



US 20120101002A1

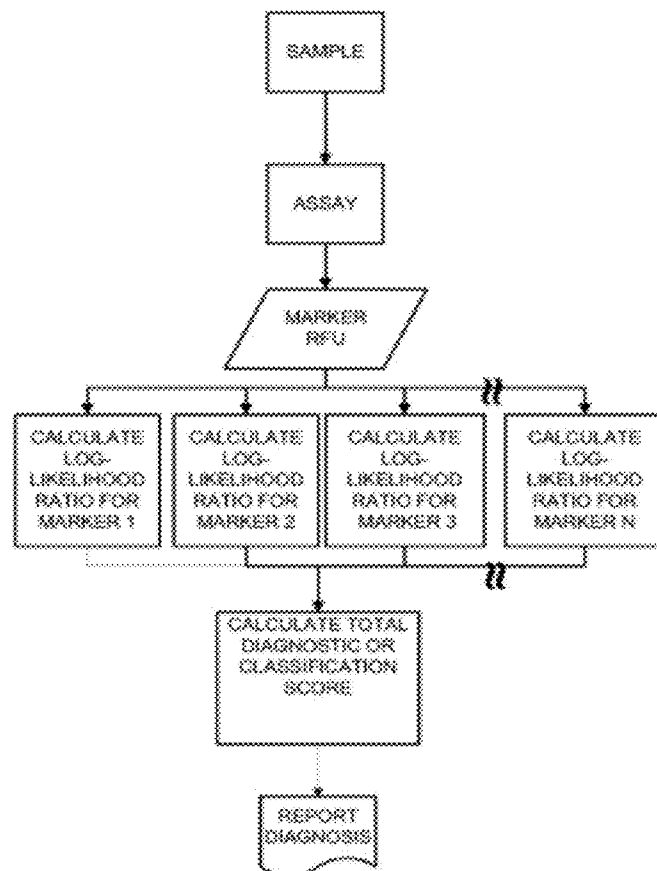
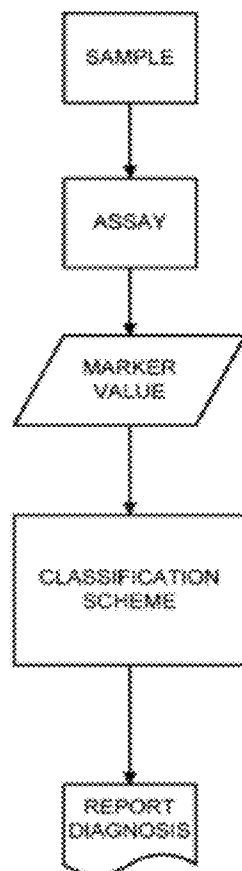
(19) **United States**(12) **Patent Application Publication**
Riel-Mehan et al.(10) **Pub. No.: US 2012/0101002 A1**(43) **Pub. Date: Apr. 26, 2012**(54) **LUNG CANCER BIOMARKERS AND USES THEREOF****Publication Classification**(75) Inventors: **Michael Riel-Mehan**, Louisville, CO (US); **Alex A.E. Stewart**, Waltham, MA (US); **Rachel M. Ostroff**, Westminster, CO (US); **Stephen Alaric Williams**, Boulder, CO (US); **Edward N. Brody**, Boulder, CO (US)(51) **Int. Cl.**
C40B 30/04 (2006.01)
G06F 19/24 (2011.01)
C40B 40/10 (2006.01)(52) **U.S. Cl. 506/9; 506/18; 702/19**(73) Assignee: **SomaLogic, Inc.**, Boulder, CO (US)(21) Appl. No.: **13/279,990**(22) Filed: **Oct. 24, 2011**(57) **ABSTRACT**

The present disclosure includes biomarkers, methods, devices, reagents, systems, and kits for the detection and diagnosis of non-small cell lung cancer and general cancer. In one aspect, methods are provided for diagnosing non-small cell lung cancer, where the methods include detecting, in a sample, at least one biomarker value corresponding to at least one biomarker selected from the biomarkers provided in Table 1, wherein an individual is classified as having lung cancer, or the likelihood of having lung cancer is determined, based on the at least one biomarker value. In another aspect, methods are provided for diagnosing cancer, where the methods include detecting, in a sample, at least one biomarker value corresponding to at least one biomarker selected from the biomarkers provided in Table 19, wherein an individual is classified as having cancer, or the likelihood of having cancer is determined, based on the at least one biomarker value.

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/556,480, filed on Sep. 9, 2009, Continuation-in-part of application No. PCT/US2011/043595, filed on Jul. 11, 2011.

(60) Provisional application No. 61/095,593, filed on Sep. 9, 2008, provisional application No. 61/152,837, filed on Feb. 16, 2009, provisional application No. 61/363,122, filed on Jul. 9, 2010, provisional application No. 61/444,947, filed on Feb. 21, 2011.



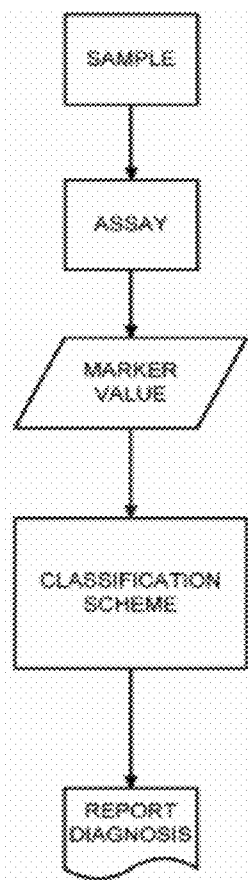


FIG. 1A

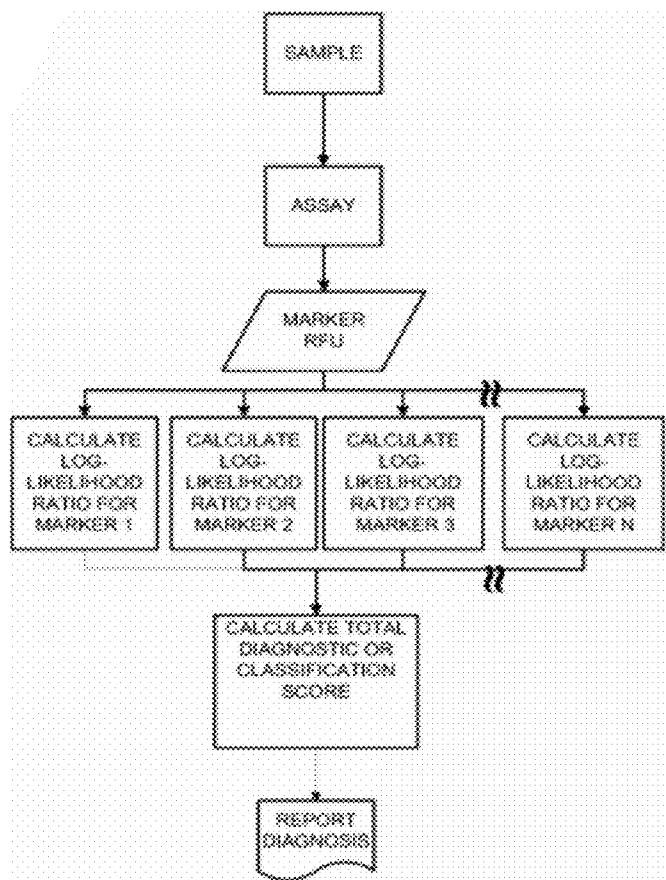
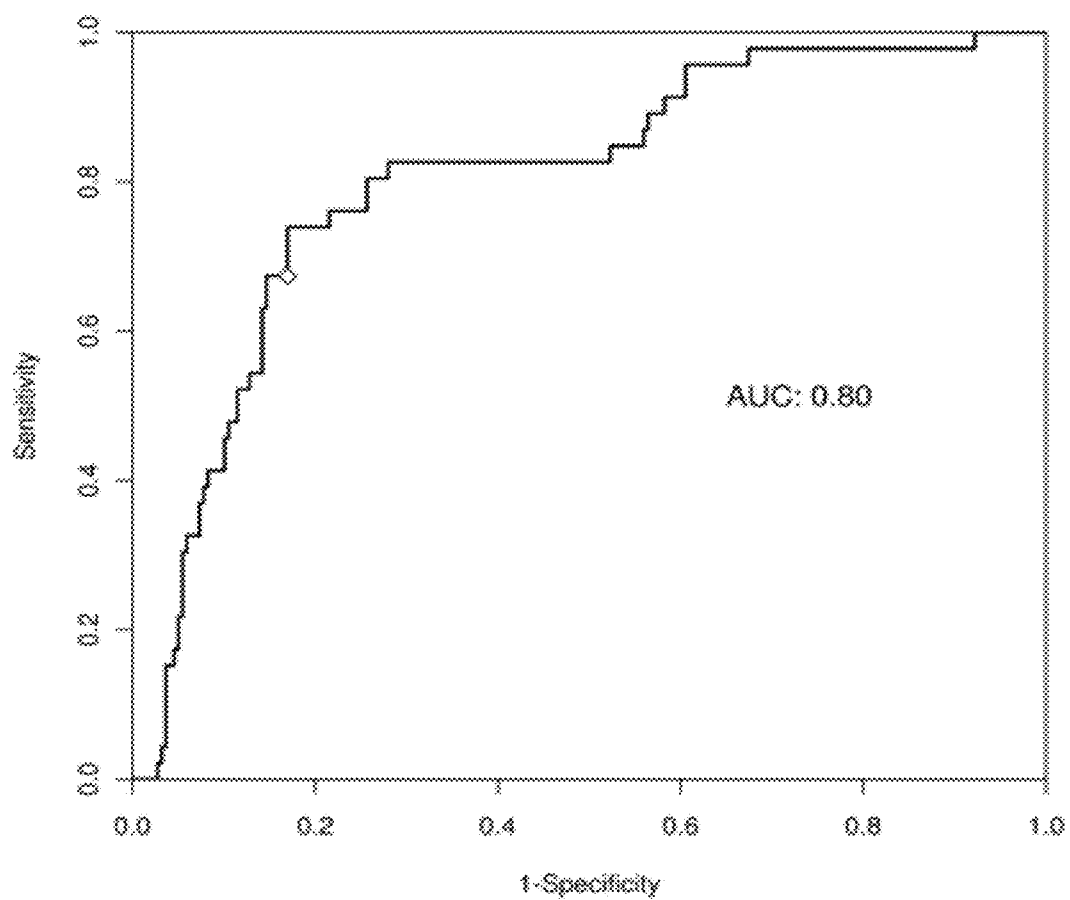


