Description

Title of Invention: PROTEIN BIOMARKER PANEL FOR DETECTING NON-SMALL CELL LUNG CANCER AND METHOD FOR DIAGNOSING A NON-SMALL CELL LUNG CANCER BY THE USE THEREOF

Technical Field

[1] The present invention relates to a Protein Biomarker Panel for Detecting Non-Small Cell Lung Cancer and method for diagnosing a non-small cell lung cancer by the use thereof, and more particularly to a Protein Biomarker Panel for Detecting Non-Small Cell Lung Cancer from a human including N of the biomarker proteins having a KIT.

[2]

Background Art

[3] The following description provides a summary of information relevant to the present application and is not an admission that any of the information provided or publications referred herein is prior art to the present application.

[4] Lung cancer remains the leading cause of cancer- related deaths in the Korea and the world.(Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, et al. (2015) Global cancer statistics, 2012. CA Cancer J Clin.). In 2011, it was estimated that there were more than 21,753 new cases of lung cancer and more than 15,000 deaths due to lung cancer in Korea.(Jung KW, Won YJ, Kong HJ, Oh CM, Lee DH, et al. (2014) Cancer statistics in Korea: incidence, mortality, survival, and pre Although lung cancer death rates are declining in developed countries, lung cancer incidence and death rates are rapidly rising in the developing world, especially in the Asian countries, where smoking prevalence continues to increase.(Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, et al. (2015) Global cancer statistics, 2012. CA Cancer J Clin.).

[5] Since most people with early stage lung cancer do not display symptoms, more than 60% of patients are diagnosed at advanced stages when a cure is unlikely (Jemal A, Center MM, DeSantis C, Ward EM (2010) Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev 19: 1893-1907; Jett JR (1993) Current Treatment of Unresectable Lung-Cancer. Mayo Clinic Proceedings 68: 603-611.). Five-year survival rate for patients with advanced disease can be less than 10%, whereas 5-year survival rate in patients with stage I disease may be greater than 70%(Hoffman PC, Mauer AM, Vokes EE (2000) Lung cancer. Lancet 355: 479-485.). Therefore, research aimed at early detection of lung cancer, which is critical to reducing mortality and morbidity, has turned to the development of lung cancer

screening.

[6] Despite the 20 % relative reduction in lung cancer related mortality shown by low dose chest CT (LDCT) screening in the National Lung Screening Trial (NLST)(Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365: 395-409.), 24.2% of screening LDCT scans were positive, with 96.4% of these nodules determined as false positives(Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365: 395-409.; Aberle DR, Abtin F, Brown K (2013) Computed Tomography Screening for Lung Cancer: Has It Finally Arrived? Implications of the National Lung Screening Trial. Journal of Clinical Oncology 31: 1002-1008.). In addition to detecting aggressive tumors, more than 18% of all lung cancers detected by LDCT in the NLST seem to be indolent, and over-diagnosis should be considered when describing the risks of LDCT screening for lung cancer.(Patz EF, Jr., Pinsky P, Gatsonis C, Sicks JD, Kramer BS, et al. (2014) Overdiagnosis in low-dose computed tomography screening for lung cancer. JAMA Intern Med 174: 269-274.) .

[7] Therefore, sensitive and specific lung cancer biomarkers, measured in a noninvasively collected biological specimen such as serum, could help guide clinical decision making in high risk subjects, particularly in patients with CT-detected indeterminate pulmonary nodules. Published data on individual serum biomarkers, most notably cytokeratin 19 fragment 21.1 (Cyfra 21-1), carcinoembryonic antigen, and tissue plasminogen activator in non-small cell lung cancer (NSCLC) show limited sensitivity and specificity, particularly in early-stage disease. (Pastor A, Menendez R, Cremades MJ, Pastor V, Llopis R, et al. (1997) Diagnostic value of SCC, CEA and CYFRA 21.1 in lung cancer: a Bayesian analysis. European Respiratory Journal 10: 603-609.; Buccheri G, Torchio P, Ferrigno D (2003) Clinical equivalence of two cytokeratin markers in non-small cell lung cancer - A study of tissue polypeptide antigen and cytokeratin 19 fragments. Chest 124: 622-632.; Lai RS, Hsu HK, Lu JY, Ger LP, Lai NS (1996) CYFRA 21-1 enzyme-linked immunosorbent assay. Evaluation as a tumor marker in non-small cell lung cancer. Chest 109: 995-1000.).

[8] Advances in molecular diagnostics and understanding of genomics have led to discovery of several lung cancer biomarkers with potential to complement the current screening standards.(Hasan N, Kumar R, Kavuru MS (2014) Lung cancer screening beyond low-dose computed tomography: the role of novel biomarkers. Lung 192: 639-648.; Bigbee WL, Gopalakrishnan V, Weissfeld JL, Wilson DO, Dacic S, et al. (2012) A multiplexed serum biomarker immunoassay panel discriminates clinical lung cancer patients from high-risk individuals found to be cancer-free by CT screening. J

Thorac Oncol 7: 698-708.; Daly S, Rinewalt D, Fhied C, Basu S, Mahon B, et al. (2013) Development and validation of a plasma biomarker panel for discerning clinical significance of indeterminate pulmonary nodules. J Thorac Oncol 8: 31-36.; Gold L, Ayers D, Bertino J, Bock C, Bock A, et al. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 5: e15004.; Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, et al. (2010) Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. PLoS One 5: e15003.; Pecot CV, Li M, Zhang XJ, Rajanbabu R, Calitri C, et al. (2012) Added value of a serum proteomic signature in the diagnostic evaluation of lung nodules. Cancer Epidemiol Biomarkers Prev 21: 786-792.) Ostroff et al. previously reported that the discovery of a protein biomarker panel for early detection of lung cancer.(Gold L, Ayers D, Bertino J, Bock C, Bock A, et al. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 5: e15004.; Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, et al. (2010) Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. PLoS One 5: e15003.).

[9] However, the epidemiology and molecular biology of lung cancer is different according to ethnic or geographic backgrounds, and there was no development of a protein biomarker and incorporation into CT imaging models.

[10] Aptamer is a novel bio-molecule and it is screened from large pool of oligo library. By series of SELEX rounds, target specific ligand aptamers are enriched and selected finally. Aptamer is single stranded nucleic acid that could be synthesized chemically, which is prone to modified with various purpose. Furthermore, aptamer can be amplified and analyzed by polymerase chain reaction method, and can be applied to high contents DNA array technology.

[11] Aptamer-based high contents quantitative assay was reported and proved to be a good platform for screening multivariate proteome signature to diagnose disease status. Over a thousand of protein could be measured in a single platform concomitantly, and this gives chances of analyzing human samples to find disease specific proteome signature.

[12]

Disclosure of Invention

Technical Problem

[13] An aspect of the present invention provides a method for diagnosing a non-small cell lung cancer in a human.

[14] Another aspect of the present invention provides a protein biomarker panel to detect non-small cell lung cancer.

[15] Objects of the present invention are not limited to the above object and other objects will be clearly understood by those skilled in the art from the following description.

[16]

Solution to Problem

[17] In order to achieve the technical subject, one embodiment of the present invention provides a Protein Biomarker Panel for the detection of lung cancer and more specifically, NSCLC. In one embodiment of the present invention, the biomarkers were identified using a multiplex aptamer-based assay which is described in detail at EXAMPLE. By using the aptamer-based biomarker identification method described herein, this application describes a list of NSCLC biomarkers that are useful for the detection and diagnosis of NSCLC. In identifying these biomarkers, small set of candidate biomarker proteins were measured in human specimen from Asians, having previously been diagnosed either as having or not having NSCLC. Each biomarker was analyzed and compared with their ability to discriminate cases and control, and smaller set of biomarker was selected with better performance as listed in table 2. By using naive Bayesian theorem, multivariate classifier was generated and analyzed as is described in detail at EXAMPLE.

[18] The objects of the present invention can be achieved by one embodiment of the present invention, a method for diagnosing a lung cancer in a human, the method comprises

[19] providing a biomarker panel which including N of biomarker proteins listed in table 2, wherein N is an integer at least 2, and detecting the biomarker proteins, in a biological sample from the human, to give biomarker values that each corresponds to one of the N biomarker proteins in the panel, wherein the lung cancer is diagnosed based on the biomarker values.

[20] The detecting the biomarker values may include performing an in vitro assay.

[21] The in vitro assay may include one capture reagent corresponding to each of the biomarker proteins, and the method may further comprise selecting the at least one capture reagent from a group consisting of aptamers, antibodies, and a nucleic acid probe.

[22] The at least one capture reagent may be an aptamer.

[23] The biological sample may be selected from a group consisting of a whole blood, a plasma, and a serum.

[24] The biological sample may be a serum

[25] The human may be an Asian.

[26] The human may be a smoker.

[27] The human may have a pulmonary malignant nodule.

[28] The N may be 3, 4, 5, 6, 7, or more.

[29]

[30] The biomarker values may be measured values of Complement C9, carbonic anhydrase 6 (CA6), C-reactive protein (CRP), epidermal growth factor receptor 1 (EGFR1), matrix metalloproteinase 7(MMP7), alpha1-antiprotease (SERPINA3), and stem cell growth factor receptor (KIT).

[31] The biomarker values may be measured from a group consisting of real-time PCR, microarray, and Luminex microbead assay.

[32] The lung cancer may be diagnosed with a statistical method.

[33] The statistical method may be selected from a group comprising a linear discriminant analysis, a logistic regression, a naive Bayesian classification, a support vector machine, and a random forest.

[34] The objects of the present invention can also be achieved by providing a protein biomarker panel to detect non-small cell lung cancer from a human, the panel comprising N of biomarker proteins listed in table 2, wherein the N is at least 2.

[35] The N may be 3, 4, 5, 6, or more.

[36] The panel may have a Complement C9, a carbonic anhydrase 6 (CA6), a C-reactive protein (CRP), a epidermal growth factor receptor 1 (EGFR1), a matrix metallo-proteinase 7 (MMP7), a alpha1-antiprotease (SERPINA3), and a stem cell growth factor receptor (KIT).

[37] The human may be an Asian.

[38] The human may be a smoker.

[39] The human may have a pulmonary malignant nodule.

[40]

Advantageous Effects of Invention

[41] The Protein Biomarker Panel can provide a Protein Biomarker Panel to Detect Non-Small Cell Lung Cancer.

[42] Advantages of the present invention are not limited to those described above and other advantages will be clearly understood by those skilled in the art from the following description.

[43]

Brief Description of Drawings

[44] Figure 1 is a study flow for algorithm training and verification.

[45] Figure 2 shows Aptamer-based multiplex assay.

[46] Figure 3 is Example of raw data along with model fit to normal CDF.

[47] figure 4a-4n shows candidate marker comparisons between cases and controls.

[48] Figure 5a-5c is ROC curves for 7-marker naive Bayes classifier.

[49] Figure 6a-6g is ROC curves for 6-marker naive Bayes classifier.

[50] Figure 7a-7b is Performance comparisons of 7-marker naive Bayes Classifier with Cyfra21-1.

[51]

Mode for the Invention

[52] Reference will now be made in detail to representative embodiments of the invention. While the invention will be described in conjunction with the enumerated embodiments, it will be understood that the invention is not to be limited to those embodiments. On the contrary, the invention is intended to cover all alternatives, modifications, and equivalents that may be included within the scope of the present invention as defined by the claims.

[53] One skilled in the art will recognize many methods and materials similar or equivalent to those described herein, which could be used in and are within the scope of the practice of the present invention. The present invention is in no way limited to the methods and materials described.

[54] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods, devices, and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

[55] All publications, published patent documents, and patent applications cited in this application are indicative of the level of skill in the art(s) to which the application pertains. All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

[56]

[57] As used in this application, including the appended claims, the singular forms "a," "an," and "the" include plural references, unless the content clearly dictates otherwise, and are used interchangeably with "at least one" and "one or more." Thus, reference to "an aptamer" includes mixtures of aptamers, reference to "a probe" includes mixtures of probes, and the like.

[58] As used herein, the term "about" represents an insignificant modification or variation of the numerical value such that the basic function of the item to which the numerical value relates is unchanged.

[59] As used herein, the terms "comprises," "comprising," "includes," "including," "contains," "containing," and any variations thereof, are intended to cover a non-

exclusive inclusion, such that a process, method, product-by-process, or composition of matter that comprises, includes, or contains an element or list of elements does not include only those elements but may include other elements not expressly listed or inherent to such process, method, product-by-process, or composition of matter.

[60] In one embodiment, the number of biomarkers useful for a biomarker subset or panel is based on the sensitivity and specificity value for the particular combination of biomarker values. The terms "sensitivity" and "specificity" are used herein with respect to the ability to correctly classify an individual, based on one or more biomarker values detected in their biological sample, as having NSCLC or not having NSCLC. "Sensitivity" indicates the performance of the biomarker(s) with respect to correctly classifying humans that have NSCLC. "Specificity" indicates the performance of the biomarker(s) with respect to correctly classifying humans who do not have NSCLC.

[61] The present application includes biomarkers, methods, devices, reagents, systems, and kits for the detection and diagnosis of NSCLC and cancer more generally.

[62] As used herein, "lung" may be interchangeably referred to as "pulmonary".

[63] As used herein, "smoker" refers to an individual who has a history of tobacco smoke inhalation.

[64] "Biological sample", "sample", and "test sample" are used interchangeably herein to refer to any material, biological fluid, tissue, or cell obtained or otherwise derived from an individual. This includes blood (including whole blood, leukocytes, peripheral blood mononuclear cells, buffy coat, plasma, and serum), sputum, tears, mucus, nasal washes, nasal aspirate, breath, urine, semen, saliva, peritoneal washings, cystic fluid, meningeal fluid, amniotic fluid, glandular fluid, lymph fluid, cytologic fluid, ascites, pleural fluid, nipple aspirate, bronchial aspirate, bronchial brushing, synovial fluid, joint aspirate, organ secretions, cells, a cellular extract, and cerebrospinal fluid. This also includes experimentally separated fractions of all of the preceding. For example, a blood sample can be fractionated into serum, plasma or into fractions containing particular types of blood cells, such as red blood cells or white blood cells (leukocytes). If desired, a sample can be a combination of samples from an individual, such as a combination of a tissue and fluid sample. The term "biological sample" also includes materials containing homogenized solid material, such as from a stool sample, a tissue sample, or a tissue biopsy, for example. The term "biological sample" also includes materials derived from a tissue culture or a cell culture. Any suitable methods for obtaining a biological sample can be employed; exemplary methods include, e.g., phlebotomy, swab (e.g., buccal swab), and a fine needle aspirate biopsy procedure. Exemplary tissues susceptible to fine needle aspiration include lymph node, lung, lung washes, BAL (bronchoalveolar lavage), pleura, thyroid, breast, pancreas and liver. Samples can also be collected, e.g., by micro dissection (e.g., laser capture micro

dissection (LCM) or laser micro dissection (LMD)), bladder wash, smear (e.g., a PAP smear), or ductal lavage. A "biological sample" obtained or derived from an individual includes any such sample that has been processed in any suitable manner after being obtained from the individual.

