oncol*;5(2)**:** 206-10, 2010.

Kim HY, Shim YM, Lee KS, Han J, Yi CA, Kim YK. Persistent pulmonary nodular ground-glass opacity at thin-section CT: histopathologic comparisons. *Radiology*;245(1)**:** 267-75, 2007.

Lee HY, Lee KS. Ground-glass opacity nodules: histopathology, imaging evaluation, and clinical implications. *J Thorac Imaging*; 26(2)**:** 106-18, 2011.

Ljujic, M., Nikolic, A., Divac, A., Djordjevic, V., and Radojkovic, D. Screening of alpha-1-antitrypsin gene by denaturing gradient gel electrophoresis (DGGE). *J Biochem Biophys Methods*, 68*:* 167-173, 2006.

MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*, 237(2): 395-400, 2005.

Marcus, P. M., Bergstralh, E. J., Fagerstrom, R. M., Williams, D. E., Fontana, R., Taylor, W. F., and Prorok, P. C. Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J Natl Cancer Inst*, 92*:* 1308-1316, 2000.

Midthun DE, Swensen SJ, Jett JR. Clinical strategies for solitary pulmonary nodule. *Annu Rev Med.*; 43: 195-208, 1992.

Patz EF, Lowe VJ, Hoffman JM, et al. Focal pulmonary abnormalities: Evaluation with F-18 fluorodeoxyflucose PET scanning. *Radiology*; 188: 487-490, 1993.

Patz EF, Jr., Campa MJ, Gottlin EB, Kusmartseva I, Guan XR, Herndon JE, 2nd. Panel of serum biomarkers for the diagnosis of lung cancer. *J Clin Oncol*; 25(35): 5578-83, 2007.

PCT International Publication No. WO 2008/144034.

Petersen, R. P., Campa, M. J., Sperlazza, J., Conlon, D., Joshi, M. B., Harpole, D. H., Jr., and Patz, E. F., Jr. Tumor infiltrating Foxp3+ regulatory T-cells are associated with recurrence in pathologic stage I NSCLC patients. *Cancer*, 107: 2866-2872, 2006.

U.S. Patent No. 4,981,785.

U.S. Patent No. 5,358,691.

U.S. Patent No. 5,599,677.

U.S. Patent No. 5,672,480.

U.S. Patent No. 5,885,530.

U.S. Patent No. 6,159,750.

Weir G, Kos S, Burrill J, Salat P, Ho S, Liu D. Diagnosis and management of the solitary pulmonary nodule. *Bmj*; 343: d4866, 2011.

Welsh, T. J., Green, R. H., Richardson, D., Waller, D. A., O'Byrne, K. J., and Bradding, P. Macrophage and mast-cell invasion of tumor cell islets confers a marked survival advantage in non-small-cell lung cancer. *J Clin Oncol.*, 23: 8959-8967, 2005.

Zelvyte, I., Wallmark, A., Piitulainen, E., Westin, U., and Janciauskiene, S. Increased plasma levels of serine proteinase inhibitors in lung cancer patients. *Anticancer Res*, 24: 241-247, 2004.

Zhong, L., Coe, S. P., Stromberg, A. J., Khattar, N. H., Jett, J. R., and Hirschowitz, E. A. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J Thorac Oncol.*, 1: 513-519, 2006.

It will be understood that various details of the presently disclosed subject matter may be changed without departing from the scope of the presently disclosed subject matter. Furthermore, the foregoing description is for the purpose of illustration only, and not for the purpose of limitation.

CLAIMS

What is claimed is:

1. A method for assigning a subject having an indeterminate pulmonary nodule or abnormality to a group having higher or lower probability of lung cancer, the method comprising:

a) providing a subject having an indeterminate pulmonary nodule or abnormality, the indeterminate pulmonary nodule or abnormality having a size;

b) determining an amount of each member of a panel of biomarkers in a sample from the subject, wherein the panel of biomarkers comprises at least two of the proteins alpha-1-antitrypsin, carcinoembryonic antigen, squamous cell carcinoma antigen, retinol binding protein, transferrin and haptoglobin; and

c) assigning the subject to a group having a higher or lower probability of lung cancer based on the determined amount of each biomarker member in the panel and based on the size of the indeterminate pulmonary nodule or abnormality.