FIG. 1B

**FIG. 2**

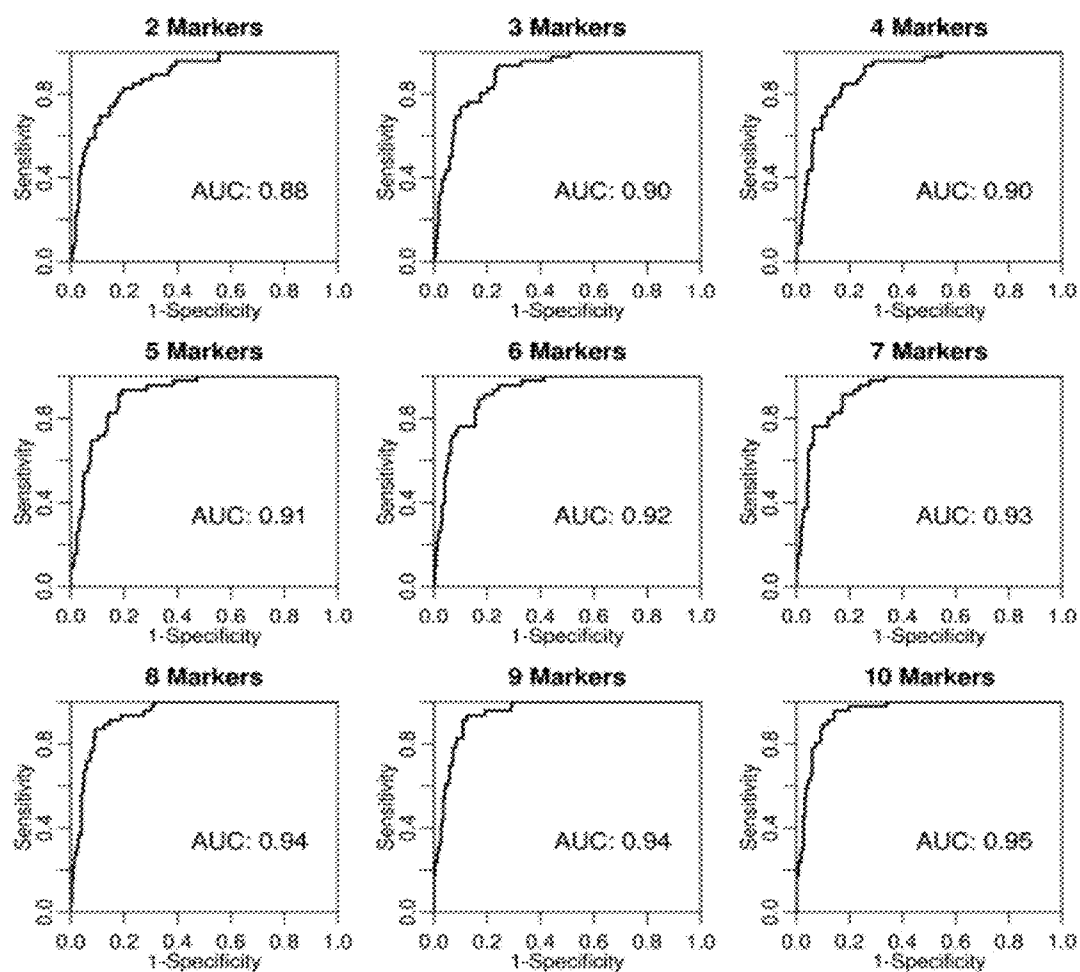


FIG. 3

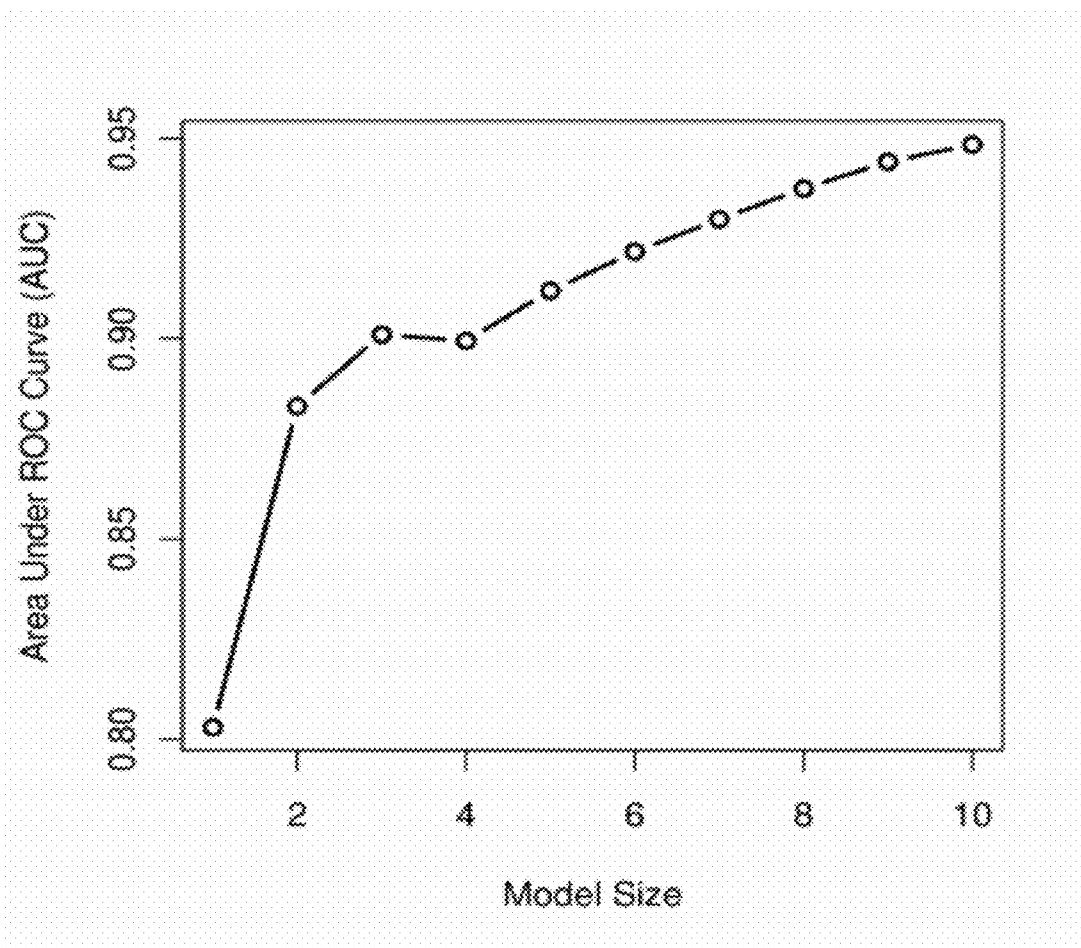


FIG. 4

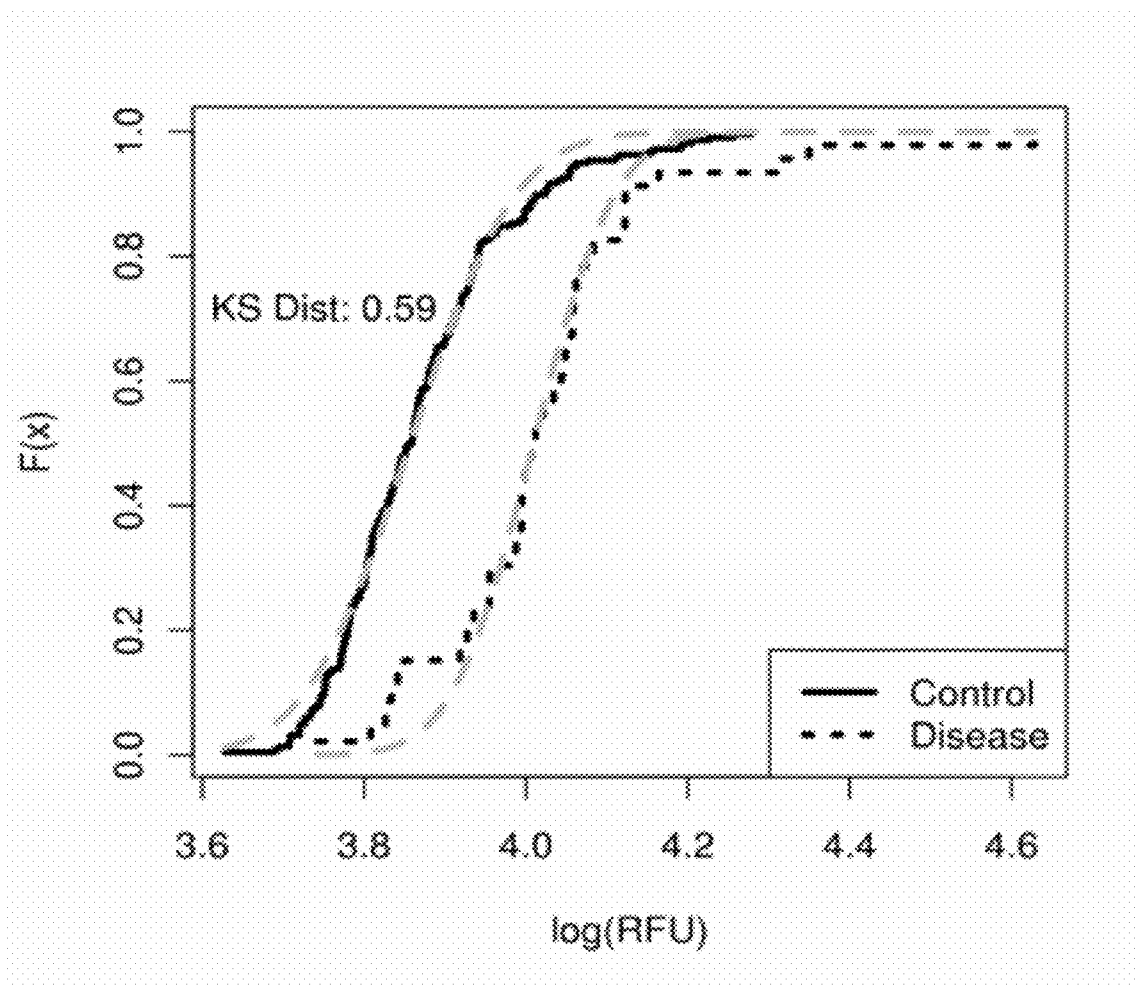


FIG. 5

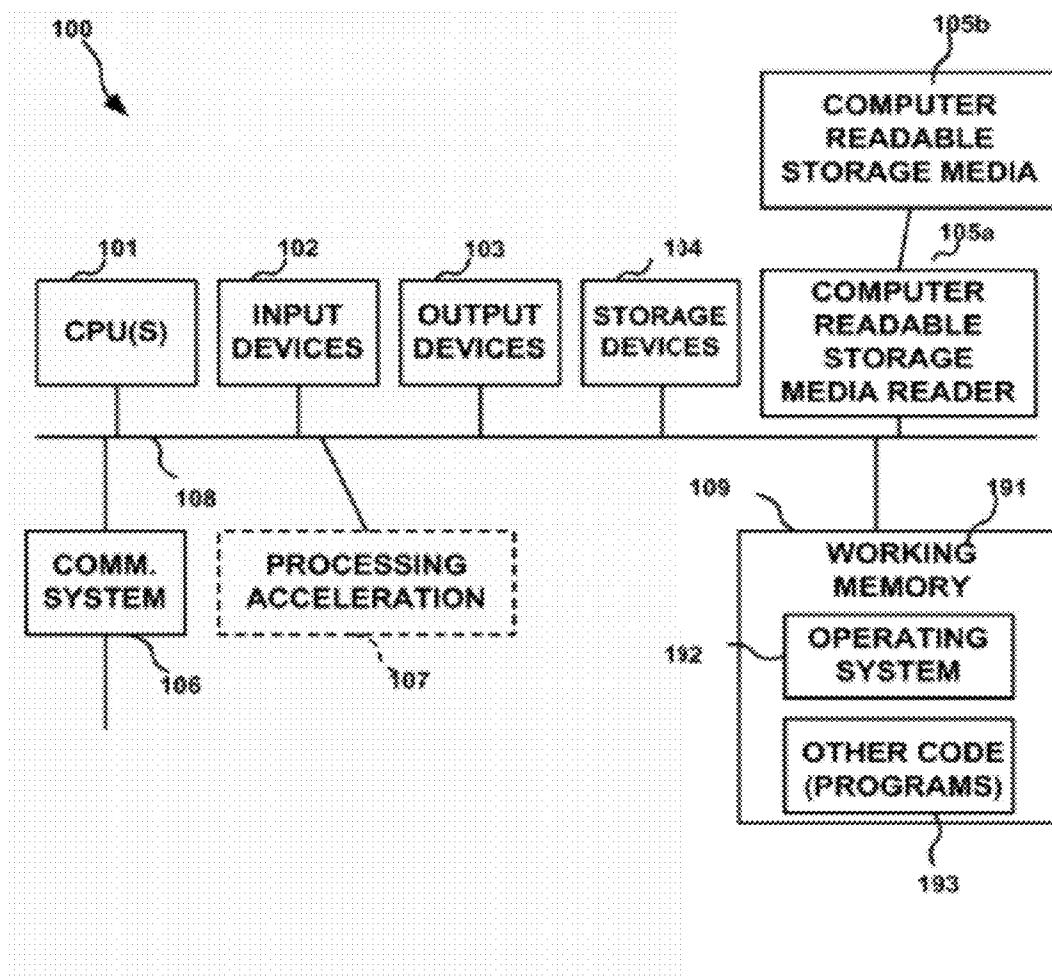


FIG. 6

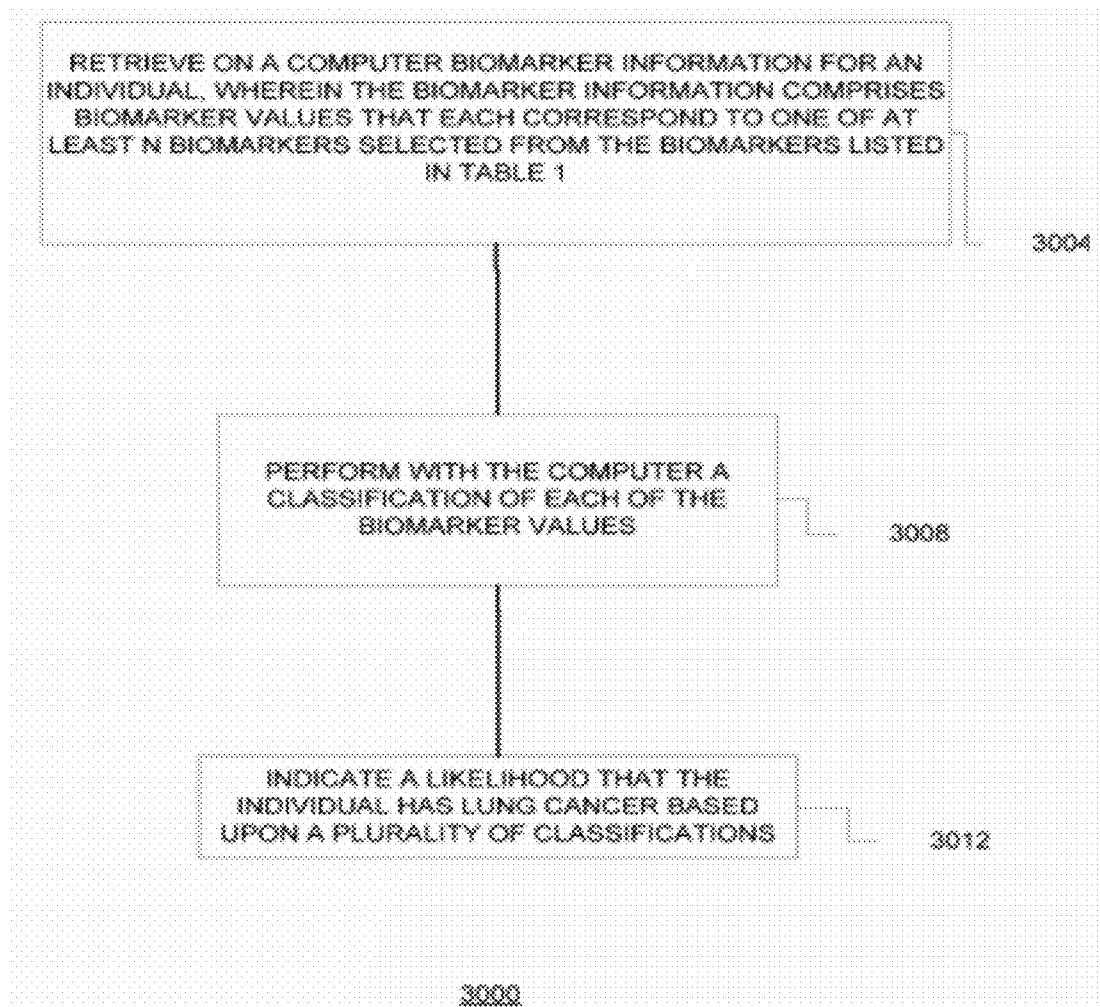


FIG. 7

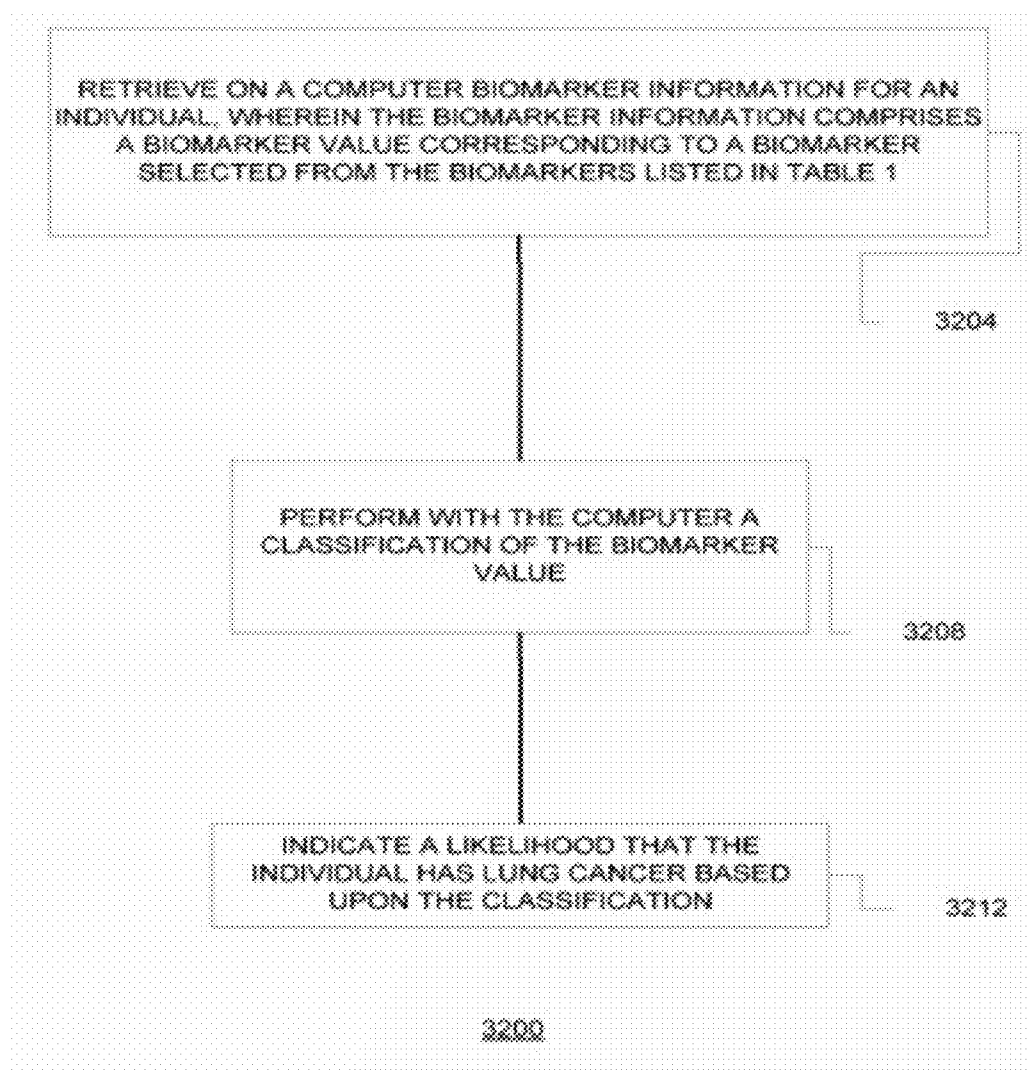


FIG. 8

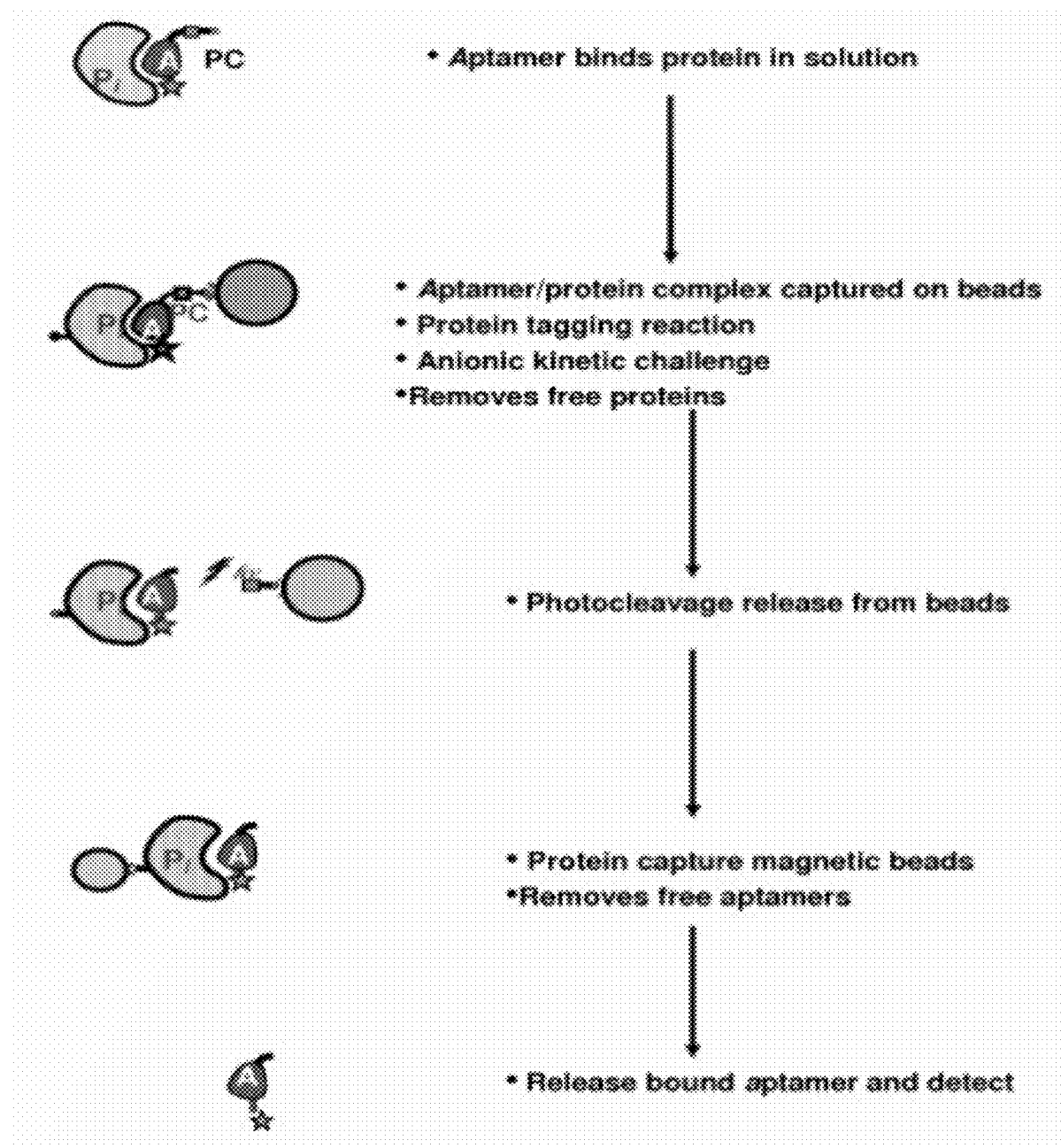


FIG. 9

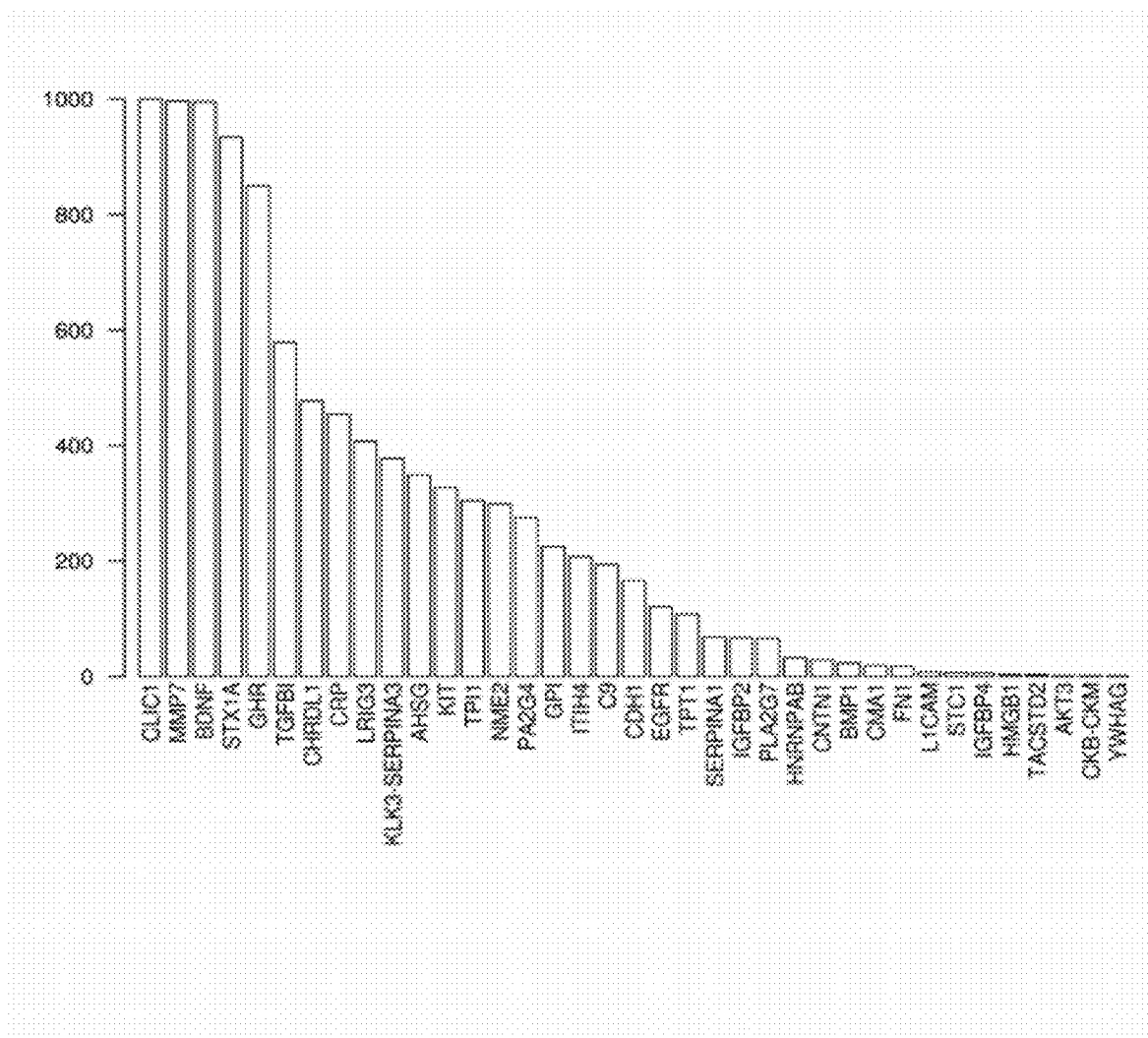


FIG. 10

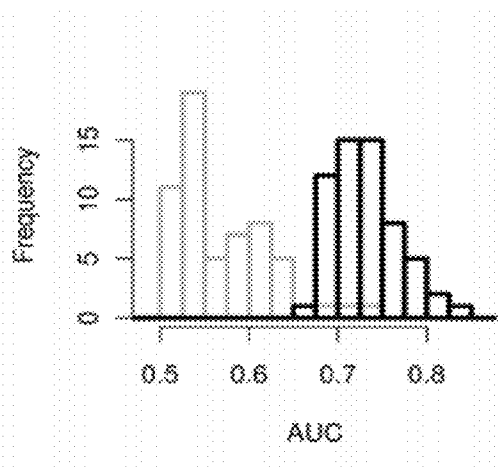


FIG. 11A

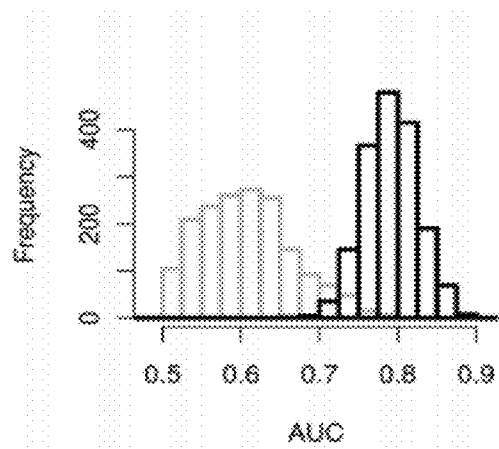


FIG. 11B

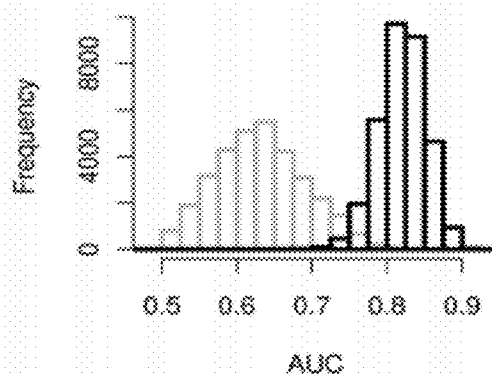


FIG. 11C

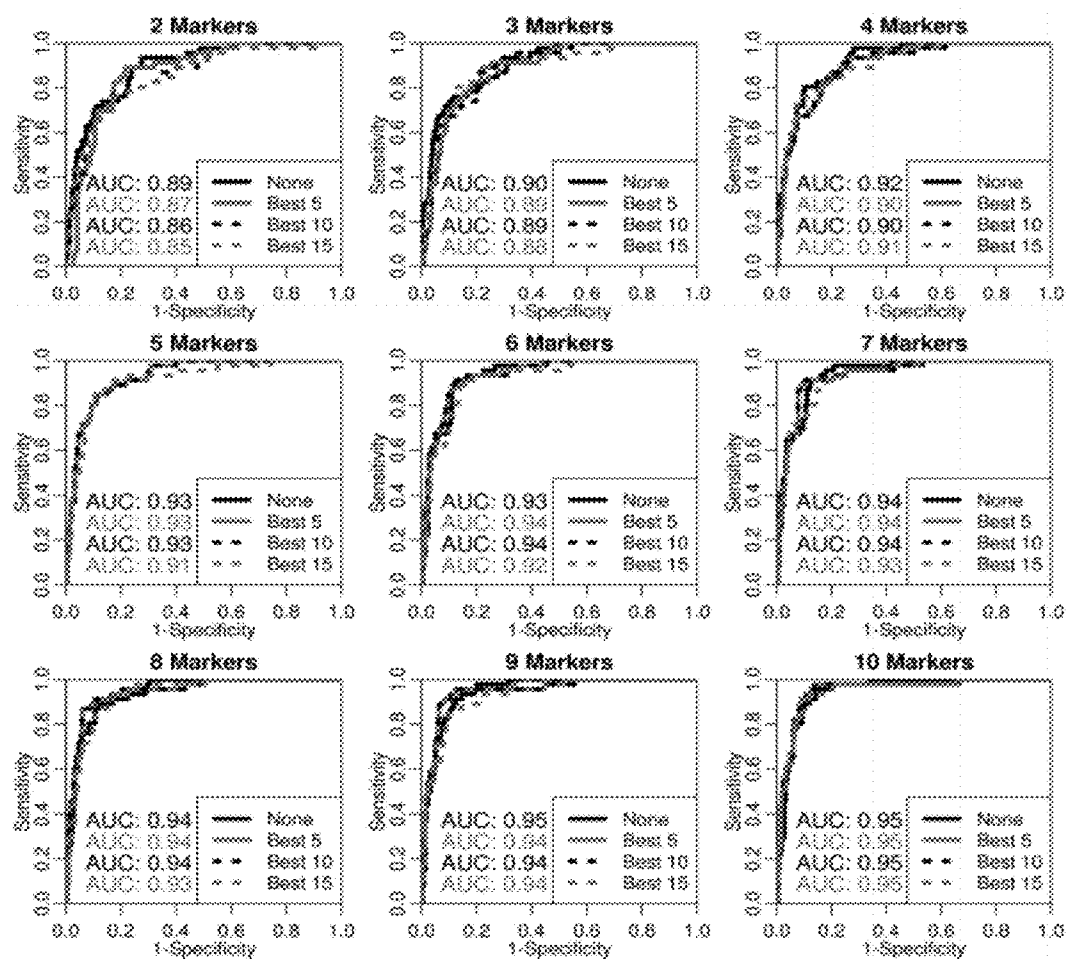


FIG. 12

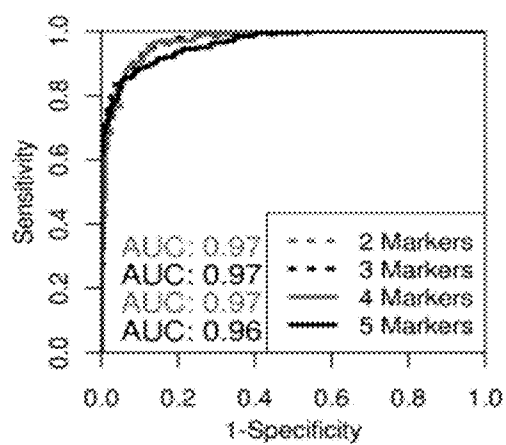


FIG. 13A

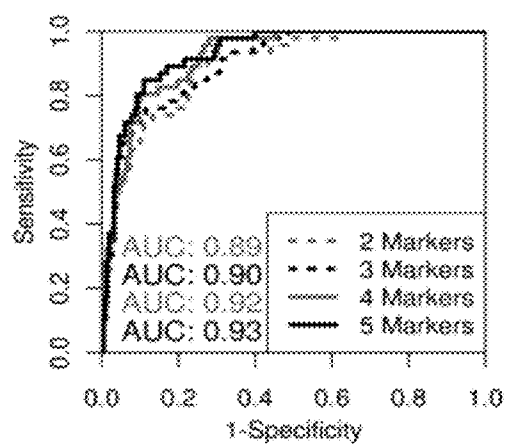


FIG. 13B

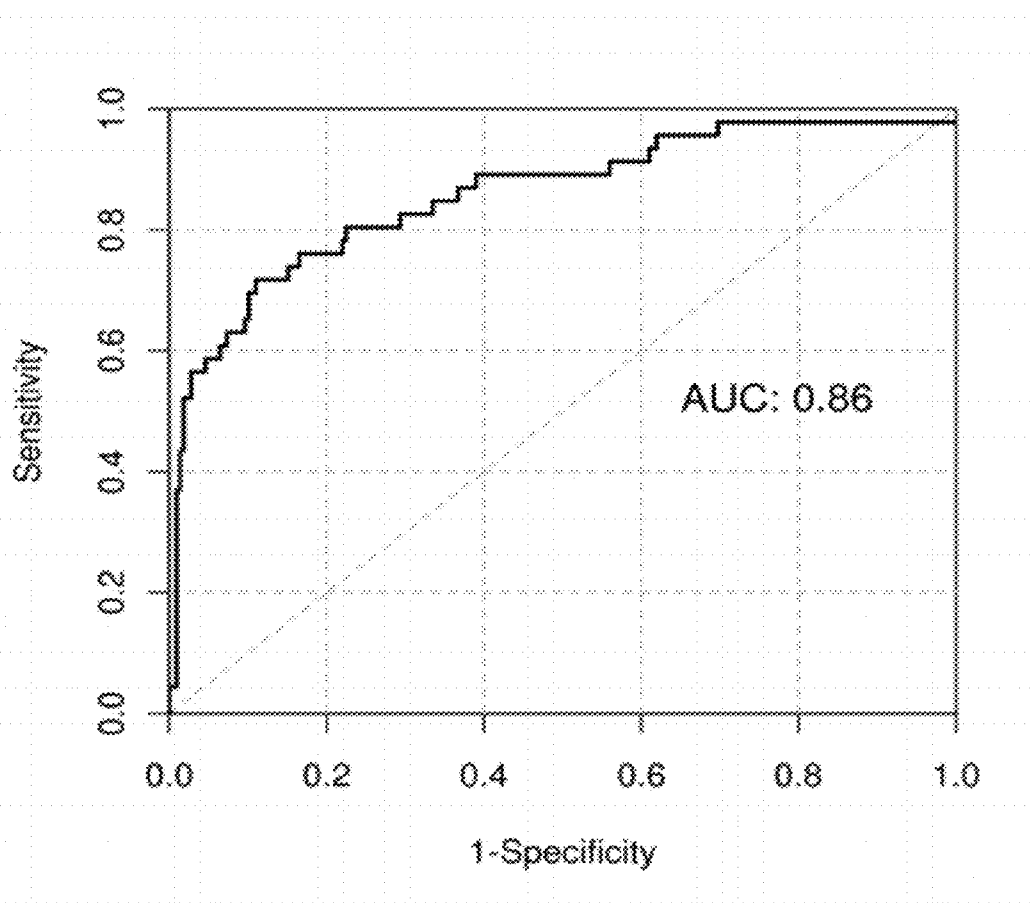


FIG. 14

15/16

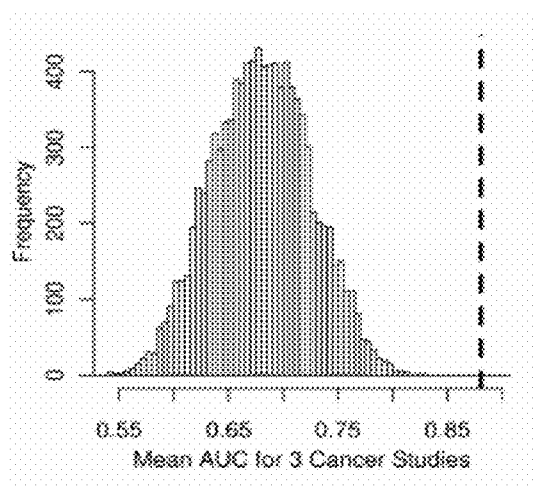


FIG. 15A

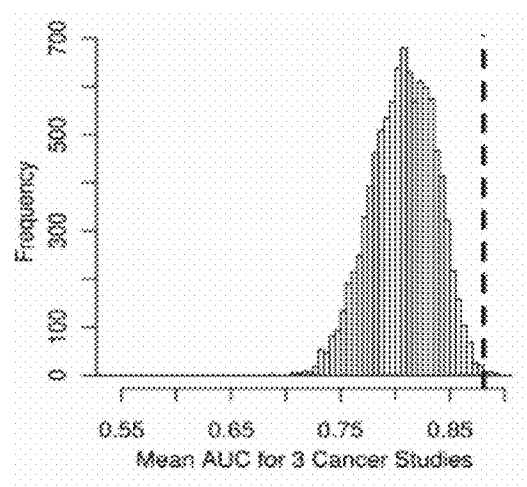


FIG. 15B

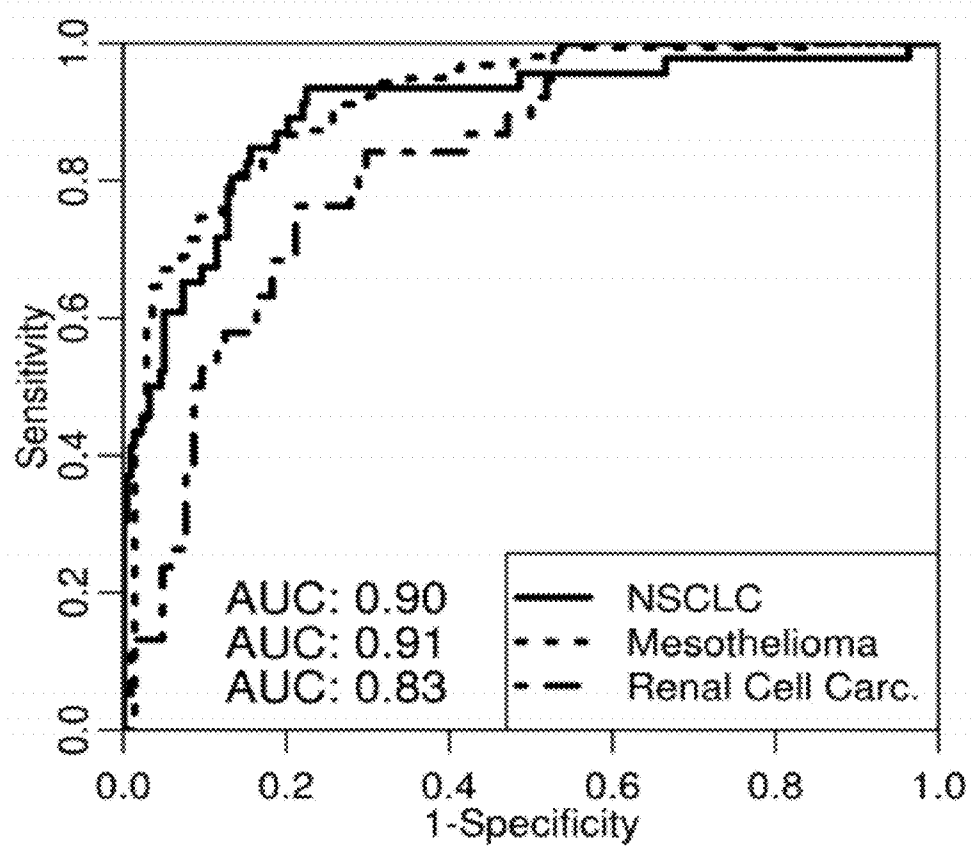


FIG. 16

LUNG CANCER BIOMARKERS AND USES THEREOF

RELATED APPLICATIONS

[0001] This application is a continuation in part of U.S. application Ser. No. 12/556,480, filed Sep. 9, 2009. This application also claims the benefit of U.S. Provisional Application Ser. No. 61/095,593, filed Sep. 9, 2008 and U.S. Provisional Application Ser. No. 61/152,837, filed Feb. 16, 2009. This application is also a continuation in part of International Application Serial No. PCT/US2011/043595, filed Jul. 11, 2011, which claims the benefit of U.S. Provisional Application Ser. No. 61/363,122, filed Jul. 9, 2010 and U.S. Provisional Application Ser. No. 61/444,947, filed Feb. 21, 2011. Each of these applications is incorporated herein by reference in its entirety for all purposes.

FIELD OF THE INVENTION

[0002] The present application relates generally to the detection of biomarkers and the diagnosis of cancer in an individual and, more specifically, to one or more biomarkers, methods, devices, reagents, systems, and kits for diagnosing cancer, more particularly lung cancer, in an individual.

BACKGROUND

[0003] 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 referenced herein is prior art to the present application.

[0004] More people die from lung cancer than any other type of cancer. This is true for both men and women. Lung cancer accounts for more deaths than breast cancer, prostate cancer, and colon cancer combined. Lung cancer accounted for an estimated 157,300 deaths, or 28% of all cancer deaths in the United States in 2010. It is estimated that in 2010, 116,750 men and 105,770 women will be diagnosed with lung cancer, and 86,220 men and 71,080 women will die from lung cancer (Jemal, C A Cancer J Clin 2010; 60:277). Among men in the United States, lung cancer is the second most common cancer among white, black, Asian/Pacific Islander, American Indian/Alaska Native, and Hispanic men. Among women in the United States, lung cancer is the second most common cancer among white, black, and American Indian/Alaska Native women, and the third most common cancer among Asian/Pacific Islander and Hispanic women. For those who do not quit smoking, the probability of death from lung cancer is 15% and remains above 5% even for those who quit at age 50-59. The annual healthcare cost of lung cancer in the U.S. alone is \$95 billion.

[0005] Ninety-one percent of lung cancer caused by smoking is non-small cell lung cancer (NSCLC), which represents approximately 85% of all lung cancers. The remaining 15% of all lung cancers are small cell lung cancers, although mixed-cell lung cancers do occur. Because small cell lung cancer is rare and rapidly fatal, the opportunity for early detection is small.

[0006] There are three main types of NSCLC: squamous cell carcinoma, large cell carcinoma, and adenocarcinoma. Adenocarcinoma is the most common form of lung cancer (30%-65%) and is the lung cancer most frequently found in both smokers and non-smokers. Squamous cell carcinoma accounts for 25-30% of all lung cancers and is generally found in a proximal bronchus. Early stage NSCLC tends to be

localized, and if detected early it can often be treated by surgery with a favorable outcome and improved survival. Other treatment options include radiation treatment, drug therapy, and a combination of these methods.

[0007] NSCLC is staged by the size of the tumor and its presence in other tissues including lymph nodes. In the occult stage, cancer cells may be found in sputum samples or lavage samples and no tumor is detectable in the lungs. In stage 0, only the innermost lining of the lungs exhibit cancer cells and the tumor has not grown through the lining. In stage IA, the cancer is considered locally invasive and has grown deep into the lung tissue but the tumor is less than 3 cm across. In this stage, the tumor is not found in the main bronchus or lymph nodes. In stage IB, the tumor is either larger than 3 cm across or has grown into the bronchus or pleura, but has not grown into the lymph nodes. In stage IIA, the tumor is less than 7 cm across and may have grown into the lymph nodes. In stage IIB, the tumor has either been found in the lymph nodes and is greater than 5 cm across or grown into the bronchus or pleura; or the cancer is not in the lymph nodes but is found in the chest wall, diaphragm, pleura, bronchus, or tissue that surrounds the heart, or separate tumor nodules are present in the same lobe of the lung. In stage IIIA, cancer cells are found in the lymph nodes near the lung and bronchi and in those between the lungs but on the side of the chest where the tumor is located. Stage IIIB, cancer cells are located on the opposite side of the chest from the tumor or in the neck. Other organs near the lungs may also have cancer cells and multiple tumors may be found in one lobe of the lungs. In stage IV, tumors are found in more than one lobe of the same lung or both lungs and cancer cells are found in other parts of the body.

[0008] Current methods of diagnosis for lung cancer include testing sputum for cancerous cells, chest x-ray, fiber optic evaluation and biopsy of airways, and low dose spiral computed tomography (CT). Sputum cytology has a very low sensitivity. Chest X-ray is also relatively insensitive, requiring lesions to be greater than 1 cm in size to be visible. Bronchoscopy requires that the tumor is visible inside airways accessible to the bronchoscope. The most widely recognized diagnostic method is low dose chest CT, but in common with X-ray, the use of CT involves ionizing radiation, which itself can cause cancer. CT also has significant limitations: the scans require a high level of technical skill to interpret and many of the observed abnormalities are not in fact lung cancer and substantial healthcare costs are incurred in following up CT findings. The most common incidental finding is a benign lung nodule.

[0009] Lung nodules are relatively round lesions, or areas of abnormal tissue, located within the lung and may vary in size. Lung nodules may be benign or cancerous, but most are benign. If a nodule is below 4 mm the prevalence is only 1.5%, if 4-8 mm the prevalence is approximately 6%, and if above 20 mm the incidence is approximately 20%. For small and medium-sized nodules, the patient is advised to undergo a repeat scan within three months to a year. For many large nodules, the patient receives a biopsy (which is invasive and may lead to complications) even though most of these are benign.

[0010] Therefore, diagnostic methods that can replace or complement CT are needed to reduce the number of surgical procedures conducted and minimize the risk of surgical complications. In addition, even when lung nodules are absent or unknown, methods are needed to detect lung cancer at its early stages to improve patient outcomes. Only 16% of lung

cancer cases are diagnosed as localized, early stage cancer, where the 5-year survival rate is 46%, compared to 84% of those diagnosed at late stage, where the 5-year survival rate is only 13%. This demonstrates that relying on symptoms for diagnosis is not useful because many of them are common to other lung diseases and often present only in the later stages of lung cancer. These symptoms include a persistent cough, bloody sputum, chest pain, and recurring bronchitis or pneumonia.

[0011] Where methods of early diagnosis in cancer exist, the benefits are generally accepted by the medical community. Cancers that have widely utilized screening protocols have the highest 5-year survival rates, such as breast cancer (88%) and colon cancer (65%) versus 16% for lung cancer. However, up to 88% of lung cancer patients survive ten years or longer if the cancer is diagnosed at Stage I through screening. This demonstrates the clear need for diagnostic methods that can reliably detect early-stage NSCLC.

[0012] Biomarker selection for a specific disease state involves first the identification of markers that have a measurable and statistically significant difference in a disease population compared to a control population for a specific medical application. Biomarkers can include secreted or shed molecules that parallel disease development or progression and readily diffuse into the blood stream from lung tissue or from distal tissues in response to a lesion. They may also include proteins made by cells in response to the tumor. The biomarker or set of biomarkers identified are generally clinically validated or shown to be a reliable indicator for the original intended use for which it was selected. Biomarkers can include small molecules, metabolites, peptides, proteins, and nucleic acids. Some of the key issues that affect the identification of biomarkers include over-fitting of the available data and bias in the data.

[0013] A variety of methods have been utilized in an attempt to identify biomarkers and diagnose disease. For protein-based markers, these include two-dimensional electrophoresis, mass spectrometry, and immunoassay methods. For nucleic acid markers, these include mRNA expression profiles, microRNA profiles, FISH, serial analysis of gene expression (SAGE), and large scale gene expression arrays.

[0014] The utility of two-dimensional electrophoresis is limited by low detection sensitivity; issues with protein solubility, charge, and hydrophobicity; gel reproducibility; and the possibility of a single spot representing multiple proteins. For mass spectrometry, depending on the format used, limitations revolve around the sample processing and separation, sensitivity to low abundance proteins, signal to noise considerations, and inability to immediately identify the detected protein. Limitations in immunoassay approaches to biomarker discovery are centered on the inability of antibody-based multiplex assays to measure a large number of analytes. One might simply print an array of high-quality antibodies and, without sandwiches, measure the analytes bound to those antibodies. (This would be the formal equivalent of using a whole genome of nucleic acid sequences to measure by hybridization all DNA or RNA sequences in an organism or a cell. The hybridization experiment works because hybridization can be a stringent test for identity. Even very good antibodies are not stringent enough in selecting their binding partners to work in the context of blood or even cell extracts because the protein ensemble in those matrices have extremely different abundances.) Thus, one must use a different approach with immunoassay-based approaches to

biomarker discovery one would need to use multiplexed ELISA assays (that is, sandwiches) to get sufficient stringency to measure many analytes simultaneously to decide which analytes are indeed biomarkers. Sandwich immunoassays do not scale to high content, and thus biomarker discovery using stringent sandwich immunoassays is not possible using standard array formats. Lastly, antibody reagents are subject to substantial lot variability and reagent instability. The instant platform for protein biomarker discovery overcomes this problem.

[0015] Many of these methods rely on or require some type of sample fractionation prior to the analysis. Thus the sample preparation required to run a sufficiently powered study designed to identify/discover statistically relevant biomarkers in a series of well-defined sample populations is extremely difficult, costly, and time consuming. During fractionation, a wide range of variability can be introduced into the various samples. For example, a potential marker could be unstable to the process, the concentration of the marker could be changed, inappropriate aggregation or disaggregation could occur, and inadvertent sample contamination could occur and thus obscure the subtle changes anticipated in early disease.

[0016] It is widely accepted that biomarker discovery and detection methods using these technologies have serious limitations for the identification of diagnostic biomarkers. These limitations include an inability to detect low-abundance biomarkers, an inability to consistently cover the entire dynamic range of the proteome, irreproducibility in sample processing and fractionation, and overall irreproducibility and lack of robustness of the method. Further, these studies have introduced biases into the data and not adequately addressed the complexity of the sample populations, including appropriate controls, in terms of the distribution and randomization required to identify and validate biomarkers within a target disease population.

[0017] Although efforts aimed at the discovery of new and effective biomarkers have gone on for several decades, the efforts have been largely unsuccessful. Biomarkers for various diseases typically have been identified in academic laboratories, usually through an accidental discovery while doing basic research on some disease process. Based on the discovery and with small amounts of clinical data, papers were published that suggested the identification of a new biomarker. Most of these proposed biomarkers, however, have not been confirmed as real or useful biomarkers, primarily because the small number of clinical samples tested provide only weak statistical proof that an effective biomarker has in fact been found. That is, the initial identification was not rigorous with respect to the basic elements of statistics. In each of the years 1994 through 2003, a search of the scientific literature shows that thousands of references directed to biomarkers were published. During that same time frame, however, the FDA approved for diagnostic use, at most, three new protein biomarkers a year, and in several years no new protein biomarkers were approved.

[0018] Based on the history of failed biomarker discovery efforts, mathematical theories have been proposed that further promote the general understanding that biomarkers for disease are rare and difficult to find. Biomarker research based on 2D gels or mass spectrometry supports these notions. Very few useful biomarkers have been identified through these approaches. However, it is usually overlooked that 2D gel and mass spectrometry measure proteins that are present in blood at approximately 1 nM concentrations and

higher, and that this ensemble of proteins may well be the least likely to change with disease. Other than the instant biomarker discovery platform, proteomic biomarker discovery platforms that are able to accurately measure protein expression levels at much lower concentrations do not exist.

[0019] Much is known about biochemical pathways for complex human biology. Many biochemical pathways culminate in or are started by secreted proteins that work locally within the pathology, for example growth factors are secreted to stimulate the replication of other cells in the pathology, and other factors are secreted to ward off the immune system, and so on. While many of these secreted proteins work in a paracrine fashion, some operate distally in the body. One skilled in the art with a basic understanding of biochemical pathways would understand that many pathology-specific proteins ought to exist in blood at concentrations below (even far below) the detection limits of 2D gels and mass spectrometry. What must precede the identification of this relatively abundant number of disease biomarkers is a proteomic platform that can analyze proteins at concentrations below those detectable by 2D gels or mass spectrometry.

[0020] Accordingly, a need exists for biomarkers, methods, devices, reagents, systems, and kits that enable (a) screening high risk smokers for lung cancer (b) the differentiation of benign pulmonary nodules from malignant pulmonary nodules; (c) the detection of lung cancer biomarkers; and (d) the diagnosis of lung cancer.

SUMMARY

[0021] The present application includes biomarkers, methods, reagents, devices, systems, and kits for the detection and diagnosis of cancer and more particularly, NSCLC. The biomarkers of the present application were identified using a multiplex aptamer-based assay which is described in detail in Example 1. By using the aptamer-based biomarker identification method described herein, this application describes a surprisingly large number of NSCLC biomarkers that are useful for the detection and diagnosis of NSCLC as well as a large number of cancer biomarkers that are useful for the detection and diagnosis of cancer more generally. In identifying these biomarkers, over 1000 proteins from hundreds of individual samples were measured, some of which were at concentrations in the low femtomolar range. This is about four orders of magnitude lower than biomarker discovery experiments done with 2D gels and/or mass spectrometry.

[0022] While certain of the described NSCLC biomarkers are useful alone for detecting and diagnosing NSCLC, methods are described herein for the grouping of multiple subsets of the NSCLC biomarkers that are useful as a panel of biomarkers. Once an individual biomarker or subset of biomarkers has been identified, the detection or diagnosis of NSCLC in an individual can be accomplished using any assay platform or format that is capable of measuring differences in the levels of the selected biomarker or biomarkers in a biological sample.

[0023] However, it was only by using the aptamer-based biomarker identification method described herein, wherein over 1000 separate potential biomarker values were individually screened from a large number of individuals having previously been diagnosed either as having or not having NSCLC that it was possible to identify the NSCLC biomarkers disclosed herein. This discovery approach is in stark contrast to biomarker discovery from conditioned media or

lysed cells as it queries a more patient-relevant system that requires no translation to human pathology.

[0024] Thus, in one aspect of the instant application, one or more biomarkers are provided for use either alone or in various combinations to diagnose NSCLC or permit the differential diagnosis of NSCLC from benign conditions such as those found in individuals 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. Exemplary embodiments include the biomarkers provided in Table 1, which as noted above, were identified using a multiplex aptamer-based assay, as described generally in Example 1 and more specifically in Example 2 and 5. The markers provided in Table 1 are useful in diagnosing NSCLC in a high risk population and for distinguishing benign pulmonary diseases in individuals with indeterminate pulmonary nodules from NSCLC.

[0025] 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 two or more biomarkers. Thus, various embodiments of the instant application provide combinations comprising N biomarkers, wherein N is at least two biomarkers. In other embodiments, N is selected to be any number from 2-59 biomarkers.

[0026] In yet other embodiments, N is selected to be any number from 2-5, 2-10, 2-15, 2-20, 2-25, 2-30, 2-35, 2-40, 2-45, 2-50, 2-55, or 2-59. In other embodiments, N is selected to be any number from 3-5, 3-10, 3-15, 3-20, 3-25, 3-30, 3-35, 3-40, 3-45, 3-50, 3-55, or 3-59. In other embodiments, N is selected to be any number from 4-5, 4-10, 4-15, 4-20, 4-25, 4-30, 4-35, 4-40, 4-45, 4-50, 4-55, or 4-59. In other embodiments, N is selected to be any number from 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5-50, 5-55, or 5-59. In other embodiments, N is selected to be any number from 6-10, 6-15, 6-20, 6-25, 6-30, 6-35, 6-40, 6-45, 6-50, 6-55, or 6-59. In other embodiments, N is selected to be any number from 7-10, 7-15, 7-20, 7-25, 7-30, 7-35, 7-40, 7-45, 7-50, 7-55, or 7-59. In other embodiments, N is selected to be any number from 8-10, 8-15, 8-20, 8-25, 8-30, 8-35, 8-40, 8-45, 8-50, 8-55, or 8-59. In other embodiments, N is selected to be any number from 9-10, 9-15, 9-20, 9-25, 9-30, 9-35, 9-40, 9-45, 9-50, 9-55, or 9-59. In other embodiments, N is selected to be any number from 10-15, 10-20, 10-25, 10-30, 10-35, 10-40, 10-45, 10-50, 10-55, or 10-59. It will be appreciated that N can be selected to encompass similar, but higher order, ranges.

[0027] In another aspect, a method is provided for diagnosing NSCLC in an individual, the method including detecting, in a biological sample from an individual, at least one biomarker value corresponding to at least one biomarker selected from the group of biomarkers provided in Table 1, wherein the individual is classified as having NSCLC based on the at least one biomarker value.

[0028] In another aspect, a method is provided for diagnosing NSCLC in an individual, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 1, wherein the likelihood of the individual having NSCLC is determined based on the biomarker values.

[0029] In another aspect, a method is provided for diagnosing NSCLC in an individual, the method including detecting,

in a biological sample from an individual, biomarker values that each correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 1, wherein the individual is classified as having NSCLC based on the biomarker values, and wherein N=2-10.

[0030] In another aspect, a method is provided for diagnosing NSCLC in an individual, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 1, wherein the likelihood of the individual having NSCLC is determined based on the biomarker values, and wherein N=2-10.

[0031] In another aspect, a method is provided for diagnosing that an individual does not have NSCLC, the method including detecting, in a biological sample from an individual, at least one biomarker value corresponding to at least one biomarker selected from the group of biomarkers set forth in Table 1, wherein the individual is classified as not having NSCLC based on the at least one biomarker value.

[0032] In another aspect, a method is provided for diagnosing that an individual does not have NSCLC, the method including detecting, in a biological sample from an individual, biomarker values that each corresponding to one of at least N biomarkers selected from the group of biomarkers set forth in Table 1, wherein the individual is classified as not having NSCLC based on the biomarker values, and wherein N=2-10.

[0033] In another aspect, a method is provided for diagnosing NSCLC, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of N biomarkers, wherein the biomarkers are selected from the group of biomarkers set forth in Table 1, wherein a classification of the biomarker values indicates that the individual has NSCLC, and wherein N=3-10.

[0034] In another aspect, a method is provided for diagnosing NSCLC, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of biomarkers selected from the group of panels set forth in Tables 2-11, wherein a classification of the biomarker values indicates that the individual has NSCLC.

[0035] In another aspect, a method is provided for diagnosing an absence of NSCLC, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of N biomarkers, wherein the biomarkers are selected from the group of biomarkers set forth in Table 1, wherein a classification of the biomarker values indicates an absence of NSCLC in the individual, and wherein N=3-10.

[0036] In another aspect, a method is provided for diagnosing an absence of NSCLC, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of N biomarkers, wherein the biomarkers are selected from the group of biomarkers set forth in Table 1, wherein a classification of the biomarker values indicates an absence of NSCLC in the individual, and wherein N=3-10.

[0037] In another aspect, a method is provided for diagnosing an absence of NSCLC, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of biomarkers selected from the group of panels provided in Tables 2-11,

wherein a classification of the biomarker values indicates an absence of NSCLC in the individual.

[0038] In another aspect, a method is provided for diagnosing NSCLC in an individual, the method including detecting, in a biological sample from an individual, biomarker values that correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 1, wherein the individual is classified as having NSCLC based on a classification score that deviates from a predetermined threshold, and wherein N=2-10.

[0039] In another aspect, a method is provided for diagnosing an absence of NSCLC in an individual, the method including detecting, in a biological sample from an individual, biomarker values that correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 1, wherein said individual is classified as not having NSCLC based on a classification score that deviates from a predetermined threshold, and wherein N=2-10.

[0040] In another aspect, a computer-implemented method is provided for indicating a likelihood of NSCLC. The method comprises: retrieving on a computer biomarker information for an individual, wherein the biomarker information comprises biomarker values that each correspond to one of at least N biomarkers, wherein N is as defined above, selected from the group of biomarkers set forth in Table 1; performing with the computer a classification of each of the biomarker values; and indicating a likelihood that the individual has NSCLC based upon a plurality of classifications.

[0041] In another aspect, a computer-implemented method is provided for classifying an individual as either having or not having NSCLC. The method comprises: retrieving on a computer biomarker information for an individual, wherein the biomarker information comprises biomarker values that each correspond to one of at least N biomarkers selected from the group of biomarkers provided in Table 1; performing with the computer a classification of each of the biomarker values; and indicating whether the individual has NSCLC based upon a plurality of classifications.

[0042] In another aspect, a computer program product is provided for indicating a likelihood of NSCLC. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises biomarker values that each correspond to one of at least N biomarkers, wherein N is as defined above, in the biological sample selected from the group of biomarkers set forth in Table 1; and code that executes a classification method that indicates a likelihood that the individual has NSCLC as a function of the biomarker values.

[0043] In another aspect, a computer program product is provided for indicating a NSCLC status of an individual. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises biomarker values that each correspond to one of at least N biomarkers in the biological sample selected from the group of biomarkers provided in Table 1; and code that executes a classification method that indicates a NSCLC status of the individual as a function of the biomarker values.

[0044] In another aspect, a computer-implemented method is provided for indicating a likelihood of NSCLC. The method comprises retrieving on a computer biomarker information for an individual, wherein the biomarker information comprises a biomarker value corresponding to a biomarker selected from the group of biomarkers set forth in Table 1; performing with the computer a classification of the biomarker value; and indicating a likelihood that the individual has NSCLC based upon the classification.

[0045] In another aspect, a computer-implemented method is provided for classifying an individual as either having or not having NSCLC. The method comprises retrieving from a computer biomarker information for an individual, wherein the biomarker information comprises a biomarker value corresponding to a biomarker selected from the group of biomarkers provided in Table 1; performing with the computer a classification of the biomarker value; and indicating whether the individual has NSCLC based upon the classification.

[0046] In still another aspect, a computer program product is provided for indicating a likelihood of NSCLC. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises a biomarker value corresponding to a biomarker in the biological sample selected from the group of biomarkers set forth in Table 1; and code that executes a classification method that indicates a likelihood that the individual has NSCLC as a function of the biomarker value.

[0047] In still another aspect, a computer program product is provided for indicating a NSCLC status of an individual. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises a biomarker value corresponding to a biomarker in the biological sample selected from the group of biomarkers provided in Table 1; and code that executes a classification method that indicates a NSCLC status of the individual as a function of the biomarker value.

[0048] While certain of the described biomarkers are also useful alone for detecting and diagnosing general cancer, methods are described herein for the grouping of multiple subsets of the biomarkers that are useful as a panel of biomarkers for detecting and diagnosing cancer in general. Once an individual biomarker or subset of biomarkers has been identified, the detection or diagnosis of cancer in an individual can be accomplished using any assay platform or format that is capable of measuring differences in the levels of the selected biomarker or biomarkers in a biological sample.

[0049] However, it was only by using the aptamer-based biomarker identification method described herein, wherein over 1000 separate potential biomarker values were individually screened from a large number of individuals having previously been diagnosed either as having or not having cancer that it was possible to identify the cancer biomarkers disclosed herein. This discovery approach is in stark contrast to biomarker discovery from conditioned media or lysed cells as it queries a more patient-relevant system that requires no translation to human pathology.

[0050] Thus, in one aspect of the instant application, one or more biomarkers are provided for use either alone or in vari-

ous combinations to diagnose cancer. Exemplary embodiments include the biomarkers provided in Table 19, which were identified using a multiplex aptamer-based assay, as described generally in Example 1 and more specifically in Example 6. The markers provided in Table 19 are useful in distinguishing individuals who have cancer from those who do not have cancer.

[0051] While certain of the described cancer biomarkers are useful alone for detecting and diagnosing cancer, methods are also described herein for the grouping of multiple subsets of the cancer 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 3-23 biomarkers.

[0052] In yet other embodiments, N is selected to be any number from 2-5, 2-10, 2-15, 2-20, or 2-23. In other embodiments, N is selected to be any number from 3-5, 3-10, 3-15, 3-20, or 3-23. In other embodiments, N is selected to be any number from 4-5, 4-10, 4-15, 4-20, or 4-23. In other embodiments, N is selected to be any number from 5-10, 5-15, 5-20, or 5-23. In other embodiments, N is selected to be any number from 6-10, 6-15, 6-20, or 6-23. In other embodiments, N is selected to be any number from 7-10, 7-15, 7-20, or 7-23. In other embodiments, N is selected to be any number from 8-10, 8-15, 8-20, or 8-23. In other embodiments, N is selected to be any number from 9-10, 9-15, 9-20, or 9-23. In other embodiments, N is selected to be any number from 10-15, 10-20, or 10-23. It will be appreciated that N can be selected to encompass similar, but higher order, ranges.

[0053] In another aspect, a method is provided for diagnosing cancer in an individual, the method including detecting, in a biological sample from an individual, at least one biomarker value corresponding to at least one biomarker selected from the group of biomarkers provided in Table 19, wherein the individual is classified as having cancer based on the at least one biomarker value.

[0054] In another aspect, a method is provided for diagnosing cancer in an individual, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 19, wherein the likelihood of the individual having cancer is determined based on the biomarker values.

[0055] In another aspect, a method is provided for diagnosing cancer in an individual, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 19, wherein the individual is classified as having cancer based on the biomarker values, and wherein N=3-10.

[0056] In another aspect, a method is provided for diagnosing cancer in an individual, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 19, wherein the likelihood of the individual having cancer is determined based on the biomarker values, and wherein N=3-10.

[0057] In another aspect, a method is provided for diagnosing that an individual does not have cancer, the method including detecting, in a biological sample from an individual, at least one biomarker value corresponding to at least one biomarker selected from the group of biomarkers set forth

in Table 19, wherein the individual is classified as not having cancer based on the at least one biomarker value.

[0058] In another aspect, a method is provided for diagnosing that an individual does not have cancer, the method including detecting, in a biological sample from an individual, biomarker values that each corresponding to one of at least N biomarkers selected from the group of biomarkers set forth in Table 19, wherein the individual is classified as not having cancer based on the biomarker values, and wherein $N=3-10$.

[0059] In another aspect, a method is provided for diagnosing cancer, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of N biomarkers, wherein the biomarkers are selected from the group of biomarkers set forth in Table 19, wherein a classification of the biomarker values indicates that the individual has cancer, and wherein $N=3-10$.

[0060] In another aspect, a method is provided for diagnosing cancer, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of biomarkers selected from the group of panels set forth in Tables 20-29 wherein a classification of the biomarker values indicates that the individual has cancer.

[0061] In another aspect, a method is provided for diagnosing an absence of cancer, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of N biomarkers, wherein the biomarkers are selected from the group of biomarkers set forth in Table 19, wherein a classification of the biomarker values indicates an absence of cancer in the individual, and wherein $N=3-10$.

[0062] In another aspect, a method is provided for diagnosing an absence of cancer, the method including detecting, in a biological sample from an individual, biomarker values that each correspond to a biomarker on a panel of biomarkers selected from the group of panels provided in Tables 20-29, wherein a classification of the biomarker values indicates an absence of cancer in the individual.

[0063] In another aspect, a method is provided for diagnosing cancer in an individual, the method including detecting, in a biological sample from an individual, biomarker values that correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 19, wherein the individual is classified as having cancer based on a classification score that deviates from a predetermined threshold, and wherein $N=3-10$.

[0064] In another aspect, a method is provided for diagnosing an absence of cancer in an individual, the method including detecting, in a biological sample from an individual, biomarker values that correspond to one of at least N biomarkers selected from the group of biomarkers set forth in Table 19, wherein said individual is classified as not having cancer based on a classification score that deviates from a predetermined threshold, and wherein $N=3-10$.

[0065] In another aspect, a computer-implemented method is provided for indicating a likelihood of cancer. The method comprises: retrieving on a computer biomarker information for an individual, wherein the biomarker information comprises biomarker values that each correspond to one of at least N biomarkers, wherein N is as defined above, selected from the group of biomarkers set forth in Table 19; performing with the computer a classification of each of the biomarker values;

and indicating a likelihood that the individual has cancer based upon a plurality of classifications.

[0066] In another aspect, a computer-implemented method is provided for classifying an individual as either having or not having cancer. The method comprises: retrieving on a computer biomarker information for an individual, wherein the biomarker information comprises biomarker values that each correspond to one of at least N biomarkers selected from the group of biomarkers provided in Table 19; performing with the computer a classification of each of the biomarker values; and indicating whether the individual has cancer based upon a plurality of classifications.

[0067] In another aspect, a computer program product is provided for indicating a likelihood of cancer. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises biomarker values that each correspond to one of at least N biomarkers, wherein N is as defined above, in the biological sample selected from the group of biomarkers set forth in Table 19; and code that executes a classification method that indicates a likelihood that the individual has cancer as a function of the biomarker values.

[0068] In another aspect, a computer program product is provided for indicating a cancer status of an individual. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises biomarker values that each correspond to one of at least N biomarkers in the biological sample selected from the group of biomarkers provided in Table 19; and code that executes a classification method that indicates a cancer status of the individual as a function of the biomarker values.

[0069] In another aspect, a computer-implemented method is provided for indicating a likelihood of cancer. The method comprises retrieving on a computer biomarker information for an individual, wherein the biomarker information comprises a biomarker value corresponding to a biomarker selected from the group of biomarkers set forth in Table 19; performing with the computer a classification of the biomarker value; and indicating a likelihood that the individual has cancer based upon the classification.

[0070] In another aspect, a computer-implemented method is provided for classifying an individual as either having or not having cancer. The method comprises retrieving from a computer biomarker information for an individual, wherein the biomarker information comprises a biomarker value corresponding to a biomarker selected from the group of biomarkers provided in Table 19; performing with the computer a classification of the biomarker value; and indicating whether the individual has cancer based upon the classification.

[0071] In still another aspect, a computer program product is provided for indicating a likelihood of cancer. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises a biomarker value corresponding to a biomarker in the biological sample selected from the group of biomarkers set forth in Table 19;

and code that executes a classification method that indicates a likelihood that the individual has cancer as a function of the biomarker value.

[0072] In still another aspect, a computer program product is provided for indicating a cancer status of an individual. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises a biomarker value corresponding to a biomarker in the biological sample selected from the group of biomarkers provided in Table 19; and code that executes a classification method that indicates a cancer status of the individual as a function of the biomarker value.

BRIEF DESCRIPTION OF THE DRAWINGS

[0073] FIG. 1A is a flowchart for an exemplary method for detecting NSCLC in a biological sample.

[0074] FIG. 1B is a flowchart for an exemplary method for detecting NSCLC in a biological sample using a naïve Bayes classification method.

[0075] FIG. 2 shows a ROC curve for a single biomarker, MMP7, using a naïve Bayes classifier for a test that detects NSCLC.

[0076] FIG. 3 shows ROC curves for biomarker panels of from two to ten biomarkers using naïve Bayes classifiers for a test that detects NSCLC.

[0077] FIG. 4 illustrates the increase in the classification score (AUC) as the number of biomarkers is increased from one to ten using naïve Bayes classification for a NSCLC panel.

[0078] FIG. 5 shows the measured biomarker distributions for MMP7 as a cumulative distribution function (cdf) in log-transformed RFU for the smokers and benign pulmonary nodules controls combined (solid line) and the NSCLC disease group (dotted line) along with their curve fits to a normal cdf (dashed lines) used to train the naïve Bayes classifiers.

[0079] FIG. 6 illustrates an exemplary computer system for use with various computer-implemented methods described herein.

[0080] FIG. 7 is a flowchart for a method of indicating the likelihood that an individual has NSCLC in accordance with one embodiment.

[0081] FIG. 8 is a flowchart for a method of indicating the likelihood that an individual has NSCLC in accordance with one embodiment.

[0082] FIG. 9 illustrates an exemplary aptamer assay that can be used to detect one or more NSCLC biomarkers in a biological sample.

[0083] FIG. 10 shows a histogram of frequencies for which biomarkers were used in building classifiers to distinguish between NSCLC and the smokers and benign pulmonary nodules control group from an aggregated set of potential biomarkers.

[0084] FIG. 11A shows a pair of histograms summarizing all possible single protein naïve Bayes classifier scores (AUC) using the biomarkers set forth in Table 1 (black) and a set of random markers (grey).

[0085] FIG. 11B shows a pair of histograms summarizing all possible two-protein protein naïve Bayes classifier scores (AUC) using the biomarkers set forth in Table 1 (black) and a set of random markers (grey).

[0086] FIG. 11C shows a pair of histograms summarizing all possible three-protein naïve Bayes classifier scores (AUC) using the biomarkers set forth in Table 1 (black) and a set of random markers (grey).

[0087] FIG. 12 shows the AUC for naïve Bayes classifiers using from 2-10 markers selected from the full panel and the scores obtained by dropping the best 5, 10, and 15 markers during classifier generation.

[0088] FIG. 13A shows a set of ROC curves modeled from the data in Table 14 for panels of from two to five markers.

[0089] FIG. 13B shows a set of ROC curves computed from the training data for panels of from two to five markers as in FIG. 12A.

[0090] FIG. 14 shows a ROC curve computed from the clinical biomarker panel described in Example 5.

[0091] FIGS. 15A and 15B show a comparison of performance between ten cancer biomarkers selected by a greedy selection procedure described in Example 6 (Table 19) and 1,000 randomly sampled sets of ten “non marker” biomarkers. The mean AUC for the ten cancer biomarkers in Table 19 is shown as a dotted vertical line. In FIG. 15A, sets of ten “nonmarkers” were randomly selected that were not selected by the greedy procedure described in Example 6. In FIG. 15B, the same procedure as 15A was used; however, the sampling was restricted to the remaining 49 NSCLC biomarkers from Table 1 that were not selected by the greedy procedure described in Example 6.

[0092] FIG. 16 shows receiver operating characteristic (ROC) curves for the 3 naïve Bayes classifiers set forth in Table 31. For each study, the area under the curve (AUC) is also displayed next to the legend.

DETAILED DESCRIPTION

[0093] 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 intended 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.

[0094] 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.

[0095] 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.

[0096] 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.

[0097] 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.

[0098] 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.

[0099] 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.

[0100] The present application includes biomarkers, methods, devices, reagents, systems, and kits for the detection and diagnosis of NSCLC and cancer more generally.

[0101] In one aspect, one or more biomarkers are provided for use either alone or in various combinations to diagnose NSCLC, permit the differential diagnosis of NSCLC from non-malignant conditions found in individuals 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, monitor NSCLC recurrence, or address other clinical indications. As described in detail below, exemplary embodiments include the biomarkers provided in Table 1, which were identified using a multiplex aptamer-based assay that is described generally in Example 1 and more specifically in Example 2.

[0102] Table 1 sets forth the findings obtained from analyzing hundreds of individual blood samples from NSCLC cases, and hundreds of equivalent individual control blood samples from high risk smokers and benign pulmonary nodules. The smokers and benign pulmonary nodules control group was designed to match the populations with which a NSCLC diagnostic test can have the most benefit, including asymptomatic individuals and symptomatic individuals. These cases and controls were obtained from multiple clinical sites to mimic the range of real world conditions under which such a test can be applied. The potential biomarkers were measured in individual samples rather than pooling the disease and control blood; this allowed a better understanding of the individual and group variations in the phenotypes associated with the presence and absence of disease (in this case NSCLC). Since over 1000 protein measurements were made on each sample, and several hundred samples from each of the disease and the control populations were individually measured, Table 1, resulted from an analysis of an uncommonly large set of data. The measurements were analyzed using the methods described in the section, “Classification of Biomarkers and Calculation of Disease Scores” herein. Table 1 lists the 59 biomarkers found to be useful in distinguishing samples obtained from individuals with NSCLC from “control” samples obtained from smokers and benign pulmonary nodules.

[0103] 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, where each grouping or subset selection is useful as a panel of three or more biomar-

kers, interchangeably referred to herein as a “biomarker panel” and a panel. Thus, various embodiments of the instant application provide combinations comprising N biomarkers, wherein N is at least two biomarkers. In other embodiments, N is selected from 2-59 biomarkers.