[65] Further, it should be realized that a biological sample can be derived by taking biological samples from a number of humans and pooling them or pooling an aliquot of each individual's biological sample. The pooled sample can be treated as a sample from a single individual and if the presence of cancer is established in the pooled sample, then each individual biological sample can be re-tested to determine which individual(s) have NSCLC.

[66] For purposes of this specification, the phrase "data attributed to a biological sample from an individual" is intended to mean that the data in some form derived from, or were generated using, the biological sample of the individual. The data may have been reformatted, revised, or mathematically altered to some degree after having been generated, such as by conversion from units in one measurement system to units in another measurement system; but, the data are understood to have been derived from, or were generated using, the biological sample.

[67] "Target", "target molecule", and "analyte" are used interchangeably herein to refer to any molecule of interest that may be present in a biological sample. A "molecule of interest" includes any minor variation of a particular molecule, such as, in the case of a protein, for example, minor variations in amino acid sequence, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling component, which does not substantially alter the identity of the molecule. A "target molecule", "target", or "analyte" is a set of copies of one type or species of molecule or multi-molecular structure. "Target molecules", "targets", and "analytes" refer to more than one such set of molecules. Exemplary target molecules include proteins, polypeptides, nucleic acids, carbohydrates, lipids, polysaccharides, glycoproteins, hormones, receptors, antigens, antibodies, affybodies, autoantibodies, antibody mimics, viruses, pathogens, toxic substances, substrates, metabolites, transition state analogs, cofactors, inhibitors, drugs, dyes, nutrients, growth factors, cells, tissues, and any fragment or portion of any of the foregoing.

[68] As used herein, "polypeptide," "peptide," and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification,

such as conjugation with a labeling component. Also included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), as well as other modifications known in the art. Polypeptides can be single chains or associated chains. Also included within the definition are preproteins and intact mature proteins; peptides or polypeptides derived from a mature protein; fragments of a protein; splice variants; recombinant forms of a protein; protein variants with amino acid modifications, deletions, or substitutions; digests; and post-translational modifications, such as glycosylation, acetylation, phosphorylation, and the like.

[69] As used herein, "marker" and "biomarker" are used interchangeably to refer to a target molecule that indicates or is a sign of a normal or abnormal process in an individual or of a disease or other condition in an individual. More specifically, a "marker" or "biomarker" is an anatomic, physiologic, biochemical, or molecular parameter associated with the presence of a specific physiological state or process, whether normal or abnormal, and, if abnormal, whether chronic or acute. Biomarkers are detectable and measurable by a variety of methods including laboratory assays and medical imaging. When a biomarker is a protein, it is also possible to use the expression of the corresponding gene as a surrogate measure of the amount or presence or absence of the corresponding protein biomarker in a biological sample or methylation state of the gene encoding the biomarker or proteins that control expression of the biomarker.

[70] As used herein, "biomarker value", "value", "biomarker level", and "level" are used interchangeably to refer to a measurement that is made using any analytical method for detecting the biomarker in a biological sample and that indicates the presence, absence, absolute amount or concentration, relative amount or concentration, titer, a level, an expression level, a ratio of measured levels, or the like, of, for, or corresponding to the biomarker in the biological sample. The exact nature of the "value" or "level" depends on the specific design and components of the particular analytical method employed to detect the biomarker.

[71] When a biomarker indicates or is a sign of an abnormal process or a disease or other condition in an individual, that biomarker is generally described as being either over-expressed or under-expressed as compared to an expression level or value of the biomarker that indicates or is a sign of a normal process or an absence of a disease or other condition in an individual. "Up-regulation", "up-regulated", "over-expression", "over-expressed", and any variations thereof are used interchangeably to refer to a value or level of a biomarker in a biological sample that is greater than a value or level (or range of values or levels) of the biomarker that is typically detected in similar biological samples from healthy or normal humans. The terms may also refer to a value

or level of a biomarker in a biological sample that is greater than a value or level (or range of values or levels) of the biomarker that may be detected at a different stage of a particular disease.

[72] "Down-regulation", "down-regulated", "under-expression", "under-expressed", and any variations thereof are used interchangeably to refer to a value or level of a biomarker in a biological sample that is less than a value or level (or range of values or levels) of the biomarker that is typically detected in similar biological samples from healthy or normal humans. The terms may also refer to a value or level of a biomarker in a biological sample that is less than a value or level (or range of values or levels) of the biomarker that may be detected at a different stage of a particular disease.

[73] Further, a biomarker that is either over-expressed or under-expressed can also be referred to as being "differentially expressed" or as having a "differential level" or "differential value" as compared to a "normal" expression level or value of the biomarker that indicates or is a sign of a normal process or an absence of a disease or other condition in an individual. Thus, "differential expression" of a biomarker can also be referred to as a variation from a "normal" expression level of the biomarker.

[74] The term "differential gene expression" and "differential expression" are used interchangeably to refer to a gene (or its corresponding protein expression product) whose expression is activated to a higher or lower level in a subject suffering from a specific disease, relative to its expression in a normal or control subject. The terms also include genes (or the corresponding protein expression products) whose expression is activated to a higher or lower level at different stages of the same disease. It is also understood that a differentially expressed gene may be either activated or inhibited at the nucleic acid level or protein level, or may be subject to alternative splicing to result in a different polypeptide product. Such differences may be evidenced by a variety of changes including mRNA levels, surface expression, secretion or other partitioning of a polypeptide. Differential gene expression may include a comparison of expression between two or more genes or their gene products; or a comparison of the ratios of the expression between two or more genes or their gene products; or even a comparison of two differently processed products of the same gene, which differ between normal subjects and subjects suffering from a disease; or between various stages of the same disease. Differential expression includes both quantitative, as well as qualitative, differences in the temporal or cellular expression pattern in a gene or its expression products among, for example, normal and diseased cells, or among cells which have undergone different disease events or disease stages.

[75] "Diagnose", "diagnosing", "diagnosis", and variations thereof refer to the detection, determination, or recognition of a health status or condition of an individual on the basis of one or more signs, symptoms, data, or other information pertaining to that in-

dividual. The health status of an individual can be diagnosed as healthy/normal (i.e., a diagnosis of the absence of a disease or condition) or diagnosed as ill/abnormal (i.e., a diagnosis of the presence, or an assessment of the characteristics, of a disease or condition). The terms "diagnose", "diagnosing", "diagnosis", etc., encompass, with respect to a particular disease or condition, the initial detection of the disease; the characterization or classification of the disease; the detection of the progression, remission, or recurrence of the disease; and the detection of disease response after the administration of a treatment or therapy to the individual. The diagnosis of NSCLC includes distinguishing humans who have cancer from humans who do not. It further includes distinguishing smokers and benign pulmonary nodules from NSCLC.

[76] "Prognose", "prognosing", "prognosis", and variations thereof refer to the prediction of a future course of a disease or condition in an individual who has the disease or condition (e.g., predicting patient survival), and such terms encompass the evaluation of disease response after the administration of a treatment or therapy to the individual.

[77] "Evaluate", "evaluating", "evaluation", and variations thereof encompass both "diagnose" and "prognose" and also encompass determinations or predictions about the future course of a disease or condition in an individual who does not have the disease as well as determinations or predictions regarding the likelihood that a disease or condition will recur in an individual who apparently has been cured of the disease. The term "evaluate" also encompasses assessing an individual's response to a therapy, such as, for example, predicting whether an individual is likely to respond favorably to a therapeutic agent or is unlikely to respond to a therapeutic agent (or will experience toxic or other undesirable side effects, for example), selecting a therapeutic agent for administration to an individual, or monitoring or determining an individual's response to a therapy that has been administered to the individual. Thus, "evaluating" NSCLC can include, for example, any of the following: prognosing the future course of NSCLC in an individual; predicting the recurrence of NSCLC in an individual who apparently has been cured of NSCLC; or determining or predicting an individual's response to a NSCLC treatment or selecting a NSCLC treatment to administer to an individual based upon a determination of the biomarker values derived from the individual's biological sample.

[78] Any of the following examples may be referred to as either "diagnosing" or "evaluating" NSCLC: initially detecting the presence or absence of NSCLC; determining a specific stage, type or sub-type, or other classification or characteristic of NSCLC; determining whether a suspicious lung nodule or mass is benign or malignant NSCLC; or detecting/monitoring NSCLC progression (e.g., monitoring tumor growth or metastatic spread), remission, or recurrence.

[79] As used herein, "additional biomedical information" refers to one or more

evaluations of an individual, other than using any of the biomarkers described herein, that are associated with cancer risk or, more specifically, NSCLC risk. "Additional biomedical information" includes any of the following: physical descriptors of an individual, physical descriptors of a pulmonary nodule observed by CT imaging, the height and/or weight of an individual, the gender of an individual, the ethnicity of an individual, smoking history, occupational history, exposure to known carcinogens (e.g., exposure to any of asbestos, radon gas, chemicals, smoke from fires, and air pollution, which can include emissions from stationary or mobile sources such as industrial/factory or auto/marine/aircraft emissions), exposure to second-hand smoke, family history of NSCLC (or other cancer), the presence of pulmonary nodules, size of nodules, location of nodules, morphology of nodules (e.g., as observed through CT imaging, ground glass opacity (GGO), solid, non-solid), edge characteristics of the nodule (e.g., smooth, lobulated, sharp and smooth, spiculated, infiltrating), and the like. Smoking history is usually quantified in terms of "pack years", which refers to the number of years a person has smoked multiplied by the average number of packs smoked per day. For example, a person who has smoked, on average, one pack of cigarettes per day for 35 years is referred to as having 35 pack years of smoking history. Additional biomedical information can be obtained from an individual using routine techniques known in the art, such as from the individual themselves by use of a routine patient questionnaire or health history questionnaire, etc., or from a medical practitioner, etc. Alternately, additional biomedical information can be obtained from routine imaging techniques, including CT imaging (e.g., low-dose CT imaging) and X-ray. Testing of biomarker levels in combination with an evaluation of any additional biomedical information may, for example, improve sensitivity, specificity, and/or AUC for detecting NSCLC (or other NSCLC-related uses) as compared to biomarker testing alone or evaluating any particular item of additional biomedical information alone (e.g., CT imaging alone).

[80] The term "area under the curve" or "AUC" refers to the area under the curve of a receiver operating characteristic (ROC) curve, both of which are well known in the art. AUC measures are useful for comparing the accuracy of a classifier across the complete data range. Classifiers with a greater AUC have a greater capacity to classify unknowns correctly between two groups of interest (e.g., NSCLC samples and normal or control samples). ROC curves are useful for plotting the performance of a particular feature (e.g., any of the biomarkers described herein and/or any item of additional biomedical information) in distinguishing between two populations (e.g., cases having NSCLC and controls without NSCLC). Typically, the feature data across the entire population (e.g., the cases and controls) are sorted in ascending order based on the value of a single feature. Then, for each value for that feature, the true positive and

false positive rates for the data are calculated. The true positive rate is determined by counting the number of cases above the value for that feature and then dividing by the total number of cases. The false positive rate is determined by counting the number of controls above the value for that feature and then dividing by the total number of controls. Although this definition refers to scenarios in which a feature is elevated in cases compared to controls, this definition also applies to scenarios in which a feature is lower in cases compared to the controls (in such a scenario, samples below the value for that feature would be counted). ROC curves can be generated for a single feature as well as for other single outputs, for example, a combination of two or more features can be mathematically combined (e.g., added, subtracted, multiplied, etc.) to provide a single sum value, and this single sum value can be plotted in a ROC curve. Additionally, any combination of multiple features, in which the combination derives a single output value, can be plotted in a ROC curve. These combinations of features may comprise a test. The ROC curve is the plot of the true positive rate (sensitivity) of a test against the false positive rate (1-specificity) of the test.

[81] As used herein, "detecting" or "determining" with respect to a biomarker value includes the use of both the instrument required to observe and record a signal corresponding to a biomarker value and the material/s required to generate that signal. In various embodiments, the biomarker value is detected using any suitable method, including fluorescence, chemiluminescence, surface plasmon resonance, surface acoustic waves, mass spectrometry, infrared spectroscopy, Raman spectroscopy, atomic force microscopy, scanning tunneling microscopy, electrochemical detection methods, nuclear magnetic resonance, quantum dots, and the like.

[82] "Solid support" refers herein to any substrate having a surface to which molecules may be attached, directly or indirectly, through either covalent or non-covalent bonds. A "solid support" can have a variety of physical formats, which can include, for example, a membrane; a chip (e.g., a protein chip); a slide (e.g., a glass slide or coverslip); a column; a hollow, solid, semi-solid, pore- or cavity-containing particle, such as, for example, a bead; a gel; a fiber, including a fiber optic material; a matrix; and a sample receptacle. Exemplary sample receptacles include sample wells, tubes, capillaries, vials, and any other vessel, groove or indentation capable of holding a sample. A sample receptacle can be contained on a multi-sample platform, such as a microtiter plate, slide, microfluidics device, and the like. A support can be composed of a natural or synthetic material, an organic or inorganic material. The composition of the solid support on which capture reagents are attached generally depends on the method of attachment (e.g., covalent attachment). Other exemplary receptacles include microdroplets and microfluidic controlled or bulk oil/aqueous emulsions within which assays and related manipulations can occur. Suitable solid supports include, for

example, plastics, resins, polysaccharides, silica or silica-based materials, functionalized glass, modified silicon, carbon, metals, inorganic glasses, membranes, nylon, natural fibers (such as, for example, silk, wool and cotton), polymers, and the like. The material composing the solid support can include reactive groups such as, for example, carboxy, amino, or hydroxyl groups, which are used for attachment of the capture reagents. Polymeric solid supports can include, e.g., polystyrene, polyethylene glycol tetraphthalate, polyvinyl acetate, polyvinyl chloride, polyvinyl pyrrolidone, polyacrylonitrile, polymethyl methacrylate, polytetrafluoroethylene, butyl rubber, styrenebutadiene rubber, natural rubber, polyethylene, polypropylene, (poly)tetrafluoroethylene, (poly)vinylidenefluoride, polycarbonate, and polymethylpentene. Suitable solid support particles that can be used include, e.g., encoded particles, such as Luminex-type encoded particles, magnetic particles, and glass particles.

[83]

[84] Exemplary Uses of Biomarkers

[85] In various exemplary embodiments, methods are provided for diagnosing NSCLC in a human (individual) by detecting one or more biomarker values corresponding to one or more biomarkers that are present in the circulation of an individual, such as in serum or plasma, by any number of analytical methods, including any of the analytical methods described herein. These biomarkers are, for example, differentially expressed in humans with NSCLC as compared to humans without NSCLC. Detection of the differential expression of a biomarker in an individual can be used, for example, to permit the early diagnosis of NSCLC, to distinguish between a benign and malignant pulmonary nodule (such as, for example, a nodule observed on a computed tomography (CT) scan), to monitor NSCLC recurrence, or for other clinical indications.