2. The method of claim 1, wherein the panel of biomarkers comprises at least alpha-1-antitrypsin, carcinoembryonic antigen, and squamous cell carcinoma antigen.

3. The method of claim 1, wherein the panel of biomarkers consists of alpha-1-antitrypsin, carcinoembryonic antigen, and squamous cell carcinoma antigen.

4. The method of claim 1, wherein the method further comprises assigning the nodule or abnormality to a size category selected from the group consisting of a small size category, an intermediate size category, and a large size category.

5. The method of claim 4, wherein the nodule or abnormality is assigned to the small size category when the largest of the length, width and

height of the nodule or abnormality is less than one centimeter; assigned to the intermediate size category when the largest of the length, width and height of the nodule or abnormality is at least one centimeter and less than three centimeters; or assigned to the large size category when the largest of the length, width and height of the nodule or abnormality is three centimeters or more.

6. The method of claim 1, comprising applying a predetermined algorithm to the determined amount of each biomarker in the panel and the size of the indeterminate pulmonary nodule or abnormality; and using the results of the algorithm to assign the subject to a group having a higher or lower probability of lung cancer.

7. The method of claim 6, wherein the algorithm is a logistic regression analysis or a classification and regression tree (CART) analysis.

8. The method of claim 6, wherein the predetermined algorithm is selected based on the size of the indeterminate pulmonary nodule or abnormality.

9. The method of claim 8, wherein a CART analysis algorithm is selected when the indeterminate pulmonary nodule or abnormality has a largest dimension that is 3 centimeters or more.

10. The method of claim 8, wherein a logistic regression analysis algorithm is selected when the indeterminate pulmonary nodule or abnormality has a largest dimension that is less than 3 centimeters.

11. The method of claim 10, wherein the logistic regression analysis algorithm is a weighted logistic regression analysis algorithm when the indeterminate pulmonary nodule or abnormality has a largest dimension that is at least one centimeter and less than 3 centimeters.

12. The method of claim 1, wherein the sample is a serum sample.

13. The method of claim 1, wherein the subject is a human subject.

14. A method for managing treatment of a subject with potential lung cancer, the method comprising:

a) providing a subject having an indeterminate pulmonary nodule or abnormality, the indeterminate pulmonary nodule or abnormality having a size;

b) determining an amount of each member of a panel of biomarkers in a sample from a subject with an indeterminate pulmonary nodule or abnormality, wherein the panel of biomarkers comprises at least two of the proteins alpha-1-antitrypsin, carcinoembryonic antigen, squamous cell carcinoma antigen, retinol binding protein, transferrin, and haptoglobin;

c) assigning the subject to a group having a higher or lower probability of lung cancer based on the determined amount of each biomarker member in the panel and based on the size of the indeterminate pulmonary nodule or abnormality; and

d) managing the treatment of the subject with potential lung cancer based on the group to which the subject is assigned and based on the size of the indeterminate pulmonary nodule or abnormality.

15. The method of claim 14, wherein the managing comprises one or more of scheduling a computed tomography scan, scheduling a biopsy, and scheduling a positron emission tomography (PET) scan.

16. The method of claim 14, wherein the panel of biomarkers comprises at least alpha-1-antitrypsin, carcinoembryonic antigen, and squamous cell carcinoma antigen.

17. The method of claim 14, wherein the panel of biomarkers consists of alpha-1-antitrypsin, carcinoembryonic antigen, and squamous cell carcinoma.