[0104] In yet other embodiments, N is selected to be any number from 2-5, 2-10, 2-15, 2-20, 2-25, 2-30, 2-35, 2-40, 2-45, 2-50, 2-55, or 2-59. In other embodiments, N is selected to be any number from 3-5, 3-10, 3-15, 3-20, 3-25, 3-30, 3-35, 3-40, 3-45, 3-50, 3-55, or 3-59. In other embodiments, N is selected to be any number from 4-5, 4-10, 4-15, 4-20, 4-25, 4-30, 4-35, 4-40, 4-45, 4-50, 4-55, or 4-59. In other embodiments, N is selected to be any number from 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5-50, 5-55, or 5-59. In other embodiments, N is selected to be any number from 6-10, 6-15, 6-20, 6-25, 6-30, 6-35, 6-40, 6-45, 6-50, 6-55, or 6-59. In other embodiments, N is selected to be any number from 7-10, 7-15, 7-20, 7-25, 7-30, 7-35, 7-40, 7-45, 7-50, 7-55, or 7-59. In other embodiments, N is selected to be any number from 8-10, 8-15, 8-20, 8-25, 8-30, 8-35, 8-40, 8-45, 8-50, 8-55, or 8-59. In other embodiments, N is selected to be any number from 9-10, 9-15, 9-20, 9-25, 9-30, 9-35, 9-40, 9-45, 9-50, 9-55, or 9-59. In other embodiments, N is selected to be any number from 10-15, 10-20, 10-25, 10-30, 10-35, 10-40, 10-45, 10-50, 10-55, or 10-59. It will be appreciated that N can be selected to encompass similar, but higher order, ranges.

[0105] 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 individuals that have NSCLC. “Specificity” indicates the performance of the biomarker(s) with respect to correctly classifying individuals who do not have NSCLC. For example, 85% specificity and 90% sensitivity for a panel of markers used to test a set of control samples and NSCLC samples indicates that 85% of the control samples were correctly classified as control samples by the panel, and 90% of the NSCLC samples were correctly classified as NSCLC samples by the panel. The desired or preferred minimum value can be determined as described in Example 3. Representative panels are set forth in Tables 4-11, which set forth a series of 100 different panels of 3-10 biomarkers, which have the indicated levels of specificity and sensitivity for each panel. The total number of occurrences of each marker in each of these panels is indicated in Table 12.

[0106] In one aspect, NSCLC is detected or diagnosed in an individual by conducting an assay on a biological sample from the individual and detecting biomarker values that each correspond to at least one of the biomarkers MMP7, CLIC1 or STX1A and at least N additional biomarkers selected from the list of biomarkers in Table 1, wherein N equals 2, 3, 4, 5, 6, 7, 8, or 9. In a further aspect, NSCLC is detected or diagnosed in an individual by conducting an assay on a biological sample from the individual and detecting biomarker values that each correspond to the biomarkers MMP7, CLIC1 or STX1A and one of at least N additional biomarkers selected from the list of biomarkers in Table 1, wherein N equals 1, 2, 3, 4, 5, 6, or 7. In a further aspect, NSCLC is

detected or diagnosed in an individual by conducting an assay on a biological sample from the individual and detecting biomarker values that each correspond to the biomarker MMP7 and one of at least N additional biomarkers selected from the list of biomarkers in Table 1, wherein N equals 2, 3, 4, 5, 6, 7, 8, or 9. In a further aspect, NSCLC is detected or diagnosed in an individual by conducting an assay on a biological sample from the individual and detecting biomarker values that each correspond to the biomarker CLIC1 and one of at least N additional biomarkers selected from the list of biomarkers in Table 1, wherein N equals 2, 3, 4, 5, 6, 7, 8, or 9. In a further aspect, NSCLC is detected or diagnosed in an individual by conducting an assay on a biological sample from the individual and detecting biomarker values that each correspond to the biomarker STX1A and one of at least N additional biomarkers selected from the list of biomarkers in Table 1, wherein N equals 2, 3, 4, 5, 6, 7, 8, or 9.

[0107] The NSCLC biomarkers identified herein represent a relatively large number of choices for subsets or panels of biomarkers that can be used to effectively detect or diagnose NSCLC. Selection of the desired number of such biomarkers depends on the specific combination of biomarkers chosen. It is important to remember that panels of biomarkers for detecting or diagnosing NSCLC may also include biomarkers not found in Table 1, and that the inclusion of additional biomarkers not found in Table 1 may reduce the number of biomarkers in the particular subset or panel that is selected from Table 1. The number of biomarkers from Table 1 used in a subset or panel may also be reduced if additional biomedical information is used in conjunction with the biomarker values to establish acceptable sensitivity and specificity values for a given assay.

[0108] Another factor that can affect the number of biomarkers to be used in a subset or panel of biomarkers is the procedures used to obtain biological samples from individuals who are being diagnosed for NSCLC. In a carefully controlled sample procurement environment, the number of biomarkers necessary to meet desired sensitivity and specificity values will be lower than in a situation where there can be more variation in sample collection, handling and storage. In developing the list of biomarkers set forth in Table 1, multiple sample collection sites were utilized to collect data for classifier training. This provides for more robust biomarkers that are less sensitive to variations in sample collection, handling and storage, but can also require that the number of biomarkers in a subset or panel be larger than if the training data were all obtained under very similar conditions.

[0109] One aspect of the instant application can be described generally with reference to FIGS. 1A and 1B. A biological sample is obtained from an individual or individuals of interest. The biological sample is then assayed to detect the presence of one or more (N) biomarkers of interest and to determine a biomarker value for each of said N biomarkers (referred to in FIG. 1B as marker RFU). Once a biomarker has been detected and a biomarker value assigned each marker is scored or classified as described in detail herein. The marker scores are then combined to provide a total diagnostic score, which indicates the likelihood that the individual from whom the sample was obtained has NSCLC.

[0110] As used herein, “lung” may be interchangeably referred to as “pulmonary”.

[0111] As used herein, “smoker” refers to an individual who has a history of tobacco smoke inhalation.

[0112] “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.

[0113] Further, it should be realized that a biological sample can be derived by taking biological samples from a number of individuals 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.

[0114] 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.

[0115] “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, affibodies, 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.

[0116] 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.

[0117] 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.

[0118] 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.

[0119] 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 overexpressed 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 individuals. 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.

[0120] “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 individuals. 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.

[0121] 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.

[0122] 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.

[0123] As used herein, “individual” refers to a test subject or patient. The individual can be a mammal or a non-mammal. In various embodiments, the individual is a mammal. A mammalian individual can be a human or non-human. In various embodiments, the individual is a human. A healthy or normal individual is an individual in which the disease or condition of interest (including, for example, lung diseases,

lung-associated diseases, or other lung conditions) is not detectable by conventional diagnostic methods.

[0124] “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 individual. 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 individuals who have cancer from individuals who do not. It further includes distinguishing smokers and benign pulmonary nodules from NSCLC.

[0125] “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.

[0126] “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.

[0127] 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.

[0128] 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).

[0129] 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.

[0130] 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.

[0131] “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 tetrakisphthalate, polyvinyl acetate, polyvinyl chloride, polyvinyl pyrrolidone, polyacrylonitrile, polymethyl methacrylate, polytetrafluoroethylene, butyl rubber, styrenebutadiene rubber, natural rubber, polyethylene, polypropylene, (poly)tetrafluoroethylene, (poly)vinylidene fluoride, 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.

Exemplary Uses of Biomarkers

[0132] In various exemplary embodiments, methods are provided for diagnosing NSCLC in an 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 individuals with NSCLC as compared to individuals 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.

[0133] 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 individuals for computed tomography (CT) screening (e.g., identifying those individuals 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 additional 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 individuals 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 individuals 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.

[0134] 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.

[0135] 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.

[0136] 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.

[0137] 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)).

[0138] 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 individuals 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.

[0139] 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.

[0140] In addition to testing biomarker levels in conjunction with radiologic screening in high risk individuals (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.

[0141] 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.

Detection and Determination of Biomarkers and Biomarker Values

[0142] 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, affibodies, nanobodies, imprinted polymers, avimers, peptidomimetics, a hormone receptor, a cytokine receptor, and synthetic receptors, and modifications and fragments of these.

[0143] In some embodiments, a biomarker value is detected using a biomarker/capture reagent complex.

[0144] 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.

[0145] In some embodiments, the biomarker value is detected directly from the biomarker in a biological sample.

[0146] 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.

[0147] 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.

[0148] 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.

[0149] 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.

[0150] 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-trihydroxybenzene), Lucigenin, peroxyoxalates, Aryl oxalates, Acridinium esters, dioxetanes, and others.

[0151] 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.

[0152] In yet other embodiments, the detection method can be a combination of fluorescence, chemiluminescence, radio-nuclide or enzyme/substrate combinations that generate a measurable signal. Multimodal signaling could have unique and advantageous characteristics in biomarker assay formats.

[0153] 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.

Determination of Biomarker Values Using Aptamer-Based Assays

[0154] 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.

[0155] 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.

[0156] 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.

[0157] 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."

[0158] 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.

[0159] 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."

[0160] 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.

[0161] 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.

[0162] 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.

[0163] 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.

[0164] 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.

[0165] 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.

[0166] 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. For NSCLC, the molecular capture reagents would be an aptamer or an antibody or the like and the specific target would be a NSCLC biomarker of Table 1.

[0167] 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.

[0168] 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.

[0169] 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.

[0170] 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.

[0171] 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.

[0172] 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.

Determination of Biomarker Values Using Immunoassays

[0173] 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.

[0174] 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.

[0175] 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).

[0176] 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.

[0177] 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.

[0178] 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.

Determination of Biomarker Values Using Gene Expression Profiling

[0179] 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.

[0180] 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.

[0181] 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 wide-spread 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.

Detection of Biomarkers Using In Vivo Molecular Imaging Technologies

[0182] Any of the described biomarkers (see Table 1) 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.

[0183] 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.

[0184] 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 imag-

ing 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.

[0185] 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.

[0186] 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.

[0187] 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.

[0188] 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.

[0189] 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 1 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.

[0190] 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 1 (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.

[0191] 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.

[0192] 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.

[0193] 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.

[0194] For a review of other techniques, see N. Blow, *Nature Methods*, 6, 465-469, 2009.

Determination of Biomarker Values Using Histology/Cytology Methods

[0195] 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 cytology. 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 (Table 1) in the individuals with NSCLC can be used to stain a histological specimen as an indication of disease.

[0196] 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.

[0197] 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.

[0198] 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.

[0199] 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.

[0200] 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.

[0201] 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.

[0202] 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.

[0203] 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".

[0204] 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.

[0205] Cell block sections can be stained with hematoxylin-eosin for cytomorphological examination while additional sections are used for examination for specific markers.

[0206] Whether the process is cytological 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 affixation 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.

[0207] 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 placed 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.

[0208] A wide range of dyes can be used to differentially highlight and contrast or "stain" cellular, sub-cellular, and tissue features or morphological structures. Hematoxylin 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.

[0209] 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.

[0210] 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.

[0211] 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 bright-field, phase contrast, fluorescence, and differential interference contrast.

[0212] 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.

[0213] 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.

[0214] 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).

[0215] 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.

[0216] 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 rehydrated 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.

[0217] Laser capture micro-dissection allows the isolation of a subset of cells for further analysis from a tissue section.

[0218] 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.

[0219] 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.

[0220] 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.

[0221] 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 2×2 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 2×2 min. Additional blocking can be performed, if required for the specific application, followed by additional rinsing with PBS Tween 20 for 3×2 min and then finally the immunostaining protocol completed.

[0222] 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.

[0223] 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.

[0224] 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.

Determination of Biomarker Values Using Mass Spectrometry Methods

[0225] 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)).

[0226] 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.

[0227] 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, affibodies, 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.

Determination of Biomarker Values Using a Proximity Ligation Assay

[0228] 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.

[0229] 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.

[0230] 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 1, 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.

[0231] 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 1,

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.

Classification of Biomarkers and Calculation of Disease Scores

[0232] 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 individuals 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 individuals 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.

[0233] 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.

[0234] 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 individuals in a control population and individuals in a particular disease population can constitute training data to develop a classifier that can classify unknown samples (or, more particularly, the individuals 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 naïve 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).

[0235] 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.

[0236] An illustrative example of the development of a diagnostic test using a set of biomarkers includes the application of a naïve 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 individual class-dependent pdfs for each biomarker. Training a naïve 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.

[0237] Specifically, the class-dependent probability of measuring a value x_i for biomarker i in the disease class is written as $p(x_i|d)$ and the overall naïve Bayes probability of observing n markers with values $\tilde{x}=(x_1, x_2, \dots, x_n)$ is written as $p(\tilde{x}|d)=\prod_{i=1}^n p(x_i|d)$ where the individual x_i s 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 $p(d|\tilde{x})$ having measured $\tilde{x}$ compared to the probability of being disease free (control) $p(c|\tilde{x})$ for the same measured values. The ratio of these probabilities is computed from the class-dependent pdfs by application of Bayes theorem, i.e.,

$$\frac{p(d|\tilde{x})}{p(c|\tilde{x})} = \frac{p(\tilde{x}|d)p(d)}{p(\tilde{x}|c)(1-p(d))}$$

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 naïve Bayes class-dependent probabilities from above gives

$$\ln\left(\frac{p(d|\tilde{x})}{p(c|\tilde{x})}\right) = \sum_{i=1}^n \ln\left(\frac{p(x_i|d)}{p(x_i|c)}\right) + \ln\left(\frac{p(d)}{1-p(d)}\right).$$

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 individual log likelihood ratios of the n individual biomarkers. In its simplest form, an unknown sample (or, more particularly, the individual 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.

[0238] In one exemplary embodiment, the class-dependent biomarker pdfs $p(x_i|c)$ and $p(x_i|d)$ are assumed to be normal or log-normal distributions in the measured RFU values x_i , i.e.

$$p(x_i|c) = \frac{1}{\sqrt{2\pi}\sigma_{c,i}} \exp\left(-\frac{(x_i - \mu_{c,i})^2}{2\sigma_{c,i}^2}\right),$$

with a similar expression for $p(x_i|d)$ with $\mu_{d,i}$ and $\sigma_{d,i}$. Parameterization of the model requires estimation of two parameters for each class-dependent pdf, a mean μ and a variance σ^2 , 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:

$$\ln\left(\frac{p(d|\tilde{x})}{p(c|\tilde{x})}\right) = \sum_{i=1}^n \ln\left(\frac{\sigma_{c,i}}{\sigma_{d,i}}\right) - \frac{1}{2} \sum_{i=1}^n \left[\left(\frac{x_i - \mu_{d,i}}{\sigma_{d,i}} \right)^2 - \left(\frac{x_i - \mu_{c,i}}{\sigma_{c,i}} \right)^2 \right] + \ln\left(\frac{p(d)}{1-p(d)}\right)$$

[0239] Once a set of μ s and σ^2 s 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 $\tilde{x}$.

[0240] The performance of the naïve 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, $D_{n,m} = \sup_x |F_{A,n}(x) - F_{B,m}(x)|$, which is the largest difference between two empirical cumulative distribution functions (cdf). The empirical cdf for a set A of n observations X_i is defined as,

$$F_{A,n}(x) = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x},$$

where $I_{X_i \leq x}$ is the indicator function which is equal to 1 if $X_i \leq x$ 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 empirical distributions do not overlap.

[0241] 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.)

[0242] The algorithm approach used here is described in detail in Example 4. Briefly, 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 thousand 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.

[0243] Various other techniques are known in the art and may be employed to generate many potential classifiers from a list of biomarkers using a naïve 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.

[0244] Exemplary embodiments use any number of the NSCLC biomarkers listed in Table 1 in various combinations to produce diagnostic tests for detecting NSCLC (see Example 2 for a detailed description of how these biomarkers were identified). In one embodiment, a method for diagnosing NSCLC uses a naïve Bayes classification method in conjunction with any number of the NSCLC biomarkers listed in Table 1. In an illustrative example (Example 3), the simplest test for detecting NSCLC from a population of smokers and benign pulmonary nodules can be constructed using a single biomarker, for example, MMP7 which is differentially expressed in NSCLC with a KS-distance of 0.59. Using the

parameters, $\mu_{c,i}$, $\sigma_{c,i}$, and $\sigma_{d,i}$ for MMP7 from Table 16 and the equation for the log-likelihood described above, a diagnostic test with an AUC of 0.803 can be derived, see Table 15. The ROC curve for this test is displayed in FIG. 2.

[0245] Addition of biomarker CLIC1, for example, with a KS-distance of 0.53, significantly improves the classifier performance to an AUC of 0.883. Note that the score for a classifier constructed of two biomarkers is not a simple sum of the KS-distances; KS-distances are not additive when combining biomarkers and it takes many more weak markers to achieve the same level of performance as a strong marker. Adding a third marker, STX1A, for example, boosts the classifier performance to an AUC of 0.901. Adding additional biomarkers, such as, for example, CHRDL1, PA2G4, SERPINA1, BDNF, GHR, TGFBI, and NME2, produces a series of NSCLC tests summarized in Table 15 and displayed as a series of ROC curves in FIG. 3. The score of the classifiers as a function of the number of analytes used in classifier construction is displayed in FIG. 4. The AUC of this exemplary ten-marker classifier is 0.948.

[0246] The markers listed in Table 1 can be combined in many ways to produce classifiers for diagnosing NSCLC. In some embodiments, panels of biomarkers are comprised of different numbers of analytes depending on a specific diagnostic performance criterion that is selected. For example, certain combinations of biomarkers will produce tests that are more sensitive (or more specific) than other combinations.

[0247] Once a panel is defined to include a particular set of biomarkers from Table 1 and a classifier is constructed from a set of training data, the definition of the diagnostic test is complete. In one embodiment, the procedure used to classify an unknown sample is outlined in FIG. 1A. In another embodiment the procedure used to classify an unknown sample is outlined in FIG. 1B. The biological sample is appropriately diluted and then run in one or more assays to produce the relevant quantitative biomarker levels used for classification. The measured biomarker levels are used as input for the classification method that outputs a classification and an optional score for the sample that reflects the confidence of the class assignment.

[0248] Table 1 identifies 59 biomarkers that are useful for diagnosing NSCLC. This is a surprisingly larger number than expected when compared to what is typically found during biomarker discovery efforts and may be attributable to the scale of the described study, which encompassed over 1000 proteins measured in hundreds of individual samples, in some cases at concentrations in the low femtomolar range. Presumably, the large number of discovered biomarkers reflects the diverse biochemical pathways implicated in both tumor biology and the body's response to the tumor's presence; each pathway and process involves many proteins. The results show that no single protein of a small group of proteins is uniquely informative about such complex processes; rather, that multiple proteins are involved in relevant processes, such as apoptosis or extracellular matrix repair, for example.

[0249] Given the numerous biomarkers identified during the described study, one would expect to be able to derive large numbers of high-performing classifiers that can be used in various diagnostic methods. To test this notion, tens of thousands of classifiers were evaluated using the biomarkers in Table 1. As described in Example 4, many subsets of the biomarkers presented in Table 1 can be combined to generate useful classifiers. By way of example, descriptions are provided for classifiers containing 1, 2, and 3 biomarkers for

detection of NSCLC. As described in Example 4, all classifiers that were built using the biomarkers in Table 1 perform distinctly better than classifiers that were built using "non-markers".

[0250] The performance of classifiers obtained by randomly excluding some of the markers in Table 1, which resulted in smaller subsets from which to build the classifiers, was also tested. As described in Example 4, the classifiers that were built from random subsets of the markers in Table 1 performed similarly to optimal classifiers that were built using the full list of markers in Table 1.

[0251] The performance of ten-marker classifiers obtained by excluding the "best" individual markers from the ten-marker aggregation was also tested. As described in Example 4, classifiers constructed without the "best" markers in Table 1 also performed well. Many subsets of the biomarkers listed in Table 1 performed close to optimally, even after removing the top 15 of the markers listed in the Table. This implies that the performance characteristics of any particular classifier are likely not due to some small core group of biomarkers and that the disease process likely impacts numerous biochemical pathways, which alters the expression level of many proteins.

[0252] The results from Example 4 suggest certain possible conclusions: First, the identification of a large number of biomarkers enables their aggregation into a vast number of classifiers that offer similarly high performance. Second, classifiers can be constructed such that particular biomarkers may be substituted for other biomarkers in a manner that reflects the redundancies that undoubtedly pervade the complexities of the underlying disease processes. That is to say, the information about the disease contributed by any individual biomarker identified in Table 1 overlaps with the information contributed by other biomarkers, such that it may be that no particular biomarker or small group of biomarkers in Table 1 must be included in any classifier.

[0253] Exemplary embodiments use naïve Bayes classifiers constructed from the data in Table 16 to classify an unknown sample. The procedure is outlined in FIGS. 1A and 1B. In one embodiment, the biological sample is optionally diluted and run in a multiplexed aptamer assay. The data from the assay are normalized and calibrated as outlined in Example 3, and the resulting biomarker levels are used as input to a Bayes classification scheme. The loglikelihood ratio is computed for each measured biomarker individually and then summed to produce a final classification score, which is also referred to as a diagnostic score. The resulting assignment as well as the overall classification score can be reported. Optionally, the individual log-likelihood risk factors computed for each biomarker level can be reported as well. The details of the classification score calculation are presented in Example 3.

Kits

[0254] Any combination of the biomarkers of Table 1 (as well as additional biomedical information) can be detected using a suitable kit, such as for use in performing the methods disclosed herein. Furthermore, any kit can contain one or more detectable labels as described herein, such as a fluorescent moiety, etc.

[0255] In one embodiment, a kit includes (a) one or more capture reagents (such as, for example, at least one aptamer or antibody) for detecting one or more biomarkers in a biological sample, wherein the biomarkers include any of the biomarkers set forth in Table 1, and optionally (b) one or more

software or computer program products for classifying the individual from whom the biological sample was obtained as either having or not having NSCLC or for determining the likelihood that the individual has NSCLC, as further described herein. Alternatively, rather than one or more computer program products, one or more instructions for manually performing the above steps by a human can be provided.

[0256] The combination of a solid support with a corresponding capture reagent and a signal generating material is referred to herein as a “detection device” or “kit”. The kit can also include instructions for using the devices and reagents, handling the sample, and analyzing the data. Further the kit may be used with a computer system or software to analyze and report the result of the analysis of the biological sample.

[0257] The kits can also contain one or more reagents (e.g., solubilization buffers, detergents, washes, or buffers) for processing a biological sample. Any of the kits described herein can also include, e.g., buffers, blocking agents, mass spectrometry matrix materials, antibody capture agents, positive control samples, negative control samples, software and information such as protocols, guidance and reference data.

[0258] In one aspect, the invention provides kits for the analysis of NSCLC status. The kits include PCR primers for one or more biomarkers selected from Table 1. The kit may further include instructions for use and correlation of the biomarkers with NSCLC. The kit may also include a DNA array containing the complement of one or more of the biomarkers selected from Table 1, reagents, and/or enzymes for amplifying or isolating sample DNA. The kits may include reagents for real-time PCR, for example, TaqMan probes and/or primers, and enzymes.

[0259] For example, a kit can comprise (a) reagents comprising at least capture reagent for quantifying one or more biomarkers in a test sample, wherein said biomarkers comprise the biomarkers set forth in Table 1, or any other biomarkers or biomarkers panels described herein, and optionally (b) one or more algorithms or computer programs for performing the steps of comparing the amount of each biomarker quantified in the test sample to one or more predetermined cutoffs and assigning a score for each biomarker quantified based on said comparison, combining the assigned scores for each biomarker quantified to obtain a total score, comparing the total score with a predetermined score, and using said comparison to determine whether an individual has NSCLC. Alternatively, rather than one or more algorithms or computer programs, one or more instructions for manually performing the above steps by a human can be provided.

Computer Methods and Software

[0260] Once a biomarker or biomarker panel is selected, a method for diagnosing an individual can comprise the following: 1) collect or otherwise obtain a biological sample; 2) perform an analytical method to detect and measure the biomarker or biomarkers in the panel in the biological sample; 3) perform any data normalization or standardization required for the method used to collect biomarker values; 4) calculate the marker score; 5) combine the marker scores to obtain a total diagnostic score; and 6) report the individual's diagnostic score. In this approach, the diagnostic score may be a single number determined from the sum of all the marker calculations that is compared to a preset threshold value that is an indication of the presence or absence of disease. Or the diagnostic score may be a series of bars that each represent a

biomarker value and the pattern of the responses may be compared to a pre-set pattern for determination of the presence or absence of disease.

[0261] At least some embodiments of the methods described herein can be implemented with the use of a computer. An example of a computer system 100 is shown in FIG. 6. With reference to FIG. 6, system 100 is shown comprised of hardware elements that are electrically coupled via bus 108, including a processor 101, input device 102, output device 103, storage device 104, computer-readable storage media reader 105a, communications system 106 processing acceleration (e.g., DSP or special-purpose processors) 107 and memory 109. Computer-readable storage media reader 105a is further coupled to computer-readable storage media 105b, the combination comprehensively representing remote, local, fixed and/or removable storage devices plus storage media, memory, etc. for temporarily and/or more permanently containing computer-readable information, which can include storage device 104, memory 109 and/or any other such accessible system 100 resource. System 100 also comprises software elements (shown as being currently located within working memory 191) including an operating system 192 and other code 193, such as programs, data and the like.

[0262] With respect to FIG. 6, system 100 has extensive flexibility and configurability. Thus, for example, a single architecture might be utilized to implement one or more servers that can be further configured in accordance with currently desirable protocols, protocol variations, extensions, etc. However, it will be apparent to those skilled in the art that embodiments may well be utilized in accordance with more specific application requirements. For example, one or more system elements might be implemented as sub-elements within a system 100 component (e.g., within communications system 106). Customized hardware might also be utilized and/or particular elements might be implemented in hardware, software or both. Further, while connection to other computing devices such as network input/output devices (not shown) may be employed, it is to be understood that wired, wireless, modem, and/or other connection or connections to other computing devices might also be utilized.

[0263] In one aspect, the system can comprise a database containing features of biomarkers characteristic of NSCLC. The biomarker data (or biomarker information) can be utilized as an input to the computer for use as part of a computer implemented method. The biomarker data can include the data as described herein.

[0264] In one aspect, the system further comprises one or more devices for providing input data to the one or more processors.

[0265] The system further comprises a memory for storing a data set of ranked data elements.

[0266] In another aspect, the device for providing input data comprises a detector for detecting the characteristic of the data element, e.g., such as a mass spectrometer or gene chip reader.

[0267] The system additionally may comprise a database management system. User requests or queries can be formatted in an appropriate language understood by the database management system that processes the query to extract the relevant information from the database of training sets.

[0268] The system may be connectable to a network to which a network server and one or more clients are connected. The network may be a local area network (LAN) or a wide area network (WAN), as is known in the art. Preferably, the

server includes the hardware necessary for running computer program products (e.g., software) to access database data for processing user requests.

[0269] The system may include an operating system (e.g., UNIX or Linux) for executing instructions from a database management system. In one aspect, the operating system can operate on a global communications network, such as the Internet, and utilize a global communications network server to connect to such a network.

[0270] The system may include one or more devices that comprise a graphical display interface comprising interface elements such as buttons, pull down menus, scroll bars, fields for entering text, and the like as are routinely found in graphical user interfaces known in the art. Requests entered on a user interface can be transmitted to an application program in the system for formatting to search for relevant information in one or more of the system databases. Requests or queries entered by a user may be constructed in any suitable database language.

[0271] The graphical user interface may be generated by a graphical user interface code as part of the operating system and can be used to input data and/or to display inputted data. The result of processed data can be displayed in the interface, printed on a printer in communication with the system, saved in a memory device, and/or transmitted over the network or can be provided in the form of the computer readable medium.

[0272] The system can be in communication with an input device for providing data regarding data elements to the system (e.g., expression values). In one aspect, the input device can include a gene expression profiling system including, e.g., a mass spectrometer, gene chip or array reader, and the like.

[0273] The methods and apparatus for analyzing NSCLC biomarker information according to various embodiments may be implemented in any suitable manner, for example, using a computer program operating on a computer system. A conventional computer system comprising a processor and a random access memory, such as a remotely-accessible application server, network server, personal computer or workstation may be used. Additional computer system components may include memory devices or information storage systems, such as a mass storage system and a user interface, for example a conventional monitor, keyboard and tracking device. The computer system may be a stand-alone system or part of a network of computers including a server and one or more databases.

[0274] The NSCLC biomarker analysis system can provide functions and operations to complete data analysis, such as data gathering, processing, analysis, reporting and/or diagnosis. For example, in one embodiment, the computer system can execute the computer program that may receive, store, search, analyze, and report information relating to the NSCLC biomarkers. The computer program may comprise multiple modules performing various functions or operations, such as a processing module for processing raw data and generating supplemental data and an analysis module for analyzing raw data and supplemental data to generate a NSCLC status and/or diagnosis. Diagnosing NSCLC status may comprise generating or collecting any other information, including additional biomedical information, regarding the condition of the individual relative to the disease, identifying whether further tests may be desirable, or otherwise evaluating the health status of the individual.

[0275] Referring now to FIG. 7, an example of a method of utilizing a computer in accordance with principles of a disclosed embodiment can be seen. In FIG. 7, a flowchart 3000 is shown. In block 3004, biomarker information can be retrieved for an individual. The biomarker information can be retrieved from a computer database, for example, after testing of the individual's biological sample is performed. The biomarker information can comprise biomarker values that each correspond to one of at least N biomarkers selected from a group consisting of the biomarkers provided in Table 1, wherein N=2-59. In block 3008, a computer can be utilized to classify each of the biomarker values. And, in block 3012, a determination can be made as to the likelihood that an individual has NSCLC based upon a plurality of classifications. The indication can be output to a display or other indicating device so that it is viewable by a person. Thus, for example, it can be displayed on a display screen of a computer or other output device.

[0276] Referring now to FIG. 8, an alternative method of utilizing a computer in accordance with another embodiment can be illustrated via flowchart 3200. In block 3204, a computer can be utilized to retrieve biomarker information for an individual. The biomarker information comprises a biomarker value corresponding to a biomarker selected from the group of biomarkers provided in Table 1. In block 3208, a classification of the biomarker value can be performed with the computer. And, in block 3212, an indication can be made as to the likelihood that the individual has NSCLC based upon the classification. The indication can be output to a display or other indicating device so that it is viewable by a person. Thus, for example, it can be displayed on a display screen of a computer or other output device.

[0277] Some embodiments described herein can be implemented so as to include a computer program product. A computer program product may include a computer readable medium having computer readable program code embodied in the medium for causing an application program to execute on a computer with a database.

[0278] As used herein, a "computer program product" refers to an organized set of instructions in the form of natural or programming language statements that are contained on a physical media of any nature (e.g., written, electronic, magnetic, optical or otherwise) and that may be used with a computer or other automated data processing system. Such programming language statements, when executed by a computer or data processing system, cause the computer or data processing system to act in accordance with the particular content of the statements. Computer program products include without limitation: programs in source and object code and/or test or data libraries embedded in a computer readable medium. Furthermore, the computer program product that enables a computer system or data processing equipment device to act in pre-selected ways may be provided in a number of forms, including, but not limited to, original source code, assembly code, object code, machine language, encrypted or compressed versions of the foregoing and any and all equivalents.

[0279] In one aspect, a computer program product is provided for indicating a likelihood of NSCLC. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises biomarker values

that each correspond to one of at least N biomarkers in the biological sample selected from the group of biomarkers provided in Table 1, wherein N=2-59; and code that executes a classification method that indicates a NSCLC status of the individual as a function of the biomarker values.

[0280] In still another aspect, a computer program product is provided for indicating a likelihood of NSCLC. The computer program product includes a computer readable medium embodying program code executable by a processor of a computing device or system, the program code comprising: code that retrieves data attributed to a biological sample from an individual, wherein the data comprises a biomarker value corresponding to a biomarker in the biological sample selected from the group of biomarkers provided in Table 1; and code that executes a classification method that indicates a NSCLC status of the individual as a function of the biomarker value.

[0281] While various embodiments have been described as methods or apparatuses, it should be understood that embodiments can be implemented through code coupled with a computer, e.g., code resident on a computer or accessible by the computer. For example, software and databases could be utilized to implement many of the methods discussed above. Thus, in addition to embodiments accomplished by hardware, it is also noted that these embodiments can be accomplished through the use of an article of manufacture comprised of a computer usable medium having a computer readable program code embodied therein, which causes the enablement of the functions disclosed in this description. Therefore, it is desired that embodiments also be considered protected by this patent in their program code means as well. Furthermore, the embodiments may be embodied as code stored in a computer-readable memory of virtually any kind including, without limitation, RAM, ROM, magnetic media, optical media, or magneto-optical media. Even more generally, the embodiments could be implemented in software, or in hardware, or any combination thereof including, but not limited to, software running on a general purpose processor, microcode, PLAs, or ASICs.

[0282] It is also envisioned that embodiments could be accomplished as computer signals embodied in a carrier wave, as well as signals (e.g., electrical and optical) propagated through a transmission medium. Thus, the various types of information discussed above could be formatted in a structure, such as a data structure, and transmitted as an electrical signal through a transmission medium or stored on a computer readable medium.

[0283] It is also noted that many of the structures, materials, and acts recited herein can be recited as means for performing a function or step for performing a function. Therefore, it should be understood that such language is entitled to cover all such structures, materials, or acts disclosed within this specification and their equivalents, including the matter incorporated by reference.

[0284] The biomarker identification process, the utilization of the biomarkers disclosed herein, and the various methods for determining biomarker values are described in detail above with respect to NSCLC. However, the application of the process, the use of identified biomarkers, and the methods for determining biomarker values are fully applicable to other specific types of cancer, to cancer generally, to any other disease or medical condition, or to the identification of individuals who may or may not be benefited by an ancillary medical treatment. Except when referring to specific results

related to NSCLC, as is clear from the context, references herein to NSCLC may be understood to include other types of cancer, cancer generally, or any other disease or medical condition.

EXAMPLES

[0285] 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).

Example 1

Multiplexed Aptamer Analysis of Samples

[0286] This example describes the multiplex aptamer assay used to analyze the samples and controls for the identification of the biomarkers set forth in Table 1 (see FIG. 9) and the identification of the cancer biomarkers set forth in Table 19. For the NSCLC, mesothelioma, and renal cell carcinoma studies, the multiplexed analysis utilized 1,034 aptamers, each unique to a specific target.

[0287] In this method, pipette tips were changed for each solution addition.

[0288] Also, unless otherwise indicated, most solution transfers and wash additions used the 96-well head of a Beckman Biomek FxP. Method steps manually pipetted used a twelve channel P200 Pipettman (Rainin Instruments, LLC, Oakland, Calif.), unless otherwise indicated. A custom buffer referred to as SB17 was prepared in-house, comprising 40 mM HEPES, 100 mM NaCl, 5 mM KCl, 5 mM MgCl₂, 1 mM EDTA at pH 7.5. A custom buffer referred to as SB18 was prepared in-house, comprising 40 mM HEPES, 100 mM NaCl, 5 mM KCl, 5 mM MgCl₂ at pH 7.5. All steps were performed at room temperature unless otherwise indicated.

[0289] 1. Preparation of Aptamer Stock Solution

[0290] Custom stock aptamer solutions for 5%, 0.316% and 0.01% serum were prepared at 2× concentration in 1×SB17, 0.05% Tween-20.

[0291] These 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 μL of each 2× aptamer mix was manually pipetted into a 96-well Hybaid plate and the plate foil sealed. The final result was three, 96-well, foil-sealed Hybaid plates with 5%, 0.316% or 0.01% aptamer mixes. The individual aptamer concentration was 2× final or 1 nM.

[0292] 2. Assay Sample Preparation

[0293] Frozen aliquots of 100% serum or plasma, stored at −80° C., were placed in 25° C. water bath for 10 minutes. Thawed samples were placed on ice, gently vortexed (set on 4) for 8 seconds and then replaced on ice.

[0294] A 10% sample solution (2× final) was prepared by transferring 8 μL of sample using a 50 μL 8-channel spanning pipettor into 96-well Hybaid plates, each well containing 72 μL of the appropriate sample diluent at 4° C. (1×SB17 for serum or 0.8×SB18 for plasma, plus 0.06% Tween-20, 11.1

μM Z-block₂, 0.44 mM MgCl_2 , 2.2 mM AEBSF, 1.1 mM EGTA, 55.6 μM EDTA). This plate was stored on ice until the next sample dilution steps were initiated on the Biomek FxP robot.

[0295] To commence sample and aptamer equilibration, the 10% sample plate was briefly centrifuged and placed on the Beckman FX where it was mixed by pipetting up and down with the 96-well pipettor. A 0.632% sample plate (2 $\times$ final) was then prepared by diluting 6 μL of the 10% sample into 89 μL of 1 $\times$ SB17, 0.05% Tween-20 with 2 mM AEBSF. Next, dilution of 6 μL of the resultant 0.632% sample into 184 μL of 1 $\times$ SB17, 0.05% Tween-20 made a 0.02% sample plate (2 $\times$ final). Dilutions were done on the Beckman Biomek FxP. After each transfer, the solutions were mixed by pipetting up and down. The 3 sample dilution plates were then transferred to their respective aptamer solutions by adding 55 μL of the sample to 55 μL of the appropriate 2 $\times$ aptamer mix. The sample and aptamer solutions were mixed on the robot by pipetting up and down.

[0296] 3. Sample Equilibration Binding

[0297] The sample/aptamer plates were foil sealed and placed into a 37° C. incubator for 3.5 hours before proceeding to the Catch 1 step.

[0298] 4. Preparation of Catch 2 Bead Plate

[0299] An 11 mL aliquot of MyOne (Invitrogen Corp., Carlsbad, Calif.) Streptavidin C1 beads (10 mg/mL) was washed 2 times with equal volumes of 20 mM NaOH (5 minute incubation for each wash), 3 times with equal volumes of 1 $\times$ SB17, 0.05% Tween-20 and resuspended in 11 mL 1 $\times$ SB17, 0.05% Tween-20. Using a 12-span multichannel pipettor, 50 μL of this solution was manually pipetted into each well of a 96-well Hybaid plate. The plate was then covered with foil and stored at 4° C. for use in the assay.

[0300] 5. Preparation of Catch 1 Bead Plates

[0301] Three 0.45 μm Millipore HV plates (Durapore membrane, Cat# MAHVN4550) were equilibrated with 100 μL of 1 $\times$ SB17, 0.05% Tween-20 for at least 10 minutes. The equilibration buffer was then filtered through the plate and 133.3 μL of a 7.5% streptavidin-agarose bead slurry (in 1 $\times$ SB17, 0.05% Tween-20) was added into each well. To keep the streptavidin-agarose beads suspended while transferring them into the filter plate, the bead solution was manually mixed with a 200 μL , 12-channel pipettor, at least 6 times between pipetting events. After the beads were distributed across the 3 filter plates, a vacuum was applied to remove the bead supernatant. Finally, the beads were washed in the filter plates with 200 μL 1 $\times$ SB17, 0.05% Tween-20 and then resuspended in 200 μL 1 $\times$ SB17, 0.05% Tween-20. The bottoms of the filter plates were blotted and the plates stored for use in the assay.

[0302] 6. Loading the Cytomat

[0303] The cytomat was loaded with all tips, plates, all reagents in troughs (except NHS-biotin reagent which was prepared fresh right before addition to the plates), 3 prepared Catch 1 filter plates and 1 prepared MyOne plate.

[0304] 7. Catch 1

[0305] After a 3.5 hour equilibration time, the sample/aptamer plates were removed from the incubator, centrifuged for about 1 minute, cover removed, and placed on the deck of the Beckman Biomek FxP. The Beckman Biomek FxP program was initiated. All subsequent steps in Catch 1 were performed by the Beckman Biomek FxP robot unless otherwise noted. Within the program, the vacuum was applied to the Catch 1 filter plates to remove the bead supernatant. One

hundred microlitres of each of the 5%, 0.316% and 0.01% equilibration binding reactions were added to their respective Catch 1 filtration plates, and each plate was mixed using an on-deck orbital shaker at 800 rpm for 10 minutes.

[0306] Unbound solution was removed via vacuum filtration. The Catch 1 beads were washed with 190 μL of 100 μM biotin in 1 $\times$ SB17, 0.05% Tween-20 followed by 5 $\times$ 190 μL of 1 $\times$ SB17, 0.05% Tween-20 by dispensing the solution and immediately drawing a vacuum to filter the solution through the plate.

[0307] 8. Tagging

[0308] A 100 mM NHS-PEO4-biotin aliquot in anhydrous DMSO was thawed at 37° C. for 6 minutes and then diluted 1:100 with tagging buffer (SB17 at pH 7.25, 0.05% Tween-20). Upon a robot prompt, the diluted NHS-PEO4-biotin reagent was manually added to an on-deck trough and the robot program was manually re-initiated to dispense 100 μL of the NHS-PEO4-biotin into each well of each Catch 1 filter plate. This solution was allowed to incubate with Catch 1 beads shaking at 800 rpm for 5 minutes on the orbital shakers.

[0309] 9. Kinetic Challenge and Photo-Cleavage

[0310] The tagging reaction was removed by vacuum filtration and quenched by the addition of 150 μL of 20 mM glycine in 1 $\times$ SB17, 0.05% Tween-20 to the Catch 1 plates. The NHS-tag/glycine solution was removed via vacuum filtration. Next, 1500 μL 20 mM glycine (1 $\times$ SB17, 0.05% Tween-20) was added to each plate and incubated for 1 minute on orbital shakers at 800 rpm before removal by vacuum filtration.

[0311] The wells of the Catch 1 plates were subsequently washed three times by adding 190 μL 1 $\times$ SB17, 0.05% Tween-20, followed by vacuum filtration and then by adding 190 μL 1 $\times$ SB17, 0.05% Tween-20 with shaking for 1 minute at 800 rpm followed by vacuum filtration. After the last wash the plates were placed on top of a 1 mL deep-well plate and removed from the deck. The Catch 1 plates were centrifuged at 1000 rpm for 1 minute to remove as much extraneous volume from the agarose beads before elution as possible.

[0312] The plates were placed back onto the Beckman Biomek FxP and 85 μL of 10 mM DxSO_4 in 1 $\times$ SB17, 0.05% Tween-20 was added to each well of the filter plates.

[0313] The filter plates were removed from the deck, placed onto a Variomag Thermoshaker (Thermo Fisher Scientific, Inc., Waltham, Mass.) under the BlackRay (Ted Pella, Inc., Redding, Calif.) light sources, and irradiated for 5 minutes while shaking at 800 rpm. After the 5 minute incubation the plates were rotated 180 degrees and irradiated with shaking for 5 minutes more.

[0314] The photocleaved solutions were sequentially eluted from each Catch 1 plate into a common deep well plate by first placing the 5% Catch 1 filter plate on top of a 1 mL deep-well plate and centrifuging at 1000 rpm for 1 minute. The 0.316% and 0.01% Catch 1 plates were then sequentially centrifuged into the same deep well plate.

[0315] 10. Catch 2 Bead Capture

[0316] The 1 mL deep well block containing the combined eluates of Catch 1 was placed on the deck of the Beckman Biomek FxP for Catch 2.

[0317] The robot transferred all of the photo-cleaved eluate from the 1 mL deep-well plate onto the Hybaid plate containing the previously prepared Catch 2 MyOne magnetic beads (after removal of the MyOne buffer via magnetic separation).

[0318] The solution was incubated while shaking at 1350 rpm for 5 minutes at 25° C. on a Variomag Thermoshaker (Thermo Fisher Scientific, Inc., Waltham, Mass.).

[0319] The robot transferred the plate to the on deck magnetic separator station. The plate was incubated on the magnet for 90 seconds before removal and discarding of the supernatant.

[0320] 11. 37° C. 30% Glycerol Washes

[0321] The Catch 2 plate was moved to the on-deck thermal shaker and 75 μ L of 1 $\times$ SB17, 0.05% Tween-20 was transferred to each well. The plate was mixed for 1 minute at 1350 rpm and 37° C. to resuspend and warm the beads. To each well of the Catch 2 plate, 75 μ L of 60% glycerol at 37° C. was transferred and the plate continued to mix for another minute at 1350 rpm and 37° C. The robot transferred the plate to the 37° C. magnetic separator where it was incubated on the magnet for 2 minutes and then the robot removed and discarded the supernatant. These washes were repeated two more times.

[0322] After removal of the third 30% glycerol wash from the Catch 2 beads, 150 μ L of 1 $\times$ SB17, 0.05% Tween-20 was added to each well and incubated at 37° C., shaking at 1350 rpm for 1 minute, before removal by magnetic separation on the 37° C. magnet.

[0323] The Catch 2 beads were washed a final time using 150 μ L 1 $\times$ SB17, 0.05% Tween-20 with incubation for 1 minute while shaking at 1350 rpm at 25° C. prior to magnetic separation.

[0324] 12. Catch 2 Bead Elution and Neutralization

[0325] The aptamers were eluted from Catch 2 beads by adding 105 μ L of 100 mM CAPSO with 1M NaCl, 0.05% Tween-20 to each well. The beads were incubated with this solution with shaking at 1300 rpm for 5 minutes.

[0326] The Catch 2 plate was then placed onto the magnetic separator for 90 seconds prior to transferring 63 μ L of the eluate to a new 96-well plate containing 7 μ L of 500 mM HCl, 500 mM HEPES, 0.05% Tween-20 in each well. After transfer, the solution was mixed robotically by pipetting 60 μ L up and down five times.