[86] Any of the biomarkers described herein may be used in a variety of clinical indications for NSCLC, including any of the following: detection of NSCLC (such as in a high-risk individual or population); characterizing NSCLC (e.g., determining NSCLC type, sub-type, or stage), such as by distinguishing between non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) and/or between adenocarcinoma and squamous cell carcinoma (or otherwise facilitating histopathology); determining whether a lung nodule is a benign nodule or a malignant lung tumor; determining NSCLC prognosis; monitoring NSCLC progression or remission; monitoring for NSCLC recurrence; monitoring metastasis; treatment selection; monitoring response to a therapeutic agent or other treatment; stratification of humans for computed tomography (CT) screening (e.g., identifying those humans at greater risk of NSCLC and thereby most likely to benefit from spiral-CT screening, thus increasing the positive predictive value of CT); combining biomarker testing with ad-

ditional biomedical information, such as smoking history, etc., or with nodule size, morphology, etc. (such as to provide an assay with increased diagnostic performance compared to CT testing or biomarker testing alone); facilitating the diagnosis of a pulmonary nodule as malignant or benign; facilitating clinical decision making once a pulmonary nodule is observed on CT (e.g., ordering repeat CT scans if the nodule is deemed to be low risk, such as if a biomarker-based test is negative, with or without categorization of nodule size, or considering biopsy if the nodule is deemed medium to high risk, such as if a biomarker-based test is positive, with or without categorization of nodule size); and facilitating decisions regarding clinical follow-up (e.g., whether to implement repeat CT scans, fine needle biopsy, nodule resection or thoracotomy after observing a non-calcified nodule on CT). Biomarker testing may improve positive predictive value (PPV) over CT or chest X-ray screening of high risk humans alone. In addition to their utilities in conjunction with CT screening, the biomarkers described herein can also be used in conjunction with any other imaging modalities used for NSCLC, such as chest X-ray, bronchoscopy or fluorescent bronchoscopy, MRI or PET scan. Furthermore, the described biomarkers may also be useful in permitting certain of these uses before indications of NSCLC are detected by imaging modalities or other clinical correlates, or before symptoms appear. It further includes distinguishing humans with indeterminate pulmonary nodules identified with a CT scan or other imaging method, screening of high risk smokers for NSCLC, and diagnosing an individual with NSCLC.

[87] As an example of the manner in which any of the biomarkers described herein can be used to diagnose NSCLC, differential expression of one or more of the described biomarkers in an individual who is not known to have NSCLC may indicate that the individual has NSCLC, thereby enabling detection of NSCLC at an early stage of the disease when treatment is most effective, perhaps before the NSCLC is detected by other means or before symptoms appear. Over-expression of one or more of the biomarkers during the course of NSCLC may be indicative of NSCLC progression, e.g., a NSCLC tumor is growing and/or metastasizing (and thus indicate a poor prognosis), whereas a decrease in the degree to which one or more of the biomarkers is differentially expressed (i.e., in subsequent biomarker tests, the expression level in the individual is moving toward or approaching a "normal" expression level) may be indicative of NSCLC remission, e.g., a NSCLC tumor is shrinking (and thus indicate a good or better prognosis). Similarly, an increase in the degree to which one or more of the biomarkers is differentially expressed (i.e., in subsequent biomarker tests, the expression level in the individual is moving further away from a "normal" expression level) during the course of NSCLC treatment may indicate that the NSCLC is progressing and therefore indicate that the treatment is ineffective, whereas a decrease in

differential expression of one or more of the biomarkers during the course of NSCLC treatment may be indicative of NSCLC remission and therefore indicate that the treatment is working successfully. Additionally, an increase or decrease in the differential expression of one or more of the biomarkers after an individual has apparently been cured of NSCLC may be indicative of NSCLC recurrence. In a situation such as this, for example, the individual can be re-started on therapy (or the therapeutic regimen modified such as to increase dosage amount and/or frequency, if the individual has maintained therapy) at an earlier stage than if the recurrence of NSCLC was not detected until later. Furthermore, a differential expression level of one or more of the biomarkers in an individual may be predictive of the individual's response to a particular therapeutic agent. In monitoring for NSCLC recurrence or progression, changes in the biomarker expression levels may indicate the need for repeat imaging (e.g., repeat CT scanning), such as to determine NSCLC activity or to determine the need for changes in treatment.

[88] Detection of any of the biomarkers described herein may be particularly useful following, or in conjunction with, NSCLC treatment, such as to evaluate the success of the treatment or to monitor NSCLC remission, recurrence, and/or progression (including metastasis) following treatment. NSCLC treatment may include, for example, administration of a therapeutic agent to the individual, performance of surgery (e.g., surgical resection of at least a portion of a NSCLC tumor or removal of NSCLC and surrounding tissue), administration of radiation therapy, or any other type of NSCLC treatment used in the art, and any combination of these treatments. Lung cancer treatment may include, for example, administration of a therapeutic agent to the individual, performance of surgery (e.g., surgical resection of at least a portion of a lung tumor), administration of radiation therapy, or any other type of NSCLC treatment used in the art, and any combination of these treatments. For example, siRNA molecules are synthetic double stranded RNA molecules that inhibit gene expression and may serve as targeted lung cancer therapeutics. For example, any of the biomarkers may be detected at least once after treatment or may be detected multiple times after treatment (such as at periodic intervals), or may be detected both before and after treatment. Differential expression levels of any of the biomarkers in an individual over time may be indicative of NSCLC progression, remission, or recurrence, examples of which include any of the following: an increase or decrease in the expression level of the biomarkers after treatment compared with the expression level of the biomarker before treatment; an increase or decrease in the expression level of the biomarker at a later time point after treatment compared with the expression level of the biomarker at an earlier time point after treatment; and a differential expression level of the biomarker at a single time point after treatment compared with normal

levels of the biomarker.

[89] As a specific example, the biomarker levels for any of the biomarkers described herein can be determined in pre-surgery and post-surgery (e.g., 2-16 weeks after surgery) serum or plasma samples. An increase in the biomarker expression level(s) in the post-surgery sample compared with the pre-surgery sample can indicate progression of NSCLC (e.g., unsuccessful surgery), whereas a decrease in the biomarker expression level(s) in the post-surgery sample compared with the pre-surgery sample can indicate regression of NSCLC (e.g., the surgery successfully removed the lung tumor). Similar analyses of the biomarker levels can be carried out before and after other forms of treatment, such as before and after radiation therapy or administration of a therapeutic agent or cancer vaccine.

[90] In addition to testing biomarker levels as a stand-alone diagnostic test, biomarker levels can also be done in conjunction with determination of SNPs or other genetic lesions or variability that are indicative of increased risk of susceptibility of disease. (See, e.g., Amos et al., Nature Genetics 40, 616-622 (2009)).

[91] In addition to testing biomarker levels as a stand-alone diagnostic test, biomarker levels can also be done in conjunction with radiologic screening, such as CT screening. For example, the biomarkers may facilitate the medical and economic justification for implementing CT screening, such as for screening large asymptomatic populations at risk for NSCLC (e.g., smokers). For example, a "pre-CT" test of biomarker levels could be used to stratify high-risk humans for CT screening, such as for identifying those who are at highest risk for NSCLC based on their biomarker levels and who should be prioritized for CT screening. If a CT test is implemented, biomarker levels (e.g., as determined by an aptamer assay of serum or plasma samples) of one or more biomarkers can be measured and the diagnostic score could be evaluated in conjunction with additional biomedical information (e.g., tumor parameters determined by CT testing) to enhance positive predictive value (PPV) over CT or biomarker testing alone. A "post-CT" aptamer panel for determining biomarker levels can be used to determine the likelihood that a pulmonary nodule observed by CT (or other imaging modality) is malignant or benign.

[92] Detection of any of the biomarkers described herein may be useful for post-CT testing. For example, biomarker testing may eliminate or reduce a significant number of false positive tests over CT alone. Further, biomarker testing may facilitate treatment of patients. By way of example, if a lung nodule is less than 5 mm in size, results of biomarker testing may advance patients from "watch and wait" to biopsy at an earlier time; if a lung nodule is 5-9 mm, biomarker testing may eliminate the use of a biopsy or thoracotomy on false positive scans; and if a lung nodule is larger than 10 mm, biomarker testing may eliminate surgery for a sub-population of these patients

with benign nodules. Eliminating the need for biopsy in some patients based on biomarker testing would be beneficial because there is significant morbidity associated with nodule biopsy and difficulty in obtaining nodule tissue depending on the location of nodule. Similarly, eliminating the need for surgery in some patients, such as those whose nodules are actually benign, would avoid unnecessary risks and costs associated with surgery.

[93] In addition to testing biomarker levels in conjunction with radiologic screening in high risk humans (e.g., assessing biomarker levels in conjunction with size or other characteristics of a lung nodule or mass observed on an imaging scan), information regarding the biomarkers can also be evaluated in conjunction with other types of data, particularly data that indicates an individual's risk for NSCLC (e.g., patient clinical history, occupational exposure history, symptoms, family history of cancer, risk factors such as whether or not the individual was a smoker, and/or status of other biomarkers, etc.). These various data can be assessed by automated methods, such as a computer program/software, which can be embodied in a computer or other apparatus/device.

[94] Any of the described biomarkers may also be used in imaging tests. For example, an imaging agent can be coupled to any of the described biomarkers, which can be used to aid in NSCLC diagnosis, to monitor disease progression/remission or metastasis, to monitor for disease recurrence, or to monitor response to therapy, among other uses.

[95]

[96] Detection and Determination of Biomarkers and Biomarker Values

[97] A biomarker value for the biomarkers described herein can be detected using any of a variety of known analytical methods. In one embodiment, a biomarker value is detected using a capture reagent. As used herein, a "capture agent" or "capture reagent" refers to a molecule that is capable of binding specifically to a biomarker. In various embodiments, the capture reagent can be exposed to the biomarker in solution or can be exposed to the biomarker while the capture reagent is immobilized on a solid support. In other embodiments, the capture reagent contains a feature that is reactive with a secondary feature on a solid support. In these embodiments, the capture reagent can be exposed to the biomarker in solution, and then the feature on the capture reagent can be used in conjunction with the secondary feature on the solid support to immobilize the biomarker on the solid support. The capture reagent is selected based on the type of analysis to be conducted. Capture reagents include but are not limited to aptamers, antibodies, antigens, adnectins, ankyrins, other antibody mimetics and other protein scaffolds, autoantibodies, chimeras, small molecules, an F(ab')2 fragment, a single chain antibody fragment, an Fv fragment, a single chain Fv fragment, a nucleic acid, a lectin, a ligand-binding receptor, affybodies, nanobodies, imprinted polymers, avimers, peptidomimetics, a hormone receptor, a cytokine receptor, and synthetic

receptors, and modifications and fragments of these.

[98] In some embodiments, a biomarker value is detected using a biomarker/capture reagent complex.

[99] In other embodiments, the biomarker value is derived from the biomarker/capture reagent complex and is detected indirectly, such as, for example, as a result of a reaction that is subsequent to the biomarker/capture reagent interaction, but is dependent on the formation of the biomarker/capture reagent complex.

[100] In some embodiments, the biomarker value is detected directly from the biomarker in a biological sample.

[101] In one embodiment, the biomarkers are detected using a multiplexed format that allows for the simultaneous detection of two or more biomarkers in a biological sample. In one embodiment of the multiplexed format, capture reagents are immobilized, directly or indirectly, covalently or non-covalently, in discrete locations on a solid support. In another embodiment, a multiplexed format uses discrete solid supports where each solid support has a unique capture reagent associated with that solid support, such as, for example quantum dots. In another embodiment, an individual device is used for the detection of each one of multiple biomarkers to be detected in a biological sample. Individual devices can be configured to permit each biomarker in the biological sample to be processed simultaneously. For example, a microtiter plate can be used such that each well in the plate is used to uniquely analyze one of multiple biomarkers to be detected in a biological sample.

[102] In one or more of the foregoing embodiments, a fluorescent tag can be used to label a component of the biomarker/capture complex to enable the detection of the biomarker value. In various embodiments, the fluorescent label can be conjugated to a capture reagent specific to any of the biomarkers described herein using known techniques, and the fluorescent label can then be used to detect the corresponding biomarker value. Suitable fluorescent labels include rare earth chelates, fluorescein and its derivatives, rhodamine and its derivatives, dansyl, allophycocyanin, PBXL-3, Qdot 605, Lissamine, phycoerythrin, Texas Red, and other such compounds.

[103] In one embodiment, the fluorescent label is a fluorescent dye molecule. In some embodiments, the fluorescent dye molecule includes at least one substituted indolium ring system in which the substituent on the 3-carbon of the indolium ring contains a chemically reactive group or a conjugated substance. In some embodiments, the dye molecule includes an AlexaFluor molecule, such as, for example, AlexaFluor 488, AlexaFluor 532, AlexaFluor 647, AlexaFluor 680, or AlexaFluor 700. In other embodiments, the dye molecule includes a first type and a second type of dye molecule, such as, e.g., two different AlexaFluor molecules. In other embodiments, the dye molecule includes a first type and a second type of dye molecule, and the two dye

molecules have different emission spectra.

[104] Fluorescence can be measured with a variety of instrumentation compatible with a wide range of assay formats. For example, spectrofluorimeters have been designed to analyze microtiter plates, microscope slides, printed arrays, cuvettes, etc. See Principles of Fluorescence Spectroscopy, by J. R. Lakowicz, Springer Science+Business Media, Inc., 2004. See Bioluminescence & Chemiluminescence: Progress & Current Applications; Philip E. Stanley and Larry J. Kricka editors, World Scientific Publishing Company, January 2002.

[105] In one or more of the foregoing embodiments, a chemiluminescence tag can optionally be used to label a component of the biomarker/capture complex to enable the detection of a biomarker value. Suitable chemiluminescent materials include any of oxalyl chloride, Rodamin 6G, Ru(bipy)32+, TMAE (tetrakis(dimethylamino)ethylene), Pyrogallol (1,2,3-trihydroxibenzene), Lucigenin, peroxyoxalates, Aryl oxalates, Acridinium esters, dioxetanes, and others.

[106] In yet other embodiments, the detection method includes an enzyme/substrate combination that generates a detectable signal that corresponds to the biomarker value. Generally, the enzyme catalyzes a chemical alteration of the chromogenic substrate which can be measured using various techniques, including spectrophotometry, fluorescence, and chemiluminescence. Suitable enzymes include, for example, luciferases, luciferin, malate dehydrogenase, urease, horseradish peroxidase (HRPO), alkaline phosphatase, beta-galactosidase, glucoamylase, lysozyme, glucose oxidase, galactose oxidase, and glucose-6-phosphate dehydrogenase, uricase, xanthine oxidase, lactoperoxidase, microperoxidase, and the like.

[107] In yet other embodiments, the detection method can be a combination of fluorescence, chemiluminescence, radionuclide or enzyme/substrate combinations that generate a measurable signal. Multimodal signaling could have unique and advantageous characteristics in biomarker assay formats.

[108] More specifically, the biomarker values for the biomarkers described herein can be detected using known analytical methods including, singleplex aptamer assays, multiplexed aptamer assays, singleplex or multiplexed immunoassays, mRNA expression profiling, miRNA expression profiling, mass spectrometric analysis, histological/cytological methods, etc. as detailed below.

