

METHODS OF DETERMINING IMPAIRED GLUCOSE TOLERANCE**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims the benefit of priority of United States Provisional Application No. 62/959,660, filed January 10, 2020, which is incorporated by reference herein in its entirety for any purpose.

FIELD OF THE INVENTION

[0002] The present application relates generally to the detection of biomarkers and the characterization of glucose tolerance, for example, to identify subjects having or likely to have impaired glucose tolerance, which is an indication of pre-diabetes, or diabetes. In various embodiments, the invention relates to one or more biomarkers, methods, devices, reagents, systems, and kits for characterizing impaired glucose tolerance, which may indicate pre-diabetes, and/or diabetes in an individual.

BACKGROUND

[0003] Current methods for determining glucose tolerance include measuring the 2-hour plasma glucose level in individuals participating in an oral glucose tolerance test (OGTT). A typical OGTT consists of the ingestion of a 75-gram oral glucose solution, following an overnight fast, with plasma glucose measurements at baseline (“fasting plasma glucose”) and at 2-hours (“2hr-OGTT glucose value”) post ingestion. Both fasting plasma glucose and 2-hour OGTT plasma glucose, along with HbA1c values, are currently accepted by the American Diabetes Association for the clinical diagnosis of Type 2 diabetes. Prevention of Type 2 diabetes has primarily been evaluated based on individuals with impaired glucose tolerance, defined as 2-hour OGTT plasma glucose levels ≥ 7.8 mmol/L. *See, e.g.,* Knowler et al. *N. Engl. J. Med.* 2002; 346:393-403 and Tuomilehto et al. *N. Engl. J. Med.* 2001; 344:1343-1350. The development of a proteomic model representative of the OGTT plasma glucose threshold for impaired glucose tolerance, without the need for fasting or glucose administration, , would be highly desirable.

SUMMARY

[0004] In some embodiments, methods of determining whether a subject has or likely has impaired glucose tolerance are provided. In some embodiments, methods of identifying subjects with pre-diabetes or likely to develop pre-diabetes are provided. In some embodiments, methods of identifying subjects likely to develop diabetes are provided.

[0005] In some embodiments, the method herein is for determining whether a subject has impaired glucose tolerance or likely has impaired glucose tolerance, which is an indication of pre-diabetes, or diabetes, comprising obtaining a sample from the subject, forming a biomarker panel having N biomarker proteins, and detecting the level of each of the N biomarker proteins in a sample from the subject, wherein N is at least 3, and wherein at least one of the N biomarker protein is selected from ACY1, COL1A1, RTN4R, CRLF1:CLCF1 complex, CBX7, KIN, SERPINA11, PELI2, TFF3, FABP12, GAD1, SVEP1, SOCS7, F9, STC1, MYOC, WFDC11, CALB1, CCL16, SMCO2, CCL23, OSTM1, RNASE10, ITIH1, ZNF134, CFAP45, and SFTPD. In some embodiments, N is 3 to 41, N is 4 to 41, N is 5 to 41, or N is 6 to 41, or N is 7 to 41, or N is 8 to 41, or N is 9 to 41, or N is 10 to 41, or N is 11 to 41, or N is 12 to 41, or N is 13 to 41, or N is 14 to 41, or N is 15 to 41, or N is 16 to 41, or N is at least 4, or N is at least 5, or N is at least 6, or N is at least 7, or N is at least 8, or N is at least 9, or N is at least 10, or N is at least 11, or N is at least 12, or N is at least 13, or N is at least 14, or N is at least 15, or N is at least 16. In some embodiments, N is 3, or N is 4, or N is 5, or N is 6, or N is 7, or N is 8, or N is 9, or N is 10, or N is 11, or N is 12, or N is 13, or N is 14, or N is 15, or N is 16, or N is 17, or N is 18, or N is 19, or N is 20, or N is 21, or N is 22, or N is 23, or N is 24, or N is 25, or N is 26, or N is 27, or N is 28, or N is 29, or N is 30, or N is 31, or N is 32, or N is 33, or N is 34, or N is 35, or N is 36, or N is 37, or N is 38, or N is 39, or N is 40, or N is 41. In some such embodiments, the subject has impaired glucose tolerance. In some embodiments, the subject likely has impaired glucose tolerance. In some embodiments, the subject is likely to develop pre-diabetes. In some embodiments, the subject is pre-diabetic. In some embodiments, the subject is likely to develop diabetes. In some embodiments, the subject has impaired glucose tolerance and has diabetes. In some embodiments, a subject likely to develop diabetes undertakes preventive measures or undergoes preventive treatment to reduce the likelihood of developing diabetes.