[0327] 13. Hybridization

[0328] The Beckman Biomek FxP transferred 20 μ L of the neutralized Catch 2 eluate to a fresh Hybaid plate, and 6 μ L of 10 $\times$ Agilent Block, containing a 10 $\times$ spike of hybridization controls, was added to each well. Next, 30 μ L of 2 $\times$ Agilent Hybridization buffer was manually pipetted to the each well of the plate containing the neutralized samples and blocking buffer and the solution was mixed by manually pipetting 25 μ L up and down 15 times slowly to avoid extensive bubble formation. The plate was spun at 1000 rpm for 1 minute.

[0329] Custom Agilent microarray slides (Agilent Technologies, Inc., Santa Clara, Calif.) were designed to contain probes complementary to the aptamer random region plus some primer region. For the majority of the aptamers, the optimal length of the complementary sequence was empirically determined and ranged between 40-50 nucleotides. For later aptamers a 46-mer complementary region was chosen by default. The probes were linked to the slide surface with a poly-T linker for a total probe length of 60 nucleotides.

[0330] A gasket slide was placed into an Agilent hybridization chamber and 40 μ L of each of the samples containing hybridization and blocking solution was manually pipetted into each gasket. An 8-channel variable spanning pipettor was used in a manner intended to minimize bubble formation. Custom Agilent microarray slides (Agilent Technologies,

Inc., Santa Clara, Calif.), with their Number Barcode facing up, were then slowly lowered onto the gasket slides (see Agilent manual for detailed description).

[0331] The top of the hybridization chambers were placed onto the slide/backing sandwich and clamping brackets slid over the whole assembly. These assemblies were tightly clamped by turning the screws securely.

[0332] Each slide/backing slide sandwich was visually inspected to assure the solution bubble could move freely within the sample. If the bubble did not move freely the hybridization chamber assembly was gently tapped to disengage bubbles lodged near the gasket.

[0333] The assembled hybridization chambers were incubated in an Agilent hybridization oven for 19 hours at 60° C. rotating at 20 rpm.

[0334] 14. Post Hybridization Washing

[0335] Approximately 400 mL Agilent Wash Buffer 1 was placed into each of two separate glass staining dishes. One of the staining dishes was placed on a magnetic stir plate and a slide rack and stir bar were placed into the buffer.

[0336] A staining dish for Agilent Wash 2 was prepared by placing a stir bar into an empty glass staining dish.

[0337] A fourth glass staining dish was set aside for the final acetonitrile wash.

[0338] Each of six hybridization chambers was disassembled. One-by-one, the slide/backing sandwich was removed from its hybridization chamber and submerged into the staining dish containing Wash 1. The slide/backing sandwich was pried apart using a pair of tweezers, while still submerging the microarray slide. The slide was quickly transferred into the slide rack in the Wash 1 staining dish on the magnetic stir plate.

[0339] The slide rack was gently raised and lowered 5 times. The magnetic stirrer was turned on at a low setting and the slides incubated for 5 minutes.

[0340] When one minute was remaining for Wash 1, Wash Buffer 2 pre-warmed to 37° C. in an incubator was added to the second prepared staining dish. The slide rack was quickly transferred to Wash Buffer 2 and any excess buffer on the bottom of the rack was removed by scraping it on the top of the stain dish. The slide rack was gently raised and lowered 5 times. The magnetic stirrer was turned on at a low setting and the slides incubated for 5 minutes.

[0341] The slide rack was slowly pulled out of Wash 2, taking approximately 15 seconds to remove the slides from the solution.

[0342] With one minute remaining in Wash 2 acetonitrile (ACN) was added to the fourth staining dish. The slide rack was transferred to the acetonitrile stain dish. The slide rack was gently raised and lowered 5 times. The magnetic stirrer was turned on at a low setting and the slides incubated for 5 minutes.

[0343] The slide rack was slowly pulled out of the ACN stain dish and placed on an absorbent towel. The bottom edges of the slides were quickly dried and the slide was placed into a clean slide box.

[0344] 15. Microarray Imaging

[0345] The microarray slides were placed into Agilent scanner slide holders and loaded into the Agilent Microarray scanner according to the manufacturers instructions.

[0346] The slides were imaged in the Cy3-channel at 5 μ m resolution at the 100% PMT setting and the XRD option

enabled at 0.05. The resulting tiff images were processed using Agilent feature extraction software version 10.5.

Example 2

Biomarker Identification

[0347] The identification of potential NSCLC biomarkers was performed for diagnosis of NSCLC in individuals 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. Enrollment criteria for this study were smokers, age 18 or older, able to give informed consent, and blood sample and documented diagnosis of NSCLC or benign findings. For cases, blood samples collected prior to treatment or surgery and subsequently diagnosed with NSCLC. Exclusion criteria included prior diagnosis or treatment of cancer (excluding squamous cell carcinoma of the skin) within 5 years of the blood draw. Serum samples were collected from 3 different sites and included 46 NSCLC samples and 218 control group samples as described in Table 17. The multiplexed aptamer affinity assay as described in Example 1 was used to measure and report the RFU value for 1,034 analytes in each of these 264 samples.

[0348] Each of the case and control populations were separately compared by generating class-dependent cumulative distribution functions (cdfs) for each of the 1,034 analytes. The KS-distance (Kolmogorov-Smirnov statistic) between values from two sets of samples is a non parametric measurement of the extent to which the empirical distribution of the values from one set (Set A) differs from the distribution of values from the other set (Set B). For any value of a threshold T some proportion of the values from Set A will be less than T, and some proportion of the values from Set B will be less than T. The KS-distance measures the maximum (unsigned) difference between the proportion of the values from the two sets for any choice of T.

[0349] This set of potential biomarkers can be used to build classifiers that assign samples to either a control or disease group. In fact, many such classifiers were produced from these sets of biomarkers and the frequency with which any biomarker was used in good scoring classifiers determined. Those biomarkers that occurred most frequently among the top scoring classifiers were the most useful for creating a diagnostic test. In this example, Bayesian classifiers were used to explore the classification space but many other supervised learning techniques may be employed for this purpose. The scoring fitness of any individual classifier was gauged by the area under the receiver operating characteristic curve (AUC of the ROC) of the classifier at the Bayesian surface assuming a disease prevalence of 0.5. This scoring metric varies from zero to one, with one being an error-free classifier. The details of constructing a Bayesian classifier from biomarker population measurements are described in Example 3.

[0350] Using the 59 analytes in Table 1, a total of 964 10-analyte classifiers were found with an AUC of 0.94 for diagnosing NSCLC from the control group. From this set of classifiers, a total of 12 biomarkers were found to be present in 30% or more of the high scoring classifiers. Table 13 provides a list of these potential biomarkers and FIG. 10 is a frequency plot for the identified biomarkers.

Example 3

Naïve Bayesian Classification for NSCLC

[0351] From the list of biomarkers identified as useful for discriminating between NSCLC and controls, a panel of ten

biomarkers was selected and a naïve Bayes classifier was constructed, see Tables 16 and 18. The class-dependent probability density functions (pdfs), $p(x_i|c)$ and $p(x_i|d)$, where x_i is the log of the measured RFU value for biomarker i , and c and d refer to the control and disease populations, were modeled as log-normal distribution functions characterized by a mean μ and variance σ^2 . The parameters for pdfs of the ten biomarkers are listed in Table 16 and an example of the raw data along with the model fit to a normal pdf is displayed in FIG. 5. The underlying assumption appears to fit the data quite well as evidenced by FIG. 5.

[0352] The naïve Bayes classification for such a model is given by the following equation, where $p(d)$ is the prevalence of the disease in the population,

$$\ln\left(\frac{p(d|\bar{x})}{p(c|\bar{x})}\right) = \sum_{i=1}^n \ln\left(\frac{\sigma_{c,i}}{\sigma_{d,i}}\right) - \frac{1}{2} \sum_{i=1}^n \left[\left(\frac{x_i - \mu_{d,i}}{\sigma_{d,i}} \right)^2 - \left(\frac{x_i - \mu_{c,i}}{\sigma_{c,i}} \right)^2 \right] + \ln\left(\frac{p(d)}{1-p(d)}\right)$$

appropriate to the test and $n=10$. Each of the terms in the summation is a log-likelihood ratio for an individual marker and the total log-likelihood ratio of a sample $\bar{x}$ being free from the disease of interest (i.e. in this case, NSCLC) versus having the disease is simply the sum of these individual terms plus a term that accounts for the prevalence of the disease. For simplicity, we assume $p(d)=0.5$ so that

$$\ln\left(\frac{p(d)}{1-p(d)}\right) = 0.$$

[0353] Given an unknown sample measurement in log (RFU) for each of the ten biomarkers of 6.9, 8.7, 7.9, 9.8, 8.4, 10.6, 7.3, 6.3, 7.3, 8.1, the calculation of the classification is detailed in Table 16. The individual components comprising the log likelihood ratio for disease versus control class are tabulated and can be computed from the parameters in Table 16 and the values of $\bar{x}$. The sum of the individual log likelihood ratios is -11.584 , or a likelihood of being free from the disease versus having the disease of $107,386$, where likelihood $e^{11.584}=107,386$. The first 3 biomarker values have likelihoods more consistent with the disease group (log likelihood >0) but the remaining 7 biomarkers are all consistently found to favor the control group. Multiplying the likelihoods together gives the same results as that shown above; a likelihood of $107,386$ that the unknown sample is free from the disease. In fact, this sample came from the control population in the training set.

Example 4

Greedy Algorithm for Selecting Biomarker Panels for Classifiers

[0354] This example describes the selection of biomarkers from Table 1 to form panels that can be used as classifiers in any of the methods described herein. Subsets of the biomarkers in Table 1 were selected to construct classifiers with good performance. This method was also used to determine which potential markers were included as biomarkers in Example 2.

[0355] The measure of classifier performance used here is the AUC; a performance of 0.5 is the baseline expectation for a random (coin toss) classifier, a classifier worse than random would score between 0.0 and 0.5, a classifier with better than random performance would score between 0.5 and 1.0. A perfect classifier with no errors would have a sensitivity of 1.0 and a specificity of 1.0. One can apply the methods described in Example 4 to other common measures of performance such as the F-measure, the sum of sensitivity and specificity, or the product of sensitivity and specificity. Specifically one might want to treat sensitivity and specificity with differing weight, so as to select those classifiers which perform with higher specificity at the expense of some sensitivity, or to select those classifiers which perform with higher sensitivity at the expense of some specificity. Since the method described here only involves a measure of “performance”, any weighting scheme which results in a single performance measure can be used. Different applications will have different benefits for true positive and true negative findings, and also different costs associated with false positive findings from false negative findings. For example, screening asymptomatic smokers and the differential diagnosis of benign nodules found on CT will not in general have the same optimal trade-off between specificity and sensitivity. The different demands of the two tests will in general require setting different weighting to positive and negative misclassifications, reflected in the performance measure. Changing the performance measure will in general change the exact subset of markers selected from Table 1 for a given set of data.

[0356] For the Bayesian approach to the discrimination of NSCLC samples from control samples described in Example 3, the classifier was completely parameterized by the distributions of biomarkers in the disease and benign training samples, and the list of biomarkers was chosen from Table 1; that is to say, the subset of markers chosen for inclusion determined a classifier in a one-to-one manner given a set of training data.

[0357] The greedy method employed here was used to search for the optimal subset of markers from Table 1. For small numbers of markers or classifiers with relatively few markers, every possible subset of markers was enumerated and evaluated in terms of the performance of the classifier constructed with that particular set of markers (see Example 4, Part 2). (This approach is well known in the field of statistics as “best subset selection”; see, e.g., Hastie et al.). However, for the classifiers described herein, the number of combinations of multiple markers can be very large, and it was not feasible to evaluate every possible set of 10 markers, as there are 30,045,015 possible combinations that can be generated from a list of only 30 total analytes. Because of the impracticality of searching through every subset of markers, the single optimal subset may not be found; however, by using this approach, many excellent subsets were found, and, in many cases, any of these subsets may represent an optimal one.

[0358] Instead of evaluating every possible set of markers, a “greedy” forward stepwise approach may be followed (see, e.g., Dabney A R, Storey J D (2007) Optimality Driven Nearest Centroid Classification from Genomic Data. *PLoS ONE* 2(10): e1002. doi:10.1371/journal.pone.0001002). Using this method, a classifier is started with the best single marker (based on KS-distance for the individual markers) and is grown at each step by trying, in turn, each member of a marker list that is not currently a member of the set of markers in the

classifier. The one marker which scores best in combination with the existing classifier is added to the classifier. This is repeated until no further improvement in performance is achieved. Unfortunately, this approach may miss valuable combinations of markers for which some of the individual markers are not all chosen before the process stops.

[0359] The greedy procedure used here was an elaboration of the preceding forward stepwise approach, in that, to broaden the search, rather than keeping just a single candidate classifier (marker subset) at each step, a list of candidate classifiers was kept. The list was seeded with every single marker subset (using every marker in the table on its own). The list was expanded in steps by deriving new classifiers (marker subsets) from the ones currently on the list and adding them to the list. Each marker subset currently on the list was extended by adding any marker from Table 1 not already part of that classifier, and which would not, on its addition to the subset, duplicate an existing subset (these are termed “permissible markers”). Every existing marker subset was extended by every permissible marker from the list. Clearly, such a process would eventually generate every possible subset, and the list would run out of space. Therefore, all the generated classifiers were kept only while the list was less than some predetermined size (often enough to hold all three marker subsets). Once the list reached the predetermined size limit, it became elitist; that is, only those classifiers which showed a certain level of performance were kept on the list, and the others fell off the end of the list and were lost. This was achieved by keeping the list sorted in order of classifier performance; new classifiers which were at least as good as the worst classifier currently on the list were inserted, forcing the expulsion of the current bottom underachiever. One further implementation detail is that the list was completely replaced on each generational step; therefore, every classifier on the list had the same number of markers, and at each step the number of markers per classifier grew by one.

[0360] Since this method produced a list of candidate classifiers using different combinations of markers, one may ask if the classifiers can be combined in order to avoid errors which might be made by the best single classifier, or by minority groups of the best classifiers. Such “ensemble” and “committee of experts” methods are well known in the fields of statistical and machine learning and include, for example, “Averaging”, “Voting”, “Stacking”, “Bagging” and “Boosting” (see, e.g., Hastie et al.). These combinations of simple classifiers provide a method for reducing the variance in the classifications due to noise in any particular set of markers by including several different classifiers and therefore information from a larger set of the markers from the biomarker table, effectively averaging between the classifiers. An example of the usefulness of this approach is that it can prevent outliers in a single marker from adversely affecting the classification of a single sample. The requirement to measure a larger number of signals may be impractical in conventional “one marker at a time” antibody assays but has no downside for a fully multiplexed aptamer assay. Techniques such as these benefit from a more extensive table of biomarkers and use the multiple sources of information concerning the disease processes to provide a more robust classification.

[0361] The biomarkers selected in Table 1 gave rise to classifiers which perform better than classifiers built with “non-markers” (i.e., proteins having signals that did not meet the criteria for inclusion in Table 1 (as described in Example 2)).

[0362] For classifiers containing only one, two, and three markers, all possible classifiers obtained using the biomarkers in Table 1 were enumerated and examined for the distribution of performance compared to classifiers built from a similar table of randomly selected non-markers signals.

[0363] In FIG. 11, the AUC was used as the measure of performance; a performance of 0.5 is the baseline expectation for a random (coin toss) classifier. The histogram of classifier performance was compared with the histogram of performance from a similar exhaustive enumeration of classifiers built from a “non-marker” table of 59 non-marker signals; the 59 signals were randomly chosen from aptamers that did not demonstrate differential signaling between control and disease populations.

[0364] FIG. 11 shows histograms of the performance of all possible one, two, and three-marker classifiers built from the biomarker parameters in Table 14 for biomarkers that can discriminate between the control group and NSCLC and compares these classifiers with all possible one, two, and three-marker classifiers built using the 59 “non-marker” aptamer RFU signals. FIG. 11A shows the histograms of single marker classifier performance, FIG. 11B shows the histogram of two marker classifier performance, and FIG. 11C shows the histogram of three marker classifier performance.

[0365] In FIG. 11, the solid lines represent the histograms of the classifier performance of all one, two, and three-marker classifiers using the biomarker data for smokers and benign pulmonary nodules and NSCLC in Table 14. The dotted lines are the histograms of the classifier performance of all one, two, and three-marker classifiers using the data for controls and NSCLC but using the set of random non-marker signals.

[0366] The classifiers built from the markers listed in Table 1 form a distinct histogram, well separated from the classifiers built with signals from the “non-markers” for all one-marker, two-marker, and three-marker comparisons. The performance and AUC score of the classifiers built from the biomarkers in Table 1 also increase faster with the number of markers than do the classifiers built from the non-markers, the separation increases between the marker and non-marker classifiers as the number of markers per classifier increases. All classifiers built using the biomarkers listed in Table 14 perform distinctly better than classifiers built using the “non-markers”.

[0367] The distributions of classifier performance show that there are many possible multiple-marker classifiers that can be derived from the set of analytes in Table 1. Although some biomarkers are better than others on their own, as evidenced by the distribution of classifier scores and AUCs for single analytes, it was desirable to determine whether such biomarkers are required to construct high performing classifiers. To make this determination, the behavior of classifier performance was examined by leaving out some number of the best biomarkers. FIG. 12 compares the performance of classifiers built with the full list of biomarkers in Table 1 with the performance of classifiers built with subsets of biomarkers from Table 1 that excluded top-ranked markers.

[0368] FIG. 12 demonstrates that classifiers constructed without the best markers perform well, implying that the performance of the classifiers was not due to some small core group of markers and that the changes in the underlying processes associated with disease are reflected in the activities of many proteins. Many subsets of the biomarkers in Table 1 performed close to optimally, even after removing the top 15 of the 59 markers from Table 1. After dropping the 15 top-ranked markers (ranked by KS-distance) from Table 1, the classifier performance increased with the number of

markers selected from the table to reach an AUC of almost 0.93, close to the performance of the optimal classifier score of 0.948 selected from the full list of biomarkers.

[0369] Finally, FIG. 13 shows how the ROC performance of typical classifiers constructed from the list of parameters in Table 14 according to Example 3. A five analyte classifier was constructed with MMP7, CLIC1, STX1A, CHRDL1, and PA2G4. FIG. 13A shows the performance of the model, assuming independence of these markers, as in Example 3, and FIG. 13B shows the empirical ROC curves generated from the study data set used to define the parameters in Table 14. It can be seen that the performance for a given number of selected markers was qualitatively in agreement, and that quantitative agreement was generally quite good, as evidenced by the AUCs, although the model calculation tends to overestimate classifier performance. This is consistent with the notion that the information contributed by any particular biomarker concerning the disease processes is redundant with the information contributed by other biomarkers provided in Table 1 while the model calculation assumes complete independence. FIG. 13 thus demonstrates that Table 1 in combination with the methods described in Example 3 enable the construction and evaluation of a great many classifiers useful for the discrimination of NSCLC from the control group.

Example 5

Clinical Biomarker Panel

[0370] A random forest classifier was built from a panel of biomarkers selected that may be the most appropriate for use in a clinical diagnostic test. Unlike the models selected by the naive Bayes greedy forward algorithm, the random forest classifier does not assume that the biomarker measurements are randomly distributed. Therefore this model can utilize biomarkers from Table 1 that are not effective in the naïve Bayes classifier.

[0371] The panel was selected using a backward elimination procedure that utilized the gini importance measure provided by the random forest classifier. The gini importance is a measure of the effectiveness of a biomarker at correctly classifying samples in the training set.

[0372] This measure of biomarker importance can be used to eliminate markers that are less vital to the performance of the classifier. The backward elimination procedure was initiated by building a random forest classifier that included all 59 in Table 1. The least important biomarker was then eliminated and a new model was built with the remaining biomarkers. This procedure continued until only single biomarkers remained.

[0373] The final panel that was selected provided the best balance between the greatest AUC and the lowest number of markers in the model. The panel of 8 biomarkers that satisfied these criteria is composed of the following analytes, MMP12, MMP7, KLK3-SERPINA3, CRP, C9, CNDP1, CA6, and EGFR. A plot of the ROC curve for this biomarker panel is shown in FIG. 14. The sensitivity of this model is 0.70 with a corresponding specificity of 0.89.

Example 6

Biomarkers for the Diagnosis of Cancer

[0374] The identification of potential biomarkers for the general diagnosis of cancer was performed. Both case and control samples were evaluated from 3 different types of cancer (lung cancer, mesothelioma, and renal cell carcinoma). Across the sites, inclusion criteria were at least 18

years old with signed informed consent. Both cases and controls were excluded for known malignancy other than the cancer in question.

[0375] Lung Cancer. Case and control samples were obtained as described in Example 2. A total of 46 cases and 218 controls were used in this Example.

[0376] Pleural Mesothelioma. Case and control samples were obtained from an academic cancer center biorepository to identify potential markers for the differential diagnosis of pleural mesothelioma from benign lung disease, including suspicious radiology findings that were later diagnosed as non-malignant. A total of 124 mesothelioma cases and 138 asbestos exposed controls were used in this Example.

[0377] Renal Cell Carcinoma. Case and control samples were obtained from an academic cancer center biorepository from patients with renal cell carcinoma (RCC) and benign masses (BEN). Pre-surgical samples (TP1) were obtained for all subjects. The primary analysis compared outcome data (as recorded in the SEER database field CA Status 1) for the RCC patients with "Evidence of Disease" (EVD) vs "No Evidence of Disease" (NED) documented through clinical follow-up. A total of 38 EVD cases and 104 NED controls were used in this Example.

[0378] A final list of cancer biomarkers was identified by combining the sets of biomarkers considered for each of the 3 different cancer studies. Bayesian classifiers that used biomarker sets of increasing size were successively constructed using a greedy algorithm (as described in greater detail in Section 6.2 of this Example). The sets (or panels) of biomarkers that were useful for diagnosing cancer in general among the different sites and types of cancer were compiled as a function of set (or panel) size and analyzed for their performance. This analysis resulted in the list of 23 cancer biomarkers shown in Table 19, each of which was present in at least one of these successive marker sets, which ranged in size from three to ten markers. As an illustrative example, we describe the generation of a specific panel composed of ten cancer biomarkers, which is shown in Table 32.

6.1 Naïve Bayesian Classification for Cancer

[0379] From the list of biomarkers in Table 1, a panel of ten potential cancer biomarkers was selected using a greedy algorithm for biomarker selection, as outlined in Section 6.2 of this Example. A distinct naïve Bayes classifier was constructed for each of the 3. The class-dependent probability density functions (pdfs), $p(x_i|c)$ and $p(x_i|d)$, where x_i is the log of the measured RFU value for biomarker i , and c and d refer to the control and disease populations, were modeled as log-normal distribution functions characterized by a mean μ and variance σ^2 . The parameters for pdfs of the 3 models composed of the ten potential biomarkers are listed in Table 31.

[0380] The naïve Bayes classification for such a model is given by the following equation, where $p(d)$ is the prevalence of the disease in the population

$$\ln\left(\frac{p(d|\bar{x})}{p(c|\bar{x})}\right) = \sum_{i=1}^n \ln\left(\frac{\sigma_{c,i}}{\sigma_{d,i}}\right) - \frac{1}{2} \sum_{i=1}^n \left[\left(\frac{x_i - \mu_{d,i}}{\sigma_{d,i}} \right)^2 - \left(\frac{x_i - \mu_{c,i}}{\sigma_{c,i}} \right)^2 \right] + \ln\left(\frac{p(d)}{1-p(d)}\right)$$

appropriate to the test and $n=10$. Each of the terms in the summation is a log-likelihood ratio for an individual marker and the total log-likelihood ratio of a sample being free from the disease interest (i.e., in this case, each particular disease from the 3 different cancer types) versus having the disease is

simply the sum of these individual terms plus a term that accounts for the prevalence of the disease. For simplicity, we assume $p(d)=0.5$ so that

$$\ln\left(\frac{p(d)}{1-p(d)}\right) = 0.$$

[0381] Given an unknown sample measurement in log (RFU) for each of the ten biomarkers of 9.5, 8.8, 7.8, 8.3, 9.4, 7.0, 7.9, 6.3, 7.7, 10.6, the calculation of the classification is detailed in Table 32. The individual components comprising the log likelihood ratio for disease versus control class are tabulated and can be computed from the parameters in Table 31 and the values of The sum of the individual log likelihood ratios is -3.326 , or a likelihood of being free from the disease versus having the disease of 28, where likelihood $e^{-3.326}=28$. The first 4 biomarker values have likelihoods more consistent with the disease group (log likelihood >0) but the remaining 6 biomarkers are all consistently found to favor the control group. Multiplying the likelihoods together gives the same results as that shown above; a likelihood of 28 that the unknown sample is free from the disease. In fact, this sample came from the control population in the renal cell carcinoma training set.

6.1 Naïve Bayesian Classification for Cancer

[0382] From the list of biomarkers in Table 1, a panel of ten potential cancer biomarkers was selected using a greedy algorithm for biomarker selection, as outlined in Section 6.2 of this Example. A distinct naïve Bayes classifier was constructed for each of the 3 different cancer types. The class-dependent probability density functions (pdfs), $p(x_i|c)$ and $p(x_i|d)$, where x_i is the log of the measured RFU value for biomarker i , and c and d refer to the control and disease populations, were modeled as log-normal distribution functions characterized by a mean μ and variance σ^2 . The parameters for pdfs of the 3 models composed of the ten potential biomarkers are listed in Table 31.

[0383] The naïve Bayes classification for such a model is given by the following equation, where $p(d)$ is the prevalence of the disease in the population,

$$\ln\left(\frac{p(d|\bar{x})}{p(c|\bar{x})}\right) = \sum_{i=1}^n \ln\left(\frac{\sigma_{c,i}}{\sigma_{d,i}}\right) - \frac{1}{2} \sum_{i=1}^n \left[\left(\frac{x_i - \mu_{d,i}}{\sigma_{d,i}} \right)^2 - \left(\frac{x_i - \mu_{c,i}}{\sigma_{c,i}} \right)^2 \right] + \ln\left(\frac{p(d)}{1-p(d)}\right)$$

appropriate to the test and $n=10$. Each of the terms in the summation is a log-likelihood ratio for an individual marker and the total log-likelihood ratio of a sample x being free from the disease interest (i.e., in this case, each particular disease from the 3 different cancer types) versus having the disease is simply the sum of these individual terms plus a term that accounts for the prevalence of the disease. For simplicity, we assume $p(d)=0.5$ so that

$$\ln\left(\frac{p(d)}{1-p(d)}\right) = 0.$$

[0384] Given an unknown sample measurement in log (RFU) for each of the ten biomarkers of 9.5, 8.8, 7.8, 8.3, 9.4, 7.0, 7.9, 6.3, 7.7, 10.6, the calculation of the classification is detailed in Table 32. The individual components comprising the log likelihood ratio for disease versus control class are tabulated and can be computed from the parameters in Table 31 and the values of x . The sum of the individual log likelihood ratios is -3.326 , or a likelihood of being free from the disease versus having the disease of 28, where likelihood $e^{3.326}=28$. Only 4 of the biomarker values have likelihoods more consistent with the disease group (log likelihood >0) but the remaining 6 biomarkers are all consistently found to favor the control group. Multiplying the likelihoods together gives the same results as that shown above; a likelihood of 28 that the unknown sample is free from the disease. In fact, this sample came from the control population in the NSCLC training set.

6.2 Greedy Algorithm for Selecting Cancer Biomarker Panels for Classifiers

Part 1

[0385] Subsets of the biomarkers in Table 1 were selected to construct potential classifiers that could be used to determine which of the markers could be used as general cancer biomarkers to detect cancer.

[0386] Given a set of markers, a distinct model was trained for each of the 3 cancer studies, so a global measure of performance was required to select a set of biomarkers that was able to classify simultaneously many different types of cancer. The measure of classifier performance used here was the mean of the area under ROC curve across all naïve Bayes classifiers. The ROC curve is a plot of a single classifier true positive rate (sensitivity) versus the false positive rate (1-specificity). The area under the ROC curve (AUC) ranges from 0 to 1.0, where an AUC of 1.0 corresponds to perfect classification and an AUC of 0.5 corresponds to random (coin toss) classifier. One can apply other common measures of performance such as the F-measure or the sum or product of sensitivity and specificity. Specifically, one might want to treat sensitivity and specificity with differing weight, in order to select those classifiers that perform with higher specificity at the expense of some sensitivity, or to select those classifiers which perform with higher sensitivity at the expense of specificity. We chose to use the AUC because it encompasses all combinations of sensitivity and specificity in a single measure. Different applications will have different benefits for true positive and true negative findings, and will have different costs associated with false positive findings from false negative findings. Changing the performance measure may change the exact subset of markers selected for a given set of data.

[0387] For the Bayesian approach to the discrimination of cancer samples from control samples described in Section 6.1 of this Example, the classifier was completely parameterized by the distributions of biomarkers in each of the 3 cancer studies, and the list of biomarkers was chosen from Table 19. That is to say, the subset of markers chosen for inclusion determined a classifier in a one-to-one manner given a set of training data.

[0388] The greedy method employed here was used to search for the optimal subset of markers from Table 1. For small numbers of markers or classifiers with relatively few markers, every possible subset of markers was enumerated

and evaluated in terms of the performance of the classifier constructed with that particular set of markers (see Example 4). (This approach is well known in the field of statistics as “best subset selection”; see, e.g., Hastie et al). However, for the classifiers described herein, the number of combinations of multiple markers can be very large, and it was not feasible to evaluate every possible set of 10 markers, as there are 30,045,015 possible combinations that can be generated from a list of only 30 total analytes. Because of the impracticality of searching through every subset of markers, the single optimal subset may not be found; however, by using this approach, many excellent subsets were found, and, in many cases, any of these subsets may represent an optimal one.

[0389] Instead of evaluating every possible set of markers, a “greedy” forward stepwise approach may be followed (see, e.g., Dabney A R, Storey J D (2007) Optimality Driven Nearest Centroid Classification from Genomic Data. *PLoS ONE* 2(10): e1002. doi:10.1371/journal.pone.0001002). Using this method, a classifier is started with the best single marker (based on KS-distance for the individual markers) and is grown at each step by trying, in turn, each member of a marker list that is not currently a member of the set of markers in the classifier. The one marker that scores the best in combination with the existing classifier is added to the classifier. This is repeated until no further improvement in performance is achieved. Unfortunately, this approach may miss valuable combinations of markers for which some of the individual markers are not all chosen before the process stops.

[0390] The greedy procedure used here was an elaboration of the preceding forward stepwise approach, in that, to broaden the search, rather than keeping just a single marker subset at each step, a list of candidate marker sets was kept. The list was seeded with a list of single markers. The list was expanded in steps by deriving new marker subsets from the ones currently on the list and adding them to the list. Each marker subset currently on the list was extended by adding any marker from Table 1 not already part of that classifier, and which would not, on its addition to the subset, duplicate an existing subset (these are termed “permissible markers”). Each time a new set of markers was defined, a set of classifiers composed of one for each cancer study was trained using these markers, and the global performance was measured via the mean AUC across all 3 studies. To avoid potential over fitting, the AUC for each cancer study model was calculated via a ten-fold cross validation procedure. Every existing marker subset was extended by every permissible marker from the list. Clearly, such a process would eventually generate every possible subset, and the list would run out of space. Therefore, all the generated marker sets were kept only while the list was less than some predetermined size. Once the list reached the predetermined size limit, it became elitist; that is, only those classifier sets which showed a certain level of performance were kept on the list, and the others fell off the end of the list and were lost. This was achieved by keeping the list sorted in order of classifier set performance; new marker sets whose classifiers were globally at least as good as the worst set of classifiers currently on the list were inserted, forcing the expulsion of the current bottom underachieving classifier sets. One further implementation detail is that the list was completely replaced on each generational step; therefore, every marker set on the list had the same number of markers, and at each step the number of markers per classifier grew by one.

[0391] In one embodiment, the set (or panel) of biomarkers useful for constructing classifiers for diagnosing general cancer from non-cancer is based on the mean AUC for the particular combination of biomarkers used in the classification scheme. We identified many combinations of biomarkers derived from the markers in Table 19 that were able to effectively classify different cancer samples from controls. Representative panels are set forth in Tables 22-29, which set forth a series of 100 different panels of 3-10 biomarkers, which have the indicated mean cross validation (CV) AUC for each panel. The total number of occurrences of each marker in each of these panels is indicated at the bottom of each table.

[0392] The biomarkers selected in Table 19 gave rise to classifiers that perform better than classifiers built with “non-markers.” In FIG. 15, we display the performance of our ten biomarker classifiers compared to the performance of other possible classifiers.

[0393] FIG. 15A shows the distribution of mean AUCs for classifiers built from randomly sampled sets of ten “non-markers” taken from the entire set of 23 present in all 3 studies, excluding the ten markers in Table 19. The performance of the ten potential cancer biomarkers is displayed as a vertical dashed line. This plot clearly shows that the performance of the ten potential biomarkers is well beyond the distribution of other marker combinations.

[0394] FIG. 15B displays a similar distribution as FIG. 15A, however the randomly sampled sets were restricted to the 49 biomarkers from Table 1 that were not selected by the greedy biomarker selection procedure for ten analyte classi-

fiers. This plot demonstrates that the ten markers chosen by the greedy algorithm represent a subset of biomarkers that generalize to other types of cancer far better than classifiers built with the remaining 49 biomarkers.

[0395] Finally, FIG. 16 shows the classifier ROC curve for each of the 3 cancer studies classifiers. The foregoing embodiments and examples are intended only as examples. No particular embodiment, example, or element of a particular embodiment or example is to be construed as a critical, required, or essential element or feature of any of the claims. Further, no element described herein is required for the practice of the appended claims unless expressly described as “essential” or “critical.” Various alterations, modifications, substitutions, and other variations can be made to the disclosed embodiments without departing from the scope of the present application, which is defined by the appended claims. The specification, including the figures and examples, is to be regarded in an illustrative manner, rather than a restrictive one, and all such modifications and substitutions are intended to be included within the scope of the application. Accordingly, the scope of the application should be determined by the appended claims and their legal equivalents, rather than by the examples given above. For example, steps recited in any of the method or process claims may be executed in any feasible order and are not limited to an order presented in any of the embodiments, the examples, or the claims. Further, in any of the aforementioned methods, one or more biomarkers of Table 1 or Table 19 can be specifically excluded either as an individual biomarker or as a biomarker from any panel.

TABLE 1

Cancer Biomarkers					
Column #1 Biomarker #	Column #2 Biomarker Designation Entrez Gene Symbol(s)	Column #3 Entrez Gene ID	Column #4 SwissProt ID	Column #5 Public Name	Column #6 Direction
1	AHSG	197	P02765	α 2-HS-Glycoprotein	Down
2	AKR7A2	8574	O43488	Aflatoxin B1 aldehyde reductase	Up
3	AKT3	10000	Q9Y243	PKB γ	Up
4	ASGR1	432	P07306	ASGPR1	Down
5	BDNF	627	P23560	BDNF	Down
6	BMP1	649	P13497	BMP-1	Down
7	BMPER	168667	Q8N8U9	BMPER	Down
8	C9	735	P02748	C9	Up
9	CA6	765	P23280	Carbonic anhydrase VI	Down
10	CAPG	822	P40121	CapG	Down
11	CDH1	999	P12830	Cadherin-1	Down
12	CHRD1	91851	Q9BU40	Chordin-Like 1	Up
13	CKB-CKM-	1152; 1158	P12277; P06732	CK-MB	Down
14	CLIC1	1192	O00299	chloride intracellular channel 1	Up
15	CMA1	1215	P23946	Chymase	Down
16	CNTN1	1272	Q12860	Contactin-1	Down
17	COL18A1	80781	P39060	Endostatin	Up
18	CRP	1401	P02741	CRP	Up
19	CTSL2	1515	O60911	Cathepsin V	Down
20	DDC	1644	P20711	dopa decarboxylase	Down
21	EGFR	1956	P00533	ERBB1	Down
22	FGA-FGB-FGG	2243; 2244; 2266	P02671; P02675; P02679	D-dimer	Up
23	FN1	2335	P02751	Fibronectin FN1.4	Down
24	GHR	2690	P10912	Growth hormone receptor	Down
25	GPI	2821	P06744	glucose phosphate isomerase	Up
26	HMG1	3146	P09429	HMG-1	Up
27	HNRNPAB	3182	Q99729	hnRNP A/B	Up
28	HP	3240	P00738	Haptoglobin, Mixed Type	Up
29	HSP90AA1	3320	P07900	HSP 90 α	Up
30	HSPA1A	3303	P08107	HSP 70	Up

TABLE 1-continued

Cancer Biomarkers					
Column #1 Biomarker #	Column #2 Biomarker Designation Entrez Gene Symbol(s)	Column #3 Entrez Gene ID	Column #4 SwissProt ID	Column #5 Public Name	Column #6 Direction
31	IGFBP2	3485	P18065	IGFBP-2	Up
32	IGFBP4	3487	P22692	IGFBP-4	Up
33	IL12B-IL23A	3593; 51561	P29460; Q9NPF7	IL-23	Up
34	ITIH4	3700	Q14624	Inter- α -trypsin inhibitor heavy chain H4	Up
35	KIT	3815	P10721	SCF sR	Down
36	KLK3-SERPINA3	354; 12	P07288; P01011	PSA-ACT	Up
37	L1CAM	3897	P32004	NCAM-L1	Down
38	LRIG3	121227	Q6UXM1	LRIG3	Down
39	MMP12	4321	P39900	MMP-12	Up
40	MMP7	4316	P09237	MMP-7	Up
41	NME2	4831	P22392	NDP kinase B	Up
42	PA2G4	5036	Q9UQ80	ErbB3 binding protein Ebp1	Up
43	PLA2G7	7941	Q13093	LpPLA2/PAFAH	Down
44	PLAUR	5329	Q03405	suPAR	Up
45	PRKACA	5566	P17612	PRKA C- α	Up
46	PRKCB	5579	P05771	PKC- β -II	Down
47	PROK1	84432	P58294	EG-VEGF	Down
48	PRSS2	5645	P07478	Trypsin-2	Up
49	PTN	5764	P21246	Pleiotrophin	Up
50	SERPINA1	5265	P01009	α 1-Antitrypsin	Up
51	STC1	6781	P52823	Stanniocalcin-1	Up
52	STX1A	6804	Q16623	Syntaxin 1A	Down
53	TACSTD2	4070	P09758	GA733-1 protein	Down
54	TFF3	7033	Q07654	Trefoil factor 3	Up
55	TGFBI	7045	Q15582	β IGH3	Down
56	TPI1	7167	P60174	Triosephosphate isomerase	Up
57	TPT1	7178	P13693	Fortilin	Up
58	YWHAG	7532	P61981	14-3-3 protein γ	Up
59	YWHAH	7533	Q04917	14-3-3 protein eta	Up

TABLE 2

Panels of 1 Biomarker		
Markers	CV AUC	
1 YWHAG	0.840	
2 MMP7	0.804	
3 CLIC1	0.803	
4 MMP12	0.773	
5 STX1A	0.771	
6 C9	0.769	
7 LRIG3	0.769	
8 EGFR	0.767	
9 TPT1	0.760	
10 CMA1	0.758	
11 YWHAH	0.756	
12 GPI	0.752	
13 BMP1	0.751	
14 DDC	0.747	
15 NME2	0.745	
16 IGFBP2	0.743	
17 FGA-FGB-FGG	0.741	
18 CAPG	0.738	
19 AKR7A2	0.733	
20 HNRNPAB	0.730	
21 CDH1	0.728	
22 HSP90AA1	0.726	
23 CKB-CKM	0.724	
24 CRP	0.724	
25 PTN	0.723	
26 BMPER	0.721	
27 TPI1	0.720	

TABLE 2-continued

Panels of 1 Biomarker		
Markers	CV AUC	
28 TGFBI	0.720	
29 KIT	0.717	
30 HP	0.715	
31 KLK3-SERPINA3	0.713	
32 PLAUR	0.711	
33 GHR	0.705	
34 CA6	0.705	
35 PRKACA	0.704	
36 COL18A1	0.701	
37 HMGB1	0.700	
38 IGFBP4	0.698	
39 AKT3	0.697	
40 AHSG	0.697	
41 CTSL2	0.694	
42 TACSTD2	0.690	
43 FN1	0.690	
44 IL12B-IL23A	0.690	
45 BDNF	0.689	
46 L1CAM	0.688	
47 SERPINA1	0.688	
48 PROK11	0.684	
49 PRKCB	0.684	
50 STC1	0.682	
51 CHRDL1	0.679	
52 TFF3	0.678	
53 PRSS2	0.663	
54 ASGR1	0.660	

TABLE 2-continued

Panels of 1 Biomarker		
Markers		CV AUC
55	HSPA1A	0.658
56	PA2G4	0.655
57	CNTN1	0.648
58	ITIH4	0.635
59	PLA2G7	0.631

TABLE 3

Panels of 2 Biomarkers			
Markers			CV AUC
1	MMP7	YWHAG	0.878
2	C9	YWHAG	0.876
3	STX1A	YWHAG	0.874
4	MMP7	CLIC1	0.874
5	LRIG3	YWHAG	0.871
6	KLK3-SERPINA3	YWHAG	0.867
7	YWHAG	CRP	0.867
8	BMP1	YWHAG	0.866
9	MMP12	CLIC1	0.865
10	TGFB1	YWHAG	0.864
11	KLK3-SERPINA3	CLIC1	0.863
12	YWHAG	L1CAM	0.863
13	STX1A	CLIC1	0.863
14	SERPINA1	YWHAG	0.862
15	CMA1	YWHAG	0.862
16	NME2	FGA-FGB-FGG	0.861
17	CA6	YWHAG	0.859
18	MMP7	AKR7A2	0.859
19	DDC	YWHAG	0.858
20	C9	CLIC1	0.857
21	MMP7	NME2	0.857
22	CKB-CKM	YWHAG	0.857
23	FGA-FGB-FGG	CLIC1	0.856
24	BMP1	CLIC1	0.856
25	EGFR	YWHAG	0.856
26	AHSG	YWHAG	0.855
27	YWHAG	MMP12	0.855
28	MMP7	TPI1	0.855
29	KIT	YWHAG	0.855
30	LRIG3	CLIC1	0.854
31	HP	YWHAG	0.854
32	PLAUR	YWHAG	0.854
33	CMA1	CLIC1	0.853
34	BDNF	YWHAG	0.853
35	EGFR	CLIC1	0.853
36	MMP7	TPT1	0.852
37	YWHAG	CLIC1	0.851
38	PTN	YWHAG	0.850
39	BDNF	CLIC1	0.849
40	IGFBP2	YWHAG	0.849
41	MMP7	GPI	0.849
42	CNTN1	YWHAG	0.849
43	BMPER	YWHAG	0.848
44	YWHAG	FGA-FGB-FGG	0.847
45	MMP7	HNRNPAB	0.847
46	C9	GPI	0.847
47	YWHAG	GPI	0.846
48	L1CAM	MMP12	0.846
49	YWHAG	ITIH4	0.846
50	GHR	YWHAG	0.846
51	YWHAG	HNRNPAB	0.846
52	MMP7	CMA1	0.846
53	C9	NME2	0.845
54	MMP7	LRIG3	0.845
55	IGFBP2	CLIC1	0.845
56	COL18A1	YWHAG	0.845
57	CHRD1	CLIC1	0.845

TABLE 3-continued

Panels of 2 Biomarkers			
Markers			CV AUC
58	CDH1	MMP7	0.844
59	PLAUR	CLIC1	0.844
60	TPI1	FGA-FGB-FGG	0.844
61	CHRD1	YWHAG	0.844
62	MMP7	PRKACA	0.844
63	C9	AKR7A2	0.843
64	YWHAG	PLA2G7	0.843
65	KLK3-SERPINA3	TPT1	0.843
66	BMP1	GPI	0.843
67	KLK3-SERPINA3	MMP7	0.842
68	C9	TPT1	0.842
69	COL18A1	CLIC1	0.842
70	YWHAG	AKR7A2	0.842
71	YWHAG	STC1	0.842
72	MMP7	TGFB1	0.842
73	AKR7A2	MMP12	0.842
74	MMP7	YWHAG	0.842
75	HMGB1	MMP7	0.841
76	TPT1	FGA-FGB-FGG	0.841
77	GHR	CLIC1	0.841
78	KLK3-SERPINA3	STX1A	0.840
79	LRIG3	TPT1	0.840
80	STX1A	MMP12	0.840
81	YWHAG	PRSS2	0.840
82	DDC	CLIC1	0.840
83	CRP	CLIC1	0.840
84	HMGB1	YWHAG	0.840
85	STX1A	TPT1	0.839
86	CDH1	YWHAG	0.839
87	STX1A	GPI	0.839
88	KLK3-SERPINA3	NME2	0.838
89	LRIG3	YWHAG	0.838
90	AKR7A2	FGA-FGB-FGG	0.838
91	C9	HNRNPAB	0.837
92	TACSTD2	YWHAG	0.837
93	YWHAG	TPI1	0.837
94	STX1A	NME2	0.836
95	KLK3-SERPINA3	AKR7A2	0.836
96	LRIG3	AKR7A2	0.836
97	NME2	MMP12	0.836
98	CAPG	CLIC1	0.836
99	YWHAG	NME2	0.836
100	MMP7	STX1A	0.835