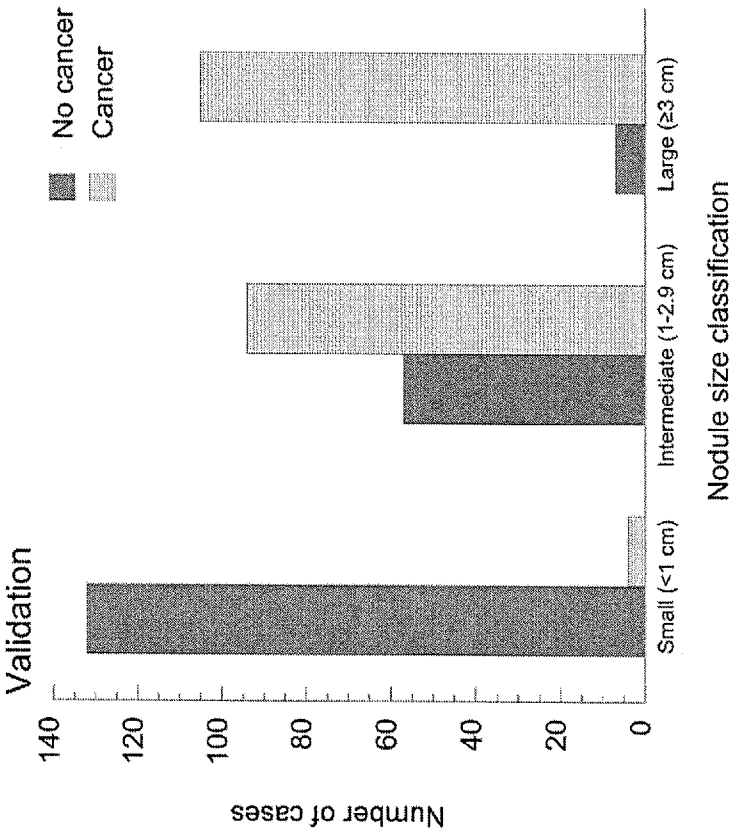

Validation
No cancer
Cancer
140
120
100
80
60
40
20
0
Number of cases
Small (<1 cm)
Intermediate (1-2.9 cm)
Large (≥3 cm)
Nodule size classification

FIG. 1B

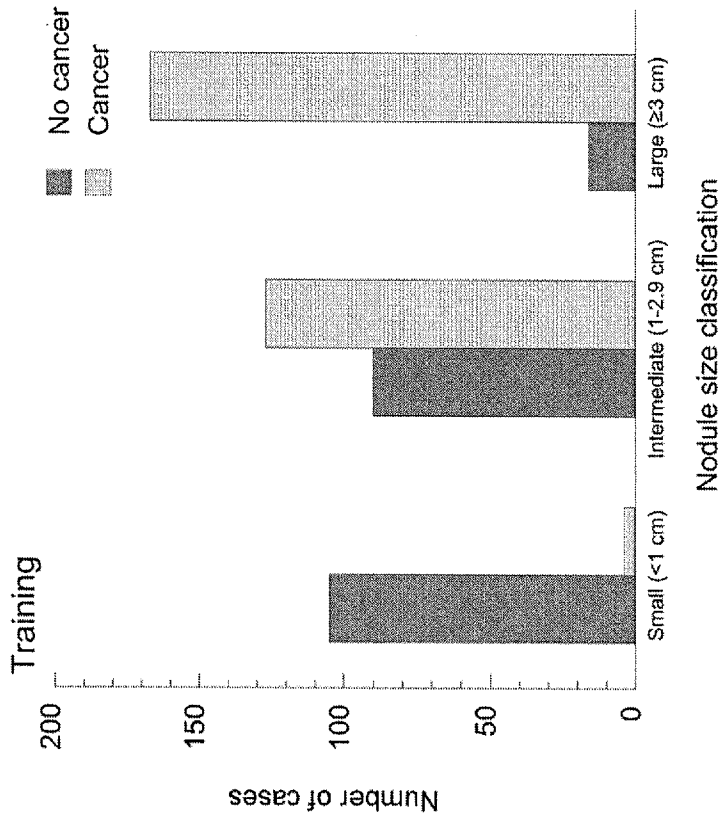

Training
No cancer
Cancer
200
150
100
50
0
Number of cases
Small (<1 cm)
Intermediate (1-2.9 cm)
Large (≥3 cm)
Nodule size classification