[109]

[110] Determination of Biomarker Values Using Aptamer-Based Assays

[111] Assays directed to the detection and quantification of physiologically significant molecules in biological samples and other samples are important tools in scientific research and in the health care field. One class of such assays involves the use of a microarray that includes one or more aptamers immobilized on a solid support. The

aptamers are each capable of binding to a target molecule in a highly specific manner and with very high affinity. See, e.g., U.S. Pat. No. 5,475,096 entitled "Nucleic Acid Ligands"; see also, e.g., U.S. Pat. No. 6,242,246, U.S. Pat. No. 6,458,543, and U.S. Pat. No. 6,503,715, each of which is entitled "Nucleic Acid Ligand Diagnostic Biochip". Once the microarray is contacted with a sample, the aptamers bind to their respective target molecules present in the sample and thereby enable a determination of a biomarker value corresponding to a biomarker.

[112] As used herein, an "aptamer" refers to a nucleic acid that has a specific binding affinity for a target molecule. It is recognized that affinity interactions are a matter of degree; however, in this context, the "specific binding affinity" of an aptamer for its target means that the aptamer binds to its target generally with a much higher degree of affinity than it binds to other components in a test sample. An "aptamer" is a set of copies of one type or species of nucleic acid molecule that has a particular nucleotide sequence. An aptamer can include any suitable number of nucleotides, including any number of chemically modified nucleotides. "Aptamers" refers to more than one such set of molecules. Different aptamers can have either the same or different numbers of nucleotides. Aptamers can be DNA or RNA or chemically modified nucleic acids and can be single stranded, double stranded, or contain double stranded regions, and can include higher ordered structures. An aptamer can also be a photoaptamer, where a photoreactive or chemically reactive functional group is included in the aptamer to allow it to be covalently linked to its corresponding target. Any of the aptamer methods disclosed herein can include the use of two or more aptamers that specifically bind the same target molecule. As further described below, an aptamer may include a tag. If an aptamer includes a tag, all copies of the aptamer need not have the same tag. Moreover, if different aptamers each include a tag, these different aptamers can have either the same tag or a different tag.

[113] An aptamer can be identified using any known method, including the SELEX process. Once identified, an aptamer can be prepared or synthesized in accordance with any known method, including chemical synthetic methods and enzymatic synthetic methods.

[114] As used herein, a "SOMAmer" or Slow Off-Rate Modified Aptamer refers to an aptamer having improved off-rate characteristics. SOMAmers can be generated using the improved SELEX methods described in U.S. Publication No. 2009/0004667, entitled "Method for Generating Aptamers with Improved Off-Rates."

[115] The terms "SELEX" and "SELEX process" are used interchangeably herein to refer generally to a combination of (1) the selection of aptamers that interact with a target molecule in a desirable manner, for example binding with high affinity to a protein, with (2) the amplification of those selected nucleic acids. The SELEX process can be

used to identify aptamers with high affinity to a specific target or biomarker.

[116] SELEX generally includes preparing a candidate mixture of nucleic acids, binding of the candidate mixture to the desired target molecule to form an affinity complex, separating the affinity complexes from the unbound candidate nucleic acids, separating and isolating the nucleic acid from the affinity complex, purifying the nucleic acid, and identifying a specific aptamer sequence. The process may include multiple rounds to further refine the affinity of the selected aptamer. The process can include amplification steps at one or more points in the process. See, e.g., U.S. Pat. No. 5,475,096, entitled "Nucleic Acid Ligands". The SELEX process can be used to generate an aptamer that covalently binds its target as well as an aptamer that non-covalently binds its target. See, e.g., U.S. Pat. No. 5,705,337 entitled "Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment: ChemiSELEX."

[117] The SELEX process can be used to identify high-affinity aptamers containing modified nucleotides that confer improved characteristics on the aptamer, such as, for example, improved in vivo stability or improved delivery characteristics. Examples of such modifications include chemical substitutions at the ribose and/or phosphate and/or base positions. SELEX process-identified aptamers containing modified nucleotides are described in U.S. Pat. No. 5,660,985, entitled "High Affinity Nucleic Acid Ligands Containing Modified Nucleotides", which describes oligonucleotides containing nucleotide derivatives chemically modified at the 5'- and 2'-positions of pyrimidines. U.S. Pat. No. 5,580,737, see supra, describes highly specific aptamers containing one or more nucleotides modified with 2'-amino (2'-NH2), 2'-fluoro (2'-F), and/or 2'-O-methyl (2'-OMe). See also, U.S. Patent Application Publication 2009/0098549, entitled "SELEX and PHOTOSELEX", which describes nucleic acid libraries having expanded physical and chemical properties and their use in SELEX and photoSELEX.

[118] SELEX can also be used to identify aptamers that have desirable off-rate characteristics. See U.S. Patent Application Publication 2009/0004667, entitled "Method for Generating Aptamers with Improved Off-Rates", which describes improved SELEX methods for generating aptamers that can bind to target molecules. Methods for producing aptamers and photoaptamers having slower rates of dissociation from their respective target molecules are described. The methods involve contacting the candidate mixture with the target molecule, allowing the formation of nucleic acid-target complexes to occur, and performing a slow off-rate enrichment process wherein nucleic acid-target complexes with fast dissociation rates will dissociate and not reform, while complexes with slow dissociation rates will remain intact. Additionally, the methods include the use of modified nucleotides in the production of candidate nucleic acid mixtures to generate aptamers with improved off-rate performance.

[119] A variation of this assay employs aptamers that include photoreactive functional

groups that enable the aptamers to covalently bind or "photocrosslink" their target molecules. See, e.g., U.S. Pat. No. 6,544,776 entitled "Nucleic Acid Ligand Diagnostic Biochip". These photoreactive aptamers are also referred to as photoaptamers. See, e.g., U.S. Pat. No. 5,763,177, U.S. Pat. No. 6,001,577, and U.S. Pat. No. 6,291,184, each of which is entitled "Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment: Photoselection of Nucleic Acid Ligands and Solution SELEX": see also, e.g., U.S. Pat. No. 6,458,539, entitled "Photoselection of Nucleic Acid Ligands". After the microarray is contacted with the sample and the photoaptamers have had an opportunity to bind to their target molecules, the photoaptamers are photoactivated, and the solid support is washed to remove any non-specifically bound molecules. Harsh wash conditions may be used, since target molecules that are bound to the photoaptamers are generally not removed, due to the covalent bonds created by the photoactivated functional group(s) on the photoaptamers. In this manner, the assay enables the detection of a biomarker value corresponding to a biomarker in the test sample.

[120] In both of these assay formats, the aptamers are immobilized on the solid support prior to being contacted with the sample. Under certain circumstances, however, immobilization of the aptamers prior to contact with the sample may not provide an optimal assay. For example, pre-immobilization of the aptamers may result in inefficient mixing of the aptamers with the target molecules on the surface of the solid support, perhaps leading to lengthy reaction times and, therefore, extended incubation periods to permit efficient binding of the aptamers to their target molecules. Further, when photoaptamers are employed in the assay and depending upon the material utilized as a solid support, the solid support may tend to scatter or absorb the light used to effect the formation of covalent bonds between the photoaptamers and their target molecules. Moreover, depending upon the method employed, detection of target molecules bound to their aptamers can be subject to imprecision, since the surface of the solid support may also be exposed to and affected by any labeling agents that are used. Finally, immobilization of the aptamers on the solid support generally involves an aptamer-preparation step (i.e., the immobilization) prior to exposure of the aptamers to the sample, and this preparation step may affect the activity or functionality of the aptamers.

[121] Aptamer assays that permit an aptamer to capture its target in solution and then employ separation steps that are designed to remove specific components of the aptamer-target mixture prior to detection have also been described (see U.S. Patent Application Publication 2009/0042206, entitled "Multiplexed Analyses of Test Samples"). The described aptamer assay methods enable the detection and quantification of a non-nucleic acid target (e.g., a protein target) in a test sample by detecting and quantifying

a nucleic acid (i.e., an aptamer). The described methods create a nucleic acid surrogate (i.e, the aptamer) for detecting and quantifying a non-nucleic acid target, thus allowing the wide variety of nucleic acid technologies, including amplification, to be applied to a broader range of desired targets, including protein targets.

[122] Aptamers can be constructed to facilitate the separation of the assay components from an aptamer biomarker complex (or photoaptamer biomarker covalent complex) and permit isolation of the aptamer for detection and/or quantification. In one embodiment, these constructs can include a cleavable or releasable element within the aptamer sequence. In other embodiments, additional functionality can be introduced into the aptamer, for example, a labeled or detectable component, a spacer component, or a specific binding tag or immobilization element. For example, the aptamer can include a tag connected to the aptamer via a cleavable moiety, a label, a spacer component separating the label, and the cleavable moiety. In one embodiment, a cleavable element is a photocleavable linker. The photocleavable linker can be attached to a biotin moiety and a spacer section, can include an NHS group for derivatization of amines, and can be used to introduce a biotin group to an aptamer, thereby allowing for the release of the aptamer later in an assay method.

[123] Homogenous assays, done with all assay components in solution, do not require separation of sample and reagents prior to the detection of signal. These methods are rapid and easy to use. These methods generate signal based on a molecular capture or binding reagent that reacts with its specific target.

[124] In one embodiment, a method for signal generation takes advantage of anisotropy signal change due to the interaction of a fluorophore-labeled capture reagent with its specific biomarker target. When the labeled capture reacts with its target, the increased molecular weight causes the rotational motion of the fluorophore attached to the complex to become much slower changing the anisotropy value. By monitoring the anisotropy change, binding events may be used to quantitatively measure the biomarkers in solutions. Other methods include fluorescence polarization assays, molecular beacon methods, time resolved fluorescence quenching, chemiluminescence, fluorescence resonance energy transfer, and the like.

[125] An exemplary solution-based aptamer assay that can be used to detect a biomarker value corresponding to a biomarker in a biological sample includes the following: (a) preparing a mixture by contacting the biological sample with an aptamer that includes a first tag and has a specific affinity for the biomarker, wherein an aptamer affinity complex is formed when the biomarker is present in the sample; (b) exposing the mixture to a first solid support including a first capture element, and allowing the first tag to associate with the first capture element; (c) removing any components of the mixture not associated with the first solid support; (d) attaching a second tag to the

biomarker component of the aptamer affinity complex; (e) releasing the aptamer affinity complex from the first solid support; (f) exposing the released aptamer affinity complex to a second solid support that includes a second capture element and allowing the second tag to associate with the second capture element; (g) removing any non-complexed aptamer from the mixture by partitioning the non-complexed aptamer from the aptamer affinity complex; (h) eluting the aptamer from the solid support; and (i) detecting the biomarker by detecting the aptamer component of the aptamer affinity complex.

[126] Any means known in the art can be used to detect a biomarker value by detecting the aptamer component of an aptamer affinity complex. A number of different detection methods can be used to detect the aptamer component of an affinity complex, such as, for example, hybridization assays, mass spectroscopy, or QPCR. In some embodiments, nucleic acid sequencing methods can be used to detect the aptamer component of an aptamer affinity complex and thereby detect a biomarker value. Briefly, a test sample can be subjected to any kind of nucleic acid sequencing method to identify and quantify the sequence or sequences of one or more aptamers present in the test sample. In some embodiments, the sequence includes the entire aptamer molecule or any portion of the molecule that may be used to uniquely identify the molecule. In other embodiments, the identifying sequencing is a specific sequence added to the aptamer; such sequences are often referred to as "tags," "barcodes," or "zipcodes." In some embodiments, the sequencing method includes enzymatic steps to amplify the aptamer sequence or to convert any kind of nucleic acid, including RNA and DNA that contain chemical modifications to any position, to any other kind of nucleic acid appropriate for sequencing.

[127] In some embodiments, the sequencing method includes one or more cloning steps. In other embodiments the sequencing method includes a direct sequencing method without cloning.

[128] In some embodiments, the sequencing method includes a directed approach with specific primers that target one or more aptamers in the test sample. In other embodiments, the sequencing method includes a shotgun approach that targets all aptamers in the test sample.

[129] In some embodiments, the sequencing method includes enzymatic steps to amplify the molecule targeted for sequencing. In other embodiments, the sequencing method directly sequences single molecules. An exemplary nucleic acid sequencing-based method that can be used to detect a biomarker value corresponding to a biomarker in a biological sample includes the following: (a) converting a mixture of aptamers that contain chemically modified nucleotides to unmodified nucleic acids with an enzymatic step; (b) shotgun sequencing the resulting unmodified nucleic acids with a

massively parallel sequencing platform such as, for example, the 454 Sequencing System (454 Life Sciences/Roche), the Illumina Sequencing System (Illumina), the ABI SOLID Sequencing System (Applied Biosystems), the HeliScope Single Molecule Sequencer (Helicon Biosciences), or the Pacific Biosciences Real Time Single-Molecule Sequencing System (Pacific BioSciences) or the Polonator G Sequencing System (Dover Systems); and (c) identifying and quantifying the aptamers present in the mixture by specific sequence and sequence count.

[130]

[131] Determination of Biomarker Values Using Immunoassays

[132] Immunoassay methods are based on the reaction of an antibody to its corresponding target or analyte and can detect the analyte in a sample depending on the specific assay format. To improve specificity and sensitivity of an assay method based on immuno-reactivity, monoclonal antibodies are often used because of their specific epitope recognition. Polyclonal antibodies have also been successfully used in various immunoassays because of their increased affinity for the target as compared to monoclonal antibodies. Immunoassays have been designed for use with a wide range of biological sample matrices. Immunoassay formats have been designed to provide qualitative, semi-quantitative, and quantitative results.

[133] Quantitative results are generated through the use of a standard curve created with known concentrations of the specific analyte to be detected. The response or signal from an unknown sample is plotted onto the standard curve, and a quantity or value corresponding to the target in the unknown sample is established.

[134] Numerous immunoassay formats have been designed. ELISA or EIA can be quantitative for the detection of an analyte. This method relies on attachment of a label to either the analyte or the antibody and the label component includes, either directly or indirectly, an enzyme. ELISA tests may be formatted for direct, indirect, competitive, or sandwich detection of the analyte. Other methods rely on labels such as, for example, radioisotopes (1125) or fluorescence. Additional techniques include, for example, agglutination, nephelometry, turbidimetry, Western blot, immunoprecipitation, immunocytochemistry, immunohistochemistry, flow cytometry, serology, Luminex assay, and others (see ImmunoAssay: A Practical Guide, edited by Brian Law, published by Taylor & Francis, Ltd., 2005 edition).

[135] Exemplary assay formats include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay, fluorescent, chemiluminescence, and fluorescence resonance energy transfer (FRET) or time resolved-FRET (TR-FRET) immunoassays. Examples of procedures for detecting biomarkers include biomarker immunoprecipitation followed by quantitative methods that allow size and peptide level discrimination, such as gel electrophoresis, capillary electrophoresis, planar electrochromatography, and the like.