TABLE 4

Panels of 3 Biomarkers				
Markers				CV AUC
1	KLK3-SERPINA3	MMP7	CLIC1	0.896
2	KLK3-SERPINA3	STX1A	CLIC1	0.895
3	KLK3-SERPINA3	STX1A	YWHAG	0.895
4	MMP7	C9	YWHAG	0.895
5	MMP7	YWHAG	CLIC1	0.894
6	C9	STX1A	YWHAG	0.893
7	MMP7	LRIG3	YWHAG	0.893
8	MMP7	TGFB1	YWHAG	0.893
9	MMP7	CMA1	CLIC1	0.893
10	BDNF	MMP7	CLIC1	0.892
11	MMP7	GHR	CLIC1	0.892
12	CDH1	MMP7	YWHAG	0.892
13	BDNF	C9	CLIC1	0.892
14	STX1A	YWHAG	CRP	0.892
15	MMP7	YWHAG	TPI1	0.892
16	MMP7	STX1A	YWHAG	0.892
17	TGFB1	STX1A	YWHAG	0.891
18	LRIG3	YWHAG	CRP	0.891
19	MMP7	YWHAG	L1CAM	0.891

TABLE 4-continued

Panels of 3 Biomarkers				
Markers			CV AUC	
20	MMP7	YWHAG	PA2G4	0.891
21	C9	LRIG3	YWHAG	0.890
22	STX1A	MMP12	CLIC1	0.890
23	MMP7	LRIG3	CLIC1	0.890
24	KLK3-SERPINA3	MMP7	YWHAG	0.890
25	MMP7	BMP1	CLIC1	0.890
26	BDNF	STX1A	CLIC1	0.890
27	MMP7	STX1A	CLIC1	0.889
28	MMP7	BMP1	YWHAG	0.889
29	HMGB1	MMP7	YWHAG	0.889
30	SERPINA1	STX1A	YWHAG	0.889
31	MMP7	YWHAG	GPI	0.889
32	MMP7	CMA1	YWHAG	0.889
33	MMP7	YWHAG	NME2	0.889
34	MMP7	C9	CLIC1	0.889
35	C9	CMA1	YWHAG	0.888
36	MMP7	YWHAG	CRP	0.888
37	KLK3-SERPINA3	CNTN1	YWHAG	0.888
38	MMP7	YWHAG	AKR7A2	0.887
39	MMP7	ITIH4	CLIC1	0.887
40	CDH1	MMP7	CLIC1	0.887
41	KLK3-SERPINA3	MMP7	AKR7A2	0.887
42	MMP7	GHR	YWHAG	0.887
43	KLK3-SERPINA3	CHRD1	CLIC1	0.887
44	KLK3-SERPINA3	LRIG3	YWHAG	0.887
45	BMP1	STX1A	CLIC1	0.887
46	C9	STX1A	CLIC1	0.887
47	MMP7	GPI	CLIC1	0.887
48	TGFB1	LRIG3	YWHAG	0.886
49	IGFBP2	MMP7	YWHAG	0.886
50	MMP7	CKB-CKM	YWHAG	0.886
51	LRIG3	STX1A	YWHAG	0.886
52	GHR	STX1A	CLIC1	0.886
53	MMP7	DDC	YWHAG	0.886
54	BMP1	STX1A	YWHAG	0.886
55	MMP7	DDC	CLIC1	0.886
56	C9	CHRD1	CLIC1	0.885
57	MMP7	C9	AKR7A2	0.885
58	BDNF	MMP7	YWHAG	0.885
59	KIT	MMP7	YWHAG	0.885
60	MMP7	TGFB1	CLIC1	0.885
61	BDNF	IGFBP2	CLIC1	0.885
62	MMP7	YWHAG	ITIH4	0.885
63	MMP7	YWHAG	HNRNPAB	0.885
64	KLK3-SERPINA3	LRIG3	CLIC1	0.885
65	MMP7	HP	YWHAG	0.885
66	HMGB1	MMP7	CLIC1	0.885
67	MMP7	YWHAG	PLA2G7	0.885
68	CHRD1	CMA1	CLIC1	0.885
69	STX1A	YWHAG	L1CAM	0.885
70	MMP7	CMA1	NME2	0.885
71	BMP1	MMP12	CLIC1	0.884
72	C9	CHRD1	YWHAG	0.884
73	KLK3-SERPINA3	CMA1	CLIC1	0.884
74	EGFR	MMP7	CLIC1	0.884
75	STX1A	YWHAG	CLIC1	0.884
76	MMP7	AHSG	YWHAG	0.884
77	IGFBP2	MMP7	CLIC1	0.884
78	MMP7	TPT1	YWHAG	0.884
79	KLK3-SERPINA3	COL18A1	CLIC1	0.884
80	EGFR	MMP7	YWHAG	0.884
81	C9	YWHAG	L1CAM	0.884
82	KLK3-SERPINA3	MMP7	TPI1	0.884
83	KLK3-SERPINA3	BDNF	CLIC1	0.884
84	MMP7	CA6	YWHAG	0.884
85	BMP1	YWHAG	CRP	0.883
86	MMP7	CMA1	TPI1	0.883
87	KLK3-SERPINA3	MMP7	NME2	0.883
88	BDNF	C9	YWHAG	0.883
89	AHSG	STX1A	YWHAG	0.883
90	C9	MMP12	CLIC1	0.883
91	C9	BMP1	YWHAG	0.883

TABLE 4-continued

Panels of 3 Biomarkers				
Markers			CV AUC	
92	KLK3-SERPINA3	STX1A	TPT1	0.883
93	CNTN1	C9	YWHAG	0.883
94	C9	CA6	YWHAG	0.883
95	CA6	STX1A	YWHAG	0.883
96	MMP7	CNTN1	YWHAG	0.883
97	KLK3-SERPINA3	STX1A	NME2	0.883
98	MMP7	HNRNPAB	CLIC1	0.883
99	MMP7	SERPINA1	YWHAG	0.883
100	TGFB1	CMA1	YWHAG	0.883

TABLE 5

Panels of 4 Biomarkers					
Markers				CV AUC	
1	KLK3-SERPINA3	MMP7	STX1A	CLIC1	0.911
2	KLK3-SERPINA3	BDNF	STX1A	CLIC1	0.910
3	BDNF	C9	CHRD1	CLIC1	0.909
4	BDNF	C9	STX1A	CLIC1	0.908
5	MMP7	C9	YWHAG	TP1	0.908
6	MMP7	C9	YWHAG	CLIC1	0.908
7	MMP7	GHR	STX1A	CLIC1	0.907
8	KLK3-SERPINA3	MMP7	CMA1	CLIC1	0.907
9	KLK3-SERPINA3	BDNF	MMP7	CLIC1	0.907
10	MMP7	C9	CMA1	CLIC1	0.907
11	BDNF	MMP7	YWHAG	CLIC1	0.907
12	CDH1	MMP7	C9	YWHAG	0.907
13	KLK3-SERPINA3	MMP7	LRIG3	CLIC1	0.906
14	MMP7	GHR	CMA1	CLIC1	0.906
15	MMP7	C9	YWHAG	NME2	0.906
16	CDH1	MMP7	STX1A	YWHAG	0.906
17	MMP7	C9	LRIG3	YWHAG	0.905
18	MMP7	C9	YWHAG	GPI	0.905
19	CDH1	MMP7	STX1A	CLIC1	0.905
20	BDNF	MMP7	GHR	CLIC1	0.905
21	MMP7	STX1A	YWHAG	CLIC1	0.905
22	BDNF	MMP7	LRIG3	CLIC1	0.905
23	BDNF	MMP7	STX1A	CLIC1	0.905
24	MMP7	LRIG3	YWHAG	CLIC1	0.905
25	BDNF	MMP7	CMA1	CLIC1	0.905
26	MMP7	C9	TGFB1	YWHAG	0.904
27	CDH1	MMP7	LRIG3	YWHAG	0.904
28	KLK3-SERPINA3	CHRD1	CMA1	CLIC1	0.904
29	TGFB1	STX1A	YWHAG	CRP	0.904
30	BDNF	MMP7	C9	CLIC1	0.904
31	KLK3-SERPINA3	CHRD1	STX1A	CLIC1	0.904
32	KLK3-SERPINA3	MMP7	STX1A	YWHAG	0.904
33	KLK3-SERPINA3	BMP1	STX1A	CLIC1	0.904
34	MMP7	STX1A	YWHAG	NME2	0.904
35	BDNF	MMP7	TGFB1	CLIC1	0.904
36	MMP7	C9	YWHAG	L1CAM	0.904
37	MMP7	TGFB1	LRIG3	YWHAG	0.904
38	KLK3-SERPINA3	BDNF	CHRD1	CLIC1	0.904
39	KLK3-SERPINA3	GHR	STX1A	CLIC1	0.904
40	KLK3-SERPINA3	LRIG3	CHRD1	CLIC1	0.904

TABLE 5-continued

Panels of 4 Biomarkers					
Markers				CV AUC	
41	KLK3-SERPINA3	MMP7	LRIG3	YWHAG	0.904
42	KLK3-SERPINA3	LRIG3	STX1A	CLIC1	0.904
43	MMP7	GHR	BMP1	CLIC1	0.904
44	CDH1	MMP7	CMA1	CLIC1	0.904
45	LRIG3	STX1A	YWHAG	CRP	0.904
46	MMP7	GHR	YWHAG	CLIC1	0.904
47	BDNF	GHR	STX1A	CLIC1	0.904
48	MMP7	C9	CMA1	YWHAG	0.904
49	MMP7	LRIG3	GPI	CLIC1	0.904
50	MMP7	C9	STX1A	YWHAG	0.903
51	BDNF	MMP7	GPI	CLIC1	0.903
52	KLK3-SERPINA3	MMP7	YWHAG	CLIC1	0.903
53	MMP7	TGFB1	STX1A	YWHAG	0.903
54	KLK3-SERPINA3	COL18A1	STX1A	CLIC1	0.903
55	MMP7	TGFB1	CMA1	CLIC1	0.903
56	MMP7	C9	YWHAG	PA2G4	0.903
57	MMP7	C9	YWHAG	AKR7A2	0.903
58	KLK3-SERPINA3	MMP7	BMP1	CLIC1	0.903
59	MMP7	GHR	LRIG3	CLIC1	0.903
60	MMP7	GHR	C9	CLIC1	0.903
61	MMP7	BMP1	YWHAG	CLIC1	0.903
62	KLK3-SERPINA3	MMP7	GHR	CLIC1	0.903
63	BDNF	STX1A	MMP12	CLIC1	0.903
64	MMP7	LRIG3	YWHAG	CRP	0.903
65	BDNF	IGFBP2	MMP7	CLIC1	0.903
66	GHR	STX1A	CRP	CLIC1	0.903
67	BDNF	STX1A	CRP	CLIC1	0.902
68	KLK3-SERPINA3	CNTN1	BMP1	CLIC1	0.902
69	BDNF	MMP7	C9	YWHAG	0.902
70	CDH1	MMP7	TGFB1	YWHAG	0.902
71	BDNF	IGFBP2	STX1A	CLIC1	0.902

TABLE 5-continued

Panels of 4 Biomarkers					
	Markers			CV AUC	
72	KLK3-SERPINA3	MMP7	NME2	CLIC1	0.902
73	KLK3-SERPINA3	MMP7	TPI1	CLIC1	0.902
74	MMP7	LRIG3	YWHAG	NME2	0.902
75	KLK3-SERPINA3	EGFR	STX1A	CLIC1	0.902
76	BDNF	IGFBP2	LRIG3	CLIC1	0.902
77	MMP7	CMA1	YWHAG	CLIC1	0.902
78	MMP7	GHR	STX1A	YWHAG	0.902
79	HMGB1	MMP7	C9	YWHAG	0.902
80	IGFBP2	MMP7	CMA1	CLIC1	0.902
81	MMP7	GHR	GPI	CLIC1	0.902
82	KLK3-SERPINA3	STX1A	YWHAG	CLIC1	0.902
83	KLK3-SERPINA3	SERPINA1	STX1A	YWHAG	0.902
84	BDNF	PLAUR	LRIG3	CLIC1	0.902
85	BDNF	TGFB1	STX1A	CLIC1	0.902
86	BDNF	MMP7	ITIH4	CLIC1	0.902
87	MMP7	LRIG3	YWHAG	GPI	0.902
88	MMP7	BMP1	YWHAG	GPI	0.902
89	C9	CHRD1	CMA1	CLIC1	0.902
90	MMP7	BMP1	CMA1	CLIC1	0.902
91	KLK3-SERPINA3	MMP7	CNTN1	CLIC1	0.902
92	MMP7	CMA1	HNRNPAB	CLIC1	0.902
93	KLK3-SERPINA3	LRIG3	STX1A	YWHAG	0.902
94	BDNF	LRIG3	STX1A	CLIC1	0.902
95	MMP7	TGFB1	CMA1	YWHAG	0.902
96	MMP7	LRIG3	YWHAG	TPI1	0.902
97	MMP7	CMA1	NME2	CLIC1	0.902
98	MMP7	GHR	CRP	CLIC1	0.902
99	C9	LRIG3	CHRD1	CLIC1	0.902
100	MMP7	LRIG3	STX1A	CLIC1	0.902

TABLE 6

Panels of 5 Biomarkers						
Markers						CV AUC
1	TGFB1	LRIG3	CHRD1	NME2	CRP	0.922
2	KLK3-SERPINA3	BDNF	MMP7	STX1A	CLIC1	0.920
3	BDNF	MMP7	GHR	STX1A	CLIC1	0.919
4	BDNF	MMP7	C9	YWHAG	CLIC1	0.918
5	KLK3-SERPINA3	MMP7	GHR	STX1A	CLIC1	0.918
6	BDNF	C9	CHRD1	AHSG	CLIC1	0.918
7	CDH1	MMP7	GHR	STX1A	CLIC1	0.918
8	KLK3-SERPINA3	MMP7	STX1A	NME2	CLIC1	0.918
9	MMP7	GHR	STX1A	YWHAG	CLIC1	0.918
10	MMP7	GHR	STX1A	GPI	CLIC1	0.918
11	KLK3-SERPINA3	MMP7	LRIG3	STX1A	CLIC1	0.917
12	BDNF	MMP7	GHR	GPI	CLIC1	0.917
13	BDNF	MMP7	STX1A	YWHAG	CLIC1	0.917
14	BDNF	TGFB1	LRIG3	CHRD1	CLIC1	0.917
15	KLK3-SERPINA3	BDNF	LRIG3	STX1A	CLIC1	0.917
16	KLK3-SERPINA3	BDNF	C9	STX1A	CLIC1	0.917
17	BDNF	MMP7	LRIG3	YWHAG	CLIC1	0.917
18	BDNF	GHR	C9	STX1A	CLIC1	0.916
19	BDNF	IGFBP2	LRIG3	CRP	CLIC1	0.916
20	KLK3-SERPINA3	BDNF	CHRD1	STX1A	CLIC1	0.916
21	KLK3-SERPINA3	CDH1	MMP7	STX1A	CLIC1	0.916
22	MMP7	GHR	STX1A	CRP	CLIC1	0.916
23	BDNF	MMP7	TGFB1	STX1A	CLIC1	0.916
24	MMP7	GHR	TGFB1	STX1A	CLIC1	0.916

TABLE 6-continued

Panels of 5 Biomarkers						
Markers					CV AUC	
25	MMP7	GHR	C9	STX1A	CLIC1	0.916
26	BDNF	MMP7	GHR	TGFB1	CLIC1	0.916
27	MMP7	GHR	STX1A	NME2	CLIC1	0.916
28	KLK3-SERPINA3	HMGB1	MMP7	STX1A	CLIC1	0.916
29	MMP7	C9	STX1A	YWHAG	NME2	0.916
30	BDNF	MMP7	LRIG3	STX1A	CLIC1	0.916
31	MMP7	C9	STX1A	YWHAG	CLIC1	0.916
32	BDNF	CDH1	MMP7	STX1A	CLIC1	0.916
33	BDNF	C9	TGFB1	CHRD1	CLIC1	0.915
34	MMP7	C9	LRIG3	YWHAG	TPI1	0.915
35	KLK3-SERPINA3	BDNF	MMP7	LRIG3	CLIC1	0.915
36	BDNF	C9	LRIG3	CHRD1	CLIC1	0.915
37	KLK3-SERPINA3	BDNF	MMP7	CMA1	CLIC1	0.915
38	BDNF	LRIG3	CHRD1	CRP	CLIC1	0.915
39	BDNF	MMP7	STX1A	ITIH4	CLIC1	0.915
40	BDNF	MMP7	GHR	C9	CLIC1	0.915
41	BDNF	MMP7	C9	GPI	CLIC1	0.915
42	HMGB1	MMP7	GHR	STX1A	CLIC1	0.915
43	BDNF	MMP7	LRIG3	GPI	CLIC1	0.915
44	GHR	BMP1	STX1A	CRP	CLIC1	0.915
45	BDNF	MMP7	BMP1	GPI	CLIC1	0.915
46	KLK3-SERPINA3	MMP7	STX1A	YWHAG	CLIC1	0.915
47	KLK3-SERPINA3	CNTN1	BMP1	CHRD1	CLIC1	0.915
48	BDNF	GHR	STX1A	CRP	CLIC1	0.915
49	KLK3-SERPINA3	BDNF	TGFB1	STX1A	CLIC1	0.915
50	KLK3-SERPINA3	BDNF	MMP7	PA2G4	CLIC1	0.915
51	CDH1	MMP7	TGFB1	STX1A	YWHAG	0.915
52	BDNF	MMP7	C9	STX1A	CLIC1	0.915
53	MMP7	GHR	TGFB1	CMA1	CLIC1	0.915
54	BDNF	MMP7	TGFB1	CMA1	CLIC1	0.915
55	CDH1	MMP7	C9	TGFB1	YWHAG	0.915
56	MMP7	C9	LRIG3	YWHAG	NME2	0.915
57	BDNF	MMP7	STX1A	NME2	CLIC1	0.915
58	BDNF	EGFR	TGFB1	STX1A	CLIC1	0.915
59	KLK3-SERPINA3	MMP7	LRIG3	GPI	CLIC1	0.915
60	BDNF	MMP7	STX1A	GPI	CLIC1	0.915
61	MMP7	C9	LRIG3	YWHAG	GPI	0.915
62	KLK3-SERPINA3	MMP7	CMA1	TPI1	CLIC1	0.915
63	CDH1	MMP7	C9	STX1A	YWHAG	0.915
64	KLK3-SERPINA3	BDNF	CNTN1	CHRD1	CLIC1	0.915
65	KLK3-SERPINA3	BDNF	LRIG3	CHRD1	CLIC1	0.915
66	BDNF	MMP7	GHR	LRIG3	CLIC1	0.914
67	KLK3-SERPINA3	BDNF	MMP7	NME2	CLIC1	0.914
68	BDNF	IGFBP2	MMP7	GPI	CLIC1	0.914
69	KLK3-SERPINA3	BDNF	STX1A	CLIC1	PLA2G7	0.914
70	CDH1	MMP7	GHR	CMA1	CLIC1	0.914
71	MMP7	C9	LRIG3	GPI	CLIC1	0.914
72	MMP7	GHR	STX1A	PA2G4	CLIC1	0.914
73	KLK3-SERPINA3	MMP7	STX1A	PA2G4	CLIC1	0.914
74	KLK3-SERPINA3	MMP7	STX1A	TPI1	CLIC1	0.914
75	KLK3-SERPINA3	MMP7	STX1A	HNRNPAB	CLIC1	0.914
76	MMP7	GHR	LRIG3	GPI	CLIC1	0.914
77	MMP7	GHR	CMA1	GPI	CLIC1	0.914
78	BDNF	IGFBP2	MMP7	LRIG3	CLIC1	0.914
79	KLK3-SERPINA3	BDNF	MMP7	TPI1	CLIC1	0.914
80	BDNF	MMP7	STX1A	TPT1	CLIC1	0.914
81	BDNF	LRIG3	STX1A	CRP	CLIC1	0.914
82	BDNF	MMP7	STX1A	CLIC1	PLA2G7	0.914
83	KLK3-SERPINA3	BDNF	AHSG	STX1A	CLIC1	0.914
84	KLK3-SERPINA3	MMP7	CNTN1	STX1A	CLIC1	0.914
85	BDNF	GHR	TGFB1	STX1A	CLIC1	0.914
86	BDNF	MMP7	NME2	ITIH4	CLIC1	0.914
87	KLK3-SERPINA3	CNTN1	BMP1	STX1A	CLIC1	0.914
88	MMP7	C9	CMA1	NME2	CLIC1	0.914
89	BDNF	MMP7	LRIG3	NME2	CLIC1	0.914
90	BDNF	TGFB1	LRIG3	STX1A	CLIC1	0.914
91	KLK3-SERPINA3	CDH1	MMP7	STX1A	YWHAG	0.914
92	MMP7	C9	LRIG3	YWHAG	CLIC1	0.914
93	BDNF	MMP7	TGFB1	LRIG3	CLIC1	0.914
94	KLK3-SERPINA3	BDNF	STX1A	CRP	CLIC1	0.914
95	BDNF	MMP7	BMP1	YWHAG	CLIC1	0.914
96	KLK3-SERPINA3	MMP7	LRIG3	CMA1	CLIC1	0.914

TABLE 6-continued

Panels of 5 Biomarkers						
Markers						CV AUC
97	KLK3-SERPINA3	MMP7	BMP1	STX1A	CLIC1	0.914
98	BDNF	IGFBP2	MMP7	STX1A	CLIC1	0.914
99	KLK3-SERPINA3	MMP7	STX1A	YWHAG	GPI	0.914
100	MMP7	LRIG3	STX1A	YWHAG	CLIC1	0.914

TABLE 7

Panels of 6 Biomarkers						
Markers						CV AUC
1	BDNF CLIC1	MMP7	GHR	STX1A	GPI	0.928
2	BDNF CLIC1	TGFB1	LRIG3	CHRD1	CRP	0.928
3	KLK3-SERPINA3 CLIC1	BDNF	MMP7	STX1A	NME2	0.928
4	KLK3-SERPINA3 CLIC1	BDNF	MMP7	GHR	STX1A	0.927
5	BDNF CLIC1	MMP7	GHR	TGFB1	STX1A	0.927
6	TGFB1 CRP	LRIG3	CHRD1	AHSG	NME2	0.927
7	KLK3-SERPINA3 CLIC1	BDNF	MMP7	TGFB1	STX1A	0.927
8	BDNF CLIC1	MMP7	C9	STX1A	YWHAG	0.926
9	KLK3-SERPINA3 CLIC1	BDNF	MMP7	STX1A	TPT1	0.926
10	BDNF CLIC1	MMP7	GHR	STX1A	PA2G4	0.926
11	KLK3-SERPINA3 CLIC1	BDNF	MMP7	LRIG3	STX1A	0.925
12	BDNF CLIC1	MMP7	C9	LRIG3	YWHAG	0.925
13	KLK3-SERPINA3 CLIC1	MMP7	GHR	STX1A	TP11	0.925
14	KLK3-SERPINA3 CLIC1	BDNF	KIT	MMP7	STX1A	0.925
15	KLK3-SERPINA3 CLIC1	BDNF	MMP7	STX1A	PA2G4	0.925
16	BDNF CLIC1	MMP7	GHR	STX1A	NME2	0.925
17	BDNF CLIC1	IGFBP2	MMP7	LRIG3	NME2	0.925
18	BDNF CLIC1	GHR	C9	AHSG	STX1A	0.925
19	KLK3-SERPINA3 CLIC1	BDNF	MMP7	STX1A	TP11	0.925
20	BDNF CLIC1	MMP7	GHR	C9	STX1A	0.925
21	BDNF CLIC1	MMP7	GHR	STX1A	CRP	0.925
22	BDNF CLIC1	MMP7	GHR	LRIG3	GPI	0.925
23	KLK3-SERPINA3 CLIC1	BDNF	CDH1	MMP7	STX1A	0.925
24	MMP7 CLIC1	GHR	C9	STX1A	YWHAG	0.925
25	MMP7 CLIC1	GHR	C9	STX1A	HNRNPAB	0.925
26	KLK3-SERPINA3 CLIC1	BDNF	TGFB1	CHRD1	STX1A	0.925
27	KLK3-SERPINA3 PLA2G7	BDNF	MMP7	STX1A	CLIC1	0.925
28	MMP7 CLIC1	GHR	C9	STX1A	GPI	0.925
29	BDNF CLIC1	MMP7	GHR	LRIG3	YWHAG	0.925

TABLE 7-continued

Panels of 6 Biomarkers						
		Markers				CV AUC
30	KLK3-SERPINA3 CLIC1	MMP7	GHR	STX1A	NME2	0.925
31	BDNF PLA2G7	MMP7	GHR	STX1A	CLIC1	0.925
32	BDNF CLIC1	MMP7	GHR	STX1A	TPT1	0.925
33	BDNF CLIC1	MMP7	C9	STX1A	NME2	0.924
34	KLK3-SERPINA3 CLIC1	BDNF	MMP7	LRIG3	NME2	0.924
35	BDNF CLIC1	MMP7	LRIG3	STX1A	GPI	0.924
36	BDNF CLIC1	MMP7	GHR	AHSG	STX1A	0.924
37	BDNF CLIC1	MMP7	GHR	C9	YWHAG	0.924
38	CDH1 CLIC1	MMP7	GHR	STX1A	CRP	0.924
39	BDNF CLIC1	IGFBP2	MMP7	LRIG3	GPI	0.924
40	KLK3-SERPINA3 CLIC1	BDNF	MMP7	STX1A	YWHAG	0.924
41	KLK3-SERPINA3 CLIC1	BDNF	MMP7	STX1A	GPI	0.924
42	BDNF CLIC1	CDH1	MMP7	GHR	STX1A	0.924
43	BDNF CLIC1	IGFBP2	MMP7	TP11	ITIH4	0.924
44	BDNF CLIC1	MMP7	STX1A	NME2	ITIH4	0.924
45	BDNF CLIC1	MMP7	GHR	STX1A	YWHAG	0.924
46	KLK3-SERPINA3 CLIC1	BDNF	CNTN1	TGFBI	CHRD1	0.924
47	KLK3-SERPINA3 CLIC1	CDH1	MMP7	LRIG3	STX1A	0.924
48	KLK3-SERPINA3 CLIC1	MMP7	LRIG3	STX1A	NME2	0.924
49	KLK3-SERPINA3 CLIC1	BDNF	TGFBI	LRIG3	STX1A	0.923
50	BDNF CLIC1	MMP7	TGFBI	LRIG3	GPI	0.923
51	BDNF CLIC1	TGFBI	LRIG3	STX1A	CRP	0.923
52	KLK3-SERPINA3 CLIC1	CDH1	MMP7	GHR	STX1A	0.923
53	BDNF CLIC1	MMP7	GHR	TGFBI	GPI	0.923
54	BDNF CLIC1	MMP7	C9	CMA1	NME2	0.923
55	KLK3-SERPINA3 CLIC1	BDNF	MMP7	AHSG	STX1A	0.923
56	KLK3-SERPINA3 CLIC1	MMP7	GHR	STX1A	GPI	0.923
57	BDNF CLIC1	MMP7	C9	STX1A	TPT1	0.923
58	BDNF CLIC1	MMP7	GHR	CNTN1	TGFBI	0.923
59	MMP7 CLIC1	GHR	TGFBI	STX1A	CRP	0.923
60	KLK3-SERPINA3 CLIC1	MMP7	GHR	STX1A	YWHAG	0.923
61	TGFBI CRP	LRIG3	CHRD1	STX1A	NME2	0.923
62	BDNF CLIC1	MMP7	C9	STX1A	GPI	0.923
63	BDNF CLIC1	IGFBP2	MMP7	TGFBI	STX1A	0.923
64	KLK3-SERPINA3 CLIC1	BDNF	MMP7	STX1A	HNRNPAB	0.923
65	MMP7 CLIC1	GHR	C9	STX1A	NME2	0.923

TABLE 7-continued

Panels of 6 Biomarkers						
Markers						CV AUC
66	CDH1 CLIC1	MMP7	GHR	TGFBI	STX1A	0.923
67	KLK3-SERPINA3 CLIC1	MMP7	GHR	STX1A	PA2G4	0.923
68	BDNF CLIC1	MMP7	TGFBI	STX1A	GPI	0.923
69	BDNF CLIC1	MMP7	STX1A	YWHAG	ITIH4	0.923
70	BDNF CLIC1	MMP7	GHR	LRIG3	STX1A	0.923
71	BDNF CLIC1	KIT	MMP7	GHR	STX1A	0.923
72	MMP7 CLIC1	GHR	TGFBI	STX1A	GPI	0.923
73	BDNF CLIC1	MMP7	STX1A	TPI1	ITIH4	0.923
74	BDNF CLIC1	MMP7	TGFBI	LRIG3	STX1A	0.923
75	BDNF CLIC1	EGFR	TGFBI	AHSG	STX1A	0.923
76	KLK3-SERPINA3 CLIC1	BDNF	TGFBI	LRIG3	CHRD1	0.923
77	CDH1 CLIC1	MMP7	GHR	STX1A	GPI	0.923
78	BDNF CLIC1	IGFBP2	MMP7	LRIG3	TPI1	0.923
79	BDNF CLIC1	GHR	LRIG3	STX1A	CRP	0.923
80	BDNF CLIC1	CDH1	MMP7	LRIG3	STX1A	0.923
81	KLK3-SERPINA3 CLIC1	MMP7	GHR	TGFBI	STX1A	0.923
82	BDNF CLIC1	IGFBP2	LRIG3	AHSG	CRP	0.923
83	KLK3-SERPINA3 CLIC1	BDNF	MMP7	LRIG3	GPI	0.923
84	BDNF CLIC1	MMP7	GHR	LRIG3	NME2	0.923
85	KLK3-SERPINA3 CLIC1	BDNF	EGFR	TGFBI	STX1A	0.923
86	BDNF CLIC1	MMP7	GHR	TGFBI	LRIG3	0.923
87	MMP7 CLIC1	GHR	C9	CMA1	NME2	0.923
88	BDNF CLIC1	MMP7	GHR	TGFBI	CMA1	0.923
89	MMP7 CLIC1	GHR	STX1A	NME2	CRP	0.922
90	BDNF CLIC1	MMP7	C9	LRIG3	GPI	0.922
91	KLK3-SERPINA3 CLIC1	BDNF	MMP7	LRIG3	TPT1	0.922
92	BDNF CLIC1	MMP7	STX1A	TPT1	ITIH4	0.922
93	KIT TPI1	MMP7	C9	LRIG3	YWHAG	0.922
94	BDNF CLIC1	CDH1	MMP7	STX1A	ITIH4	0.922
95	MMP7 CLIC1	GHR	STX1A	TPI1	CRP	0.922
96	BDNF CLIC1	C9	TGFBI	LRIG3	CHRD1	0.922
97	KLK3-SERPINA3 CLIC1	BDNF	CNTN1	BMP1	CHRD1	0.922
98	BDNF CLIC1	GHR	TGFBI	STX1A	CRP	0.922
99	KLK3-SERPINA3 CLIC1	LRIG3	CHRD1	STX1A	CRP	0.922
100	MMP7 CLIC1	GHR	LRIG3	STX1A	YWHAG	0.922

TABLE 8

Panels of 7 Biomarkers						
Markers						CV AUC
1	BDNF GPI	MMP7 CLIC1	GHR	TGFBI	STX1A	0.933
2	KLK3-SERPINA3 NME2	BDNF CLIC1	MMP7	GHR	STX1A	0.932
3	BDNF GPI	MMP7 CLIC1	GHR	C9	STX1A	0.932
4	KLK3-SERPINA3 PA2G4	BDNF CLIC1	MMP7	GHR	STX1A	0.932
5	BDNF CRP	MMP7 CLIC1	GHR	TGFBI	STX1A	0.932
6	BDNF PA2G4	MMP7 CLIC1	GHR	TGFBI	STX1A	0.932
7	KLK3-SERPINA3 TPI1	BDNF CLIC1	MMP7	GHR	STX1A	0.932
8	BDNF TPT1	MMP7 CLIC1	GHR	C9	STX1A	0.932
9	KLK3-SERPINA3 ITIH4	BDNF CLIC1	MMP7	STX1A	NME2	0.932
10	BDNF STX1A	CDH1 CLIC1	MMP7	GHR	TGFBI	0.932
11	BDNF CRP	TGFBI CLIC1	LRIG3	CHRD1	STX1A	0.932
12	BDNF NME2	MMP7 CLIC1	GHR	TGFBI	STX1A	0.932
13	KLK3-SERPINA3 NME2	BDNF CLIC1	MMP7	LRIG3	STX1A	0.932
14	KLK3-SERPINA3 GPI	BDNF CLIC1	MMP7	LRIG3	STX1A	0.932
15	KLK3-SERPINA3 STX1A	BDNF CLIC1	KIT	MMP7	GHR	0.931
16	KLK3-SERPINA3 CHRD1	BDNF CLIC1	CNTN1	TGFBI	LRIG3	0.931
17	BDNF NME2	MMP7 CLIC1	GHR	C9	STX1A	0.931
18	KLK3-SERPINA3 TPT1	BDNF CLIC1	MMP7	LRIG3	STX1A	0.931
19	KLK3-SERPINA3 STX1A	BDNF CLIC1	MMP7	TGFBI	LRIG3	0.931
20	KLK3-SERPINA3 GPI	BDNF CLIC1	MMP7	GHR	STX1A	0.931
21	BDNF YWHAG	MMP7 CLIC1	GHR	C9	STX1A	0.931
22	BDNF ITIH4	MMP7 CLIC1	GHR	STX1A	NME2	0.931
23	BDNF STX1A	MMP7 CLIC1	GHR	TGFBI	LRIG3	0.931
24	KLK3-SERPINA3 TPT1	BDNF CLIC1	MMP7	GHR	STX1A	0.931
25	BDNF GPI	MMP7 CLIC1	GHR	AHSG	STX1A	0.931
26	KLK3-SERPINA3 ITIH4	BDNF CLIC1	MMP7	STX1A	TPI1	0.931
27	BDNF GPI	MMP7 CLIC1	GHR	STX1A	PA2G4	0.931
28	KLK3-SERPINA3 PA2G4	BDNF CLIC1	KIT	MMP7	STX1A	0.931
29	BDNF CRP	MMP7 CLIC1	GHR	STX1A	NME2	0.931
30	KLK3-SERPINA3 NME2	BDNF CLIC1	MMP7	TGFBI	STX1A	0.931
31	BDNF STX1A	MMP7 CLIC1	GHR	TGFBI	AHSG	0.931
32	BDNF STX1A	CDH1 CLIC1	MMP7	GHR	AHSG	0.931
33	KLK3-SERPINA3 NME2	BDNF CLIC1	EGFR	MMP7	STX1A	0.931
34	KLK3-SERPINA3 STX1A	BDNF CLIC1	KIT	MMP7	LRIG3	0.931
35	BDNF GPI	MMP7 CLIC1	GHR	LRIG3	STX1A	0.930
36	BDNF CRP	GHR CLIC1	TGFBI	LRIG3	CHRD1	0.930

TABLE 8-continued

Panels of 7 Biomarkers						
Markers						CV AUC
37	BDNF CLIC1	MMP7 PLA2G7	GHR	TGFB1	STX1A	0.930
38	BDNF YWHAG	MMP7 CLIC1	GHR	C9	LRIG3	0.930
39	BDNF TPT1	KIT CLIC1	MMP7	GHR	STX1A	0.930
40	BDNF ITIH4	MMP7 CLIC1	C9	STX1A	NME2	0.930
41	KLK3-SERPINA3 STX1A	BDNF CLIC1	TGFB1	LRIG3	CHRD1	0.930
42	KLK3-SERPINA3 TPI1	BDNF CLIC1	KIT	MMP7	STX1A	0.930
43	BDNF GPI	MMP7 CLIC1	TGFB1	LRIG3	STX1A	0.930
44	BDNF CRP	MMP7 CLIC1	GHR	STX1A	GPI	0.930
45	BDNF GPI	MMP7 CLIC1	GHR	TGFB1	LRIG3	0.930
46	BDNF PA2G4	MMP7 CLIC1	GHR	C9	STX1A	0.930
47	BDNF TPT1	MMP7 CLIC1	GHR	CHRD1	STX1A	0.930
48	KLK3-SERPINA3 PA2G4	BDNF CLIC1	MMP7	CHRD1	STX1A	0.930
49	BDNF PA2G4	MMP7 CLIC1	GHR	AHSG	STX1A	0.930
50	KLK3-SERPINA3 NME2	BDNF CLIC1	IGFBP2	MMP7	STX1A	0.930
51	KLK3-SERPINA3 STX1A	BDNF CLIC1	KIT	CDH1	MMP7	0.930
52	KLK3-SERPINA3 STX1A	BDNF CLIC1	CDH1	MMP7	LRIG3	0.930
53	BDNF CRP	GHR CLIC1	TGFB1	LRIG3	STX1A	0.930
54	KLK3-SERPINA3 PA2G4	BDNF CLIC1	MMP7	TGFB1	STX1A	0.930
55	KLK3-SERPINA3 NME2	BDNF CLIC1	KIT	MMP7	LRIG3	0.930
56	KLK3-SERPINA3 STX1A	BDNF CLIC1	MMP7	GHR	TGFB1	0.930
57	KLK3-SERPINA3 NME2	BDNF CLIC1	KIT	MMP7	STX1A	0.930
58	BDNF TPI1	MMP7 CLIC1	GHR	SERPINA1	STX1A	0.930
59	BDNF STX1A	EGFR CLIC1	MMP7	GHR	TGFB1	0.930
60	BDNF CRP	GHR CLIC1	TGFB1	CHRD1	STX1A	0.930
61	BDNF CLIC1	MMP7 PLA2G7	GHR	STX1A	CRP	0.930
62	BDNF CRP	MMP7 CLIC1	GHR	STX1A	TPT1	0.930
63	BDNF STX1A	KIT CLIC1	MMP7	GHR	TGFB1	0.930
64	BDNF PA2G4	KIT CLIC1	MMP7	GHR	STX1A	0.930
65	KLK3-SERPINA3 CLIC1	BDNF PLA2G7	MMP7	GHR	STX1A	0.930
66	KLK3-SERPINA3 NME2	BDNF CLIC1	IGFBP2	MMP7	LRIG3	0.930
67	BDNF NME2	KIT CRP	TGFB1	LRIG3	CHRD1	0.930
68	TGFB1 NME2	LRIG3 CRP	CHRD1	AHSG	STX1A	0.930
69	BDNF TPI1	MMP7 CLIC1	GHR	TGFB1	STX1A	0.929
70	KLK3-SERPINA3 CLIC1	BDNF PLA2G7	MMP7	STX1A	NME2	0.929
71	KLK3-SERPINA3 PA2G4	KIT CLIC1	MMP7	GHR	STX1A	0.929
72	KLK3-SERPINA3 TPI1	BDNF CLIC1	KIT	MMP7	LRIG3	0.929

TABLE 8-continued

Panels of 7 Biomarkers						
Markers						CV AUC
73	BDNF GPI	CDH1 CLIC1	MMP7	GHR	STX1A	0.929
74	BDNF TPT1	MMP7 CLIC1	GHR	TGFBI	STX1A	0.929
75	BDNF CRP	MMP7 CLIC1	GHR	LRIG3	GPI	0.929
76	KLK3-SERPINA3 NME2	BDNF CLIC1	MMP7	C9	STX1A	0.929
77	BDNF NME2	MMP7 CLIC1	C9	TGFBI	CMA1	0.929
78	CDH1 CRP	MMP7 CLIC1	GHR	TGFBI	STX1A	0.929
79	BDNF HNRNPAB	MMP7 CLIC1	GHR	C9	STX1A	0.929
80	BDNF YWHAG	MMP7 CLIC1	C9	LRIG3	STX1A	0.929
81	BDNF CRP	IGFBP2 CLIC1	TGFBI	LRIG3	STX1A	0.929
82	BDNF NME2	MMP7 CLIC1	GHR	LRIG3	STX1A	0.929
83	BDNF CRP	IGFBP2 CLIC1	MMP7	LRIG3	NME2	0.929
84	BDNF PA2G4	MMP7 CLIC1	GHR	CHRD1	STX1A	0.929
85	BDNF STX1A	MMP7 CLIC1	GHR	C9	TGFBI	0.929
86	BDNF NME2	MMP7 CLIC1	C9	TGFBI	STX1A	0.929
87	BDNF ITIH4	EGFR CLIC1	MMP7	STX1A	TPI1	0.929
88	KLK3-SERPINA3 YWHAG	BDNF CLIC1	MMP7	C9	STX1A	0.929
89	KLK3-SERPINA3 PA2G4	BDNF CLIC1	MMP7	LRIG3	STX1A	0.929
90	KLK3-SERPINA3 ITIH4	BDNF CLIC1	MMP7	STX1A	PA2G4	0.929
91	KLK3-SERPINA3 HNRNPAB	BDNF CLIC1	MMP7	LRIG3	STX1A	0.929
92	KLK3-SERPINA3 STX1A	BDNF CLIC1	EGFR	MMP7	TGFBI	0.929
93	BDNF ITIH4	MMP7 CLIC1	GHR	STX1A	TPI1	0.929
94	BDNF CRP	CDH1 CLIC1	MMP7	GHR	STX1A	0.929
95	BDNF CLIC1	MMP7 PLA2G7	GHR	STX1A	NME2	0.929
96	KLK3-SERPINA3 TPI1	BDNF CLIC1	MMP7	TGFBI	STX1A	0.929
97	BDNF ITIH4	MMP7 CLIC1	GHR	STX1A	PA2G4	0.929
98	MMP7 CRP	GHR CLIC1	BMP1	STX1A	NME2	0.929
99	KLK3-SERPINA3 CLIC1	BDNF PLA2G7	MMP7	STX1A	L1CAM	0.929
100	BDNF GPI	KIT CLIC1	MMP7	GHR	STX1A	0.929

TABLE 9

Panels of 8 Biomarkers						
Markers						CV AUC
1	KLK3-SERPINA3 STX1A	BDNF PA2G4	KIT CLIC1	MMP7	GHR	0.940
2	BDNF STX1A	TGFBI CRP	LRIG3 CLIC1	CHRD1	AHSG	0.938
3	BDNF NME2	MMP7 CRP	GHR CLIC1	TGFBI	STX1A	0.938