FIG. 1A

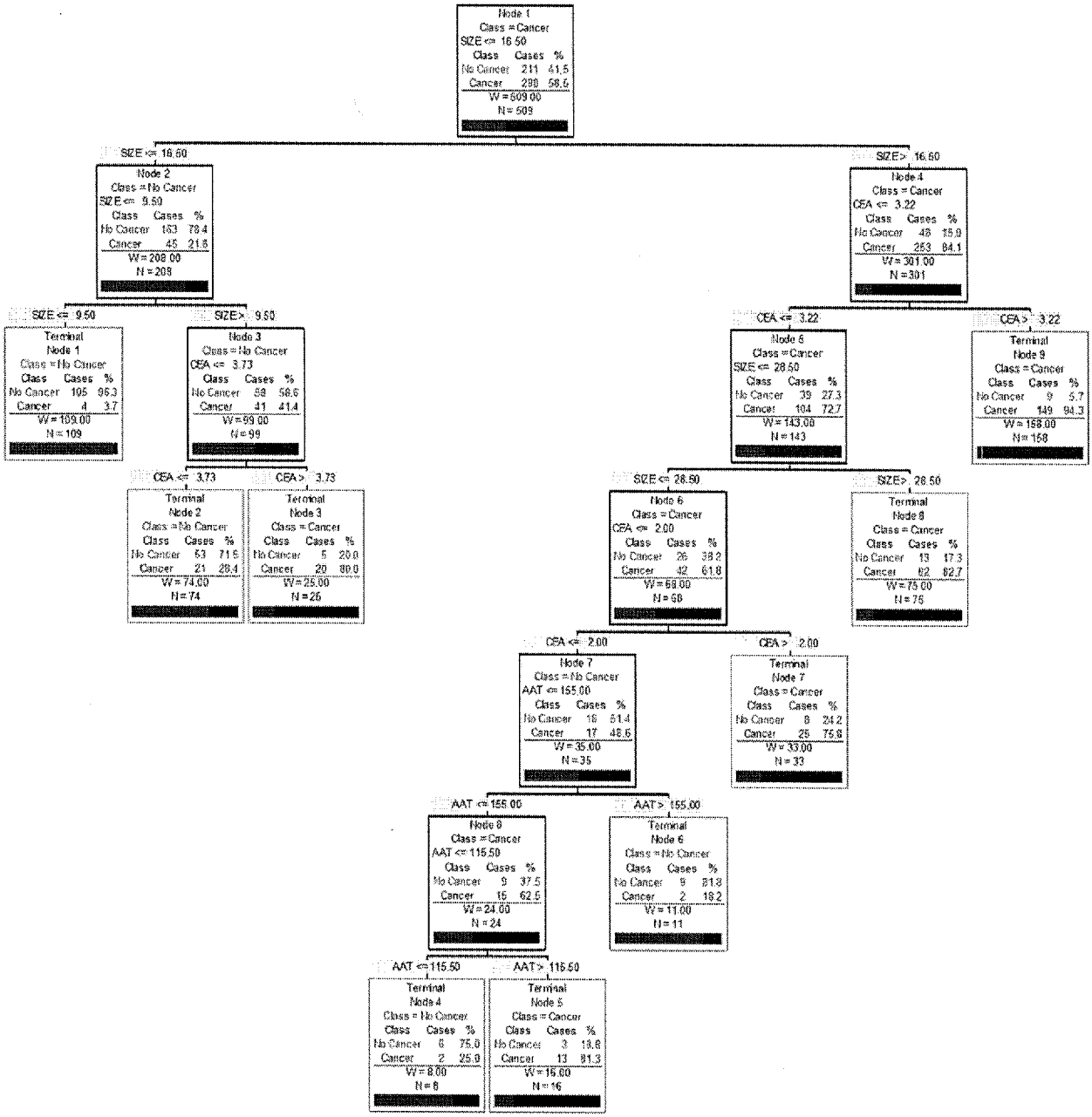

Node 1
Class = Cancer
SIZE <= 16.50
Class Cases %
No Cancer 211 41.5
Cancer 298 58.5
W = 509.00
N = 509
SIZE <= 16.50
Node 2
Class = No Cancer
SIZE <= 9.50
Class Cases %
No Cancer 163 78.4
Cancer 45 21.6
W = 208.00
N = 208
SIZE > 16.50
Node 4
Class = Cancer
CEA <= 3.22
Class Cases %
No Cancer 48 15.9
Cancer 253 84.1
W = 301.00
N = 301
SIZE <= 9.50
Terminal
Node 1
Class = No Cancer
Class Cases %
No Cancer 105 96.3
Cancer 4 3.7
W = 109.00
N = 109
SIZE > 9.50
Node 3
Class = No Cancer
CEA <= 3.73
Class Cases %
No Cancer 58 58.6
Cancer 41 41.4
W = 99.00
N = 99
CEA <= 3.22
Node 5
Class = Cancer
SIZE <= 28.50
Class Cases %
No Cancer 39 27.3
Cancer 104 72.7
W = 143.00
N = 143
CEA > 3.22
Terminal
Node 9
Class = Cancer
Class Cases %
No Cancer 9 5.7
Cancer 149 94.3
W = 158.00
N = 158
CEA <= 3.73
Terminal
Node 2
Class = No Cancer
Class Cases %
No Cancer 53 71.6
Cancer 21 28.4
W = 74.00
N = 74
CEA > 3.73
Terminal
Node 3
Class = Cancer
Class Cases %
No Cancer 5 20.0
Cancer 20 80.0
W = 25.00
N = 25
SIZE <= 28.50
Node 6
Class = Cancer
CEA <= 2.00
Class Cases %
No Cancer 26 38.2
Cancer 42 61.8
W = 68.00
N = 68
SIZE > 28.50
Terminal
Node 8
Class = Cancer
Class Cases %
No Cancer 13 17.3
Cancer 62 82.7
W = 75.00
N = 75
CEA <= 2.00
Node 7
Class = No Cancer
AAT <= 155.00
Class Cases %
No Cancer 18 51.4
Cancer 17 48.6
W = 35.00
N = 35
CEA > 2.00
Terminal
Node 7
Class = Cancer
Class Cases %
No Cancer 8 24.2
Cancer 25 75.8
W = 33.00
N = 33
AAT <= 155.00
Node 8
Class = Cancer
AAT <= 115.50
Class Cases %
No Cancer 9 37.5
Cancer 15 62.5
W = 24.00
N = 24
AAT > 155.00
Terminal
Node 6
Class = No Cancer
Class Cases %
No Cancer 9 81.8
Cancer 2 18.2
W = 11.00
N = 11
AAT <= 115.50
Terminal
Node 4
Class = No Cancer
Class Cases %