[136] Methods of detecting and/or quantifying a detectable label or signal generating material depend on the nature of the label. The products of reactions catalyzed by appropriate enzymes (where the detectable label is an enzyme; see above) can be, without limitation, fluorescent, luminescent, or radioactive or they may absorb visible or ultraviolet light. Examples of detectors suitable for detecting such detectable labels include, without limitation, x-ray film, radioactivity counters, scintillation counters, spectrophotometers, colorimeters, fluorometers, luminometers, and densitometers.

[137] Any of the methods for detection can be performed in any format that allows for any suitable preparation, processing, and analysis of the reactions. This can be, for example, in multi-well assay plates (e.g., 96 wells or 384 wells) or using any suitable array or microarray. Stock solutions for various agents can be made manually or robotically, and all subsequent pipetting, diluting, mixing, distribution, washing, incubating, sample readout, data collection and analysis can be done robotically using commercially available analysis software, robotics, and detection instrumentation capable of detecting a detectable label.

[138]

[139] Determination of Biomarker Values Using Gene Expression Profiling

[140] Measuring mRNA in a biological sample may be used as a surrogate for detection of the level of the corresponding protein in the biological sample. Thus, any of the biomarkers or biomarker panels described herein can also be detected by detecting the appropriate RNA.

[141] mRNA expression levels are measured by reverse transcription quantitative polymerase chain reaction (RT-PCR followed with qPCR). RT-PCR is used to create a cDNA from the mRNA. The cDNA may be used in a qPCR assay to produce fluorescence as the DNA amplification process progresses. By comparison to a standard curve, qPCR can produce an absolute measurement such as number of copies of mRNA per cell. Northern blots, microarrays, Invader assays, and RT-PCR combined with capillary electrophoresis have all been used to measure expression levels of mRNA in a sample. See Gene Expression Profiling: Methods and Protocols, Richard A. Shimkets, editor, Humana Press, 2004.

[142] miRNA molecules are small RNAs that are non-coding but may regulate gene expression. Any of the methods suited to the measurement of mRNA expression levels can also be used for the corresponding miRNA. Recently many laboratories have investigated the use of miRNAs as biomarkers for disease. Many diseases involve widespread transcriptional regulation, and it is not surprising that miRNAs might find a role as biomarkers. The connection between miRNA concentrations and disease is often even less clear than the connections between protein levels and disease, yet the value of miRNA biomarkers might be substantial. Of course, as with any RNA expressed

differentially during disease, the problems facing the development of an in vitro diagnostic product will include the requirement that the miRNAs survive in the diseased cell and are easily extracted for analysis, or that the miRNAs are released into blood or other matrices where they must survive long enough to be measured. Protein biomarkers have similar requirements, although many potential protein biomarkers are secreted intentionally at the site of pathology and function, during disease, in a paracrine fashion. Many potential protein biomarkers are designed to function outside the cells within which those proteins are synthesized.

[143]

[144] Detection of Biomarkers Using In Vivo Molecular Imaging Technologies

[145] Any of the described biomarkers may also be used in molecular imaging tests. For example, an imaging agent can be coupled to any of the described biomarkers, which can be used to aid in NSCLC diagnosis, to monitor disease progression/remission or metastasis, to monitor for disease recurrence, or to monitor response to therapy, among other uses.

[146] In vivo imaging technologies provide non-invasive methods for determining the state of a particular disease in the body of an individual. For example, entire portions of the body, or even the entire body, may be viewed as a three dimensional image, thereby providing valuable information concerning morphology and structures in the body. Such technologies may be combined with the detection of the biomarkers described herein to provide information concerning the cancer status, in particular the NSCLC status, of an individual.

[147] The use of in vivo molecular imaging technologies is expanding due to various advances in technology. These advances include the development of new contrast agents or labels, such as radiolabels and/or fluorescent labels, which can provide strong signals within the body; and the development of powerful new imaging technology, which can detect and analyze these signals from outside the body, with sufficient sensitivity and accuracy to provide useful information. The contrast agent can be visualized in an appropriate imaging system, thereby providing an image of the portion or portions of the body in which the contrast agent is located. The contrast agent may be bound to or associated with a capture reagent, such as an aptamer or an antibody, for example, and/or with a peptide or protein, or an oligonucleotide (for example, for the detection of gene expression), or a complex containing any of these with one or more macromolecules and/or other particulate forms.

[148] The contrast agent may also feature a radioactive atom that is useful in imaging. Suitable radioactive atoms include technetium-99m or iodine-123 for scintigraphic studies. Other readily detectable moieties include, for example, spin labels for magnetic resonance imaging (MRI) such as, for example, iodine-123 again, iodine-

131, indium-111, fluorine-19, carbon-13, nitrogen-15, oxygen-17, gadolinium, manganese or iron. Such labels are well known in the art and could easily be selected by one of ordinary skill in the art.

[149] Standard imaging techniques include but are not limited to magnetic resonance imaging, computed tomography scanning, positron emission tomography (PET), single photon emission computed tomography (SPECT), and the like. For diagnostic in vivo imaging, the type of detection instrument available is a major factor in selecting a given contrast agent, such as a given radionuclide and the particular biomarker that it is used to target (protein, mRNA, and the like). The radionuclide chosen typically has a type of decay that is detectable by a given type of instrument. Also, when selecting a radionuclide for in vivo diagnosis, its half-life should be long enough to enable detection at the time of maximum uptake by the target tissue but short enough that deleterious radiation of the host is minimized.

[150] Exemplary imaging techniques include but are not limited to PET and SPECT, which are imaging techniques in which a radionuclide is synthetically or locally administered to an individual. The subsequent uptake of the radiotracer is measured over time and used to obtain information about the targeted tissue and the biomarker. Because of the high-energy (gamma-ray) emissions of the specific isotopes employed and the sensitivity and sophistication of the instruments used to detect them, the two-dimensional distribution of radioactivity may be inferred from outside of the body.

[151] Commonly used positron-emitting nuclides in PET include, for example, carbon-11, nitrogen-13, oxygen-15, and fluorine-18. Isotopes that decay by electron capture and/or gamma-emission are used in SPECT and include, for example iodine-123 and technetium-99m. An exemplary method for labeling amino acids with technetium-99m is the reduction of pertechnetate ion in the presence of a chelating precursor to form the labile technetium-99m-precursor complex, which, in turn, reacts with the metal binding group of a bifunctionally modified chemotactic peptide to form a technetium-99m-chemotactic peptide conjugate.

[152] Antibodies are frequently used for such in vivo imaging diagnostic methods. The preparation and use of antibodies for in vivo diagnosis is well known in the art. Labeled antibodies which specifically bind any of the biomarkers in Table 2 can be injected into an individual suspected of having a certain type of cancer (e.g., NSCLC), detectable according to the particular biomarker used, for the purpose of diagnosing or evaluating the disease status of the individual. The label used will be selected in accordance with the imaging modality to be used, as previously described. Localization of the label permits determination of the spread of the cancer. The amount of label within an organ or tissue also allows determination of the presence or absence of cancer in that organ or tissue.

[153] Similarly, aptamers may be used for such in vivo imaging diagnostic methods. For example, an aptamer that was used to identify a particular biomarker described in Table 2 (and therefore binds specifically to that particular biomarker) may be appropriately labeled and injected into an individual suspected of having NSCLC, detectable according to the particular biomarker, for the purpose of diagnosing or evaluating the NSCLC status of the individual. The label used will be selected in accordance with the imaging modality to be used, as previously described. Localization of the label permits determination of the spread of the cancer. The amount of label within an organ or tissue also allows determination of the presence or absence of cancer in that organ or tissue. Aptamer-directed imaging agents could have unique and advantageous characteristics relating to tissue penetration, tissue distribution, kinetics, elimination, potency, and selectivity as compared to other imaging agents.

[154] Such techniques may also optionally be performed with labeled oligonucleotides, for example, for detection of gene expression through imaging with antisense oligonucleotides. These methods are used for in situ hybridization, for example, with fluorescent molecules or radionuclides as the label. Other methods for detection of gene expression include, for example, detection of the activity of a reporter gene.

[155] Another general type of imaging technology is optical imaging, in which fluorescent signals within the subject are detected by an optical device that is external to the subject. These signals may be due to actual fluorescence and/or to bioluminescence. Improvements in the sensitivity of optical detection devices have increased the usefulness of optical imaging for in vivo diagnostic assays.

[156] The use of in vivo molecular biomarker imaging is increasing, including for clinical trials, for example, to more rapidly measure clinical efficacy in trials for new cancer therapies and/or to avoid prolonged treatment with a placebo for those diseases, such as multiple sclerosis, in which such prolonged treatment may be considered to be ethically questionable.

[157] For a review of other techniques, see N. Blow, Nature Methods, 6, 465-469, 2009.

[158]

[159] Determination of Biomarker Values Using Histology/Cytology Methods

[160] For evaluation of NSCLC, a variety of tissue samples may be used in histological or cytological methods. Sample selection depends on the primary tumor location and sites of metastases. For example, endo- and trans-bronchial biopsies, fine needle aspirates, cutting needles, and core biopsies can be used for histology. Bronchial washing and brushing, pleural aspiration, pleural fluid, and sputum, can be used for cyotology. While cytological analysis is still used in the diagnosis of NSCLC, histological methods are known to provide better sensitivity for the detection of cancer. Any of the biomarkers identified herein that were shown to be up-regulated in the humans with

NSCLC can be used to stain a histological specimen as an indication of disease.

[161] In one embodiment, one or more capture reagents specific to the corresponding biomarker(s) are used in a cytological evaluation of a lung tissue cell sample and may include one or more of the following: collecting a cell sample, fixing the cell sample, dehydrating, clearing, immobilizing the cell sample on a microscope slide, permeabilizing the cell sample, treating for analyte retrieval, staining, destaining, washing, blocking, and reacting with one or more capture reagent/s in a buffered solution. In another embodiment, the cell sample is produced from a cell block.

[162] In another embodiment, one or more capture reagent(s) specific to the corresponding biomarker(s) are used in a histological evaluation of a lung tissue sample and may include one or more of the following: collecting a tissue specimen, fixing the tissue sample, dehydrating, clearing, immobilizing the tissue sample on a microscope slide, permeabilizing the tissue sample, treating for analyte retrieval, staining, destaining, washing, blocking, rehydrating, and reacting with capture reagent(s) in a buffered solution. In another embodiment, fixing and dehydrating are replaced with freezing.

[163] In another embodiment, the one or more aptamer(s) specific to the corresponding biomarker(s) are reacted with the histological or cytological sample and can serve as the nucleic acid target in a nucleic acid amplification method. Suitable nucleic acid amplification methods include, for example, PCR, q-beta replicase, rolling circle amplification, strand displacement, helicase dependent amplification, loop mediated isothermal amplification, ligase chain reaction, and restriction and circularization aided rolling circle amplification.

[164] In one embodiment, the one or more capture reagent(s) specific to the corresponding biomarkers for use in the histological or cytological evaluation are mixed in a buffered solution that can include any of the following: blocking materials, competitors, detergents, stabilizers, carrier nucleic acid, polyanionic materials, etc.

[165] A "cytology protocol" generally includes sample collection, sample fixation, sample immobilization, and staining. "Cell preparation" can include several processing steps after sample collection, including the use of one or more slow off-rate aptamers for the staining of the prepared cells.

[166] Sample collection can include directly placing the sample in an untreated transport container, placing the sample in a transport container containing some type of media, or placing the sample directly onto a slide (immobilization) without any treatment or fixation.

[167] Sample immobilization can be improved by applying a portion of the collected specimen to a glass slide that is treated with polylysine, gelatin, or a silane. Slides can be prepared by smearing a thin and even layer of cells across the slide. Care is generally taken to minimize mechanical distortion and drying artifacts. Liquid

specimens can be processed in a cell block method. Or, alternatively, liquid specimens can be mixed 1:1 with the fixative solution for about 10 minutes at room temperature.

[168] Cell blocks can be prepared from residual effusions, sputum, urine sediments, gastrointestinal fluids, pulmonary fluids, cell scraping, or fine needle aspirates. Cells are concentrated or packed by centrifugation or membrane filtration. A number of methods for cell block preparation have been developed. Representative procedures include the fixed sediment, bacterial agar, or membrane filtration methods. In the fixed sediment method, the cell sediment is mixed with a fixative like Bouins, picric acid, or buffered formalin and then the mixture is centrifuged to pellet the fixed cells. The supernatant is removed, drying the cell pellet as completely as possible. The pellet is collected and wrapped in lens paper and then placed in a tissue cassette. The tissue cassette is placed in a jar with additional fixative and processed as a tissue sample. Agar method is very similar but the pellet is removed and dried on paper towel and then cut in half. The cut side is placed in a drop of melted agar on a glass slide and then the pellet is covered with agar making sure that no bubbles form in the agar. The agar is allowed to harden and then any excess agar is trimmed away. This is placed in a tissue cassette and the tissue process completed. Alternatively, the pellet may be directly suspended in 2% liquid agar at 65° C. and the sample centrifuged. The agar cell pellet is allowed to solidify for an hour at 4° C. The solid agar may be removed from the centrifuge tube and sliced in half. The agar is wrapped in filter paper and then the tissue cassette. Processing from this point forward is as described above. Centrifugation can be replaced in any these procedures with membrane filtration. Any of these processes may be used to generate a "cell block sample".

[169] Cell blocks can be prepared using specialized resin including Lowicryl resins, LR White, LR Gold, Unicryl, and MonoStep. These resins have low viscosity and can be polymerized at low temperatures and with ultra violet (UV) light. The embedding process relies on progressively cooling the sample during dehydration, transferring the sample to the resin, and polymerizing a block at the final low temperature at the appropriate UV wavelength.

[170] Cell block sections can be stained with hematoxylin-eosin for cytomorphological examination while additional sections are used for examination for specific markers.

[171] Whether the process is cytologoical or histological, the sample may be fixed prior to additional processing to prevent sample degradation. This process is called "fixation" and describes a wide range of materials and procedures that may be used interchangeably. The sample fixation protocol and reagents are best selected empirically based on the targets to be detected and the specific cell/tissue type to be analyzed. Sample fixation relies on reagents such as ethanol, polyethylene glycol, methanol, formalin, or isopropanol. The samples should be fixed as soon after collection and af-

fixation to the slide as possible. However, the fixative selected can introduce structural changes into various molecular targets making their subsequent detection more difficult. The fixation and immobilization processes and their sequence can modify the appearance of the cell and these changes must be anticipated and recognized by the cytotechnologist. Fixatives can cause shrinkage of certain cell types and cause the cytoplasm to appear granular or reticular. Many fixatives function by crosslinking cellular components. This can damage or modify specific epitopes, generate new epitopes, cause molecular associations, and reduce membrane permeability. Formalin fixation is one of the most common cytological/histological approaches. Formalin forms methyl bridges between neighboring proteins or within proteins. Precipitation or coagulation is also used for fixation and ethanol is frequently used in this type of fixation. A combination of crosslinking and precipitation can also be used for fixation. A strong fixation process is best at preserving morphological information while a weaker fixation process is best for the preservation of molecular targets.