[0006] In some embodiments, each of the N biomarkers is selected from Table 1. In some embodiments, at least one of the N biomarker proteins is selected from ACY1, COL1A1, RTN4R, CRLF1:CLCF1 complex, CBX7, and KIN. In some embodiments, one or two of the N biomarker proteins are INHBC and/or SHBG. In some embodiments, at least 2 or at least 3 of the N protein biomarkers are selected from ACY1, COL1A1, RTN4R, CRLF1:CLCF1 complex, CBX7, KIN, SERPINA11, PELI2, TFF3, FABP12, GAD1, SVEP1, SOCS7, F9, STC1, MYOC, WFDC11, CALB1, CCL16, SMCO2, CCL23, OSTM1, RNASE10, ITIH1, ZNF134, CFAP45, and SFTPD. In some embodiments, at least one of the N biomarker proteins is selected from FAM20B, COL15A1, MARCKSL1, HTRA1, CHAD, CPM, DLK1, HERC1, IL20RB, MAP2K4, GPX2, and FGFR4. In some embodiments, two of the N biomarker proteins are INHBC and ACY1, or two of the N biomarker proteins are SHBG and ACY1, or three of the N

biomarker proteins are INHBC, SHBG, and ACY1. In some embodiments, two of the N biomarker proteins are INHBC and COL1A1, or two of the N biomarker proteins are SHBG and COL1A1, or wherein three of the N biomarker proteins are INHBC, SHBG, and COL1A1. In some embodiments, two of the N biomarker proteins are INHBC and RTN4R, or two of the N biomarker proteins are SHBG and RTN4R, or wherein three of the N biomarker proteins are INHBC, SHBG, and RTN4R. In some embodiments, two of the N biomarker proteins are INHBC and CRLF1:CLCF1 complex, or two of the N biomarker proteins are SHBG and CRLF1:CLCF1 complex, or wherein three of the N biomarker proteins are INHBC, SHBG, and CRLF1:CLCF1 complex. In some embodiments, two of the N biomarker proteins are INHBC and CBX7, or two of the N biomarker proteins are SHBG and CBX7, or wherein three of the N biomarker proteins are INHBC, SHBG, and CBX7. In some embodiments, two of the N biomarker proteins are INHBC and KIN, or two of the N biomarker proteins are SHBG and KIN, or wherein three of the N biomarker proteins are INHBC, SHBG, and KIN. In some embodiments, N is at least five and five of the N biomarker proteins are INHBC, SHBG, ACY1, COL1A1, and RTN4R. In some embodiments, N is at least 16, and wherein 16 of the N biomarker proteins are ACY1, COL1A1, RTN4R, CRLF1, CBX7, KIN, SERPINA11, PELI2, TFF3, FABP12, INHBC, SHBG, FAM20B, COL15A1, MARCKSL1, and HTRA1.

[0007] In any of the embodiments described herein, the subject may be at risk of developing impaired glucose tolerance. In any of the embodiments described herein, the subject may be at risk of developing pre-diabetes. In any of the embodiments described herein, the subject may be at risk of developing diabetes. In some embodiments, the method comprises determining whether the subject has or likely has impaired glucose tolerance, which is an indication of pre-diabetes, or diabetes. In some embodiments, the method comprises determining whether the subject is pre-diabetic or is likely to develop pre-diabetes or diabetes. In some embodiments, the diabetes is type 2 diabetes. In some embodiments, the method comprises administering a treatment to the subject. In some such embodiments, the treatment comprises administering insulin and/or metformin to the subject. In some embodiments the treatment comprises implementing a weight loss program, implementing dietary restrictions, implementing caloric restrictions and/or implementing an exercise program for the subject.

[0008] In any of the embodiments described herein, each of the N biomarker proteins are different from each other. In some embodiments, the method comprises contacting biomarkers of the sample from the subject with a set of biomarker capture reagents, wherein each biomarker capture reagent of the set of biomarker capture reagents specifically binds to a different biomarker being detected. In some embodiments, each biomarker capture reagent is an antibody or an aptamer. In some embodiments, each biomarker capture reagent is an aptamer. In some

embodiments, at least one aptamer is a slow off-rate aptamer. In some embodiments, at least one slow off-rate aptamer comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least 10 nucleotides with modifications. In some embodiments, each slow off-rate aptamer binds to a biomarker protein with an off rate ($t_{1/2}$) of ≥ 30 minutes, ≥ 60 minutes, ≥ 90 minutes, ≥ 120 minutes, ≥ 150 minutes, ≥ 180 minutes, ≥ 210 minutes, or ≥ 240 minutes.

[0009] In any of the embodiments described herein, the sample may be a blood sample. In any of the embodiments described herein, the sample may be selected from a serum sample or a plasma sample. In any of the embodiments described herein, the sample is a plasma sample. In some embodiments, the subject eats a typical diet before the sample is obtained, wherein the typical diet excludes fasting longer than usual.

[0010] In some embodiments, a level of at least one biomarker selected from SHBG, COL1A1, CRLF1:CLCF1 complex, FAM20B, COL15A1, KIN, SERPINA11, PELI2, MARCKSL1, CHAD, IL20RB, MYOC, WFDC11, MAP2K4, CALB1, FGFR4, OSTM1, ITIH1, CFAP45, and SFTPD that is higher than a control level of the respective biomarker, indicates that the subject has or likely has impaired glucose tolerance, has pre-diabetes, and/or is likely to develop pre-diabetes or diabetes.