No Cancer 6 75.0
Cancer 2 25.0
W = 8.00
N = 8
AAT > 115.50
Terminal
Node 5
Class = Cancer
Class Cases %
No Cancer 3 18.8
Cancer 13 81.3
W = 16.00
N = 16

FIG. 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/35632

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G01N 33/53; C12Q 1/00 (2013.01)
USPC - 435/7.1; 436/501; 382/159, 131, 133
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): C40B 40/10; G01N 33/53, 566; C07K 14/435, 7/08; G06K 9/00; C12Q 1/00, 68 (2013.01)
USPC: 506/18, 15, 13, 8.7; 435/7.1, 7.92; 436/501; 382/128, 130, 133, 155, 159, 160, 161; 536/24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MicroPatent (US-G, US-A, EP-A, EP-B, WO, JP-bib, DE-C,B, DE-A, DE-T, DE-U, GB-A, FR-A); Google; Google Scholar; DialogPRO; PubMed; assign, assess, determine, classify, risk, stage, pulmonary, nodule, biomarker, marker, tumor, alpha-1-antitrypsin, carcinoembryonic, antigen, squamous, 'CEA,' 'SCC,' ertinol, 'RBP,' transferrin, haptoglobin, abnormal, lesion, size

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 7892760 B2 (BIRSE, C et al.), February 22, 2011; column 10, lines 4-8; column 16, lines 6-15; column 17, lines 37-67 to column 18, lines 1-9; column 19, lines 14-62; column 24, lines 18-54; Claim 14	1, 4, 6, 8, 10, 12-15 ----- 2, 3, 5, 7, 9, 11, 16, 17
Y	US 2010/0179067 A1 (PATZ, JR., EF et al.), July 15, 2010; paragraphs [0022], [0032], [0045]	2, 3, 7, 9, 16, 17
Y	US 7949169 B2 (BAE, KT et al.), May 24, 2011; figure 1; column 4, lines 45-65; column 7, lines 18-33; column 8, lines 13-52	5, 9, 11
A	US 2012/0071334 A1 (COLPITTS, T et al.), March 22, 2012; paragraph [0030]	2, 3, 16, 17

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 August 2013 (02.08.2013)	08 AUG 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)