[172] A representative fixative is 50% absolute ethanol, 2 mM polyethylene glycol (PEG), 1.85% formaldehyde. Variations on this formulation include ethanol (50% to 95%), methanol (20%-50%), and formalin (formaldehyde) only. Another common fixative is 2% PEG 1500, 50% ethanol, and 3% methanol. Slides are place in the fixative for about 10 to 15 minutes at room temperature and then removed and allowed to dry. Once slides are fixed they can be rinsed with a buffered solution like PBS.

[173] A wide range of dyes can be used to differentially highlight and contrast or "stain" cellular, sub-cellular, and tissue features or morphological structures. Hematoylin is used to stain nuclei a blue or black color. Orange G-6 and Eosin Azure both stain the cell's cytoplasm. Orange G stains keratin and glycogen containing cells yellow. Eosin Y is used to stain nucleoli, cilia, red blood cells, and superficial epithelial squamous cells. Romanowsky stains are used for air dried slides and are useful in enhancing pleomorphism and distinguishing extracellular from intracytoplasmic material.

[174] The staining process can include a treatment to increase the permeability of the cells to the stain. Treatment of the cells with a detergent can be used to increase permeability. To increase cell and tissue permeability, fixed samples can be further treated with solvents, saponins, or non-ionic detergents. Enzymatic digestion can also improve the accessibility of specific targets in a tissue sample.

[175] After staining, the sample is dehydrated using a succession of alcohol rinses with increasing alcohol concentration. The final wash is done with xylene or a xylene substitute, such as a citrus terpene, that has a refractive index close to that of the coverslip to be applied to the slide. This final step is referred to as clearing. Once the sample is dehydrated and cleared, a mounting medium is applied. The mounting medium is selected to have a refractive index close to the glass and is capable of

bonding the coverslip to the slide. It will also inhibit the additional drying, shrinking, or fading of the cell sample.

[176] Regardless of the stains or processing used, the final evaluation of the lung cytological specimen is made by some type of microscopy to permit a visual inspection of the morphology and a determination of the marker's presence or absence. Exemplary microscopic methods include brightfield, phase contrast, fluorescence, and differential interference contrast.

[177] If secondary tests are required on the sample after examination, the coverslip may be removed and the slide destained. Destaining involves using the original solvent systems used in staining the slide originally without the added dye and in a reverse order to the original staining procedure. Destaining may also be completed by soaking the slide in an acid alcohol until the cells are colorless. Once colorless the slides are rinsed well in a water bath and the second staining procedure applied.

[178] In addition, specific molecular differentiation may be possible in conjunction with the cellular morphological analysis through the use of specific molecular reagents such as antibodies or nucleic acid probes or aptamers. This improves the accuracy of diagnostic cytology. Micro-dissection can be used to isolate a subset of cells for additional evaluation, in particular, for genetic evaluation of abnormal chromosomes, gene expression, or mutations.

[179] Preparation of a tissue sample for histological evaluation involves fixation, dehydration, infiltration, embedding, and sectioning. The fixation reagents used in histology are very similar or identical to those used in cytology and have the same issues of preserving morphological features at the expense of molecular ones such as individual proteins. Time can be saved if the tissue sample is not fixed and dehydrated but instead is frozen and then sectioned while frozen. This is a more gentle processing procedure and can preserve more individual markers. However, freezing is not acceptable for long term storage of a tissue sample as subcellular information is lost due to the introduction of ice crystals. Ice in the frozen tissue sample also prevents the sectioning process from producing a very thin slice and thus some microscopic resolution and imaging of subcellular structures can be lost. In addition to formalin fixation, osmium tetroxide is used to fix and stain phospholipids (membranes).

[180] Dehydration of tissues is accomplished with successive washes of increasing alcohol concentration. Clearing employs a material that is miscible with alcohol and the embedding material and involves a stepwise process starting at 50:50 alcohol:clearing reagent and then 100% clearing agent (xylene or xylene substitute). Infiltration involves incubating the tissue with a liquid form of the embedding agent (warm wax, nitrocellulose solution) first at 50:50 embedding agent: clearing agent and the 100% embedding agent. Embedding is completed by placing the tissue in a mold or cassette

and filling with melted embedding agent such as wax, agar, or gelatin. The embedding agent is allowed to harden. The hardened tissue sample may then be sliced into thin section for staining and subsequent examination.

[181] Prior to staining, the tissue section is dewaxed and rehydrated. Xylene is used to dewax the section, one or more changes of xylene may be used, and the tissue is re-hydrated by successive washes in alcohol of decreasing concentration. Prior to dewax, the tissue section may be heat immobilized to a glass slide at about 80° C. for about 20 minutes.

[182] Laser capture micro-dissection allows the isolation of a subset of cells for further analysis from a tissue section.

[183] As in cytology, to enhance the visualization of the microscopic features, the tissue section or slice can be stained with a variety of stains. A large menu of commercially available stains can be used to enhance or identify specific features.

[184] To further increase the interaction of molecular reagents with cytological/histological samples, a number of techniques for "analyte retrieval" have been developed. The first such technique uses high temperature heating of a fixed sample. This method is also referred to as heat-induced epitope retrieval or HIER. A variety of heating techniques have been used, including steam heating, microwaving, autoclaving, water baths, and pressure cooking or a combination of these methods of heating. Analyte retrieval solutions include, for example, water, citrate, and normal saline buffers. The key to analyte retrieval is the time at high temperature but lower temperatures for longer times have also been successfully used. Another key to analyte retrieval is the pH of the heating solution. Low pH has been found to provide the best immunostaining but also gives rise to backgrounds that frequently require the use of a second tissue section as a negative control. The most consistent benefit (increased immunostaining without increase in background) is generally obtained with a high pH solution regardless of the buffer composition. The analyte retrieval process for a specific target is empirically optimized for the target using heat, time, pH, and buffer composition as variables for process optimization. Using the microwave analyte retrieval method allows for sequential staining of different targets with antibody reagents. But the time required to achieve antibody and enzyme complexes between staining steps has also been shown to degrade cell membrane analytes. Microwave heating methods have improved in situ hybridization methods as well.

[185] To initiate the analyte retrieval process, the section is first dewaxed and hydrated. The slide is then placed in 10 mM sodium citrate buffer pH 6.0 in a dish or jar. A representative procedure uses an 1100 W microwave and microwaves the slide at 100% power for 2 minutes followed by microwaving the slides using 20% power for 18 minutes after checking to be sure the slide remains covered in liquid. The slide is then

allowed to cool in the uncovered container and then rinsed with distilled water. HIER may be used in combination with an enzymatic digestion to improve the reactivity of the target to immunochemical reagents.

[186] One such enzymatic digestion protocol uses proteinase K. A 20 g/ml concentration of proteinase K is prepared in 50 mM Tris Base, 1 mM EDTA, 0.5% Triton X-100, pH 8.0 buffer. The process first involves dewaxing sections in 2 changes of xylene, 5 minutes each. Then the sample is hydrated in 2 changes of 100% ethanol for 3 minutes each, 95% and 80% ethanol for 1 minute each, and then rinsed in distilled water. Sections are covered with Proteinase K working solution and incubated 10-20 minutes at 37° C. in humidified chamber (optimal incubation time may vary depending on tissue type and degree of fixation). The sections are cooled at room temperature for 10 minutes and then rinsed in PBS Tween 20 for 2 x 2 min. If desired, sections can be blocked to eliminate potential interference from endogenous compounds and enzymes. The section is then incubated with primary antibody at appropriate dilution in primary antibody dilution buffer for 1 hour at room temperature or overnight at 4° C. The section is then rinsed with PBS Tween 20 for 2 x 2 min. Additional blocking can be performed, if required for the specific application, followed by additional rinsing with PBS Tween 20 for 3 x 2 min and then finally the immunostaining protocol completed.

[187] A simple treatment with 1% SDS at room temperature has also been demonstrated to improve immunohistochemical staining. Analyte retrieval methods have been applied to slide mounted sections as well as free floating sections. Another treatment option is to place the slide in a jar containing citric acid and 0.1 Nonident P40 at pH 6.0 and heating to 95° C. The slide is then washed with a buffer solution like PBS.

[188] For immunological staining of tissues it may be useful to block non-specific association of the antibody with tissue proteins by soaking the section in a protein solution like serum or non-fat dry milk.

[189] Blocking reactions may include the need to reduce the level of endogenous biotin; eliminate endogenous charge effects; inactivate endogenous nucleases; and/or inactivate endogenous enzymes like peroxidase and alkaline phosphatase. Endogenous nucleases may be inactivated by degradation with proteinase K, by heat treatment, use of a chelating agent such as EDTA or EGTA, the introduction of carrier DNA or RNA, treatment with a chaotrope such as urea, thiourea, guanidine hydrochloride, guanidine thiocyanate, lithium perchlorate, etc, or diethyl pyrocarbonate. Alkaline phosphatase may be inactivated by treated with 0.1N HCl for 5 minutes at room temperature or treatment with 1 mM levamisole. Peroxidase activity may be eliminated by treatment with 0.03% hydrogen peroxide. Endogenous biotin may be blocked by soaking the slide or section in an avidin (streptavidin, neutravidin may be substituted) solution for at least 15 minutes at room temperature. The slide or section is then washed for at least

10 minutes in buffer. This may be repeated at least three times. Then the slide or section is soaked in a biotin solution for 10 minutes. This may be repeated at least three times with a fresh biotin solution each time. The buffer wash procedure is repeated. Blocking protocols should be minimized to prevent damaging either the cell or tissue structure or the target or targets of interest but one or more of these protocols could be combined to "block" a slide or section prior to reaction with one or more slow off-rate aptamers. See Basic Medical Histology: the Biology of Cells, Tissues and Organs, authored by Richard G. Kessel, Oxford University Press, 1998.

[190]

[191] Determination of Biomarker Values Using Mass Spectrometry Methods

[192] A variety of configurations of mass spectrometers can be used to detect biomarker values. Several types of mass spectrometers are available or can be produced with various configurations. In general, a mass spectrometer has the following major components: a sample inlet, an ion source, a mass analyzer, a detector, a vacuum system, and instrument-control system, and a data system. Difference in the sample inlet, ion source, and mass analyzer generally define the type of instrument and its capabilities. For example, an inlet can be a capillary-column liquid chromatography source or can be a direct probe or stage such as used in matrix-assisted laser desorption. Common ion sources are, for example, electrospray, including nanospray and microspray or matrix-assisted laser desorption. Common mass analyzers include a quadrupole mass filter, ion trap mass analyzer and time-of-flight mass analyzer. Additional mass spectrometry methods are well known in the art (see Burlingame et al. Anal. Chem. 70:647 R-716R (1998); Kinter and Sherman, New York (2000)).

[193] Protein biomarkers and biomarker values can be detected and measured by any of the following: electrospray ionization mass spectrometry (ESI-MS), ESI-MS/MS, ESIMS/(MS)n, matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS), desorption/ionization on silicon (DIOS), secondary ion mass spectrometry (SIMS), quadrupole time-of-flight (Q-TOF), tandem time-of-flight (TOF/TOF) technology, called ultraflex III TOF/TOF, atmospheric pressure chemical ionization mass spectrometry (APCI-MS), APCI-MS/MS, APCI-(MS)N, atmospheric pressure photoionization mass spectrometry (APPI-MS), APPI-MS/MS, and APPI-(MS)N, quadrupole mass spectrometry, Fourier transform mass spectrometry (FTMS), quantitative mass spectrometry, and ion trap mass spectrometry.

[194] Sample preparation strategies are used to label and enrich samples before mass spectroscopic characterization of protein biomarkers and determination biomarker values. Labeling methods include but are not limited to isobaric tag for relative and absolute quantitation (iTRAQ) and stable isotope labeling with amino acids in cell culture

(SILAC). Capture reagents used to selectively enrich samples for candidate biomarker proteins prior to mass spectroscopic analysis include but are not limited to aptamers, antibodies, nucleic acid probes, chimeras, small molecules, an F(ab')2 fragment, a single chain antibody fragment, an Fv fragment, a single chain Fv fragment, a nucleic acid, a lectin, a ligand-binding receptor, affybodies, nanobodies, ankyrins, domain antibodies, alternative antibody scaffolds (e.g. diabodies etc) imprinted polymers, avimers, peptidomimetics, peptoids, peptide nucleic acids, threose nucleic acid, a hormone receptor, a cytokine receptor, and synthetic receptors, and modifications and fragments of these.

[195]

[196] Determination of Biomarker Values Using a Proximity Ligation Assay

[197] A proximity ligation assay can be used to determine biomarker values. Briefly, a test sample is contacted with a pair of affinity probes that may be a pair of antibodies or a pair of aptamers, with each member of the pair extended with an oligonucleotide. The targets for the pair of affinity probes may be two distinct determinates on one protein or one determinate on each of two different proteins, which may exist as homo- or heteromultimeric complexes. When probes bind to the target determinates, the free ends of the oligonucleotide extensions are brought into sufficiently close proximity to hybridize together. The hybridization of the oligonucleotide extensions is facilitated by a common connector oligonucleotide which serves to bridge together the oligonucleotide extensions when they are positioned in sufficient proximity. Once the oligonucleotide extensions of the probes are hybridized, the ends of the extensions are joined together by enzymatic DNA ligation.

[198] Each oligonucleotide extension comprises a primer site for PCR amplification. Once the oligonucleotide extensions are ligated together, the oligonucleotides form a continuous DNA sequence which, through PCR amplification, reveals information regarding the identity and amount of the target protein, as well as, information regarding protein-protein interactions where the target determinates are on two different proteins. Proximity ligation can provide a highly sensitive and specific assay for real-time protein concentration and interaction information through use of real-time PCR. Probes that do not bind the determinates of interest do not have the corresponding oligonucleotide extensions brought into proximity and no ligation or PCR amplification can proceed, resulting in no signal being produced.

[199] The foregoing assays enable the detection of biomarker values that are useful in methods for diagnosing NSCLC, where the methods comprise detecting, in a biological sample from an individual, at least N biomarker values that each correspond to a biomarker selected from the group consisting of the biomarkers provided in Table 2, wherein a classification, as described in detail below, using the biomarker values

indicates whether the individual has NSCLC. While certain of the described NSCLC biomarkers are useful alone for detecting and diagnosing NSCLC, methods are also described herein for the grouping of multiple subsets of the NSCLC biomarkers that are each useful as a panel of three or more biomarkers. Thus, various embodiments of the instant application provide combinations comprising N biomarkers, wherein N is at least three biomarkers. In other embodiments, N is selected to be any number from 2-59 biomarkers. It will be appreciated that N can be selected to be any number from any of the above described ranges, as well as similar, but higher order, ranges. In accordance with any of the methods described herein, biomarker values can be detected and classified individually or they can be detected and classified collectively, as for example in a multiplex assay format.

[200] In another aspect, methods are provided for detecting an absence of NSCLC, the methods comprising detecting, in a biological sample from an individual, at least N biomarker values that each correspond to a biomarker selected from the group consisting of the biomarkers provided in Table 2, wherein a classification, as described in detail below, of the biomarker values indicates an absence of NSCLC in the individual. While certain of the described NSCLC biomarkers are useful alone for detecting and diagnosing the absence of NSCLC, methods are also described herein for the grouping of multiple subsets of the NSCLC biomarkers that are each useful as a panel of three or more biomarkers. Thus, various embodiments of the instant application provide combinations comprising N biomarkers, wherein N is at least three biomarkers. In other embodiments, N is selected to be any number from 2-59 biomarkers. It will be appreciated that N can be selected to be any number from any of the above described ranges, as well as similar, but higher order, ranges. In accordance with any of the methods described herein, biomarker values can be detected and classified individually or they can be detected and classified collectively, as for example in a multiplex assay format.