TABLE 9-continued

Panels of 8 Biomarkers						
Markers						CV AUC
4	KLK3-SERPINA3 STX1A	BDNF NME2	KIT CLIC1	MMP7	LRIG3	0.937
5	BDNF STX1A	MMP7 NME2	GHR CLIC1	TGFBI	LRIG3	0.937
6	KLK3-SERPINA3 STX1A	BDNF NME2	MMP7 CLIC1	GHR	TGFBI	0.937
7	KLK3-SERPINA3 STX1A	BDNF TPT1	KIT CLIC1	MMP7	GHR	0.937
8	KLK3-SERPINA3 STX1A	BDNF NME2	MMP7 CLIC1	GHR	LRIG3	0.937
9	BDNF STX1A	MMP7 GPI	GHR CLIC1	TGFBI	LRIG3	0.937
10	BDNF GPI	MMP7 CRP	GHR CLIC1	TGFBI	STX1A	0.936
11	KLK3-SERPINA3 STX1A	BDNF TPI1	KIT CLIC1	MMP7	GHR	0.936
12	BDNF STX1A	EGFR NME2	MMP7 CLIC1	GHR	TGFBI	0.936
13	BDNF PA2G4	MMP7 GPI	GHR CLIC1	C9	STX1A	0.936
14	KLK3-SERPINA3 STX1A	BDNF PA2G4	MMP7 CLIC1	GHR	TGFBI	0.936
15	KLK3-SERPINA3 STX1A	BDNF NME2	MMP7 CLIC1	TGFBI	LRIG3	0.936
16	BDNF AHSB	CDH1 STX1A	MMP7 CLIC1	GHR	TGFBI	0.936
17	KLK3-SERPINA3 STX1A	BDNF PA2G4	MMP7 CLIC1	GHR	AHSB	0.936
18	BDNF STX1A	EGFR GPI	MMP7 CLIC1	GHR	TGFBI	0.936
19	KLK3-SERPINA3 STX1A	BDNF NME2	KIT CLIC1	MMP7	GHR	0.936
20	KLK3-SERPINA3 PA2G4	BDNF GPI	MMP7 CLIC1	GHR	STX1A	0.936
21	BDNF STX1A	GHR CRP	TGFBI CLIC1	LRIG3	AHSB	0.936
22	BDNF TPI1	MMP7 CRP	GHR CLIC1	TGFBI	STX1A	0.936
23	BDNF STX1A	KIT NME2	MMP7 CLIC1	GHR	LRIG3	0.936
24	BDNF STX1A	KIT PA2G4	MMP7 CLIC1	GHR	C9	0.936
25	KLK3-SERPINA3 STX1A	BDNF TPT1	CDH1 CLIC1	MMP7	LRIG3	0.936
26	KLK3-SERPINA3 TPI1	BDNF ITIH4	KIT CLIC1	MMP7	STX1A	0.936
27	KLK3-SERPINA3 STX1A	BDNF GPI	MMP7 CLIC1	GHR	TGFBI	0.935
28	BDNF STX1A	KIT TPT1	MMP7 CLIC1	GHR	C9	0.935
29	KLK3-SERPINA3 STX1A	BDNF GPI	MMP7 CLIC1	TGFBI	LRIG3	0.935
30	KLK3-SERPINA3 NME2	BDNF ITIH4	KIT CLIC1	MMP7	STX1A	0.935
31	BDNF STX1A	MMP7 GPI	GHR CLIC1	TGFBI	AHSB	0.935
32	BDNF PA2G4	KIT ITIH4	MMP7 CLIC1	GHR	STX1A	0.935
33	KLK3-SERPINA3 STX1A	BDNF NME2	MMP7 CLIC1	LRIG3	CHRD1	0.935
34	BDNF PA2G4	MMP7 CRP	GHR CLIC1	TGFBI	STX1A	0.935
35	BDNF STX1A	MMP7 CRP	GHR CLIC1	TGFBI	LRIG3	0.935
36	BDNF GPI	MMP7 CRP	GHR CLIC1	TGFBI	LRIG3	0.935
37	KLK3-SERPINA3 STX1A	BDNF TPI1	MMP7 CLIC1	LRIG3	CHRD1	0.935
38	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT CLIC1	CDH1	MMP7	0.935
39	KLK3-SERPINA3 STX1A	BDNF TPT1	MMP7 CLIC1	LRIG3	CHRD1	0.935

TABLE 9-continued

Panels of 8 Biomarkers						
Markers						CV AUC
40	BDNF STX1A	GHR CRP	TGFBI CLIC1	CHRD1	AHSG	0.935
41	KLK3-SERPINA3 STX1A	BDNF TPT1	KIT CLIC1	MMP7	LRIG3	0.935
42	KLK3-SERPINA3 STX1A	BDNF TPI1	MMP7 CLIC1	GHR	TGFBI	0.935
43	KLK3-SERPINA3 STX1A	BDNF PA2G4	MMP7 CLIC1	GHR	CHRD1	0.935
44	BDNF STX1A	MMP7 PA2G4	GHR CLIC1	TGFBI	CHRD1	0.935
45	KLK3-SERPINA3 STX1A	BDNF TPT1	MMP7 CLIC1	GHR	CHRD1	0.935
46	BDNF STX1A	MMP7 TPT1	GHR CLIC1	C9	TGFBI	0.935
47	KLK3-SERPINA3 TGFBI	BDNF STX1A	KIT CLIC1	MMP7	GHR	0.935
48	KLK3-SERPINA3 STX1A	BDNF TPI1	KIT CLIC1	MMP7	LRIG3	0.935
49	BDNF CRP	MMP7 ITIH4	GHR CLIC1	STX1A	TPI1	0.935
50	KLK3-SERPINA3 PA2G4	BDNF ITIH4	KIT CLIC1	MMP7	STX1A	0.935
51	BDNF STX1A	CDH1 CRP	MMP7 CLIC1	GHR	TGFBI	0.935
52	BDNF STX1A	MMP7 NME2	GHR CLIC1	C9	TGFBI	0.935
53	BDNF STX1A	KIT GPI	MMP7 CLIC1	GHR	C9	0.935
54	KLK3-SERPINA3 STX1A	BDNF TPT1	MMP7 CLIC1	GHR	LRIG3	0.935
55	BDNF STX1A	KIT PA2G4	MMP7 CLIC1	GHR	TGFBI	0.935
56	BDNF STX1A	MMP7 NME2	GHR CLIC1	C9	AHSG	0.935
57	KLK3-SERPINA3 STX1A	BDNF GPI	MMP7 CLIC1	GHR	LRIG3	0.935
58	BDNF GPI	MMP7 CRP	GHR CLIC1	STX1A	NME2	0.935
59	BDNF AHSG	GHR CRP	TGFBI CLIC1	LRIG3	CHRD1	0.935
60	KLK3-SERPINA3 NME2	BDNF CLIC1	MMP7 PLA2G7	GHR	STX1A	0.935
61	BDNF TPI1	KIT ITIH4	MMP7 CLIC1	GHR	STX1A	0.935
62	BDNF NME2	MMP7 CLIC1	GHR PLA2G7	C9	STX1A	0.935
63	BDNF NME2	MMP7 ITIH4	GHR CLIC1	C9	STX1A	0.935
64	BDNF NME2	MMP7 CRP	GHR CLIC1	LRIG3	STX1A	0.935
65	BDNF STX1A	MMP7 YWHAG	GHR CLIC1	C9	TGFBI	0.935
66	BDNF PA2G4	MMP7 CRP	GHR CLIC1	CHRD1	STX1A	0.935
67	BDNF STX1A	MMP7 PA2G4	GHR CLIC1	TGFBI	LRIG3	0.935
68	BDNF TPI1	MMP7 ITIH4	GHR CLIC1	C9	STX1A	0.935
69	BDNF AHSG	EGFR STX1A	MMP7 CLIC1	GHR	TGFBI	0.935
70	BDNF TPT1	MMP7 CRP	GHR CLIC1	CHRD1	STX1A	0.935
71	BDNF CRP	MMP7 ITIH4	GHR CLIC1	STX1A	NME2	0.935
72	KLK3-SERPINA3 LRIG3	BDNF STX1A	MMP7 CLIC1	GHR	TGFBI	0.935
73	KLK3-SERPINA3 STX1A	BDNF TPI1	MMP7 CLIC1	TGFBI	LRIG3	0.935
74	BDNF CRP	MMP7 CLIC1	GHR PLA2G7	TGFBI	STX1A	0.935
75	KLK3-SERPINA3 STX1A	BDNF PA2G4	MMP7 CLIC1	TGFBI	CHRD1	0.935

TABLE 9-continued

Panels of 8 Biomarkers						
Markers						CV AUC
76	BDNF	MMP7	GHR	TGFBI	STX1A	0.934
	PA2G4	GPI	CLIC1			
77	BDNF	MMP7	GHR	C9	TGFBI	0.934
	STX1A	GPI	CLIC1			
78	BDNF	GHR	TGFBI	LRIG3	CHRD1	0.934
	STX1A	CRP	CLIC1			
79	BDNF	MMP7	GHR	TGFBI	STX1A	0.934
	TPI1	ITIH4	CLIC1			
80	KLK3-SERPINA3	BDNF	TGFBI	LRIG3	CHRD1	0.934
	STX1A	CRP	CLIC1			
81	BDNF	MMP7	GHR	C9	CHRD1	0.934
	STX1A	TPT1	CLIC1			
82	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.934
	STX1A	HNRNPAB	CLIC1			
83	BDNF	MMP7	GHR	CHRD1	STX1A	0.934
	NME2	CRP	CLIC1			
84	KLK3-SERPINA3	BDNF	EGFR	MMP7	TGFBI	0.934
	STX1A	NME2	CLIC1			
85	BDNF	KIT	MMP7	GHR	C9	0.934
	STX1A	HNRNPAB	CLIC1			
86	BDNF	MMP7	GHR	C9	STX1A	0.934
	NME2	GPI	CLIC1			
87	BDNF	KIT	MMP7	GHR	TGFBI	0.934
	LRIG3	STX1A	CLIC1			
88	BDNF	MMP7	GHR	TGFBI	AHSG	0.934
	STX1A	CLIC1	PLA2G7			
89	KLK3-SERPINA3	BDNF	KIT	MMP7	LRIG3	0.934
	STX1A	HNRNPAB	CLIC1			
90	BDNF	KIT	MMP7	GHR	TGFBI	0.934
	STX1A	NME2	CLIC1			
91	BDNF	MMP7	GHR	STX1A	NME2	0.934
	CRP	CLIC1	PLA2G7			
92	KLK3-SERPINA3	BDNF	MMP7	GHR	AHSG	0.934
	STX1A	NME2	CLIC1			
93	KLK3-SERPINA3	BDNF	MMP7	GHR	AHSG	0.934
	STX1A	TPI1	CLIC1			
94	BDNF	MMP7	GHR	TGFBI	STX1A	0.934
	CRP	HNRNPAB	CLIC1			
95	BDNF	MMP7	GHR	CHRD1	STX1A	0.934
	PA2G4	GPI	CLIC1			
96	KLK3-SERPINA3	BDNF	KIT	MMP7	LRIG3	0.934
	STX1A	PA2G4	CLIC1			
97	BDNF	MMP7	GHR	CHRD1	AHSG	0.934
	STX1A	PA2G4	CLIC1			
98	BDNF	MMP7	GHR	TGFBI	STX1A	0.934
	GPI	CLIC1	PLA2G7			
99	KLK3-SERPINA3	BDNF	MMP7	GHR	C9	0.934
	STX1A	TPT1	CLIC1			
100	BDNF	MMP7	GHR	C9	STX1A	0.934
	TPT1	ITIH4	CLIC1			

TABLE 10

Panels of 9 Biomarkers						
Markers						CV AUC
1	BDNF	MMP7	GHR	TGFBI	LRIG3	0.941
	STX1A	NME2	CRP	CLIC1		
2	BDNF	MMP7	GHR	TGFBI	CHRD1	0.941
	STX1A	PA2G4	CRP	CLIC1		
3	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.941
	TGFBI	STX1A	TPI1	CLIC1		
4	BDNF	KIT	MMP7	GHR	LRIG3	0.941
	STX1A	NME2	CRP	CLIC1		
5	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.941
	TGFBI	STX1A	PA2G4	CLIC1		
6	BDNF	MMP7	GHR	TGFBI	CHRD1	0.941
	STX1A	TPI1	CRP	CLIC1		

TABLE 10-continued

Panels of 9 Biomarkers						
Markers					CV AUC	
7	KLK3-SERPINA3 CHRD1	BDNF STX1A	MMP7 NME2	TGFBI CLIC1	LRIG3	0.940
8	BDNF STX1A	MMP7 GPI	GHR CRP	TGFBI CLIC1	LRIG3	0.940
9	KLK3-SERPINA3 CHRD1	BDNF STX1A	MMP7 TPI1	GHR CLIC1	TGFBI	0.940
10	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT NME2	MMP7 CLIC1	GHR	0.940
11	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT TPI1	MMP7 CLIC1	GHR	0.940
12	KLK3-SERPINA3 LRIG3	BDNF STX1A	MMP7 GPI	GHR CLIC1	TGFBI	0.940
13	KLK3-SERPINA3 STX1A	BDNF PA2G4	KIT GPI	MMP7 CLIC1	GHR	0.940
14	BDNF AHS	EGFR STX1A	MMP7 NME2	GHR CLIC1	TGFBI	0.940
15	BDNF STX1A	EGFR NME2	MMP7 CRP	GHR CLIC1	TGFBI	0.940
16	BDNF STX1A	MMP7 NME2	GHR CRP	TGFBI CLIC1	CHRD1	0.940
17	BDNF STX1A	KIT PA2G4	MMP7 GPI	GHR CLIC1	C9	0.940
18	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT PA2G4	MMP7 CLIC1	GHR	0.940
19	KLK3-SERPINA3 LRIG3	BDNF STX1A	MMP7 NME2	GHR CLIC1	TGFBI	0.940
20	BDNF STX1A	MMP7 GPI	GHR CRP	TGFBI CLIC1	AHS	0.940
21	BDNF STX1A	MMP7 TPT1	GHR CRP	TGFBI CLIC1	CHRD1	0.940
22	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT TPT1	MMP7 CLIC1	GHR	0.940
23	BDNF STX1A	CDH1 NME2	MMP7 CRP	GHR CLIC1	TGFBI	0.940
24	BDNF AHS	MMP7 STX1A	GHR PA2G4	TGFBI CLIC1	CHRD1	0.940
25	BDNF NME2	MMP7 GPI	GHR CRP	TGFBI CLIC1	STX1A	0.940
26	BDNF TPI1	KIT CRP	MMP7 ITIH4	GHR CLIC1	STX1A	0.940
27	KLK3-SERPINA3 CHRD1	BDNF STX1A	MMP7 TPT1	GHR CLIC1	TGFBI	0.939
28	KLK3-SERPINA3 TGFBI	BDNF LRIG3	KIT STX1A	MMP7 CLIC1	GHR	0.939
29	BDNF STX1A	IGFBP2 NME2	MMP7 CRP	TGFBI CLIC1	LRIG3	0.939
30	KLK3-SERPINA3 AHS	BDNF STX1A	KIT PA2G4	MMP7 CLIC1	GHR	0.939
31	BDNF STX1A	MMP7 PA2G4	GHR GPI	C9 CLIC1	CHRD1	0.939
32	BDNF TPI1	MMP7 CRP	GHR ITIH4	CHRD1 CLIC1	STX1A	0.939
33	KLK3-SERPINA3 CHRD1	BDNF STX1A	MMP7 TPI1	TGFBI CLIC1	LRIG3	0.939
34	KLK3-SERPINA3 TGFBI	BDNF STX1A	KIT NME2	MMP7 CLIC1	GHR	0.939
35	KLK3-SERPINA3 AHS	BDNF STX1A	MMP7 TPI1	GHR CLIC1	CHRD1	0.939
36	BDNF PA2G4	MMP7 CRP	GHR ITIH4	TGFBI CLIC1	STX1A	0.939
37	BDNF STX1A	KIT TPI1	MMP7 ITIH4	GHR CLIC1	C9	0.939
38	KLK3-SERPINA3 STX1A	BDNF NME2	MMP7 PA2G4	GHR CLIC1	CHRD1	0.939
39	BDNF PA2G4	KIT CRP	MMP7 ITIH4	GHR CLIC1	STX1A	0.939
40	BDNF PA2G4	MMP7 GPI	GHR CRP	CHRD1 CLIC1	STX1A	0.939
41	KLK3-SERPINA3 LRIG3	BDNF STX1A	IGFBP2 NME2	MMP7 CLIC1	TGFBI	0.939
42	BDNF NME2	MMP7 CRP	GHR ITIH4	TGFBI CLIC1	STX1A	0.939

TABLE 10-continued

Panels of 9 Biomarkers						
Markers				CV AUC		
43	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT HNRNPAB	MMP7 CLIC1	GHR	0.939
44	BDNF LRIG3	KIT STX1A	MMP7 NME2	GHR CLIC1	TGFB1	0.939
45	BDNF STX1A	MMP7 NME2	GHR GPI	C9 CLIC1	TGFB1	0.939
46	KLK3-SERPINA3 STX1A	BDNF NME2	MMP7 CRP	GHR CLIC1	TGFB1	0.939
47	KLK3-SERPINA3 STX1A	BDNF GPI	MMP7 CRP	GHR CLIC1	TGFB1	0.939
48	KLK3-SERPINA3 STX1A	BDNF TPI1	MMP7 CRP	GHR CLIC1	TGFB1	0.939
49	BDNF STX1A	KIT TPI1	MMP7 CRP	GHR CLIC1	TGFB1	0.939
50	BDNF SERPINA1	MMP7 STX1A	GHR TPI1	TGFB1 CLIC1	CHRD1	0.939
51	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT TPI1	MMP7 CLIC1	TGFB1	0.939
52	BDNF STX1A	KIT NME2	MMP7 CRP	GHR CLIC1	TGFB1	0.939
53	BDNF LRIG3	KIT STX1A	MMP7 CRP	GHR CLIC1	TGFB1	0.939
54	KLK3-SERPINA3 STX1A	BDNF PA2G4	KIT ITIH4	MMP7 CLIC1	GHR	0.939
55	KLK3-SERPINA3 STX1A	BDNF PA2G4	MMP7 GPI	GHR CLIC1	CHRD1	0.939
56	BDNF AHSG	GHR STX1A	TGFB1 CRP	LRIG3 CLIC1	CHRD1	0.939
57	KLK3-SERPINA3 GHR	BDNF STX1A	KIT TPT1	CDH1 CLIC1	MMP7	0.939
58	KLK3-SERPINA3 TGFB1	BDNF STX1A	EGFR PA2G4	MMP7 CLIC1	GHR	0.939
59	KLK3-SERPINA3 AHSG	BDNF STX1A	MMP7 GPI	GHR CLIC1	TGFB1	0.939
60	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT NME2	MMP7 CLIC1	TGFB1	0.939
61	KLK3-SERPINA3 CHRD1	BDNF STX1A	MMP7 PA2G4	GHR CLIC1	TGFB1	0.939
62	BDNF LRIG3	KIT STX1A	MMP7 TPI1	GHR CLIC1	TGFB1	0.939
63	BDNF AHSG	CDH1 STX1A	MMP7 CRP	GHR CLIC1	TGFB1	0.939
64	BDNF STX1A	MMP7 PA2G4	GHR CRP	CHRD1 CLIC1	AHSG	0.939
65	BDNF STX1A	KIT TPI1	MMP7 ITIH4	GHR CLIC1	TGFB1	0.939
66	KLK3-SERPINA3 AHSG	BDNF STX1A	MMP7 PA2G4	GHR CLIC1	TGFB1	0.939
67	BDNF AHSG	EGFR STX1A	MMP7 CLIC1	GHR PLA2G7	TGFB1	0.939
68	KLK3-SERPINA3 TGFB1	BDNF STX1A	KIT TPT1	MMP7 CLIC1	GHR	0.938
69	KLK3-SERPINA3 NME2	BDNF PA2G4	MMP7 ITIH4	CHRD1 CLIC1	STX1A	0.938
70	KLK3-SERPINA3 STX1A	BDNF TPI1	KIT ITIH4	MMP7 CLIC1	LRIG3	0.938
71	BDNF LRIG3	KIT STX1A	MMP7 NME2	GHR CLIC1	C9	0.938
72	BDNF LRIG3	KIT STX1A	MMP7 GPI	GHR CLIC1	TGFB1	0.938
73	KLK3-SERPINA3 STX1A	BDNF NME2	MMP7 GPI	GHR CLIC1	TGFB1	0.938
74	KLK3-SERPINA3 GHR	BDNF STX1A	KIT PA2G4	CDH1 CLIC1	MMP7	0.938
75	KLK3-SERPINA3 LRIG3	BDNF STX1A	EGFR NME2	MMP7 CLIC1	TGFB1	0.938
76	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT GPI	MMP7 CLIC1	GHR	0.938
77	BDNF STX1A	KIT TPI1	MMP7 CRP	GHR CLIC1	LRIG3	0.938
78	KLK3-SERPINA3 STX1A	BDNF PA2G4	MMP7 GPI	GHR CLIC1	AHSG	0.938

TABLE 10-continued

Panels of 9 Biomarkers						
Markers					CV AUC	
79	BDNF STX1A	KIT PA2G4	MMP7 ITIH4	GHR CLIC1	TGFBI	0.938
80	KLK3-SERPINA3 STX1A	BDNF TPT1	MMP7 PA2G4	GHR CLIC1	CHRD1	0.938
81	BDNF STX1A	MMP7 TPI1	GHR CRP	CHRD1	AHSG	0.938
82	KLK3-SERPINA3 LRIG3	BDNF STX1A	KIT NME2	CDH1 CLIC1	MMP7	0.938
83	BDNF STX1A	KIT PA2G4	MMP7 ITIH4	GHR CLIC1	C9	0.938
84	BDNF AHSG	IGFBP2 STX1A	MMP7 TPI1	GHR CLIC1	TGFBI	0.938
85	KLK3-SERPINA3 STX1A	BDNF TPI1	KIT ITIH4	MMP7 CLIC1	GHR	0.938
86	BDNF STX1A	MMP7 TPT1	GHR CRP	CHRD1	AHSG	0.938
87	BDNF TPI1	MMP7 CRP	GHR ITIH4	TGFBI CLIC1	STX1A	0.938
88	BDNF NME2	KIT CRP	MMP7 ITIH4	GHR CLIC1	STX1A	0.938
89	KLK3-SERPINA3 AHSG	BDNF STX1A	EGFR PA2G4	MMP7 CLIC1	GHR	0.938
90	KLK3-SERPINA3 CHRD1	BDNF STX1A	MMP7 NME2	GHR CLIC1	TGFBI	0.938
91	BDNF STX1A	MMP7 NME2	GHR GPI	TGFBI CLIC1	LRIG3	0.938
92	BDNF LRIG3	CDH1 AHSG	MMP7 STX1A	GHR CLIC1	TGFBI	0.938
93	BDNF STX1A	CDH1 GPI	MMP7 CRP	GHR CLIC1	TGFBI	0.938
94	KLK3-SERPINA3 NME2	BDNF GPI	MMP7 CRP	GHR CLIC1	STX1A	0.938
95	BDNF AHSG	MMP7 STX1A	GHR TPI1	C9 CLIC1	CHRD1	0.938
96	BDNF STX1A	KIT PA2G4	MMP7 CRP	GHR CLIC1	TGFBI	0.938
97	KLK3-SERPINA3 TGFBI	BDNF STX1A	EGFR TPI1	MMP7 CLIC1	GHR	0.938
98	BDNF NME2	MMP7 CRP	GHR CLIC1	TGFBI PLA2G7	STX1A	0.938
99	BDNF STX1A	MMP7 NME2	GHR CRP	TGFBI CLIC1	BMP1	0.938
100	BDNF STX1A	EGFR GPI	MMP7 CRP	GHR CLIC1	TGFBI	0.938

TABLE 11

Panels of 10 Biomarkers						
Markers					CV AUC	
1	BDNF SERPINA1	MMP7 STX1A	GHR NME2	TGFBI PA2G4	CHRD1 CLIC1	0.944
2	BDNF AHSG	MMP7 STX1A	GHR TPI1	TGFBI CRP	CHRD1 CLIC1	0.944
3	BDNF LRIG3	KIT STX1A	MMP7 NME2	GHR CRP	TGFBI CLIC1	0.944
4	BDNF STX1A	MMP7 PA2G4	GHR CRP	TGFBI ITIH4	CHRD1 CLIC1	0.944
5	BDNF STX1A	KIT TPI1	MMP7 CRP	GHR ITIH4	TGFBI CLIC1	0.944
6	BDNF STX1A	MMP7 TPI1	GHR CRP	TGFBI ITIH4	CHRD1 CLIC1	0.943
7	KLK3-SERPINA3 CHRD1	BDNF AHSG	MMP7 STX1A	GHR TPI1	TGFBI CLIC1	0.943
8	BDNF CHRD1	MMP7 STX1A	GHR TPI1	TGFBI CRP	LRIG3 CLIC1	0.943
9	KLK3-SERPINA3 CHRD1	BDNF STX1A	MMP7 TPI1	GHR CRP	TGFBI CLIC1	0.943

TABLE 11-continued

Panels of 10 Biomarkers						
Markers						CV AUC
10	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.943
		TGFB1	LRIG3	STX1A	TP11	CLIC1
11	BDNF	MMP7	GHR	TGFB1	CHRD1	0.943
	STX1A	PA2G4	GPI	CRP	CLIC1	
12	BDNF	MMP7	GHR	TGFB1	CHRD1	0.943
	AHSG	STX1A	NME2	CRP	CLIC1	
13	BDNF	IGFBP2	MMP7	GHR	TGFB1	0.943
	LRIG3	STX1A	NME2	CRP	CLIC1	
14	BDNF	KIT	MMP7	GHR	TGFB1	0.943
	STX1A	PA2G4	CRP	ITIH4	CLIC1	
15	BDNF	KIT	MMP7	GHR	TGFB1	0.943
	LRIG3	STX1A	TP11	CRP	CLIC1	
16	BDNF	MMP7	GHR	C9	TGFB1	0.943
	CHRD1	STX1A	PA2G4	GPI	CLIC1	
17	BDNF	MMP7	GHR	TGFB1	CHRD1	0.943
	AHSG	STX1A	PA2G4	CRP	CLIC1	
18	BDNF	MMP7	GHR	TGFB1	CHRD1	0.943
	STX1A	NME2	PA2G4	CRP	CLIC1	
19	BDNF	EGFR	MMP7	GHR	TGFB1	0.943
	CHRD1	STX1A	TP11	CRP	CLIC1	
20	BDNF	MMP7	GHR	TGFB1	CHRD1	0.943
	STX1A	NME2	CRP	ITIH4	CLIC1	
21	BDNF	MMP7	GHR	CHRD1	STX1A	0.943
	NME2	PA2G4	CRP	ITIH4	CLIC1	
22	BDNF	MMP7	GHR	TGFB1	LRIG3	0.942
	CHRD1	STX1A	NME2	CRP	CLIC1	
23	KLK3-SERPINA3	BDNF	MMP7	GHR	CHRD1	0.942
	STX1A	NME2	PA2G4	CRP	CLIC1	
24	KLK3-SERPINA3	BDNF	MMP7	TGFB1	LRIG3	0.942
	CHRD1	SERPINA1	STX1A	TP11	CLIC1	
25	KLK3-SERPINA3	BDNF	MMP7	GHR	CHRD1	0.942
	AHSG	STX1A	TP11	CRP	CLIC1	
26	BDNF	KIT	MMP7	GHR	TGFB1	0.942
	LRIG3	STX1A	TPT1	CRP	CLIC1	
27	BDNF	MMP7	GHR	TGFB1	LRIG3	0.942
	AHSG	STX1A	NME2	CRP	CLIC1	
28	KLK3-SERPINA3	BDNF	MMP7	GHR	CHRD1	0.942
	AHSG	STX1A	PA2G4	GPI	CLIC1	
29	BDNF	MMP7	GHR	C9	CHRD1	0.942
	AHSG	STX1A	TPT1	PA2G4	CLIC1	
30	KLK3-SERPINA3	BDNF	EGFR	MMP7	GHR	0.942
	TGFB1	AHSG	STX1A	TP11	CLIC1	
31	KLK3-SERPINA3	BDNF	EGFR	MMP7	GHR	0.942
	TGFB1	AHSG	STX1A	NME2	CLIC1	
32	BDNF	EGFR	MMP7	GHR	TGFB1	0.942
	AHSG	STX1A	NME2	ITIH4	CLIC1	
33	BDNF	MMP7	GHR	TGFB1	LRIG3	0.942
	STX1A	NME2	GPI	CRP	CLIC1	
34	BDNF	MMP7	GHR	C9	TGFB1	0.942
	CHRD1	STX1A	NME2	PA2G4	CLIC1	
35	KLK3-SERPINA3	BDNF	MMP7	GHR	TGFB1	0.942
	CHRD1	STX1A	PA2G4	CRP	CLIC1	
36	BDNF	MMP7	GHR	TGFB1	CHRD1	0.942
	AHSG	STX1A	PA2G4	ITIH4	CLIC1	
37	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	LRIG3	AHSG	STX1A	NME2	CLIC1	
38	BDNF	KIT	EGFR	MMP7	GHR	0.942
	TGFB1	STX1A	TP11	ITIH4	CLIC1	
39	KLK3-SERPINA3	BDNF	MMP7	GHR	TGFB1	0.942
	CHRD1	AHSG	STX1A	PA2G4	CLIC1	
40	KLK3-SERPINA3	BDNF	MMP7	GHR	TGFB1	0.942
	LRIG3	STX1A	GPI	CRP	CLIC1	
41	BDNF	MMP7	GHR	CHRD1	AHSG	0.942
	STX1A	PA2G4	GPI	CRP	CLIC1	
42	BDNF	KIT	MMP7	GHR	TGFB1	0.942
	LRIG3	STX1A	TP11	ITIH4	CLIC1	
43	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	TGFB1	AHSG	STX1A	PA2G4	CLIC1	
44	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	TGFB1	LRIG3	STX1A	PA2G4	CLIC1	
45	KLK3-SERPINA3	BDNF	MMP7	GHR	TGFB1	0.942
	CHRD1	STX1A	NME2	PA2G4	CLIC1	

TABLE 11-continued

Panels of 10 Biomarkers						
Markers						CV AUC
46	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	TGFB1	LRIG3	STX1A	NME2	CLIC1	
47	BDNF	MMP7	GHR	C9	CHRD1	0.942
	AHSG	STX1A	PA2G4	GPI	CLIC1	
48	BDNF	MMP7	GHR	C9	CHRD1	0.942
	AHSG	GPI	TPI1	CRP	CLIC1	
49	BDNF	CDH1	MMP7	GHR	TGFB1	0.942
	LRIG3	STX1A	NME2	CRP	CLIC1	
50	BDNF	EGFR	MMP7	GHR	TGFB1	0.942
	AHSG	STX1A	NME2	CRP	CLIC1	
51	BDNF	MMP7	GHR	TGFB1	CHRD1	0.942
	SERPINA1	STX1A	TPI1	CRP	CLIC1	
52	BDNF	CDH1	MMP7	GHR	CHRD1	0.942
	AHSG	STX1A	TPT1	CRP	CLIC1	
53	BDNF	EGFR	MMP7	GHR	TGFB1	0.942
	STX1A	TPI1	CRP	ITIH4	CLIC1	
54	KLK3-SERPINA3	BDNF	MMP7	GHR	TGFB1	0.942
	LRIG3	CHRD1	STX1A	TPI1	CLIC1	
55	BDNF	MMP7	GHR	TGFB1	LRIG3	0.942
	CHRD1	STX1A	CRP	HNRNPAB	CLIC1	
56	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	TGFB1	LRIG3	STX1A	TPT1	CLIC1	
57	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	LRIG3	STX1A	TPI1	ITIH4	CLIC1	
58	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	TGFB1	STX1A	PA2G4	CRP	CLIC1	
59	BDNF	MMP7	GHR	TGFB1	CHRD1	0.942
	STX1A	NME2	CRP	CLIC1	PLA2G7	
60	BDNF	EGFR	MMP7	GHR	C9	0.942
	TGFB1	AHSG	STX1A	NME2	CLIC1	
61	BDNF	KIT	MMP7	GHR	TGFB1	0.942
	CHRD1	TPI1	CRP	ITIH4	CLIC1	
62	BDNF	MMP7	GHR	TGFB1	LRIG3	0.942
	CHRD1	STX1A	GPI	CRP	CLIC1	
63	BDNF	MMP7	GHR	TGFB1	LRIG3	0.942
	CHRD1	AHSG	STX1A	NME2	CLIC1	
64	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	TGFB1	STX1A	PA2G4	ITIH4	CLIC1	
65	KLK3-SERPINA3	BDNF	KIT	EGFR	MMP7	0.942
	GHR	TGFB1	STX1A	PA2G4	CLIC1	
66	KLK3-SERPINA3	BDNF	KIT	CDH1	MMP7	0.942
	GHR	LRIG3	STX1A	NME2	CLIC1	
67	BDNF	MMP7	GHR	TGFB1	CHRD1	0.942
	STX1A	PA2G4	GPI	ITIH4	CLIC1	
68	BDNF	MMP7	GHR	TGFB1	LRIG3	0.942
	CHRD1	GPI	TPI1	CRP	CLIC1	
69	KLK3-SERPINA3	BDNF	MMP7	GHR	TGFB1	0.942
	STX1A	GPI	TPI1	CRP	CLIC1	
70	BDNF	MMP7	GHR	TGFB1	CHRD1	0.942
	FN1	STX1A	TPI1	CRP	CLIC1	
71	BDNF	EGFR	MMP7	GHR	TGFB1	0.942
	STX1A	NME2	CRP	ITIH4	CLIC1	
72	BDNF	MMP7	GHR	TGFB1	CHRD1	0.942
	SERPINA1	AHSG	STX1A	TPI1	CLIC1	
73	BDNF	CDH1	MMP7	GHR	CHRD1	0.942
	AHSG	STX1A	PA2G4	CRP	CLIC1	
74	BDNF	MMP7	GHR	TGFB1	CHRD1	0.942
	STX1A	GPI	TPI1	CRP	CLIC1	
75	BDNF	HMG1	MMP7	GHR	TGFB1	0.942
	CHRD1	AHSG	STX1A	CRP	CLIC1	
76	BDNF	MMP7	GHR	TGFB1	CHRD1	0.942
	STX1A	NME2	GPI	CRP	CLIC1	
77	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.942
	TGFB1	LRIG3	CHRD1	TPI1	CLIC1	
78	BDNF	KIT	MMP7	GHR	TGFB1	0.942
	LRIG3	CHRD1	NME2	CRP	CLIC1	
79	BDNF	MMP7	GHR	C9	TGFB1	0.942
	CHRD1	AHSG	TPI1	CRP	CLIC1	
80	BDNF	MMP7	GHR	C9	CHRD1	0.942
	AHSG	STX1A	GPI	TPI1	CLIC1	
81	BDNF	EGFR	MMP7	GHR	TGFB1	0.942
	CHRD1	AHSG	STX1A	NME2	CLIC1	

TABLE 11-continued

Panels of 10 Biomarkers						
Markers						CV AUC
82	BDNF	MMP7	GHR	TGFBI	LRIG3	0.941
	STX1A	NME2	CRP	CLIC1	PLA2G7	
83	BDNF	MMP7	GHR	C9	TGFBI	0.941
	CHRD1	AHSG	STX1A	TPT1	CLIC1	
84	BDNF	MMP7	GHR	TGFBI	CHRD1	0.941
	SERPINA1	STX1A	PA2G4	GPI	CLIC1	
85	BDNF	MMP7	GHR	TGFBI	CHRD1	0.941
	AHSG	STX1A	PA2G4	GPI	CLIC1	
86	KLK3-SERPINA3	BDNF	KIT	MMP7	GHR	0.941
	LRIG3	STX1A	NME2	GPI	CLIC1	
87	KLK3-SERPINA3	BDNF	MMP7	GHR	LRIG3	0.941
	CHRD1	AHSG	STX1A	TPT1	CLIC1	
88	KLK3-SERPINA3	BDNF	MMP7	GHR	CNTN1	0.941
	TGFBI	CHRD1	STX1A	PA2G4	CLIC1	
89	BDNF	MMP7	GHR	C9	CHRD1	0.941
	STX1A	NME2	PA2G4	ITIH4	CLIC1	
90	BDNF	IGFBP2	MMP7	GHR	TGFBI	0.941
	AHSG	STX1A	TPT1	CRP	CLIC1	
91	BDNF	MMP7	GHR	C9	TGFBI	0.941
	CHRD1	GPI	TPT1	CRP	CLIC1	
92	BDNF	MMP7	GHR	TGFBI	LRIG3	0.941
	CHRD1	STX1A	TPT1	CRP	CLIC1	
93	KLK3-SERPINA3	BDNF	MMP7	GHR	TGFBI	0.941
	LRIG3	STX1A	NME2	CRP	CLIC1	
94	BDNF	EGFR	MMP7	GHR	TGFBI	0.941
	CHRD1	STX1A	NME2	CRP	CLIC1	
95	BDNF	KIT	MMP7	GHR	C9	0.941
	LRIG3	STX1A	NME2	CRP	CLIC1	
96	BDNF	CDH1	MMP7	GHR	TGFBI	0.941
	LRIG3	AHSG	STX1A	CRP	CLIC1	
97	BDNF	MMP7	GHR	CHRD1	AHSG	0.941
	STX1A	GPI	TPT1	CRP	CLIC1	
98	BDNF	KIT	MMP7	GHR	LRIG3	0.941
	STX1A	NME2	GPI	CRP	CLIC1	
99	KLK3-SERPINA3	BDNF	MMP7	GHR	TGFBI	0.941
	AHSG	STX1A	TPT1	CRP	CLIC1	
100	BDNF	MMP7	GHR	C9	CHRD1	0.941
	AHSG	STX1A	NME2	GPI	CLIC1	

TABLE 12

Counts of markers in biomarker panels								
Biomarker	Panel Size							
	3	4	5	6	7	8	9	10
AHSG	37	45	59	85	116	159	222	349
AKR7A2	87	48	23	9	3	3	1	0
AKT3	0	0	0	0	0	0	0	1
BDNF	53	129	332	583	801	953	988	995
BMP1	81	93	84	74	42	32	26	23
BMPER	13	1	0	0	0	0	0	0
C9	131	178	252	244	233	211	203	194
CA6	29	14	1	0	0	0	0	0
CAPG	6	0	0	0	0	0	0	0
CDH1	22	56	104	105	112	129	145	166
CHRD1	50	61	81	98	116	170	304	477
CKB-CKM	26	18	8	8	6	2	0	1
CLIC1	260	447	669	883	978	994	1000	1000
CMA1	84	119	189	158	99	62	37	19
CNTN1	20	52	61	59	42	30	31	29
COL18A1	25	17	7	0	1	0	0	0
CRP	74	89	95	112	153	200	308	454
CTSL2	2	0	0	0	0	0	0	0
DDC	37	23	7	5	4	0	0	0
EGFR	63	47	27	41	50	88	100	121

TABLE 12-continued

Counts of markers in biomarker panels								
Biomarker	Panel Size							
	3	4	5	6	7	8	9	10
FGA-FGB-FGG	23	0	0	0	0	0	0	0
FN1	3	0	0	2	0	2	8	18
GHR	32	67	159	315	452	587	745	850
GPI	71	79	103	147	167	183	202	225
HMGB1	15	36	11	17	19	4	6	4
HNRNPAB	46	27	35	45	60	41	38	32
HP	21	7	0	0	0	0	0	0
HSP90AA1	2	0	0	0	0	0	0	0
HSPA1A	6	2	0	0	0	0	0	0
IGFBP2	42	51	74	105	142	129	91	67
IGFBP4	19	6	1	3	2	0	5	6
ITIH4	23	46	51	64	117	163	180	208
KIT	21	26	30	51	109	203	295	327
KLK3-SERPINA3	111	188	262	287	307	338	377	378
L1CAM	41	45	44	16	9	8	3	8
LRIG3	109	161	241	293	330	367	376	407
MMP12	71	29	5	2	0	0	0	0
MMP7	270	626	782	852	916	960	982	996
NME2	83	77	112	159	189	251	282	299

TABLE 12-continued

Counts of markers in biomarker panels								
Biomarker	Panel Size							
	3	4	5	6	7	8	9	10
PA2G4	7	33	41	57	85	146	203	275
PLA2G7	17	32	28	30	47	67	70	66
PLAUR	33	22	11	5	0	0	0	0
PRKACA	8	0	0	0	0	0	0	0
PRKCB	3	0	0	0	0	0	0	0
PROK11	2	0	0	0	0	0	0	0
PRSS2	5	0	0	0	0	0	0	0
PTN	17	2	0	0	0	0	0	0
SERPINA1	51	35	23	16	29	36	43	68
STC1	17	10	7	4	4	8	8	7
STX1A	131	268	345	520	691	823	902	934
TACSTD2	7	1	2	0	1	0	0	3
TGFBI	62	98	136	191	266	339	462	579
TP11	42	64	106	124	139	187	243	305
TPT1	54	33	22	29	67	88	108	108
YWHAG	419	492	369	202	96	37	6	1
YWHAH	16	0	1	0	0	0	0	0

TABLE 13

Analytes in ten marker classifiers	
CLIC1	BDNF
MMP7	STX1A
GHR	TGFBI
CHRD1	CRP
LRIG3	KLK3-SERPINA3
AHSG	KIT

TABLE 14

Parameters derived from training set for naïve Bayes classifier.				
Biomarker	μ_c	μ_d	σ_c	σ_d
BMPER	7.450	7.323	0.108	0.164
COL18A1	8.763	8.876	0.125	0.162
CMA1	6.800	6.754	0.047	0.041
MMP7	8.881	9.232	0.235	0.182
KIT	9.603	9.503	0.139	0.141
IGFBP2	8.514	9.006	0.417	0.448
PROK11	6.196	6.154	0.042	0.058
DDC	6.746	6.711	0.034	0.043
PRKACA	7.594	7.753	0.187	0.113
FGA-FGB-FGG	9.836	10.258	0.338	0.580

TABLE 14-continued

Parameters derived from training set for naïve Bayes classifier.				
Biomarker	μ_c	μ_d	σ_c	σ_d
CNTN1	9.265	9.149	0.181	0.114
CRP	7.733	9.005	1.095	1.422
HNRNPAB	7.252	7.517	0.304	0.225
HSP90AA1	9.165	9.343	0.226	0.182
PLA2G7	10.131	9.952	0.277	0.184
BDNF	6.931	6.854	0.102	0.068
AKR7A2	6.761	7.155	0.432	0.248
IGFBP4	8.138	8.268	0.140	0.163
PLAUR	8.248	8.385	0.133	0.178
C9	11.715	11.936	0.189	0.223
SERPINA1	10.215	10.371	0.169	0.239
STC1	8.475	8.691	0.242	0.293
HP	11.848	12.057	0.222	0.196
L1CAM	7.893	7.721	0.226	0.152
ITIH4	10.596	10.738	0.121	0.227
BMP1	8.766	8.548	0.213	0.234
TFF3	8.288	8.536	0.195	0.307
PRKCB	6.817	6.780	0.051	0.060
IL12B-IL23A	6.189	6.153	0.037	0.039
CLIC1	7.907	8.260	0.259	0.230
CDH1	9.252	9.050	0.200	0.181
CHRD1	8.665	8.938	0.215	0.388
EGFR	10.578	10.428	0.119	0.135
ASGR1	6.661	6.619	0.050	0.052
TACSTD2	6.879	6.849	0.040	0.043
PRSS2	10.080	10.457	0.421	0.529
AKT3	7.816	7.886	0.074	0.068
HMGB1	8.430	8.546	0.133	0.096
CAPG	7.271	7.602	0.272	0.277
YWHAH	7.644	7.774	0.107	0.105
PTN	8.149	8.250	0.116	0.152
YWHAG	8.156	8.496	0.205	0.187
CTSL2	6.262	6.207	0.063	0.069
GHR	7.724	7.595	0.135	0.102
TGFBI	9.944	9.777	0.178	0.239
GPI	7.506	7.760	0.278	0.260
TP11	9.087	9.392	0.450	0.221
STX1A	7.186	7.143	0.035	0.033
LRIG3	7.411	7.301	0.090	0.092
TPT1	8.847	9.137	0.290	0.224
PA2G4	7.735	8.026	0.643	0.329
NME2	6.333	6.618	0.339	0.242
CKB-CKM	7.515	7.230	0.317	0.307
CA6	7.180	7.038	0.228	0.108
AHSG	11.197	11.107	0.149	0.134
KLK3-SERPINA3	8.102	8.327	0.194	0.330
FN1	9.286	9.058	0.239	0.325
MMP12	6.129	6.323	0.100	0.260
HSPA1A	8.819	9.011	0.316	0.224

TABLE 15

AUC for exemplary combinations of biomarkers									
#									AUC
1	MMP7								0.803
2	MMP7	CLIC1							0.883
3	MMP7	CLIC1	STX1A						0.901
4	MMP7	CLIC1	STX1A	CHRD1					0.899
5	MMP7	CLIC1	STX1A	CHRD1	PA2G4				0.912
6	MMP7	CLIC1	STX1A	CHRD1	PA2G4	SERPINA1			0.922
7	MMP7	CLIC1	STX1A	CHRD1	PA2G4	SERPINA1	BDNF		0.930
8	MMP7	CLIC1	STX1A	CHRD1	PA2G4	SERPINA1	BDNF	GHR	0.937
9	MMP7	CLIC1	STX1A	CHRD1	PA2G4	SERPINA1	BDNF	GHR TGFBI	0.944
10	MMP7	CLIC1	STX1A	CHRD1	PA2G4	SERPINA1	BDNF	GHR TGFBI NME2	0.948

TABLE 16

Calculations derived from training set for naïve Bayes classifier.								
Biomarker	μ_c	μ_d	σ_c	σ_d	$\tilde{x}$	$p(c \tilde{x})$	$p(d \tilde{x})$	$\ln(p(d \tilde{x})/p(c \tilde{x}))$
GHR	7.724	7.595	0.135	0.102	7.860	1.778	0.136	-2.572
SERPINA1	10.215	10.371	0.169	0.239	10.573	0.252	1.166	1.531
STX1A	7.186	7.143	0.035	0.033	7.259	1.382	0.024	-4.053
CHRD1	8.665	8.938	0.215	0.388	8.405	0.896	0.401	-0.804
CLIC1	7.907	8.260	0.259	0.230	8.068	1.267	1.226	-0.034
PA2G4	7.735	8.026	0.643	0.329	7.285	0.486	0.096	-1.622
NME2	6.333	6.618	0.339	0.242	6.322	1.175	0.783	-0.406
MMP7	8.881	9.232	0.235	0.182	8.684	1.194	0.023	-3.942
TGFB1	9.944	9.777	0.178	0.239	9.778	1.446	1.669	0.144
BDNF	6.931	6.854	0.102	0.068	6.904	3.768	4.484	0.174