[201]

[202] Classification of Biomarkers and Calculation of Disease Scores

[203] A biomarker "signature" for a given diagnostic test contains a set of markers, each marker having different levels in the populations of interest. Different levels, in this context, may refer to different means of the marker levels for the humans in two or more groups, or different variances in the two or more groups, or a combination of both. For the simplest form of a diagnostic test, these markers can be used to assign an unknown sample from an individual into one of two groups, either diseased or not diseased. The assignment of a sample into one of two or more groups is known as classification, and the procedure used to accomplish this assignment is known as a classifier or a classification method. Classification methods may also be referred to as

scoring methods. There are many classification methods that can be used to construct a diagnostic classifier from a set of biomarker values. In general, classification methods are most easily performed using supervised learning techniques where a data set is collected using samples obtained from humans within two (or more, for multiple classification states) distinct groups one wishes to distinguish. Since the class (group or population) to which each sample belongs is known in advance for each sample, the classification method can be trained to give the desired classification response. It is also possible to use unsupervised learning techniques to produce a diagnostic classifier.

[204] Common approaches for developing diagnostic classifiers include decision trees; bagging, boosting, forests and random forests; rule inference based learning; Parzen Windows; linear models; logistic; neural network methods; unsupervised clustering; K-means; hierarchical ascending/descending; semi-supervised learning; prototype methods; nearest neighbor; kernel density estimation; support vector machines; hidden Markov models; Boltzmann Learning; and classifiers may be combined either simply or in ways which minimize particular objective functions. For a review, see, e.g., Pattern Classification, R. O. Duda, et al., editors, John Wiley & Sons, 2nd edition, 2001; see also, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, T. Hastie, et al., editors, Springer Science+Business Media, LLC, 2nd edition, 2009; each of which is incorporated by reference in its entirety.

[205] To produce a classifier using supervised learning techniques, a set of samples called training data are obtained. In the context of diagnostic tests, training data includes samples from the distinct groups (classes) to which unknown samples will later be assigned. For example, samples collected from humans in a control population and humans in a particular disease population can constitute training data to develop a classifier that can classify unknown samples (or, more particularly, the humans from whom the samples were obtained) as either having the disease or being free from the disease. The development of the classifier from the training data is known as training the classifier. Specific details on classifier training depend on the nature of the supervised learning technique. For purposes of illustration, an example of training a naive Bayesian classifier will be described below (see, e.g., Pattern Classification, R. O. Duda, et al., editors, John Wiley & Sons, 2nd edition, 2001; see also, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, T. Hastie, et al., editors, Springer Science+Business Media, LLC, 2nd edition, 2009).

[206] Since typically there are many more potential biomarker values than samples in a training set, care must be used to avoid over-fitting. Over-fitting occurs when a statistical model describes random error or noise instead of the underlying relationship. Over-fitting can be avoided in a variety of way, including, for example, by limiting the number of markers used in developing the classifier, by assuming that the marker

responses are independent of one another, by limiting the complexity of the underlying statistical model employed, and by ensuring that the underlying statistical model conforms to the data.

[207] An illustrative example of the development of a diagnostic test using a set of biomarkers includes the application of a naive Bayes classifier, a simple probabilistic classifier based on Bayes theorem with strict independent treatment of the biomarkers. Each biomarker is described by a class-dependent probability density function (pdf) for the measured RFU values or log RFU (relative fluorescence units) values in each class. The joint pdfs for the set of markers in one class is assumed to be the product of the human class-dependent pdfs for each biomarker. Training a naive Bayes classifier in this context amounts to assigning parameters ("parameterization") to characterize the class dependent pdfs. Any underlying model for the class-dependent pdfs may be used, but the model should generally conform to the data observed in the training set.

[208] Specifically, the class-dependent probability of measuring a value xi for biomarker i in the disease class is written as p(xild) and the overall naive Bayes probability of observing n markers with values {tilde over (x)}=(x1, x2, . . . xn) is written as p({tilde over (x)}ld)=Πi=1np(xild) where the human xis are the measured biomarker levels in RFU or log RFU. The classification assignment for an unknown is facilitated by calculating the probability of being diseased p(dl{tilde over (x)}) having measured {tilde over (x)} compared to the probability of being disease free (control) p(cl{tilde over (x)}) for the same measured values. The ratio of these probabilities is computed from the class-dependent pdfs by application of Bayes theorem, i.e., where p(d) is the prevalence of the disease in the population appropriate to the test. Taking the logarithm of both sides of this ratio and substituting the naive Bayes class-dependent probabilities from above gives

[209]

[210] This form is known as the log likelihood ratio and simply states that the log likelihood of being free of the particular disease versus having the disease and is primarily composed of the sum of human log likelihood ratios of the n human biomarkers. In its simplest form, an unknown sample (or, more particularly, the human from whom the sample was obtained) is classified as being free of the disease if the above ratio is greater than zero and having the disease if the ratio is less than zero.

[211] In one exemplary embodiment, the class-dependent biomarker pdfs p(xilc) and p(xild) are assumed to be normal or log-normal distributions in the measured RFU values xi, i.e. with a similar expression for p(xild) with μd and σd. Parameterization of the model requires estimation of two parameters for each class-dependent pdf, a mean μ and a variance σ2, from the training data. This may be accomplished in a number of ways, including, for example, by maximum likelihood estimates, by least-squares, and

by any other methods known to one skilled in the art. Substituting the normal distributions for μ and σ into the log-likelihood ratio defined above gives the following expression:

[212]

[213] Once a set of μs and σ2s have been defined for each pdf in each class from the training data and the disease prevalence in the population is specified, the Bayes classifier is fully determined and may be used to classify unknown samples with measured values {tilde over (x)}.

[214] The performance of the naive Bayes classifier is dependent upon the number and quality of the biomarkers used to construct and train the classifier. A single biomarker will perform in accordance with its KS-distance (Kolmogorov-Smirnov), as defined in Example 3, below. If a classifier performance metric is defined as the area under the receiver operator characteristic curve (AUC), a perfect classifier will have a score of 1 and a random classifier, on average, will have a score of 0.5. The definition of the KS-distance between two sets A and B of sizes n and m is the value, Dn,m=supx|FA,n(x)FB,m(x)|, which is the largest difference between two empirical cumulative distribution functions (cdf). The empirical cdf for a set A of n observations Xi is defined as, where IXix is the indicator function which is equal to 1 if Xi<x and is otherwise equal to 0. By definition, this value is bounded between 0 and 1, where a KS-distance of 1 indicates that the emperical distributions do not overlap.

[215] The addition of subsequent markers with good KS distances (>0.3, for example) will, in general, improve the classification performance if the subsequently added markers are independent of the first marker. Using the area under the ROC curve (AUC) as a classifier score, it is straightforward to generate many high scoring classifiers with a variation of a greedy algorithm. (A greedy algorithm is any algorithm that follows the problem solving metaheuristic of making the locally optimal choice at each stage with the hope of finding the global optimum.)

[216] All single analyte classifiers are generated from a table of potential biomarkers and added to a list. Next, all possible additions of a second analyte to each of the stored single analyte classifiers is then performed, saving a predetermined number of the best scoring pairs, say, for example, a thousand, on a new list. All possible three marker classifiers are explored using this new list of the best two-marker classifiers, again saving the best thousands of these. This process continues until the score either plateaus or begins to deteriorate as additional markers are added. Those high scoring classifiers that remain after convergence can be evaluated for the desired performance for an intended use. For example, in one diagnostic application, classifiers with a high sensitivity and modest specificity may be more desirable than modest sensitivity and high specificity. In another diagnostic application, classifiers with a high specificity

and a modest sensitivity may be more desirable. The desired level of performance is generally selected based upon a trade-off that must be made between the number of false positives and false negatives that can each be tolerated for the particular diagnostic application. Such trade-offs generally depend on the medical consequences of an error, either false positive or false negative.

[217] Various other techniques are known in the art and may be employed to generate many potential classifiers from a list of biomarkers using a naive Bayes classifier. In one embodiment, what is referred to as a genetic algorithm can be used to combine different markers using the fitness score as defined above. Genetic algorithms are particularly well suited to exploring a large diverse population of potential classifiers. In another embodiment, so-called ant colony optimization can be used to generate sets of classifiers. Other strategies that are known in the art can also be employed, including, for example, other evolutionary strategies as well as simulated annealing and other stochastic search methods. Metaheuristic methods, such as, for example, harmony search may also be employed.

[218]

[219] **Clinical Sample Acquisition**

[220] In order to select biomarker, cases and controls were collected under the study design described in Figure 1 and table 1. Cases were Asians having NSCLC, form stage I to IV. Controls consist with healthy normal population and patient with non-malignant lung nodule. Each sample was draw by venipuncture and serum was gained following general protocol. Each serum was frozen and stored at -80°C or below.

[221]

[222] **Candidate** Biomarker Measurement

[223] To measure candidate markers, aptamer-based multiplex assay was performed as described in Figure 2. Biotin moiety is linked via photocleavable linker to each aptamers. Briefly, modified aptamer mixture was mixed with diluted serum and the equilibrium binding is performed by incubating at 37°C for 3 hours. Then the mixture was transferred to streptavidin coated plate and incubated for 30 minutes for capturing aptamer-protein complex via biotin tag on the aptamers. Series of washing step remove unbound proteins in the sample. Then amine-reactive biotin reagent was incubated for tagging biotin to captured proteins. Then aptamer-protein complex were released by photocleavage reaction under UV irradiation. Supernatant are transferred to fresh streptavidin coated plates, and incubated to capture complex via biotin moiety on proteins. Following series of washing step remove unbound aptamers. Then captured aptamer is released by high pH condition with elution buffer and neutralized. Eluted solution undergo hybridization reaction with detection probe and Luminex-bead conjugated capture probe. Luminex bead mixture is washed and measured by Luminex

200 instrument and data was analyzed by Luminex 3.1 software.

[224]

[225] **Biomarker** Value calibration

[226] To minimize assay variation, 3 point QC samples are measured with samples accordingly. 3 points are high, mid and low level of protein concentration in the detection range of the assay. QC sample consists with serum and spiked protein and formerly tested and nominal value was confirmed. By testing QC samples at each assay, measured protein level in QC samples are compared with nominal value and then calibration factor is generated for each protein. Calibration factor can be denoted as nominal value/testing value. Then assay result was calibrated with calibration factor.

[227]

[228] **Selection of** biomarker for NSCLC

[229] In order to select small set of biomarkers for NSCLC, calibrated MFI values of cases and controls were compared and cumulative distribution plot and ROC curves for each aptamers were plotted. (figure 4a-4n) Each marker tested Area under curves compared with other.

[230]

[231] **Establishment of naive** Bayesian classifier

[232] To make multivariate into single score, naive Bayesian classifier was generated. From the list of biomarkers for NSCLC, 7 markers were selected and a naive bayes classifer was constructed. Bayesian classifier for single protein measurement easily described as follows. The probability of having disease when the protein is measured at x, is P(dlx). And the probability of not having disease when the protein is measured at x, is P(clx). The protein level x could classify as disease if P(dlx)>P(clx). Upon bayes theorem,

[233] Posterior=likelihood x prior/evidence

[234] The class-dependent probability density function, p(xilc) and p(xild), where xi is the log of the calibrated MFI value for biomarker i,, were modeled as log-normal distribution function characterized by mean u and variance s2. The parameters for pdfs of the 7 biomarkers are listed in Table 3 and an example of the raw data along with the model fit to a normal pdf is displayed in Figure 3. The naive bayese classification for such a model is given by the following equation, where p(d) is the prevalence of the disease in the population

[235]

[236] likelihood ratio = P(dlx)/P(clx) = p(d)/(1-p(d) x Pi (p(xld)/p(xlc))

[237]

[238] Log likelihood ratio act as single score to discriminate disease state.

[239]

[240] EXAMPLES

[241] The following examples are provided for illustrative purposes only and are not intended to limit the scope of the application as defined by the appended claims. All examples described herein were carried out using standard techniques, which are well known and routine to those of skill in the art. Routine molecular biology techniques described in the following examples can be carried out as described in standard laboratory manuals, such as Sambrook et al., Molecular Cloning : A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (2001).

[242]

[243] **Example 1**

[244] Sample preparation

[245] In order to prepare the sample, whole blood should be stored at red top tube. After drawing the whole blood, test tube is incubated undisturbed at room temperature to allow the blood to clot for 30 minutes. Then the serum and clot could be separated by centrifuging at 1,000-2,000 x g for 10 minutes in a refrigerated centrifuge. After centrifugation, supernatant serum should be transferred into a clean tube immediately. The samples should be maintained at 2-8°C while handling. If the serum is not analyzed on site, the serum should be transferred into small volume aliquots, stored at -20°C or lower immediately. It is important to avoid freeze-thaw cycles because this is detrimental to many serum components. Samples which are hemolyzed, icteric or lipemic can invalidate certain tests.

[246]

[247] **Example 2**

[248] Sample analysis by multiplex aptamer assay.

[249] This example describes the multiplex aptamer assay used to analyze the samples for the identification of the NSCLC biomarkers set forth in table 2. In this methods, pipette tips were changed for each solution addition.

[250] Also, unless otherwise indicated, most solution transfers and wash additions used the BioTek EL406 washer and dispenser. Method steps manually pipetted used an eight channel P200 Pipetteman (Rainin Instruments, LLC, Oakland, Calif.), unless otherwise indicated. A custom buffer referred to as SB17 was prepared in-house, comprising 40 mM HEPES, 101mM NaCl, 5mM KCl, 5mM $MgCl_2$, 1mM EDTA at pH7.5. All steps were performed at room temperature unless otherwise indicated.

[251]

[252] 1. Preparation of aptamer stock solution

[253] Custom stock aptamer solutions for 2%, 0.1%, and 0.01% serum were prepared at 2x concentration in 1x SB17, 0.05% tween-20.

[254] These solution solutions are stored at -20°C until use. The day of the assay, each aptamer mix was thawed at 37°C for 10 minutes, placed in a boiling water bath for 10 minutes and allowed to cool to 25°C for 20 minutes with vigorous mixing in between each heating step. After heat-cool, 55 ul of each 2x aptamer mix was manually pipetted into a 96-well PCR plate and the plate foil sealed.

[255] 2. Assay sample preparation

[256] Frozen aliquots of 100% serum, stored at -80°C were placed in 25°C water bath for 5 minutes. Thawed sample were placed on ice, gently vortexed and then replaced on ice.