TABLE 17

Clinical characteristics of the training set				
Meta Data	Levels	Control	Disease	p-value
Samples		218	46	
GENDER	F	118	36	
	M	100	10	4.34e-03
AGE	Mean	57.2	67.3	
	SD	10.2	10.8	2.35e-07
CANCER STAGE	I	0	26	
	II	0	4	
	III	0	7	
	IV	0	9	NaN
TOBACCO USER	Never	1	2	
	Not Reported	3	10	
	Past	84	24	
	Current	130	10	1.27e-10

TABLE 18

Ten biomarker classifier proteins			
Biomarker	UniProt ID	Direction*	Biological Process (GO)
BDNF	P23560	Down	response to stress cell communication regulation of cell death signaling process
MMP7	P09237	Up	proteolysis
GHR	P10912	Down	regulation of cell death signaling process signaling
TGFB1	Q15582	Down	regulation of signaling pathway cell proliferation regulation of cell adhesion
CHRD1	Q9BU40	Up	signaling
SERPINA1	P01009	Up	response to stress
STX1A	Q16623	Down	cell communication signaling
NME2	P22392	Up	
PA2G4	Q9UQ80	Up	cell proliferation
CLIC1	O00299	Up	signaling process

TABLE 19

Biomarkers of general cancer	
KLK3-SERPINA3	EGFR
BMPER	FGA-FGB-FGG
C9	STX1A

TABLE 19-continued

Biomarkers of general cancer	
AKR7A2	CKB-CKM
DDC	CA6
IGFBP2	IGFBP4
FN1	BMP1
CRP	KIT
CNTN1	SERPINA1
BDNF	GHR
ITIH4	NME2
AHSG	

TABLE 20

Panels of 1 Biomarker		
Markers		Mean CV AUC
1	C9	0.792
2	KLK3-SERPINA3	0.782
3	CRP	0.763
4	BMPER	0.745
5	BMP1	0.732
6	KIT	0.729
7	AKR7A2	0.726
8	EGFR	0.726
9	ITIH4	0.721
10	IGFBP2	0.720
11	BDNF	0.720
12	STX1A	0.719
13	NME2	0.714
14	FGA-FGB-FGG	0.712
15	CNTN1	0.708
16	CKB-CKM	0.708
17	AHSG	0.707
18	GHR	0.704
19	IGFBP4	0.703
20	CA6	0.700
21	DDC	0.696
22	FN1	0.694
23	SERPINA1	0.688

TABLE 21

Panels of 2 Biomarkers			
Markers			Mean CV AUC
1	C9	AKR7A2	0.832
2	KLK3-SERPINA3	AKR7A2	0.831

TABLE 21-continued

Panels of 2 Biomarkers			
Markers			Mean CV AUC
3	KLK3-SERPINA3	NME2	0.828
4	AKR7A2	CRP	0.827
5	KLK3-SERPINA3	EGFR	0.826
6	KLK3-SERPINA3	STX1A	0.826
7	C9	NME2	0.824
8	KLK3-SERPINA3	BDNF	0.823
9	KLK3-SERPINA3	IGFBP4	0.822
10	KLK3-SERPINA3	CA6	0.819
11	KIT	C9	0.819
12	BDNF	C9	0.818
13	KLK3-SERPINA3	BMP1	0.816
14	KLK3-SERPINA3	BMPER	0.816
15	NME2	CRP	0.815
16	KLK3-SERPINA3	KIT	0.815
17	C9	BMPER	0.814
18	BMPER	NME2	0.812
19	KLK3-SERPINA3	C9	0.811
20	KLK3-SERPINA3	CRP	0.811
21	C9	STX1A	0.811
22	EGFR	C9	0.811
23	BMPER	AKR7A2	0.810
24	BMPER	CRP	0.810
25	BDNF	CRP	0.810
26	C9	DDC	0.809
27	KLK3-SERPINA3	CNTN1	0.809
28	KLK3-SERPINA3	IGFBP2	0.808
29	SERPINA1	AKR7A2	0.808
30	AKR7A2	ITIH4	0.808
31	C9	AHSG	0.807
32	IGFBP4	C9	0.807
33	KLK3-SERPINA3	DDC	0.807
34	BMP1	AKR7A2	0.806
35	CNTN1	C9	0.806
36	STX1A	CRP	0.805
37	IGFBP2	CRP	0.805
38	NME2	ITIH4	0.805
39	BMP1	CRP	0.805
40	KLK3-SERPINA3	AHSG	0.804
41	C9	CA6	0.803
42	C9	CRP	0.802
43	GHR	C9	0.802
44	BDNF	AKR7A2	0.802
45	KLK3-SERPINA3	FN1	0.801
46	BDNF	KIT	0.801
47	KLK3-SERPINA3	GHR	0.799
48	EGFR	ITIH4	0.799
49	C9	BMP1	0.798
50	KIT	CRP	0.798
51	IGFBP2	C9	0.798
52	BMP1	NME2	0.797
53	C9	ITIH4	0.797
54	EGFR	AKR7A2	0.797
55	NME2	FGA-FGB-FGG	0.796
56	EGFR	CRP	0.795
57	IGFBP2	AKR7A2	0.795
58	STX1A	ITIH4	0.795
59	SERPINA1	NME2	0.795
60	KIT	AKR7A2	0.795
61	IGFBP2	BMPER	0.794
62	CNTN1	AKR7A2	0.794
63	C9	FN1	0.794
64	AKR7A2	FGA-FGB-FGG	0.793
65	BDNF	NME2	0.793
66	GHR	CRP	0.792
67	AHSG	AKR7A2	0.792
68	CNTN1	BMPER	0.791
69	KIT	BMP1	0.791
70	CNTN1	BMP1	0.791
71	KIT	BMPER	0.790
72	KLK3-SERPINA3	ITIH4	0.790

TABLE 21-continued

Panels of 2 Biomarkers			
Markers			Mean CV AUC
73	DDC	CRP	0.789
74	CA6	CRP	0.788
75	IGFBP4	AKR7A2	0.788
76	IGFBP4	CRP	0.788
77	GHR	BMPER	0.787
78	IGFBP2	CNTN1	0.787
79	EGFR	NME2	0.787
80	BMPER	ITIH4	0.786
81	BDNF	CNTN1	0.785
82	C9	CKB-CKM	0.785
83	GHR	AKR7A2	0.785
84	FN1	CRP	0.784
85	BDNF	BMPER	0.784
86	CNTN1	CRP	0.784
87	KLK3-SERPINA3	CKB-CKM	0.784
88	EGFR	AHSG	0.783
89	EGFR	BMPER	0.783
90	STX1A	NME2	0.783
91	BMP1	BMPER	0.783
92	DDC	ITIH4	0.783
93	CA6	BMPER	0.782
94	STX1A	AKR7A2	0.781
95	CRP	ITIH4	0.781
96	BDNF	ITIH4	0.780
97	IGFBP2	ITIH4	0.780
98	AHSG	NME2	0.779
99	CNTN1	NME2	0.779
100	CA6	AKR7A2	0.778

TABLE 22

Panels of 3 Biomarkers				
Markers				Mean CV AUC
1	IGFBP2	AKR7A2	CRP	0.849
2	KLK3-SERPINA3	BMPER	NME2	0.849
3	KLK3-SERPINA3	C9	AKR7A2	0.848
4	KLK3-SERPINA3	AKR7A2	CRP	0.848
5	KLK3-SERPINA3	EGFR	AKR7A2	0.848
6	BMP1	AKR7A2	CRP	0.848
7	C9	BMPER	AKR7A2	0.848
8	C9	BMPER	NME2	0.848
9	KLK3-SERPINA3	BMP1	AKR7A2	0.847
10	C9	AKR7A2	CRP	0.847
11	KLK3-SERPINA3	BMP1	NME2	0.847
12	BDNF	KIT	C9	0.846
13	BDNF	C9	AKR7A2	0.845
14	KLK3-SERPINA3	EGFR	NME2	0.845
15	BMPER	NME2	CRP	0.845
16	BMPER	AKR7A2	CRP	0.845
17	KLK3-SERPINA3	BMPER	AKR7A2	0.845
18	KLK3-SERPINA3	BDNF	AKR7A2	0.844
19	KIT	C9	AKR7A2	0.844
20	KLK3-SERPINA3	NME2	CRP	0.844
21	EGFR	C9	AKR7A2	0.844
22	BDNF	AKR7A2	CRP	0.844
23	KLK3-SERPINA3	IGFBP4	AKR7A2	0.843
24	CNTN1	C9	AKR7A2	0.843
25	KLK3-SERPINA3	CA6	AKR7A2	0.843
26	C9	AHSG	AKR7A2	0.843
27	KLK3-SERPINA3	IGFBP2	AKR7A2	0.843
28	KLK3-SERPINA3	BDNF	KIT	0.842
29	KLK3-SERPINA3	C9	NME2	0.842
30	KLK3-SERPINA3	CNTN1	AKR7A2	0.842
31	KLK3-SERPINA3	BDNF	NME2	0.841
32	BMP1	NME2	CRP	0.841

TABLE 22-continued

Panels of 3 Biomarkers				
Markers				Mean CV AUC
33	KLK3-SERPINA3	KIT	AKR7A2	0.841
34	KIT	AKR7A2	CRP	0.841
35	BMPER	NME2	ITIH4	0.840
36	EGFR	AKR7A2	CRP	0.840
37	KLK3-SERPINA3	STX1A	AKR7A2	0.840
38	IGFBP4	C9	AKR7A2	0.839
39	KLK3-SERPINA3	IGFBP4	NME2	0.839
40	KLK3-SERPINA3	CNTN1	NME2	0.839
41	C9	DDC	AKR7A2	0.839
42	BDNF	C9	NME2	0.839
43	GHR	AKR7A2	CRP	0.839
44	C9	BMP1	AKR7A2	0.839
45	KLK3-SERPINA3	BDNF	CNTN1	0.838
46	KLK3-SERPINA3	STX1A	NME2	0.838
47	IGFBP2	C9	AKR7A2	0.838
48	GHR	C9	AKR7A2	0.838
49	C9	AKR7A2	ITIH4	0.838
50	BMP1	BMPER	NME2	0.837
51	BDNF	KIT	CRP	0.837
52	C9	STX1A	AKR7A2	0.837
53	BDNF	NME2	CRP	0.837
54	KLK3-SERPINA3	AKR7A2	ITIH4	0.837
55	C9	NME2	CRP	0.836
56	C9	NME2	ITIH4	0.836
57	BMP1	BMPER	AKR7A2	0.836
58	KLK3-SERPINA3	BDNF	C9	0.836
59	KLK3-SERPINA3	AHSG	AKR7A2	0.836
60	KLK3-SERPINA3	CA6	NME2	0.835
61	KLK3-SERPINA3	GHR	AKR7A2	0.835
62	KIT	C9	NME2	0.835
63	KLK3-SERPINA3	CNTN1	BMP1	0.835
64	C9	AHSG	NME2	0.835
65	BDNF	KIT	AKR7A2	0.835
66	KLK3-SERPINA3	IGFBP2	NME2	0.835
67	STX1A	AKR7A2	CRP	0.835
68	KLK3-SERPINA3	KIT	STX1A	0.835
69	KLK3-SERPINA3	NME2	ITIH4	0.835
70	KLK3-SERPINA3	SERPINA1	AKR7A2	0.834
71	IGFBP4	AKR7A2	CRP	0.834
72	IGFBP2	BMPER	AKR7A2	0.834
73	EGFR	C9	NME2	0.834
74	KLK3-SERPINA3	BDNF	CRP	0.834
75	KLK3-SERPINA3	STX1A	CRP	0.834
76	GHR	BMPER	AKR7A2	0.833
77	IGFBP2	NME2	CRP	0.833
78	KLK3-SERPINA3	CNTN1	BMPER	0.833
79	KLK3-SERPINA3	KIT	BMP1	0.833
80	KLK3-SERPINA3	BDNF	EGFR	0.833
81	CNTN1	C9	NME2	0.833
82	KLK3-SERPINA3	KIT	NME2	0.833
83	KLK3-SERPINA3	BDNF	STX1A	0.833
84	KLK3-SERPINA3	AHSG	NME2	0.833
85	CNTN1	AKR7A2	CRP	0.833
86	C9	SERPINA1	AKR7A2	0.833
87	KLK3-SERPINA3	C9	STX1A	0.833
88	KLK3-SERPINA3	BDNF	CA6	0.833
89	EGFR	AKR7A2	ITIH4	0.833
90	KLK3-SERPINA3	KIT	EGFR	0.833
91	C9	DDC	NME2	0.833
92	KLK3-SERPINA3	DDC	AKR7A2	0.833
93	CNTN1	BMP1	AKR7A2	0.832
94	AKR7A2	CRP	ITIH4	0.832
95	KLK3-SERPINA3	EGFR	ITIH4	0.832
96	CNTN1	BMPER	AKR7A2	0.832
97	KLK3-SERPINA3	EGFR	AHSG	0.832
98	KLK3-SERPINA3	BDNF	IGFBP4	0.832
99	IGFBP4	SERPINA1	AKR7A2	0.832
100	SERPINA1	BMPER	AKR7A2	0.832

TABLE 23

Panels of 4 Biomarkers					Mean CV AUC
Markers					
1	BDNF	KIT	AKR7A2	CRP	0.860
2	KLK3-SERPINA3	CNTN1	BMPER	NME2	0.860
3	BDNF	KIT	C9	AKR7A2	0.859
4	KLK3-SERPINA3	BMP1	BMPER	NME2	0.859
5	KIT	BMP1	AKR7A2	CRP	0.859
6	KLK3-SERPINA3	BMP1	NME2	CRP	0.858
7	KLK3-SERPINA3	CNTN1	BMP1	NME2	0.858
8	KLK3-SERPINA3	EGFR	AKR7A2	CRP	0.857
9	KLK3-SERPINA3	C9	BMPER	AKR7A2	0.857
10	KLK3-SERPINA3	KIT	C9	AKR7A2	0.857
11	KLK3-SERPINA3	BMP1	AKR7A2	CRP	0.857
12	KLK3-SERPINA3	IGFBP2	AKR7A2	CRP	0.857
13	C9	BMPER	AKR7A2	CRP	0.857
14	KLK3-SERPINA3	IGFBP4	C9	AKR7A2	0.857
15	GHR	BMPER	AKR7A2	CRP	0.857
16	CNTN1	C9	BMPER	AKR7A2	0.857
17	BDNF	IGFBP2	AKR7A2	CRP	0.857
18	KIT	C9	AKR7A2	CRP	0.857
19	IGFBP2	BMPER	AKR7A2	CRP	0.857
20	KLK3-SERPINA3	EGFR	C9	AKR7A2	0.856
21	KLK3-SERPINA3	CNTN1	BMP1	AKR7A2	0.856
22	KLK3-SERPINA3	CNTN1	C9	AKR7A2	0.856
23	KLK3-SERPINA3	IGFBP4	AKR7A2	CRP	0.856
24	KLK3-SERPINA3	C9	BMPER	NME2	0.856
25	KLK3-SERPINA3	KIT	BMP1	AKR7A2	0.856
26	KLK3-SERPINA3	BMPER	NME2	CRP	0.856
27	GHR	C9	BMPER	AKR7A2	0.856
28	CNTN1	C9	BMPER	NME2	0.856
29	GHR	BMPER	NME2	CRP	0.855
30	KLK3-SERPINA3	BDNF	KIT	AKR7A2	0.855
31	BDNF	C9	AKR7A2	CRP	0.855
32	KLK3-SERPINA3	C9	AKR7A2	CRP	0.855
33	KLK3-SERPINA3	BDNF	AKR7A2	CRP	0.855
34	IGFBP2	BMPER	NME2	CRP	0.855
35	KLK3-SERPINA3	CNTN1	BMPER	AKR7A2	0.855
36	KLK3-SERPINA3	BMPER	AKR7A2	CRP	0.855
37	BMP1	BMPER	AKR7A2	CRP	0.855
38	KLK3-SERPINA3	EGFR	BMPER	NME2	0.855
39	CNTN1	C9	BMP1	AKR7A2	0.855
40	KLK3-SERPINA3	KIT	AKR7A2	CRP	0.854
41	KLK3-SERPINA3	GHR	BMPER	NME2	0.854
42	KLK3-SERPINA3	IGFBP4	BMPER	NME2	0.854
43	IGFBP2	C9	AKR7A2	CRP	0.854
44	KLK3-SERPINA3	IGFBP2	CNTN1	AKR7A2	0.854
45	KLK3-SERPINA3	BDNF	C9	AKR7A2	0.854
46	GHR	C9	BMPER	NME2	0.854
47	KLK3-SERPINA3	BMPER	NME2	ITIH4	0.854
48	KIT	IGFBP2	AKR7A2	CRP	0.854
49	KLK3-SERPINA3	EGFR	NME2	CRP	0.854
50	KIT	C9	BMPER	AKR7A2	0.854
51	KIT	EGFR	C9	AKR7A2	0.854
52	BMP1	BMPER	NME2	CRP	0.854
53	KLK3-SERPINA3	IGFBP2	BMPER	AKR7A2	0.853
54	EGFR	C9	AHSG	AKR7A2	0.853
55	KLK3-SERPINA3	EGFR	NME2	ITIH4	0.853
56	IGFBP2	CNTN1	AKR7A2	CRP	0.853
57	C9	BMPER	NME2	ITIH4	0.853
58	IGFBP2	BMP1	AKR7A2	CRP	0.853
59	KLK3-SERPINA3	CNTN1	AKR7A2	CRP	0.853
60	KLK3-SERPINA3	IGFBP4	C9	NME2	0.853
61	KLK3-SERPINA3	IGFBP2	BMPER	NME2	0.853
62	KLK3-SERPINA3	IGFBP4	SERPINA1	AKR7A2	0.853
63	BDNF	CNTN1	C9	AKR7A2	0.853
64	CNTN1	BMP1	AKR7A2	CRP	0.853
65	KLK3-SERPINA3	BDNF	CNTN1	AKR7A2	0.853
66	BDNF	KIT	C9	NME2	0.853
67	KLK3-SERPINA3	CNTN1	C9	NME2	0.853
68	KLK3-SERPINA3	EGFR	BMPER	AKR7A2	0.853
69	KLK3-SERPINA3	IGFBP4	AKR7A2	ITIH4	0.853
70	KLK3-SERPINA3	IGFBP4	NME2	CRP	0.853

TABLE 23-continued

Panels of 4 Biomarkers					Mean CV AUC
Markers					
71	KLK3-SERPINA3	IGFBP4	BMP1	AKR7A2	0.852
72	EGFR	C9	AKR7A2	ITIH4	0.852
73	EGFR	C9	AKR7A2	CRP	0.852
74	KLK3-SERPINA3	KIT	BMP1	NME2	0.852
75	KLK3-SERPINA3	KIT	EGFR	AKR7A2	0.852
76	KLK3-SERPINA3	EGFR	AKR7A2	ITIH4	0.852
77	KLK3-SERPINA3	BDNF	NME2	CRP	0.852
78	IGFBP4	C9	AKR7A2	ITIH4	0.852
79	KLK3-SERPINA3	GHR	BMPER	AKR7A2	0.852
80	KLK3-SERPINA3	BMP1	BMPER	AKR7A2	0.852
81	IGFBP2	C9	BMPER	AKR7A2	0.852
82	BDNF	KIT	NME2	CRP	0.852
83	KLK3-SERPINA3	KIT	C9	NME2	0.852
84	IGFBP2	AKR7A2	CRP	ITIH4	0.852
85	C9	BMPER	AKR7A2	ITIH4	0.852
86	KLK3-SERPINA3	EGFR	BMP1	AKR7A2	0.852

TABLE 23-continued

Panels of 4 Biomarkers					Mean CV AUC
Markers					
87	KLK3-SERPINA3	C9	CA6	AKR7A2	0.852
88	KLK3-SERPINA3	NME2	CRP	ITIH4	0.852
89	EGFR	CNTN1	C9	AKR7A2	0.852
90	KLK3-SERPINA3	C9	STX1A	AKR7A2	0.852
91	C9	BMPER	NME2	CRP	0.852
92	KIT	CNTN1	C9	AKR7A2	0.852
93	KLK3-SERPINA3	IGFBP4	BMPER	AKR7A2	0.851
94	KIT	C9	BMP1	AKR7A2	0.851
95	KLK3-SERPINA3	KIT	BMPER	NME2	0.851
96	KLK3-SERPINA3	CNTN1	NME2	CRP	0.851
97	KLK3-SERPINA3	BDNF	KIT	NME2	0.851
98	BDNF	C9	AHSG	AKR7A2	0.851
99	KLK3-SERPINA3	BDNF	EGFR	AKR7A2	0.851
100	KIT	C9	BMPER	NME2	0.851

TABLE 24

Panels of 5 Biomarkers						
Markers					Mean CV AUC	
1	KLK3-SERPINA3	CNTN1	C9	BMPER	AKR7A2	0.866
2	BDNF	KIT	C9	AKR7A2	CRP	0.866
3	KLK3-SERPINA3	CNTN1	BMP1	BMPER	NME2	0.865
4	KLK3-SERPINA3	IGFBP2	CNTN1	AKR7A2	CRP	0.865
5	KLK3-SERPINA3	IGFBP2	CNTN1	BMPER	AKR7A2	0.865
6	BDNF	KIT	IGFBP2	AKR7A2	CRP	0.865
7	KLK3-SERPINA3	BDNF	KIT	AKR7A2	CRP	0.865
8	KLK3-SERPINA3	IGFBP2	CNTN1	BMPER	NME2	0.865
9	KLK3-SERPINA3	CNTN1	BMP1	NME2	CRP	0.865
10	KLK3-SERPINA3	KIT	CNTN1	BMP1	AKR7A2	0.864
11	KLK3-SERPINA3	KIT	C9	BMPER	AKR7A2	0.864
12	KLK3-SERPINA3	KIT	BMP1	AKR7A2	CRP	0.864
13	BDNF	KIT	BMP1	AKR7A2	CRP	0.864
14	KLK3-SERPINA3	KIT	CNTN1	BMP1	NME2	0.864
15	KLK3-SERPINA3	KIT	C9	BMPER	NME2	0.864
16	GHR	C9	BMPER	AKR7A2	CRP	0.864
17	KLK3-SERPINA3	EGFR	NME2	CRP	ITIH4	0.864
18	KLK3-SERPINA3	KIT	BMP1	BMPER	NME2	0.864
19	KLK3-SERPINA3	KIT	CNTN1	C9	AKR7A2	0.864
20	KLK3-SERPINA3	BDNF	KIT	C9	AKR7A2	0.864
21	KLK3-SERPINA3	IGFBP4	C9	BMPER	AKR7A2	0.863
22	KIT	GHR	C9	BMPER	AKR7A2	0.863
23	KLK3-SERPINA3	CNTN1	BMPER	AKR7A2	CRP	0.863
24	KLK3-SERPINA3	BDNF	KIT	CNTN1	AKR7A2	0.863
25	KLK3-SERPINA3	KIT	IGFBP4	C9	AKR7A2	0.863
26	KLK3-SERPINA3	CNTN1	BMP1	AKR7A2	CRP	0.863
27	KLK3-SERPINA3	C9	BMPER	AKR7A2	ITIH4	0.863
28	KIT	BMP1	BMPER	AKR7A2	CRP	0.863
29	KIT	CNTN1	C9	BMP1	AKR7A2	0.863
30	KLK3-SERPINA3	KIT	CNTN1	BMPER	NME2	0.863
31	KLK3-SERPINA3	IGFBP2	BMPER	AKR7A2	CRP	0.863
32	KLK3-SERPINA3	CNTN1	C9	BMPER	NME2	0.863
33	KIT	C9	BMPER	AKR7A2	CRP	0.863
34	KLK3-SERPINA3	CNTN1	BMP1	BMPER	AKR7A2	0.863
35	KLK3-SERPINA3	IGFBP4	CNTN1	C9	AKR7A2	0.862
36	KIT	GHR	BMPER	AKR7A2	CRP	0.862
37	GHR	CNTN1	C9	BMPER	AKR7A2	0.862
38	KLK3-SERPINA3	CNTN1	BMPER	NME2	CRP	0.862
39	KLK3-SERPINA3	GHR	BMPER	AKR7A2	CRP	0.862
40	BDNF	KIT	CNTN1	C9	AKR7A2	0.862
41	KLK3-SERPINA3	C9	BMPER	AKR7A2	CRP	0.862
42	KLK3-SERPINA3	GHR	C9	BMPER	AKR7A2	0.862
43	KLK3-SERPINA3	IGFBP4	C9	AKR7A2	ITIH4	0.862
44	KLK3-SERPINA3	CNTN1	C9	BMP1	AKR7A2	0.862

TABLE 24-continued

Panels of 5 Biomarkers						
Markers					Mean CV AUC	
45	KLK3-SERPINA3	KIT	CNTN1	C9	NME2	0.862
46	IGFBP2	CNTN1	C9	BMPER	AKR7A2	0.862
47	IGFBP2	CNTN1	BMPER	AKR7A2	CRP	0.862
48	KLK3-SERPINA3	KIT	IGFBP2	AKR7A2	CRP	0.862
49	KLK3-SERPINA3	IGFBP4	BMP1	NME2	CRP	0.862
50	KLK3-SERPINA3	IGFBP4	BMP1	AKR7A2	CRP	0.862
51	KIT	GHR	BMP1	AKR7A2	CRP	0.862
52	KIT	IGFBP2	C9	AKR7A2	CRP	0.862
53	KLK3-SERPINA3	BDNF	CNTN1	C9	AKR7A2	0.862
54	KLK3-SERPINA3	IGFBP2	BMPER	NME2	CRP	0.862
55	KLK3-SERPINA3	EGFR	AKR7A2	CRP	ITIH4	0.862
56	KLK3-SERPINA3	EGFR	CNTN1	C9	AKR7A2	0.862
57	KLK3-SERPINA3	KIT	BMP1	NME2	CRP	0.861
58	KLK3-SERPINA3	IGFBP4	BMPER	AKR7A2	CRP	0.861
59	KLK3-SERPINA3	KIT	C9	AKR7A2	CRP	0.861
60	KLK3-SERPINA3	KIT	EGFR	AKR7A2	CRP	0.861
61	KLK3-SERPINA3	IGFBP4	C9	BMPER	NME2	0.861
62	KLK3-SERPINA3	KIT	C9	BMP1	AKR7A2	0.861
63	KIT	GHR	C9	AKR7A2	CRP	0.861
64	KLK3-SERPINA3	C9	DDC	BMPER	AKR7A2	0.861
65	KLK3-SERPINA3	IGFBP2	CNTN1	NME2	CRP	0.861
66	KIT	CNTN1	C9	BMPER	AKR7A2	0.861
67	KLK3-SERPINA3	KIT	EGFR	C9	AKR7A2	0.861
68	KLK3-SERPINA3	CNTN1	BMPER	AKR7A2	ITIH4	0.861
69	KLK3-SERPINA3	EGFR	C9	BMPER	AKR7A2	0.861
70	CNTN1	C9	BMPER	AKR7A2	CRP	0.861
71	KIT	GHR	BMPER	NME2	CRP	0.861
72	IGFBP2	C9	BMPER	AKR7A2	CRP	0.861
73	KLK3-SERPINA3	GHR	BMPER	NME2	CRP	0.861
74	KLK3-SERPINA3	IGFBP2	CNTN1	C9	AKR7A2	0.861
75	BDNF	IGFBP2	CNTN1	AKR7A2	CRP	0.861
76	IGFBP2	CNTN1	BMP1	AKR7A2	CRP	0.861
77	BDNF	KIT	C9	BMPER	AKR7A2	0.861
78	KLK3-SERPINA3	BDNF	C9	AKR7A2	CRP	0.861
79	KIT	IGFBP2	BMP1	AKR7A2	CRP	0.861
80	KLK3-SERPINA3	BMP1	BMPER	NME2	CRP	0.861
81	KLK3-SERPINA3	BDNF	IGFBP2	AKR7A2	CRP	0.861
82	KLK3-SERPINA3	IGFBP2	BMP1	AKR7A2	CRP	0.861
83	BDNF	KIT	GHR	AKR7A2	CRP	0.861
84	KLK3-SERPINA3	IGFBP4	BMPER	NME2	ITIH4	0.861
85	KLK3-SERPINA3	KIT	BMPER	NME2	CRP	0.861
86	KLK3-SERPINA3	IGFBP2	AKR7A2	CRP	ITIH4	0.861
87	KLK3-SERPINA3	KIT	BMPER	AKR7A2	CRP	0.861
88	BDNF	KIT	C9	AHSG	AKR7A2	0.860
89	IGFBP2	BMPER	NME2	CRP	ITIH4	0.860
90	KIT	IGFBP2	BMPER	AKR7A2	CRP	0.860
91	KLK3-SERPINA3	IGFBP4	BMPER	NME2	CRP	0.860
92	KLK3-SERPINA3	KIT	IGFBP4	AKR7A2	CRP	0.860
93	KLK3-SERPINA3	IGFBP2	EGFR	AKR7A2	CRP	0.860
94	KLK3-SERPINA3	IGFBP4	CNTN1	C9	NME2	0.860
95	KLK3-SERPINA3	GHR	CNTN1	BMPER	NME2	0.860
96	KLK3-SERPINA3	IGFBP4	C9	AKR7A2	CRP	0.860
97	KLK3-SERPINA3	KIT	CNTN1	BMPER	AKR7A2	0.860
98	KIT	C9	BMP1	AKR7A2	CRP	0.860
99	KLK3-SERPINA3	IGFBP2	C9	BMPER	AKR7A2	0.860
100	KLK3-SERPINA3	EGFR	C9	AKR7A2	CRP	0.860

TABLE 25

Panels of 6 Biomarkers						
Markers					Mean CV AUC	
1	KLK3-SERPINA3	KIT	CNTN1	C9	BMPER	0.871
	AKR7A2					
2	KIT	GHR	C9	BMPER	AKR7A2	0.871
	CRP					
3	KLK3-SERPINA3	BDNF	KIT	CNTN1	C9	0.871
	AKR7A2					

TABLE 25-continued

Panels of 6 Biomarkers						
Markers						Mean CV AUC
4	KLK3-SERPINA3 AKR7A2	KIT	CNTN1	C9	BMP1	0.871
5	KLK3-SERPINA3 CRP	IGFBP2	CNTN1	BMPER	AKR7A2	0.871
6	KLK3-SERPINA3 AKR7A2	IGFBP4	CNTN1	C9	BMPER	0.871
7	KLK3-SERPINA3 NME2	KIT	CNTN1	C9	BMPER	0.870
8	KLK3-SERPINA3 AKR7A2	KIT	GHR	C9	BMPER	0.870
9	KLK3-SERPINA3 AKR7A2	IGFBP2	CNTN1	C9	BMPER	0.870
10	KLK3-SERPINA3 ITIH4	IGFBP2	CNTN1	BMPER	AKR7A2	0.870
11	KLK3-SERPINA3 AKR7A2	KIT	IGFBP4	C9	BMPER	0.870
12	KLK3-SERPINA3 NME2	KIT	CNTN1	BMP1	BMPER	0.870
13	BDNF CRP	KIT	IGFBP2	C9	AKR7A2	0.869
14	BDNF CRP	KIT	GHR	C9	AKR7A2	0.869
15	KLK3-SERPINA3 CRP	IGFBP2	CNTN1	BMP1	AKR7A2	0.869
16	KLK3-SERPINA3 CRP	KIT	CNTN1	BMP1	NME2	0.869
17	KLK3-SERPINA3 AKR7A2	KIT	CNTN1	BMP1	BMPER	0.869
18	BDNF CRP	KIT	IGFBP2	CNTN1	AKR7A2	0.869
19	KIT CRP	GHR	BMP1	BMPER	AKR7A2	0.869
20	KLK3-SERPINA3 CRP	IGFBP2	CNTN1	BMPER	NME2	0.869
21	KLK3-SERPINA3 AKR7A2	CNTN1	C9	BMP1	BMPER	0.868
22	KLK3-SERPINA3 ITIH4	IGFBP4	C9	BMPER	AKR7A2	0.868
23	KLK3-SERPINA3 AKR7A2	KIT	IGFBP2	CNTN1	BMPER	0.868
24	GHR CRP	CNTN1	C9	BMPER	AKR7A2	0.868
25	KIT AKR7A2	GHR	CNTN1	C9	BMPER	0.868
26	KLK3-SERPINA3 AKR7A2	GHR	CNTN1	C9	BMPER	0.868
27	KLK3-SERPINA3 ITIH4	KIT	IGFBP4	C9	AKR7A2	0.868
28	KLK3-SERPINA3 ITIH4	IGFBP4	CNTN1	BMPER	AKR7A2	0.868
29	BDNF AKR7A2	KIT	IGFBP2	CNTN1	C9	0.867
30	KIT AKR7A2	CNTN1	C9	BMP1	BMPER	0.867
31	KLK3-SERPINA3 CRP	BDNF	KIT	IGFBP2	AKR7A2	0.867
32	KLK3-SERPINA3 ITIH4	CNTN1	BMP1	BMPER	AKR7A2	0.867
33	KLK3-SERPINA3 CRP	KIT	GHR	BMPER	NME2	0.867
34	KLK3-SERPINA3 CRP	BDNF	KIT	C9	AKR7A2	0.867
35	KLK3-SERPINA3 NME2	KIT	CNTN1	C9	BMP1	0.867
36	KLK3-SERPINA3 CRP	BDNF	KIT	CNTN1	AKR7A2	0.867
37	KLK3-SERPINA3 ITIH4	IGFBP4	CNTN1	C9	AKR7A2	0.867
38	KLK3-SERPINA3 ITIH4	CNTN1	C9	BMPER	AKR7A2	0.867
39	KLK3-SERPINA3 NME2	BDNF	KIT	CNTN1	C9	0.867

TABLE 25-continued

Panels of 6 Biomarkers						
Markers						Mean CV AUC
40	KLK3-SERPINA3 CRP	KIT	CNTN1	BMP1	AKR7A2	0.867
41	KLK3-SERPINA3 CRP	BDNF	IGFBP2	CNTN1	AKR7A2	0.867
42	KLK3-SERPINA3 CRP	KIT	IGFBP2	CNTN1	AKR7A2	0.867
43	KLK3-SERPINA3 ITIH4	IGFBP4	CNTN1	BMPER	NME2	0.867
44	KLK3-SERPINA3 CRP	GHR	CNTN1	BMPER	AKR7A2	0.867
45	KLK3-SERPINA3 AKR7A2	EGFR	CNTN1	C9	BMPER	0.867
46	KLK3-SERPINA3 CRP	KIT	IGFBP2	BMPER	AKR7A2	0.867
47	KIT CRP	IGFBP2	C9	BMPER	AKR7A2	0.867
48	KLK3-SERPINA3 AKR7A2	KIT	EGFR	C9	BMPER	0.867
49	KLK3-SERPINA3 NME2	KIT	IGFBP2	CNTN1	BMPER	0.867
50	KLK3-SERPINA3 AKR7A2	KIT	C9	BMP1	BMPER	0.867
51	KLK3-SERPINA3 CRP	KIT	GHR	BMPER	AKR7A2	0.867
52	KLK3-SERPINA3 CRP	CNTN1	BMP1	BMPER	AKR7A2	0.867
53	BDNF CRP	KIT	C9	BMPER	AKR7A2	0.867
54	KIT CRP	GHR	BMP1	BMPER	NME2	0.867
55	KLK3-SERPINA3 ITIH4	IGFBP2	BMPER	AKR7A2	CRP	0.867
56	KIT AKR7A2	IGFBP2	CNTN1	C9	BMPER	0.867
57	KLK3-SERPINA3 ITIH4	IGFBP4	CNTN1	BMP1	AKR7A2	0.867
58	KLK3-SERPINA3 AKR7A2	KIT	IGFBP4	CNTN1	C9	0.867
59	KLK3-SERPINA3 ITIH4	IGFBP2	CNTN1	BMPER	NME2	0.867
60	IGFBP2 CRP	CNTN1	C9	BMPER	AKR7A2	0.866
61	KLK3-SERPINA3 AKR7A2	KIT	EGFR	CNTN1	C9	0.866
62	KLK3-SERPINA3 CRP	KIT	IGFBP2	BMP1	AKR7A2	0.866
63	BDNF CRP	KIT	CNTN1	C9	AKR7A2	0.866
64	KLK3-SERPINA3 CRP	KIT	C9	BMPER	AKR7A2	0.866
65	KLK3-SERPINA3 ITIH4	IGFBP4	CNTN1	BMP1	NME2	0.866
66	KLK3-SERPINA3 ITIH4	IGFBP4	BMP1	AKR7A2	CRP	0.866
67	KLK3-SERPINA3 CRP	KIT	IGFBP4	BMP1	AKR7A2	0.866
68	KLK3-SERPINA3 CRP	GHR	C9	BMPER	AKR7A2	0.866
69	KLK3-SERPINA3 CRP	KIT	BMP1	BMPER	AKR7A2	0.866
70	KLK3-SERPINA3 AKR7A2	IGFBP2	CNTN1	BMP1	BMPER	0.866
71	KLK3-SERPINA3 AKR7A2	IGFBP2	CNTN1	DDC	BMPER	0.866
72	KLK3-SERPINA3 AKR7A2	KIT	IGFBP2	CNTN1	C9	0.866
73	KLK3-SERPINA3 AKR7A2	KIT	C9	DDC	BMPER	0.866
74	KLK3-SERPINA3 CRP	KIT	CNTN1	BMPER	AKR7A2	0.866
75	KLK3-SERPINA3 CRP	IGFBP4	CNTN1	BMP1	AKR7A2	0.866

TABLE 25-continued

Panels of 6 Biomarkers						
Markers						Mean CV AUC
76	KLK3-SERPINA3 AKR7A2	KIT	IGFBP4	CNTN1	BMP1	0.866
77	KLK3-SERPINA3 CRP	BDNF	KIT	CNTN1	NME2	0.866
78	KLK3-SERPINA3 ITIH4	CNTN1	BMP1	AKR7A2	CRP	0.866
79	KIT CRP	IGFBP2	CNTN1	BMP1	AKR7A2	0.866
80	KLK3-SERPINA3 CRP	IGFBP2	CNTN1	BMP1	NME2	0.866
81	KLK3-SERPINA3 CRP	CNTN1	C9	BMPER	AKR7A2	0.866
82	KLK3-SERPINA3 AKR7A2	BDNF	KIT	CNTN1	BMP1	0.866
83	KLK3-SERPINA3 CRP	IGFBP4	CNTN1	BMP1	NME2	0.866
84	KLK3-SERPINA3 ITIH4	IGFBP4	BMPER	AKR7A2	CRP	0.866
85	KLK3-SERPINA3 CRP	IGFBP2	EGFR	CNTN1	AKR7A2	0.866
86	BDNF CRP	KIT	C9	AHSG	AKR7A2	0.866
87	KLK3-SERPINA3 AKR7A2	KIT	IGFBP2	CNTN1	BMP1	0.866
88	KLK3-SERPINA3 CRP	KIT	BMP1	BMPER	NME2	0.866
89	KLK3-SERPINA3 CRP	BDNF	CNTN1	C9	AKR7A2	0.866
90	KLK3-SERPINA3 ITIH4	KIT	CNTN1	BMPER	NME2	0.866
91	KLK3-SERPINA3 ITIH4	IGFBP4	BMPER	NME2	CRP	0.866
92	KIT CRP	IGFBP2	CNTN1	BMPER	AKR7A2	0.866
93	KLK3-SERPINA3 NME2	IGFBP4	CNTN1	C9	BMPER	0.866
94	KLK3-SERPINA3 AKR7A2	IGFBP4	CNTN1	C9	BMP1	0.866
95	KLK3-SERPINA3 CRP	KIT	CNTN1	BMPER	NME2	0.866
96	KLK3-SERPINA3 ITIH4	IGFBP2	CNTN1	AKR7A2	CRP	0.866
97	KLK3-SERPINA3 AKR7A2	BDNF	KIT	C9	BMPER	0.866
98	KLK3-SERPINA3 CRP	GHR	IGFBP4	BMPER	AKR7A2	0.866
99	BDNF CRP	KIT	IGFBP2	AHSG	AKR7A2	0.866
100	KLK3-SERPINA3 ITIH4	KIT	C9	BMPER	AKR7A2	0.866

TABLE 26

Panels of 7 Biomarkers						
Markers						Mean CV AUC
1	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR	CNTN1	C9	0.875
2	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2	CNTN1	C9	0.875
3	KLK3-SERPINA3 BMPER	KIT AKR7A2	CNTN1	C9	BMP1	0.875
4	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP2	CNTN1	BMPER	0.874
5	KLK3-SERPINA3 AKR7A2	KIT ITIH4	CNTN1	BMP1	BMPER	0.873
6	KLK3-SERPINA3 AKR7A2	IGFBP4 ITIH4	CNTN1	C9	BMPER	0.873

TABLE 26-continued

Panels of 7 Biomarkers						
Markers			Mean CV AUC			
7	KLK3-SERPINA3 AKR7A2	BDNF CRP	KIT	IGFBP2	CNTN1	0.873
8	KIT AKR7A2	GHR CRP	CNTN1	C9	BMPER	0.873
9	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4	CNTN1	C9	0.873
10	KLK3-SERPINA3 AKR7A2	KIT ITIH4	IGFBP4	C9	BMPER	0.873
11	KLK3-SERPINA3 BMPER	BDNF AKR7A2	KIT	CNTN1	C9	0.872
12	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP2	CNTN1	BMP1	0.872
13	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR	CNTN1	BMPER	0.872
14	KLK3-SERPINA3 AKR7A2	IGFBP4 ITIH4	CNTN1	BMP1	BMPER	0.872
15	KIT AKR7A2	GHR CRP	CNTN1	BMP1	BMPER	0.872
16	KLK3-SERPINA3 CRP	IGFBP2 ITIH4	CNTN1	BMPER	AKR7A2	0.872
17	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR	C9	BMPER	0.872
18	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR	IGFBP4	C9	0.872
19	KLK3-SERPINA3 AKR7A2	KIT CRP	CNTN1	BMP1	BMPER	0.872
20	KLK3-SERPINA3 BMPER	KIT NME2	GHR	CNTN1	C9	0.872
21	KLK3-SERPINA3 AKR7A2	BDNF CRP	KIT	CNTN1	C9	0.872
22	KLK3-SERPINA3 NME2	KIT ITIH4	IGFBP4	CNTN1	BMP1	0.872
23	KLK3-SERPINA3 AKR7A2	KIT ITIH4	IGFBP2	CNTN1	BMPER	0.872
24	KLK3-SERPINA3 AKR7A2	IGFBP2 ITIH4	IGFBP4	CNTN1	BMPER	0.872
25	KLK3-SERPINA3 CRP	IGFBP4 ITIH4	CNTN1	BMP1	AKR7A2	0.872
26	KIT AKR7A2	GHR CRP	C9	BMP1	BMPER	0.872
27	KLK3-SERPINA3 BMPER	IGFBP4 AKR7A2	CNTN1	C9	BMP1	0.872
28	KLK3-SERPINA3 AKR7A2	IGFBP2 CRP	GHR	CNTN1	BMPER	0.872
29	KIT BMPER	GHR AKR7A2	CNTN1	C9	BMP1	0.872
30	KLK3-SERPINA3 BMPER	KIT AKR7A2	CNTN1	C9	AHSG	0.872
31	KLK3-SERPINA3 AKR7A2	IGFBP2 ITIH4	CNTN1	DDC	BMPER	0.872
32	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR	IGFBP4	BMPER	0.872
33	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR	CNTN1	BMP1	0.872
34	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2	CNTN1	BMP1	0.872
35	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR	BMP1	BMPER	0.871
36	KLK3-SERPINA3 BMPER	GHR AKR7A2	IGFBP4	CNTN1	C9	0.871
37	BDNF AKR7A2	KIT CRP	IGFBP2	CNTN1	C9	0.871
38	BDNF AKR7A2	KIT CRP	GHR	C9	BMPER	0.871
39	KIT AKR7A2	IGFBP2 CRP	CNTN1	C9	BMPER	0.871
40	KLK3-SERPINA3 NME2	KIT CRP	IGFBP2	CNTN1	BMPER	0.871
41	KLK3-SERPINA3 BMP1	KIT AKR7A2	IGFBP4	CNTN1	C9	0.871
42	KIT AKR7A2	GHR CRP	C9	AHSG	BMPER	0.871

TABLE 26-continued

Panels of 7 Biomarkers						
Markers			Mean CV AUC			
43	KLK3-SERPINA3 AKR7A2	IGFBP2 ITIH4	CNTN1	C9	BMPER	0.871
44	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR	CNTN1	BMP1	0.871
45	KLK3-SERPINA3 AKR7A2	KIT ITIH4	IGFBP4	CNTN1	BMP1	0.871
46	KLK3-SERPINA3 AKR7A2	IGFBP2 CRP	IGFBP4	CNTN1	BMPER	0.871
47	KLK3-SERPINA3 AKR7A2	IGFBP2 CRP	CNTN1	C9	BMPER	0.871
48	KLK3-SERPINA3 CRP	GHR ITIH4	IGFBP4	BMPER	AKR7A2	0.871
49	KLK3-SERPINA3 AKR7A2	KIT ITIH4	GHR	CNTN1	BMPER	0.871
50	KLK3-SERPINA3 CRP	IGFBP2 ITIH4	CNTN1	BMP1	AKR7A2	0.871
51	KLK3-SERPINA3 CRP	IGFBP2 ITIH4	IGFBP4	BMPER	AKR7A2	0.871
52	KLK3-SERPINA3 AKR7A2	KIT ITIH4	IGFBP4	CNTN1	BMPER	0.871
53	KLK3-SERPINA3 AKR7A2	KIT ITIH4	CNTN1	C9	BMPER	0.871
54	KLK3-SERPINA3 AKR7A2	IGFBP2 CRP	CNTN1	BMP1	BMPER	0.871
55	KLK3-SERPINA3 NME2	KIT CRP	IGFBP2	CNTN1	BMP1	0.871
56	KLK3-SERPINA3 AKR7A2	KIT ITIH4	IGFBP4	CNTN1	C9	0.871
57	KLK3-SERPINA3 BMPER	KIT AKR7A2	EGFR	CNTN1	C9	0.870
58	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4	C9	BMP1	0.870
59	KLK3-SERPINA3 AKR7A2	IGFBP2 ITIH4	CNTN1	BMP1	BMPER	0.870
60	KLK3-SERPINA3 AKR7A2	GHR ITIH4	IGFBP4	CNTN1	BMPER	0.870
61	KLK3-SERPINA3 BMP1	BDNF AKR7A2	KIT	CNTN1	C9	0.870
62	KIT AKR7A2	IGFBP2 CRP	CNTN1	BMP1	BMPER	0.870
63	KLK3-SERPINA3 AKR7A2	BDNF CRP	KIT	CNTN1	BMP1	0.870
64	KLK3-SERPINA3 AKR7A2	GHR CRP	CNTN1	C9	BMPER	0.870
65	KLK3-SERPINA3 NME2	IGFBP4 ITIH4	CNTN1	BMP1	BMPER	0.870
66	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP2	C9	BMPER	0.870
67	KLK3-SERPINA3 C9	BDNF AKR7A2	KIT	IGFBP2	CNTN1	0.870
68	KLK3-SERPINA3 CRP	KIT ITIH4	IGFBP4	BMP1	AKR7A2	0.870
69	KIT AKR7A2	GHR CRP	CNTN1	C9	BMP1	0.870
70	KLK3-SERPINA3 DDC	KIT AKR7A2	CNTN1	C9	BMP1	0.870
71	KLK3-SERPINA3 AKR7A2	KIT CRP	CNTN1	C9	BMPER	0.870
72	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR	IGFBP4	C9	0.870
73	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR	IGFBP4	BMP1	0.870
74	KLK3-SERPINA3 AKR7A2	BDNF CRP	KIT	C9	BMPER	0.870
75	KLK3-SERPINA3 NME2	KIT CRP	CNTN1	BMP1	BMPER	0.870
76	KLK3-SERPINA3 CRP	IGFBP2 ITIH4	IGFBP4	CNTN1	AKR7A2	0.870
77	KLK3-SERPINA3 AKR7A2	GHR CRP	CNTN1	BMP1	BMPER	0.870
78	KLK3-SERPINA3 BMPER	KIT NME2	CNTN1	C9	BMP1	0.870

TABLE 26-continued

Panels of 7 Biomarkers							
		Markers					Mean CV AUC
79	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP2	GHR	BMPER		0.870
80	KLK3-SERPINA3 AKR7A2	KIT CRP	EGFR	CNTN1	BMP1		0.870
81	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP4	CNTN1	BMP1		0.870
82	KIT AKR7A2	EGFR CRP	GHR	C9	BMPER		0.870
83	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR	C9	AHSG		0.870
84	KLK3-SERPINA3 AKR7A2	BDNF CRP	IGFBP2	CNTN1	C9		0.870
85	KIT AKR7A2	IGFBP2 CRP	GHR	C9	BMPER		0.870
86	KLK3-SERPINA3 CRP	KIT ITIH4	IGFBP4	BMPER	AKR7A2		0.870
87	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR	C9	BMP1		0.870
88	KLK3-SERPINA3 AKR7A2	KIT CRP	CNTN1	C9	BMP1		0.870
89	KLK3-SERPINA3 BMP1	KIT AKR7A2	EGFR	CNTN1	C9		0.870
90	KLK3-SERPINA3 AKR7A2	KIT ITIH4	EGFR	C9	BMPER		0.870
91	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2	CNTN1	DDC		0.870
92	BDNF AKR7A2	KIT CRP	IGFBP2	C9	BMPER		0.870
93	KLK3-SERPINA3 BMPER	IGFBP2 AKR7A2	IGFBP4	CNTN1	C9		0.870
94	KLK3-SERPINA3 BMPER	KIT AKR7A2	CNTN1	C9	CA6		0.870
95	KLK3-SERPINA3 CRP	KIT ITIH4	GHR	IGFBP4	AKR7A2		0.870
96	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR	CNTN1	C9		0.870
97	KLK3-SERPINA3 BMPER	IGFBP2 AKR7A2	CNTN1	C9	DDC		0.870
98	KLK3-SERPINA3 BMPER	KIT AKR7A2	CNTN1	C9	DDC		0.870
99	KLK3-SERPINA3 NME2	IGFBP4 ITIH4	CNTN1	C9	BMPER		0.870
100	KIT AKR7A2	CNTN1 CRP	C9	BMP1	BMPER		0.870

TABLE 27

Panels of 8 Biomarkers							
		Markers				Mean CV AUC	
1	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	CNTN1	C9		0.877
2	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 ITIH4	CNTN1	BMP1		0.876
3	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR ITIH4	IGFBP4	BMPER		0.876
4	KLK3-SERPINA3 C9	KIT BMPER	GHR AKR7A2	IGFBP4	CNTN1		0.876
5	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	CNTN1	C9		0.876
6	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	CNTN1	BMP1		0.876
7	KLK3-SERPINA3 AKR7A2	IGFBP2 CRP	IGFBP4 ITIH4	CNTN1	BMPER		0.876
8	KIT BMPER	GHR AKR7A2	CNTN1 CRP	C9	BMP1		0.876
9	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	GHR	CNTN1		0.876

TABLE 27-continued

Panels of 8 Biomarkers						
Markers			Mean CV AUC			
10	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	CNTN1	BMP1	0.876
11	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	CNTN1	C9	0.875
12	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 ITIH4	CNTN1	C9	0.875
13	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	CNTN1	C9	0.875
14	KLK3-SERPINA3 AHSB	KIT BMPER	IGFBP2 AKR7A2	CNTN1	C9	0.875
15	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	CNTN1	C9	0.875
16	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR ITIH4	CNTN1	BMP1	0.875
17	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP4 ITIH4	CNTN1	BMP1	0.875
18	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR ITIH4	IGFBP4	C9	0.875
19	KLK3-SERPINA3 BMP1	KIT AKR7A2	GHR CRP	IGFBP4	CNTN1	0.875
20	KLK3-SERPINA3 AHSB	KIT BMPER	CNTN1 AKR7A2	C9	BMP1	0.874
21	KLK3-SERPINA3 C9	KIT BMPER	IGFBP2 AKR7A2	GHR	CNTN1	0.874
22	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4	C9	0.874
23	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	CNTN1	C9	0.874
24	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	CNTN1	C9	0.874
25	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP4 ITIH4	BMP1	BMPER	0.874
26	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 ITIH4	CNTN1	BMP1	0.874
27	KLK3-SERPINA3 BMPER	KIT NME2	IGFBP4 ITIH4	CNTN1	BMP1	0.874
28	KLK3-SERPINA3 BMPER	KIT NME2	GHR CRP	CNTN1	BMP1	0.874
29	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP2 ITIH4	CNTN1	BMPER	0.874
30	KLK3-SERPINA3 BMP1	KIT AKR7A2	GHR CRP	CNTN1	C9	0.874
31	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR ITIH4	CNTN1	BMPER	0.874
32	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4	CNTN1	0.874
33	KLK3-SERPINA3 AKR7A2	GHR CRP	IGFBP4 ITIH4	CNTN1	BMP1	0.874
34	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 ITIH4	CNTN1	C9	0.874
35	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR ITIH4	IGFBP4	CNTN1	0.874
36	KLK3-SERPINA3 BMPER	IGFBP2 AKR7A2	IGFBP4 ITIH4	CNTN1	C9	0.874
37	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 ITIH4	C9	BMP1	0.874
38	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR ITIH4	CNTN1	C9	0.874
39	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1	C9	0.874
40	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4	BMP1	0.874
41	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	C9	BMPER	0.873
42	KLK3-SERPINA3 BMPER	GHR AKR7A2	IGFBP4 ITIH4	CNTN1	C9	0.873
43	KLK3-SERPINA3 BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1	C9	0.873
44	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 CRP	CNTN1	BMP1	0.873
45	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR ITIH4	BMP1	BMPER	0.873

TABLE 27-continued

Panels of 8 Biomarkers						
Markers						Mean CV AUC
46	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 ITIH4	C9	AHSG	0.873
47	KLK3-SERPINA3 BMP1	KIT AKR7A2	IGFBP4 ITIH4	CNTN1	C9	0.873
48	KLK3-SERPINA3 BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1	BMP1	0.873
49	BDNF BMPER	KIT AKR7A2	GHR CRP	CNTN1	C9	0.873
50	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 ITIH4	C9	DDC	0.873
51	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR ITIH4	IGFBP4	BMP1	0.873
52	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 ITIH4	CNTN1	DDC	0.873
53	KLK3-SERPINA3 BMPER	KIT AKR7A2	CNTN1 ITIH4	BMP1	DDC	0.873
54	KLK3-SERPINA3 BMPER	BDNF AKR7A2	KIT CRP	CNTN1	C9	0.873
55	KLK3-SERPINA3 BMPER	BDNF AKR7A2	KIT CRP	IGFBP2	CNTN1	0.873
56	KIT BMPER	GHR AKR7A2	CNTN1 CRP	C9	FN1	0.873
57	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP2 ITIH4	IGFBP4	BMPER	0.873
58	KLK3-SERPINA3 AKR7A2	GHR CRP	IGFBP4 ITIH4	CNTN1	BMPER	0.873
59	KLK3-SERPINA3 BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	C9	AHSG	0.873
60	KLK3-SERPINA3 NME2	KIT CRP	IGFBP4 ITIH4	CNTN1	BMP1	0.873
61	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	C9	AHSG	0.873
62	KLK3-SERPINA3 BMP1	BDNF AKR7A2	KIT CRP	IGFBP2	CNTN1	0.873
63	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP2 AKR7A2	CNTN1	C9	0.873
64	KLK3-SERPINA3 AKR7A2	IGFBP2 CRP	CNTN1 ITIH4	BMP1	BMPER	0.873
65	KLK3-SERPINA3 DDC	KIT BMPER	GHR AKR7A2	CNTN1	C9	0.873
66	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	BMP1	BMPER	0.873
67	KLK3-SERPINA3 AKR7A2	BDNF CRP	KIT ITIH4	IGFBP2	CNTN1	0.873
68	KLK3-SERPINA3 C9	KIT BMPER	IGFBP2 AKR7A2	IGFBP4	CNTN1	0.873
69	KLK3-SERPINA3 C9	BDNF AKR7A2	KIT CRP	IGFBP2	CNTN1	0.873
70	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 ITIH4	IGFBP4	CNTN1	0.873
71	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP4 AKR7A2	CNTN1	C9	0.873
72	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	C9	BMP1	0.873
73	KLK3-SERPINA3 AKR7A2	IGFBP2 CRP	IGFBP4 ITIH4	CNTN1	BMP1	0.873
74	KLK3-SERPINA3 BMP1	KIT NME2	GHR CRP	IGFBP4	CNTN1	0.873
75	KLK3-SERPINA3 NME2	KIT CRP	GHR ITIH4	IGFBP4	BMPER	0.873
76	KLK3-SERPINA3 DDC	KIT AKR7A2	IGFBP2 CRP	CNTN1	BMP1	0.873
77	KLK3-SERPINA3 AKR7A2	KIT CRP	CNTN1 ITIH4	BMP1	BMPER	0.873
78	BDNF BMPER	KIT AKR7A2	IGFBP2 CRP	CNTN1	C9	0.873
79	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	CNTN1	CA6	0.873
80	KLK3-SERPINA3 BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	C9	BMP1	0.873
81	KLK3-SERPINA3 AKR7A2	KIT CRP	EGFR ITIH4	CNTN1	BMP1	0.873

TABLE 27-continued

Panels of 8 Biomarkers						
Markers			Mean CV AUC			
82	KLK3-SERPINA3 BMP1	KIT AKR7A2	IGFBP2 CRP	CNTN1	C9	0.872
83	KLK3-SERPINA3 AKR7A2	KIT CRP	IGFBP2 ITIH4	CNTN1	BMP1	0.872
84	KLK3-SERPINA3 BMPER	KIT AKR7A2	CNTN1 ITIH4	C9	BMP1	0.872
85	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR ITIH4	IGFBP4	CNTN1	0.872
86	KLK3-SERPINA3 BMPER	KIT AKR7A2	CNTN1 CRP	C9	BMP1	0.872
87	KLK3-SERPINA3 AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	BMP1	BMPER	0.872
88	KLK3-SERPINA3 C9	KIT BMPER	IGFBP2 AKR7A2	EGFR	CNTN1	0.872
89	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	CNTN1	AHSG	0.872
90	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	GHR	C9	0.872
91	KLK3-SERPINA3 AKR7A2	KIT CRP	GHR ITIH4	IGFBP4	C9	0.872
92	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	IGFBP4	C9	0.872
93	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	IGFBP4	C9	0.872
94	KLK3-SERPINA3 C9	BDNF BMPER	KIT AKR7A2	IGFBP2	CNTN1	0.872
95	KIT BMPER	GHR AKR7A2	CNTN1 CRP	C9	AHSG	0.872
96	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	CNTN1	AHSG	0.872
97	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1	BMP1	0.872
98	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR ITIH4	CNTN1	CA6	0.872
99	KLK3-SERPINA3 FN1	KIT BMPER	GHR AKR7A2	CNTN1	C9	0.872
100	KIT AHSG	GHR BMPER	CNTN1 AKR7A2	C9	BMP1	0.872

TABLE 28

Panels of 9 Biomarkers						
Markers			Mean CV AUC			
1	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	BMP1	0.878
2	KLK3-SERPINA3 BMP1	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	CNTN1	0.878
3	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	CNTN1	0.878
4	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	IGFBP4 ITIH4	CNTN1	0.878
5	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	CNTN1 CRP	C9	0.877
6	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	C9	0.877
7	KLK3-SERPINA3 BMP1	KIT AHSG	GHR BMPER	CNTN1 AKR7A2	C9	0.877
8	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1	0.877
9	KLK3-SERPINA3 C9	KIT BMP1	GHR BMPER	IGFBP4 AKR7A2	CNTN1	0.877
10	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	C9	0.877
11	KLK3-SERPINA3 C9	KIT BMPER	GHR AKR7A2	IGFBP4 ITIH4	CNTN1	0.877
12	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	IGFBP4 ITIH4	CNTN1	0.877

TABLE 28-continued

Panels of 9 Biomarkers						
Markers			Mean CV AUC			
13	KLK3-SERPINA3 C9	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1	0.877
14	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1	0.876
15	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	CNTN1 ITIH4	BMP1	0.876
16	KLK3-SERPINA3 C9	KIT BMP1	GHR AKR7A2	IGFBP4 CRP	CNTN1	0.876
17	KLK3-SERPINA3 C9	KIT AHSG	GHR BMPER	IGFBP4 AKR7A2	CNTN1	0.876
18	KLK3-SERPINA3 C9	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1	0.876
19	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	CNTN1 CRP	C9	0.876
20	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	CNTN1 ITIH4	BMP1	0.876
21	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	GHR ITIH4	CNTN1	0.876
22	KLK3-SERPINA3 BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	BMP1	0.876
23	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	IGFBP4 ITIH4	CNTN1	0.876
24	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	IGFBP4 ITIH4	CNTN1	0.876
25	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	BMP1	0.876
26	KLK3-SERPINA3 BMP1	KIT AHSG	IGFBP4 BMPER	CNTN1 AKR7A2	C9	0.876
27	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1	0.876
28	KLK3-SERPINA3 BMPER	IGFBP2 AKR7A2	GHR CRP	IGFBP4 ITIH4	CNTN1	0.876
29	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	C9	0.875
30	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	CNTN1 CRP	BMP1	0.875
31	KLK3-SERPINA3 BMP1	KIT AKR7A2	IGFBP2 CRP	IGFBP4 ITIH4	CNTN1	0.875
32	KLK3-SERPINA3 C9	KIT BMPER	IGFBP2 AKR7A2	IGFBP4 ITIH4	CNTN1	0.875
33	KLK3-SERPINA3 C9	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	CNTN1	0.875
34	KLK3-SERPINA3 BMP1	KIT BMPER	GHR NME2	IGFBP4 CRP	CNTN1	0.875
35	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	CA6	0.875
36	KLK3-SERPINA3 C9	KIT CA6	GHR BMPER	IGFBP4 AKR7A2	CNTN1	0.875
37	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	GHR ITIH4	IGFBP4	0.875
38	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	IGFBP4 ITIH4	C9	0.875
39	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	CNTN1 ITIH4	C9	0.875
40	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	CNTN1 CRP	C9	0.875
41	KLK3-SERPINA3 BMPER	KIT NME2	IGFBP4 CRP	CNTN1 ITIH4	BMP1	0.875
42	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	IGFBP4 ITIH4	CNTN1	0.875
43	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	BMP1	0.875
44	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	IGFBP4 CRP	CNTN1	0.875
45	KLK3-SERPINA3 BMP1	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	C9	0.875
46	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	AHSG	0.874
47	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	BMP1	0.874
48	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP2 AKR7A2	CNTN1 ITIH4	BMP1	0.874

TABLE 28-continued

Panels of 9 Biomarkers						
Markers					Mean CV AUC	
49	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	IGFBP4 ITIH4	C9	0.874
50	KLK3-SERPINA3 C9	KIT AHSG	EGFR BMPER	GHR AKR7A2	CNTN1	0.874
51	KLK3-SERPINA3 CNTN1	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	IGFBP4	0.874
52	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	IGFBP4 ITIH4	C9	0.874
53	KLK3-SERPINA3 C9	KIT SERPINA1	GHR BMPER	IGFBP4 AKR7A2	CNTN1	0.874
54	KLK3-SERPINA3 BMPER	IGFBP2 AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	AHSG	0.874
55	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP2 AKR7A2	CNTN1 ITIH4	C9	0.874
56	KLK3-SERPINA3 C9	KIT AHSG	IGFBP2 BMPER	GHR AKR7A2	CNTN1	0.874
57	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	CNTN1 CRP	C9	0.874
58	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP2 AKR7A2	CNTN1 CRP	C9	0.874
59	KLK3-SERPINA3 FN1	KIT BMPER	GHR AKR7A2	CNTN1 CRP	BMP1	0.874
60	KLK3-SERPINA3 BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	C9	0.874
61	KLK3-SERPINA3 BMP1	KIT AHSG	IGFBP2 BMPER	CNTN1 AKR7A2	C9	0.874
62	KLK3-SERPINA3 CNTN1	KIT C9	IGFBP2 BMPER	GHR AKR7A2	IGFBP4	0.874
63	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	C9	0.874
64	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1	0.874
65	KLK3-SERPINA3 CA6	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	CNTN1	0.874
66	KLK3-SERPINA3 BMP1	KIT FN1	GHR BMPER	CNTN1 AKR7A2	C9	0.874
67	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	IGFBP4 ITIH4	FN1	0.874
68	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	C9 ITIH4	BMP1	0.874
69	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	CNTN1 ITIH4	CA6	0.874
70	KLK3-SERPINA3 BMP1	KIT NME2	GHR CRP	IGFBP4 ITIH4	CNTN1	0.874
71	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 CRP	IGFBP4 ITIH4	BMP1	0.874
72	KLK3-SERPINA3 C9	BDNF BMPER	KIT AKR7A2	GHR CRP	CNTN1	0.874
73	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	C9	0.874
74	KLK3-SERPINA3 BMP1	KIT AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	C9	0.874
75	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	IGFBP4 ITIH4	C9	0.874
76	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	CNTN1 ITIH4	C9	0.874
77	KIT BMP1	IGFBP2 BMPER	GHR AKR7A2	CNTN1 CRP	C9	0.874
78	KLK3-SERPINA3 C9	BDNF BMPER	KIT AKR7A2	IGFBP2 CRP	CNTN1	0.874
79	KLK3-SERPINA3 C9	KIT AHSG	IGFBP2 BMPER	EGFR AKR7A2	CNTN1	0.874
80	KLK3-SERPINA3 BMPER	KIT NME2	GHR CRP	IGFBP4 ITIH4	CNTN1	0.874
81	KLK3-SERPINA3 BMP1	KIT AHSG	EGFR BMPER	CNTN1 AKR7A2	C9	0.874
82	KLK3-SERPINA3 BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	CA6	0.874
83	KLK3-SERPINA3 BMPER	KIT NME2	GHR CRP	IGFBP4 ITIH4	BMP1	0.874
84	KLK3-SERPINA3 AKR7A2	KIT NME2	IGFBP4 CRP	CNTN1 ITIH4	BMP1	0.874

TABLE 28-continued

Panels of 9 Biomarkers						
Markers				Mean CV AUC		
85	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	C9	0.874
86	KIT AHSG	IGFBP2 BMPER	GHR AKR7A2	CNTN1 CRP	C9	0.874
87	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR CRP	CNTN1 ITIH4	C9	0.874
88	KLK3-SERPINA3 BMP1	KIT DDC	GHR BMPER	CNTN1 AKR7A2	C9	0.874
89	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	C9 CRP	BMP1	0.874
90	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	CNTN1 CRP	BMP1	0.874
91	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP2 AKR7A2	CNTN1 CRP	BMP1	0.874
92	KIT AHSG	GHR BMPER	CNTN1 AKR7A2	C9 CRP	BMP1	0.874
93	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP2 AKR7A2	CNTN1 CRP	BMP1	0.874
94	KLK3-SERPINA3 DDC	KIT AKR7A2	IGFBP2 CRP	CNTN1 ITIH4	BMP1	0.874
95	KLK3-SERPINA3 BMPER	IGFBP2 AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	DDC	0.874
96	KLK3-SERPINA3 BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	AHSG	0.874
97	KLK3-SERPINA3 CA6	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1	0.874
98	KLK3-SERPINA3 BMP1	KIT SERPINA1	GHR BMPER	IGFBP4 AKR7A2	CNTN1	0.874
99	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	BMP1	0.874
100	KLK3-SERPINA3 BMP1	KIT AHSG	GHR AKR7A2	CNTN1 CRP	C9	0.874

TABLE 29

Panels of 10 Biomarkers						
Markers					Mean CV AUC	
1	KLK3-SERPINA3 CNTN1	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	IGFBP4 ITIH4	0.880
2	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.880
3	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.878
4	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.878
5	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 NME2	CNTN1 CRP	C9 ITIH4	0.878
6	KLK3-SERPINA3 C9	KIT BMP1	GHR BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	0.878
7	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	C9 ITIH4	0.878
8	KLK3-SERPINA3 C9	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.877
9	KLK3-SERPINA3 BMP1	KIT AHSG	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.877
10	KLK3-SERPINA3 BMP1	KIT BMPER	GHR NME2	IGFBP4 CRP	CNTN1 ITIH4	0.877
11	KLK3-SERPINA3 C9	KIT AHSG	IGFBP2 BMPER	GHR AKR7A2	CNTN1 CRP	0.877
12	KLK3-SERPINA3 C9	KIT CA6	GHR BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	0.877
13	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP2 AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.877
14	KLK3-SERPINA3 BMP1	KIT CA6	GHR BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	0.877
15	KLK3-SERPINA3 C9	KIT BMP1	GHR BMPER	IGFBP4 AKR7A2	CNTN1 CRP	0.877

TABLE 29-continued

Panels of 10 Biomarkers						
Markers					Mean CV AUC	
16	KLK3-SERPINA3 BMP1	KIT CA6	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.877
17	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 NME2	CNTN1 CRP	BMP1 ITIH4	0.877
18	KLK3-SERPINA3 BMP1	KIT AHSG	IGFBP4 BMPER	CNTN1 AKR7A2	C9 ITIH4	0.877
19	KLK3-SERPINA3 C9	KIT BMP1	GHR AHSG	IGFBP4 BMPER	CNTN1 AKR7A2	0.877
20	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.876
21	KLK3-SERPINA3 AHSG	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	C9 ITIH4	0.876
22	KLK3-SERPINA3 BMP1	KIT AHSG	GHR BMPER	CNTN1 AKR7A2	C9 CRP	0.876
23	KLK3-SERPINA3 C9	KIT BMP1	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.876
24	KLK3-SERPINA3 C9	KIT AHSG	GHR BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	0.876
25	KLK3-SERPINA3 BMP1	KIT AHSG	IGFBP2 BMPER	GHR AKR7A2	CNTN1 CRP	0.876
26	KLK3-SERPINA3 C9	KIT AHSG	GHR BMPER	IGFBP4 AKR7A2	CNTN1 CRP	0.876
27	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1 ITIH4	0.876
28	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP4 AKR7A2	CNTN1 NME2	C9 ITIH4	0.876
29	KLK3-SERPINA3 CA6	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1 ITIH4	0.876
30	KLK3-SERPINA3 AHSG	GHR BMPER	IGFBP4 AKR7A2	CNTN1 CRP	BMP1 ITIH4	0.876
31	KLK3-SERPINA3 CNTN1	KIT BMP1	IGFBP2 BMPER	GHR AKR7A2	IGFBP4 CRP	0.876
32	KLK3-SERPINA3 C9	KIT BMP1	IGFBP2 BMPER	GHR AKR7A2	CNTN1 CRP	0.876
33	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR NME2	IGFBP4 CRP	C9 ITIH4	0.876
34	KLK3-SERPINA3 C9	KIT CA6	GHR BMPER	IGFBP4 AKR7A2	CNTN1 CRP	0.876
35	KLK3-SERPINA3 CA6	KIT AHSG	GHR BMPER	CNTN1 AKR7A2	C9 CRP	0.876
36	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP2 AKR7A2	CNTN1 CRP	BMP1 ITIH4	0.876
37	KLK3-SERPINA3 C9	KIT CA6	GHR AHSG	IGFBP4 BMPER	CNTN1 AKR7A2	0.876
38	KLK3-SERPINA3 BMP1	KIT FN1	GHR BMPER	CNTN1 AKR7A2	C9 CRP	0.876
39	KLK3-SERPINA3 BMP1	KIT CA6	GHR BMPER	IGFBP4 AKR7A2	CNTN1 CRP	0.875
40	KLK3-SERPINA3 C9	KIT BMP1	GHR AHSG	IGFBP4 AKR7A2	CNTN1 CRP	0.875
41	KLK3-SERPINA3 C9	IGFBP2 BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.875
42	KLK3-SERPINA3 BMP1	KIT AHSG	GHR BMPER	IGFBP4 AKR7A2	CNTN1 CRP	0.875
43	KLK3-SERPINA3 C9	KIT DDC	GHR BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	0.875
44	KLK3-SERPINA3 CNTN1	KIT C9	IGFBP2 BMPER	GHR AKR7A2	IGFBP4 CRP	0.875
45	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP2 AKR7A2	CNTN1 CRP	BMP1 ITIH4	0.875
46	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1 ITIH4	0.875
47	KLK3-SERPINA3 BMP1	KIT AKR7A2	IGFBP4 NME2	CNTN1 CRP	C9 ITIH4	0.875
48	KLK3-SERPINA3 BMP1	KIT DDC	IGFBP2 AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.875
49	KLK3-SERPINA3 AHSG	GHR BMPER	IGFBP4 AKR7A2	CNTN1 CRP	C9 ITIH4	0.875
50	KLK3-SERPINA3 C9	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	IGFBP4 ITIH4	0.875

TABLE 29-continued

Panels of 10 Biomarkers						
Markers					Mean CV AUC	
51	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR NME2	CNTN1 CRP	C9 ITIH4	0.875
52	KLK3-SERPINA3 CA6	KIT BMPER	IGFBP4 AKR7A2	CNTN1 CRP	BMP1 ITIH4	0.875
53	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	BMP1 ITIH4	0.875
54	KLK3-SERPINA3 CNTN1	KIT C9	IGFBP2 BMPER	GHR AKR7A2	IGFBP4 ITIH4	0.875
55	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR NME2	IGFBP4 CRP	CNTN1 ITIH4	0.875
56	KLK3-SERPINA3 C9	KIT CA6	IGFBP2 BMPER	GHR AKR7A2	CNTN1 CRP	0.875
57	KLK3-SERPINA3 AHSB	KIT BMPER	GHR AKR7A2	CNTN1 CRP	BMP1 ITIH4	0.875
58	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 NME2	CNTN1 CRP	C9 ITIH4	0.875
59	KLK3-SERPINA3 FN1	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	C9 ITIH4	0.875
60	KLK3-SERPINA3 C9	KIT AHSB	IGFBP2 BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	0.875
61	KLK3-SERPINA3 BMP1	KIT FN1	GHR AHSB	CNTN1 BMPER	C9 AKR7A2	0.875
62	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	IGFBP4 ITIH4	0.875
63	KLK3-SERPINA3 BMP1	KIT DDC	GHR AHSB	CNTN1 BMPER	C9 AKR7A2	0.875
64	KLK3-SERPINA3 FN1	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	BMP1 ITIH4	0.875
65	KLK3-SERPINA3 BMPER	KIT AKR7A2	GHR NME2	IGFBP4 CRP	BMP1 ITIH4	0.875
66	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP2 AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.875
67	KLK3-SERPINA3 BMP1	KIT BMPER	EGFR AKR7A2	GHR CRP	CNTN1 ITIH4	0.875
68	KLK3-SERPINA3 BMP1	KIT AHSB	GHR BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	0.875
69	KLK3-SERPINA3 C9	KIT BMP1	GHR BMPER	IGFBP4 AKR7A2	CNTN1 NME2	0.875
70	KLK3-SERPINA3 C9	KIT BMPER	GHR AKR7A2	IGFBP4 NME2	CNTN1 CRP	0.875
71	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.875
72	KLK3-SERPINA3 C9	KIT SERPINA1	GHR AHSB	IGFBP4 BMPER	CNTN1 AKR7A2	0.875
73	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.875
74	KLK3-SERPINA3 AHSB	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	BMP1 ITIH4	0.875
75	KLK3-SERPINA3 C9	KIT BMPER	GHR AKR7A2	IGFBP4 NME2	CNTN1 ITIH4	0.875
76	KLK3-SERPINA3 BMP1	KIT BMPER	GHR AKR7A2	CNTN1 NME2	C9 CRP	0.875
77	KLK3-SERPINA3 C9	KIT BMPER	IGFBP2 AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.875
78	KLK3-SERPINA3 DDC	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1 ITIH4	0.875
79	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 NME2	IGFBP4 CRP	CNTN1 ITIH4	0.875
80	KLK3-SERPINA3 CNTN1	KIT CA6	IGFBP2 BMPER	GHR AKR7A2	IGFBP4 CRP	0.875
81	KLK3-SERPINA3 AHSB	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	CNTN1 ITIH4	0.875
82	KLK3-SERPINA3 C9	KIT BMP1	GHR SERPINA1	IGFBP4 BMPER	CNTN1 AKR7A2	0.875
83	KLK3-SERPINA3 BMP1	KIT DDC	IGFBP4 BMPER	CNTN1 AKR7A2	C9 ITIH4	0.875
84	KLK3-SERPINA3 BMP1	KIT AHSB	GHR BMPER	IGFBP4 AKR7A2	C9 CRP	0.875
85	KLK3-SERPINA3 BMP1	KIT FN1	IGFBP2 BMPER	GHR AKR7A2	CNTN1 CRP	0.875

TABLE 29-continued

Panels of 10 Biomarkers						
Markers					Mean CV AUC	
86	KLK3-SERPINA3 FN1	KIT AHSG	GHR BMPER	CNTN1 AKR7A2	BMP1 CRP	0.875
87	KLK3-SERPINA3 C9	KIT FN1	GHR BMPER	IGFBP4 AKR7A2	CNTN1 CRP	0.875
88	KLK3-SERPINA3 C9	KIT BMP1	IGFBP2 BMPER	IGFBP4 AKR7A2	CNTN1 ITIH4	0.875
89	KLK3-SERPINA3 CA6	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	C9 ITIH4	0.875
90	KLK3-SERPINA3 AHSG	KIT BMPER	IGFBP2 AKR7A2	IGFBP4 CRP	C9 ITIH4	0.875
91	KIT BMPER	GHR AKR7A2	IGFBP4 NME2	CNTN1 CRP	C9 ITIH4	0.875
92	KLK3-SERPINA3 BMP1	KIT AHSG	GHR AKR7A2	IGFBP4 CRP	C9 ITIH4	0.875
93	KLK3-SERPINA3 BMP1	KIT AHSG	IGFBP2 BMPER	IGFBP4 AKR7A2	CNTN1 CRP	0.875
94	KLK3-SERPINA3 BMP1	KIT FN1	GHR AKR7A2	IGFBP4 CRP	C9 ITIH4	0.875
95	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP2 NME2	IGFBP4 CRP	C9 ITIH4	0.875
96	KLK3-SERPINA3 BMPER	KIT AKR7A2	IGFBP4 NME2	C9 CRP	BMP1 ITIH4	0.875
97	KLK3-SERPINA3 DDC	KIT BMPER	GHR AKR7A2	IGFBP4 CRP	CNTN1 ITIH4	0.875
98	KLK3-SERPINA3 C9	KIT FN1	IGFBP2 BMPER	GHR AKR7A2	CNTN1 CRP	0.875
99	KLK3-SERPINA3 BMP1	KIT BMPER	IGFBP2 AKR7A2	GHR CRP	IGFBP4 ITIH4	0.875
100	KLK3-SERPINA3 DDC	KIT BMPER	GHR AKR7A2	CNTN1 CRP	BMP1 ITIH4	0.874

TABLE 30

Counts of markers in biomarker panels								
Biomarker	Panel Size							
	3	4	5	6	7	8	9	10
AHSG	118	104	104	117	135	211	284	376
AKR7A2	205	485	676	738	810	859	921	950
BDNF	143	212	185	171	162	125	113	78
BMP1	127	157	214	273	308	404	457	495
BMPER	168	205	346	471	572	673	750	820
C9	197	313	402	466	515	536	543	587
CA6	107	96	88	74	96	120	165	223
CKB-CKM	40	1	0	0	0	0	0	0
CNTN1	137	164	235	420	579	717	763	815
CRP	183	267	407	506	558	588	671	721
DDC	110	93	93	109	129	154	161	197
EGFR	135	162	190	196	193	170	177	179
FGA-FGB-FGG	34	0	0	0	0	0	0	0
FN1	90	46	13	11	18	44	70	103
GHR	107	98	126	181	261	398	513	611
IGFBP2	123	127	176	211	277	320	360	380
IGFBP4	97	112	152	198	265	356	461	570
ITIH4	143	148	214	272	379	455	542	636
KIT	147	201	290	481	626	760	836	881
KLK3-SERPINA3	213	448	592	721	809	851	916	947
NME2	177	337	365	307	245	198	215	310
SERPINA1	83	91	56	31	25	35	60	104
STX1A	116	133	76	46	38	26	22	17

TABLE 31

Parameters derived from cancer datasets set for naïve Bayes classifiers							
		Mesothelioma		NSCLC		Renal Cell Carc.	
		Control		Control		Control	
		Cancer		cer		cer	
AKR7A2	Mean	6.65	7.35	6.76	7.16	7.48	7.16
	SD	0.51	0.48	0.43	0.25	0.58	0.39
BMPER	Mean	7.31	7.06	7.45	7.32	7.33	7.21
	SD	0.21	0.25	0.11	0.16	0.11	0.20
CNTN1	Mean	9.15	8.89	9.26	9.15	9.14	8.90
	SD	0.21	0.36	0.18	0.11	0.19	0.26
CRP	Mean	7.84	9.79	7.73	9.00	8.32	10.59
	SD	1.06	1.96	1.09	1.42	1.63	1.39
GHR	Mean	7.60	7.45	7.72	7.59	7.80	7.67
	SD	0.13	0.17	0.14	0.10	0.14	0.17
IGFBP2	Mean	8.45	8.98	8.51	9.01	8.51	8.92
	SD	0.47	0.61	0.42	0.45	0.45	0.45
IGFBP4	Mean	7.89	8.05	8.14	8.27	8.15	8.36
	SD	0.15	0.24	0.14	0.16	0.20	0.22
ITIH4	Mean	10.18	10.46	10.60	10.74	10.56	10.82
	SD	0.32	0.34	0.12	0.23	0.15	0.20
KIT	Mean	9.39	9.18	9.60	9.50	9.39	9.25
	SD	0.16	0.20	0.14	0.14	0.16	0.19
KLK3-SERPINA3	Mean	8.00	8.51	8.10	8.33	8.09	8.68
	SD	0.16	0.53	0.19	0.33	0.23	0.48

TABLE 32

Calculations derived from training set for naïve Bayes classifier.								
Biomarker	μ_c	μ_d	σ_c	σ_d	$\bar{x}$	$p(c \bar{x})$	$p(d \bar{x})$	$\ln(p(d \bar{x})/p(c \bar{x}))$
BMPER	7.450	7.323	0.108	0.164	7.045	0.003	0.576	5.176
KIT	9.603	9.503	0.139	0.141	9.534	2.546	2.767	0.083
AKR7A2	6.761	7.155	0.432	0.248	6.347	0.583	0.008	-4.309
IGFBP4	8.138	8.268	0.140	0.163	8.336	1.046	2.251	0.767
GHR	7.724	7.595	0.135	0.102	7.756	2.867	1.126	-0.935
ITIH4	10.596	10.738	0.121	0.227	10.600	3.301	1.460	-0.816
IGFBP2	8.514	9.006	0.417	0.448	8.812	0.741	0.811	0.091
KLK3-SERPINA3	8.102	8.327	0.194	0.330	7.909	1.253	0.542	-0.838
CNTN1	9.265	9.149	0.181	0.114	9.410	1.602	0.252	-1.848
CRP	7.733	9.005	1.095	1.422	7.675	0.364	0.181	-0.697

1. A method for diagnosing that an individual does or does not have non-small cell lung cancer (NSCLC), the method comprising:

providing a biomarker panel comprising N of the biomarker proteins listed in Table 1; wherein N is an integer from 2 to 59; and

detecting biomarker proteins, in a biological sample from an individual, to give biomarker values that each correspond to one of said N biomarker proteins in the panel, wherein said individual is classified as having or not having lung cancer based on said biomarker values.

2. The method of claim 1, wherein detecting the biomarker values comprises performing an in vitro assay.

3. The method of claim 2, wherein said in vitro assay comprises at least one capture reagent corresponding to each of said biomarkers, and further comprising selecting said at least one capture reagent from the group consisting of aptamers, antibodies, and a nucleic acid probe.

4. The method of claim 3, wherein said at least one capture reagent is an aptamer.

5. The method of claim 2, wherein the in vitro assay is selected from the group consisting of an immunoassay, an aptamer-based assay, a histological or cytological assay, and an mRNA expression level assay.

6. The method of claim 1, wherein an individual is classified as having ovarian cancer, or the likelihood of the individual having ovarian cancer is determined, based on a classification score that deviates from on a predetermined threshold.

7. The method claim 1, wherein the biological sample is lung tissue and wherein the biomarker values derive from a histological or cytological analysis of said lung tissue.

8. The method of claim 1, wherein the biological sample is selected from the group consisting of whole blood, plasma, and serum.

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

10. The method of claim 1, wherein the individual is a human.

11. The method of claim 1, wherein N=2 to 15.

12. The method of claim 1, wherein N=2 to 10.

13. The method of claim 1, wherein N=3 to 10.

14. The method of claim 1, wherein N=4 to 10.

15. The method of claim 1, wherein N=5 to 10.

16. The method of claim 1, wherein the individual is a smoker.

17. The method of claim 1, wherein the individual has a pulmonary nodule.

18. The method of claim 1, wherein the biomarkers are selected from Table 13.

19. A computer-implemented method for indicating a likelihood of NSCLC, the method comprising:

retrieving on a computer biomarker information for an individual, wherein the biomarker information comprises biomarker values that each correspond to one of at least N biomarkers selected from Table 1;

performing with the computer a classification of each of said biomarker values; and

indicating a likelihood that said individual has lung cancer based upon a plurality of classifications, and wherein N=2-59.

20-23. (canceled)

24. A method for screening an asymptomatic high risk individual for NSCL, the method comprising:

providing a biomarker panel comprising N of the biomarker proteins listed in Table 1; wherein N is an integer from 2 to 59; and

detecting biomarker proteins, in a biological sample from an individual, to give biomarker values that each correspond to one of said N biomarkers proteins in the panel, wherein said individual is classified as having or not having NSCLC, or the likelihood of the individual having NSCLC is determined, based on said biomarker values.

25-41. (canceled)

42. A method for diagnosing that an individual does or does not have NSCLC, the method comprising:

providing a biomarker panel comprising N of the biomarker proteins listed in Table 1; and

detecting, in a biological sample from an individual, biomarker values that each correspond to one of said N biomarker proteins in the panel, wherein the biomarker panel has a AUC value of 0.80 or greater.

43. (canceled)

45. A method for diagnosing that an individual does or does not have cancer, the method comprising:

providing a biomarker panel comprising N of the biomarker proteins listed in Table 19; wherein N is an integer from 3 to 23; and

detecting biomarker proteins, in a biological sample from an individual, to give biomarker values that each correspond to one of said N biomarker proteins in the panel,

wherein said individual is classified as having or not having cancer based on said biomarker values.

46-56. (canceled)

57. A computer-implemented method for indicating a likelihood of cancer, the method comprising:

retrieving on a computer biomarker information for an individual, wherein the biomarker information comprises biomarker values that each correspond to one of at least N biomarkers selected from Table 19;

performing with the computer a classification of each of said biomarker values; and

indicating a likelihood that said individual has cancer based upon a plurality of classifications, and wherein N=3-3.

58-61. (canceled)

62. The method according to claim 1, wherein said individual is classified as having or not having NSCLC, or the likelihood of the individual having NSCLC is determined, based on said biomarker values and at least one item of additional biomedical information corresponding to said individual.

63. (canceled)

64. The method according to claim 62, wherein said at least one item of additional biomedical information is independently selected from the group consisting of:

- (a) information corresponding to physical descriptors of said individual,
- (b) information corresponding to the radiologic descriptors of a pulmonary abnormality in said individual,
- (c) information corresponding to the presence or absence of a pulmonary nodule in said individual,
- (d) information corresponding to physical descriptors of a pulmonary nodule in said individual,

(e) information corresponding to a change in height and/or weight of said individual,

(f) information corresponding to the ethnicity of said individual,

(g) information corresponding to the gender of said individual,

(h) information corresponding the smoking history of said individual,

(i) information corresponding to environmental tobacco exposure in said individual,

(j) information corresponding to alcohol use history in said individual,

(k) information corresponding to occupational history of said individual,

(l) information corresponding to family history of lung cancer or other cancer in said individual,

(m) information corresponding to the presence or absence in said individual of at least one genetic marker correlating with a higher risk of lung cancer or cancer in said individual or a family member of said individual,

(n) information corresponding to clinical symptoms of said individual,

(o) information corresponding to other laboratory tests,

(p) information corresponding to gene expression values of said individual, and

(q) information corresponding to said individual's exposure to known carcinogens.

65. A classifier comprising the biomarkers of Table 13.

66. A classifier comprising the biomarkers in Example 5, specifically MMP12, MMP7, KLK3-SERPINA3, CRP, C9, CNDP1, and EGFR.

* * * * *