[257] A 4% sample solution (2x final) was prepared by transferring 3ul of sample using a 20 ul 8-channel pipette into 96-well PCR plates, each well containing 72ul of the appropriate sample diluent at 4C (1x SB17 with 0.02% tween 20). After mixing samples with several pipetting, 4 ul of 4% sample solution was transferred into 76 ul of the appropriate sample diluent and mixed resulting 0.2% (2x final) sample solution. Finally 6 ul of the 0.2% sample solution was transferred to 54 ul of sample diluent and mixed resulting 0.02% (2x final) sample solution. After preparing sample dilution, 55 ul of each sample were transferred into a new PCR plate for equilibrium binding.

[258] 3. Equilibrium binding

[259] Three stock of heat-cooled aptamer solution were transferred to appropriate sample dilution with 55 ul of volume. Sample and aptamer solution was mixed by pipetting and covered by foil cover. Then the plate was placed into a 37°C incubator for 3 hours.

[260] 4. Catch 1

[261] After equilibrium binding step, 100 ul of aptamer-sample mixture was transferred into fresh streptavidin coated plate. Streptavidin coated plate was placed on thermomixer and incubated at 37°C with mixing (800 rpm) for 30 minutes. To avoid light, the plate should be covered throughout the catch 1 step.

[262] 5. Automated step 1 : Washing, tagging

[263] After catch1 step, the plate was placed on EL406 washer and dispenser, which had been programmed to perform the following steps: unadsorbed material is removed by aspiration, and wells are washed 4 times with 300 ul of buffer PB1 supplemented with 1mM dextran sulfate and 500 uM biotin. Wells are then washed 3 times with 300 ul of PB1 buffer.

[264] 150 uL of freshly prepared solution of 1 mM NHS-PEG4-biotin in buffer PB1 is added to wells and incubated for 5 minutes with shaking. Liquid is aspirated, and wells were washed 8 times with 300 ul buffer PB1 supplemented with 10mM glycine. 100 Ul of buffer PB1 supplemented with 1 mM dextran sulfate are added.

[265] 6. Kinetic challenge, Photocleavage and catch 2

[266] Plates were removed from EL406 washer and dispenser, and placed on a thermomixer mounted under UV light source (LED UV source) at a distance of 5 cm

for 20 min. The thermomixer was set at 800 rpm and RT. After 5 minutes irradiation, samples were manually transferred to a fresh, washed streptavidin coated plate. Catch 2 step was performed at thermomixer at 800 rpm and RT for 10 minutes.

[267] 7. Automated step 2: Washing, elution

[268] After catch 2 step, the plate was placed on EL406 washer and dispenser, which had been programmed following steps: Liquid is aspirated and wells are washed 8 times with 300 ul of PB1 supplemented with 25% of propylene glycol. Wells are washed 5 times with 300 ul PB1, and the final wash is aspirated. 100 ul of CAPSO elution buffer are added, and aptamers are eluted for 5 minutes with shaking.

[269] 8. Elution and neutralization

[270] Following these automated steps, the plate was then removed from the deck of the plate washer, and 90 ul aliquots of the samples were transferred manually to the wells of PCR plate that contained 10 uL of neutralization buffer

[271] 9. Luminex read out

[272] Microsphere stock solution were vortexed and sonicated for 60 seconds to suspend the microspheres. Suspended microspheres were diluted to 2000 microspheres per reaction in 1.5 x TMAC hybridization solutions and mixed by vortexing and sonication. 33uL per reaction of the bead mixture were transferred in to 96 well pcr plate. 7 ul of 15 nM biotinylated detection oligonucleotide stock in 1x TE buffer were added to each reaction and mixed. 10 ul of neutralized assay sample were added and the plate was sealed with a silicon cap mat seal. The plate was first incubated at 96C for 5 minutes and incubated at 50C without agitation overnight in a thermocycler. A filter plate was prewetted with 0.005x TMAC hybridization solution supplemented with 0.5% BSA. The entire sample volume from the hybridization reaction was transferred to the filter plate. The hybridization plate was rinsed with 75 ul of 0.005x TMAC solution with 0.5%BSA and any remaining material was transferred to the filter plate. Samples were filtered under slow vacuum. The filter plate was washed once with 75uL 0.005x TMAC with 0.5%BSA and the microspheres in the filter plate were re-suspended in the same buffer. The filter plated was protected from light and incubated on thermomixer for 5 minutes at 1000 rpm. The filter plate then washed with 0.005x TMAC solution containing 0.5 % BSA. 075ul of 10 ug/ml SAPE in 0.005x TMAC solution with 0.5% BSA was added to each reaction and incubated on thermomixer at 25°C a 1000rpm for 1 hour. The filter plate was washed twice with 75 ul 0.005x TMAC solution with 0.5% BSA and the microspheres in the filter plate were re-suspended with 75 ul of 0.005x TMAC with 0.5% BSA and analyzed on a Luminex 200 instrument running XPonent 3.0 software. At least 100 microspheres were counted per bead type, under high PMT calibration and doublet discriminator setting of 7500 to 18000.

[273]

[274] **Example 3**

[275] Biomarker identification

[276] To minimize assay variation, 3 point QC samples are measured with samples accordingly. 3 points are high, mid and low level of protein concentration in the detection range of the assay. QC sample consists with serum and spiked protein and formerly tested and nominal value was obtained. By testing QC samples at each assay, measured protein level in QC samples are compared with nominal value and then calibration factor is generated for each protein. Calibration factor can be denoted as nominal value/testing value. Then assay result was calibrated with calibration factor.

[277] In order to select small set of biomarkers for NSCLC, calibrated MFI values of cases and controls were compared and ROC curves for each aptamers were plotted (figure 4a-4n). Each markers tested Area under curves compared with other.

[278]

[279] **Example 4**

[280] Training naive Bayesian classifier for NSCLC

[281] To make multivariate into single score, naive Bayesian classifier was generated. From the list of biomarkers for NSCLC, 7 markers were selected and a naive bayes classifer was constructed. Bayesian classifier for single protein measurement easily described as follows. The probability of having disease when the protein is measured at x, is P(dlx). And the probability of not having disease when the protein is measured at x, is P(clx). The protein level x could classify as disease if P(dlx)>P(clx). Upon bayes theorem,

[282]

[283] Posterior=likelihood x prior/evidence

[284]

[285] The class-dependent probability density function, p(xilc) and p(xild), where xi is the log of the calibrated MFI value for biomarker i,, were modeled as log-normal distribution function characterized by mean u and variance s2. The parameters for pdfs of the 7 biomarkers are listed in Table 3, and an example of the raw data along with the model fit to a normal cdf is displayed in Figure 5a-5c. The naive Bayes classification for such a model is given by the following equation, where p(d) is the prevalence of the disease in the population Likelihood of disease, P(dlx)/P(clx) can be described as below,

[286]

[287] P(dlx)/P(clx) = p(d)/(1-p(d) x Pi (p(xld)/p(xlc))

[288]

[289] Log likelihood (p(dlx)/p(clx)) value act as single score to discriminate disease state.

[290]

[291] Parameters for 7-marker classifier were calculated as shown in table 4. And trained classifier showed disease discriminating performance as shown in Figure 5a-5c and 7.

[292]

[293] **Example 5**

[294] To optimize the minimum number of biomarkers required in the panel without any loss in diagnostic performance, a greedy backward elimination algorithm was used. Briefly, seven 6-marker classifiers were constructed but the performance declined compared to the 7 marker model. (Figure 6a-6g, and table 6) The final 7-marker model was chosen for optimal performance with the least number of proteins to simplify the panel for lung cancer diagnostics.

[295]

[296] **Example 6**

[297] Verification of 7-marker classifier

[298] To verify disease discriminating performance of classifier, untested samples were measured with aptamer-based multiplex assay. Measured value of 7 markers were applied to 7-marker classifier and showed similar ROC curve with train set. (Figure 5a-5c and 7, table 5)

[299]

[300] **Example 7**

[301] Performance comparison with conventional blood test, Cyfra 21-1

[302] To compare the performance of 7-marker classifier, total samples were tested with Cyfra 21-1 Elisa kit. The Enzymun-test Cyfra 21-1 (DRG Instruments GmbH, Germany) is based on enzyme immunoassay technology, following a typical sandwich protocol. The Cyfra 21-1 concentration was calculated from the standard curve. The sensitivity of the assay was 0.15ng/mL. Positive test was set if serum specimen detected greater than 0.15 ng/mL. The ROC curve of Cyfra21-1 was plotted and compared with that of 7-marker classifier (Figure 7a-7b).

[303] The AUC of the classifier was observed as 0.88 for all NSCLC cases and 0.83 for early stage lung cancer patients (stage I, II), while Cyfra 21-1 showed AUC of 0.72 for all cases, and 0.63 for early stage patients. Although the diagnostic performance of verification set was decreased to training set, it has still a good differential value between the lung cancer and control from Cyfra 21-1. (Figure 7a-7b)

[304]

[305] [Table 1]

	Training set			**Verification set**	
	case	control		case	Control
No. of subjects	75	75		25	25
Average age (years)	61.4	58.4		61.16	59.2
Median age (years)	63	58		59	57
Histopathology					
squamous cell carcinoma	21			8	
Adenocarcinoma	53			17	
large cell carcinoma	1				
Stage					
I	21			9	
II	8			4	
III	14			2	
IV	32			10	
Gender					
Male	46	50		19	18
Female	29	25		6	7
Smoking history					
Never smoker	27	28		7	9
Current smoker	22	26		9	7
Former smoker	26	21		9	9
Benign nodule		34			14

[306]

[307] [Table 2]

Protein	**Name**
EGFR1	Epidermal growth factor receptor 1
MMP7	Matrix metallo-proteinase 7
CA6	Carbonic anhydrase 6
PTN	Pleiotrophin
CRDL1	Chordin-like protein 1
IGFBP2	Insulin like growth factor binding protein 2
MMP12	Matrix metallo-proteinase 12
CNDP1	Carnosine dipeptidase
KIT	Stem cell growth factor receptor
SERPINF2	alpha2-antiplasmin
CRP	C-reactive protein
C9	Complement component 9
SELL	L- selectin
SERPINA3	Alpha-1-antichymotrypsin

[308]

[309] [Table 3]

Protein	**AUC** on ROC curve	**Significance**	**Direction**	**Description**
EGFR1	0.69	p<0.0001	Down	Receptor tyrosine kinase,
MMP7	0.61	p=0.02447	Up	Breakdown of extracellular matrix, positive regulation of cell proliferation, collagen catabolism, degrades fibronectin
CA6	0.7	p<0.0001	Down	Reversible hydratation of carbon dioxide, one-carbon metabolism, nitrogen metabolism
KIT	0.56	p=0.1784	Down	Decoy receptor
CRP	0.66	p<0.001	Up	Inflammatory acute phase reactant, immune effector
C9	0.73	p<0.0001	Up	Inflammatory acute phase reactant, pore-forming subunit of cytolytic MAC complex
SERPINA 3	0.66	p<0.001	Up	Serine protease inhibitor, part of acute phase response and tissue homeostasis

[310]

[311] [Table 4]

	Control			NSCLC	
Analyte	Mean (μ_c)	StDev(σ_c)		Mean(μ_d)	StDev(σ_d)
EGFR1	2.89	0.06		2.82	0.10
MMP7	2.60	0.16		2.74	0.31
CA6	2.26	0.16		2.04	0.22
KIT	1.81	0.10		1.77	0.10
CRP	2.86	0.38		3.34	0.68
C9	4.08	0.13		4.21	0.13
SERPINA3	2.84	0.11		2.96	0.17

[312]

[313] [Table 5]

		Positive	Negative			
Case	**Total case**	**75**	**25**	**100**	**75.00%**	**Sensitivity**
	Train	58	17	75	77.30%	
	Verification	17	8	25	68.00%	
	Early stage (I, II)	**26**	**16**	**42**	**61.90%**	
	Train	20	9	29	69.00%	
	Verification	6	7	13	46.20%	
Control	**Total control**	**27**	**73**	**100**	**73.00%**	**Specificity**
	Train	21	54	75	72.00%	
	Verification	6	19	25	76.00%	
	Benign nodule	**4**	**44**	**48**	**91.70%**	
	Train	4	30	34	88.20%	
	Verification	0	14	14	100%	

[314]

[315] [Table 6]

Markers	**Specificity at 75% sensitivity**
EGFR1, MMP7, CA6, KIT, CRP, C9, SERPINA3	91.70%
MMP7, CA6, KIT, CRP, C9, SERPINA3	79.20%
EGFR1, CA6, KIT, CRP, C9, SERPINA3	81.30%
EGFR1, MMP7, KIT, CRP, C9, SERPINA3	62.50%
EGFR1, MMP7, CA6, CRP, C9, SERPINA3	89.60%
EGFR1, MMP7, CA6, KIT, C9, SERPINA3	87.50%
EGFR1, MMP7, CA6, KIT, CRP, SERPINA3	89.60%
EGFR1, MMP7, CA6, KIT, CRP, C9	89.60%

[316]

Claims

[Claim 1] An *in vitro* method for diagnosing non-small cell lung cancer in a human, the method comprising:
providing a biomarker panel comprising N = 6 or 7 biomarker proteins, wherein the biomarker proteins are (i) epidermal growth factor receptor 1 (EGFR1), (ii) matrix metalloproteinase 7 (MMP7), (iii) carbonic anhydrase 6 (CA6), (iv) Complement component C9, and (v) at least two proteins selected from stem cell growth factor receptor (KIT), C-reactive protein (CRP), and alpha1-antichymotrypsin (SERPINA3); and detecting the biomarker proteins, in a biological sample from the human, to give biomarker values that each corresponds to one of the N biomarker proteins in the panel, wherein the non-small cell lung cancer is diagnosed based on the biomarker values, wherein the biological sample is selected from a group consisting of a whole blood, a plasma, and a serum.

[Claim 2] The method of claim 1, wherein the detecting the biomarker values includes performing an in vitro assay.

[Claim 3] The method of claim 2, wherein the in vitro assay includes one capture reagent corresponding to each of the biomarker proteins, and wherein the one capture reagent is an aptamer.

[Claim 4] The method of claim 1, wherein the biological sample is a serum.

[Claim 5] The method of claim 1, wherein the human is an Asian.

[Claim 6] The method of claim 1, wherein the human is a smoker.

[Claim 7] The method of claim 1, wherein the human has a pulmonary malignant nodule.

[Claim 8] The method of claim 1, wherein the N is 6.

[Claim 9] The method of claim 1, wherein the N is 7.

[Claim 10] The method of claim 1, wherein the biomarker values have measured values of Complement component C9, carbonic anhydrase 6 (CA6), C-reactive protein (CRP), epidermal growth factor receptor 1 (EGFR1), matrix metalloproteinase 7 (MMP7), alpha1-antichymotrypsin (SERPINA3), and stem cell growth factor receptor (KIT).

[Claim 11] The method of claim 3, wherein the biomarker values are measured by

detecting the aptamers using any one of (i) real-time PCR (qPCR), (ii) microarray, and (iii) Luminex microbead assay.

[Claim 12] The method of claim 1, wherein the non-small cell lung cancer is diagnosed with a statistical method.

[Claim 13] The method of claim 12, wherein the statistical method is selected from a group comprising a linear discriminant analysis, a logistic regression, a naïve Bayesian classification, a support vector machine, and a random forest.