[0011] In some embodiments, a level of at least one biomarker selected from INHBC, ACY1, RTN4R, CBX7, TFF3, HTRA1, FABP12, GAD1, CPM, SVEP1, SOCS7, F9, DLK1, HERC1, STC1, CCL16, SMC02, GPX2, CCL23, RNASE10, and ZNF134 that is lower than a control level of the respective biomarker, indicates that the subject has or likely has impaired glucose tolerance, has pre-diabetes, and/or is likely to develop pre-diabetes or diabetes.

[0012] In some embodiments, a method described herein is for the purpose of determining a medical insurance premium or life insurance premium. In some embodiments, a method further comprises determining a medical insurance premium or life insurance premium. In some embodiments, a method described herein further comprises using information resulting from the method to predict and/or manage the utilization of medical resources.

[0013] In some embodiments, kits are provided. In some embodiments, a kit comprises N biomarker protein capture reagents, wherein N is at least 3, and wherein at least one of the N biomarker protein capture reagents specifically binds to a biomarker protein selected from ACY1, COL1A1, RTN4R, CRLF1:CLCF1 complex, CBX7, KIN, SERPINA11, PELI2, TFF3, FABP12, GAD1, SVEP1, SOCS7, F9, STC1, MYOC, WFDC11, CALB1, CCL16, SMC02, CCL23, OSTM1, RNASE10, ITIH1, ZNF134, CFAP45, and SFTPD. In some embodiments, N is 3 to 41, N is 4 to 41, N is 5 to 41, or N is 6 to 41, or N is 7 to 41, or N is 8 to 41, or N is 9 to 41, or N is 10 to 41, or N is 11 to 41, or N is 12 to 41, or N is 13 to 41, or N is 14 to 41, or N is

15 to 41, or N is 16 to 41, or N is at least 4, or N is at least 5, or N is at least 6, or N is at least 7, or N is at least 8, or N is at least 9, or N is at least 10, or N is at least 11, or N is at least 12, or N is at least 13, or N is at least 14, or N is at least 15, or N is at least 16. In some embodiments, N is 3, or N is 4, or N is 5, or N is 6, or N is 7, or N is 8, or N is 9, or N is 10, or N is 11, or N is 12, or N is 13, or N is 14, or N is 15, or N is 16, or N is 17, or N is 18, or N is 19, or N is 20, or N is 21, or N is 22, or N is 23, or N is 24, or N is 25, or N is 26, or N is 27, or N is 28, or N is 29, or N is 30, or N is 31, or N is 32, or N is 33, or N is 34, or N is 35, or N is 36, or N is 37, or N is 38, or N is 39, or N is 40, or N is 41. In some such embodiments, the kit is used to detect the levels of the N biomarker proteins in a sample, wherein the sample is from a subject. In some embodiments, the kit is used to determine whether the subject has or likely has impaired glucose tolerance, has pre-diabetes, and/or is likely to develop pre-diabetes or diabetes.

[0014] In some embodiments, each of the N biomarkers is selected from Table 1. In some embodiments, at least one of the N biomarker proteins is selected from ACY1, COL1A1, RTN4R, CRLF1:CLCF1 complex, CBX7, and KIN. In some embodiments, one or two of the N biomarker proteins are INHBC and/or SHBG. In some embodiments, at least 2 or at least 3 of the N protein biomarkers are selected from ACY1, COL1A1, RTN4R, CRLF1:CLCF1 complex, CBX7, KIN, SERPINA11, PELI2, TFF3, FABP12, GAD1, SVEP1, SOCS7, F9, STC1, MYOC, WFDC11, CALB1, CCL16, SMCO2, CCL23, OSTM1, RNASE10, ITIH1, ZNF134, CFAP45, and SFTPD. In some embodiments, at least one of the N biomarker proteins is selected from FAM20B, COL15A1, MARCKSL1, HTRA1, CHAD, CPM, DLK1, HERC1, IL20RB, MAP2K4, GPX2, and FGFR4. In some embodiments, two of the N biomarker proteins are INHBC and ACY1, or two of the N biomarker proteins are SHBG and ACY1, or three of the N biomarker proteins are INHBC, SHBG, and ACY1. In some embodiments, two of the N biomarker proteins are INHBC and COL1A1, or two of the N biomarker proteins are SHBG and COL1A1, or wherein three of the N biomarker proteins are INHBC, SHBG, and COL1A1. In some embodiments, two of the N biomarker proteins are INHBC and RTN4R, or two of the N biomarker proteins are SHBG and RTN4R, or wherein three of the N biomarker proteins are INHBC, SHBG, and RTN4R. In some embodiments, two of the N biomarker proteins are INHBC and CRLF1:CLCF1 complex, or two of the N biomarker proteins are SHBG and CRLF1:CLCF1 complex, or wherein three of the N biomarker proteins are INHBC, SHBG, and CRLF1:CLCF1 complex. In some embodiments, two of the N biomarker proteins are INHBC and CBX7, or two of the N biomarker proteins are SHBG and CBX7, or wherein three of the N biomarker proteins are INHBC, SHBG, and CBX7. In some embodiments, two of the N biomarker proteins are INHBC and KIN, or two of the N biomarker proteins are SHBG and KIN, or wherein three of the N biomarker proteins are INHBC, SHBG, and KIN. In some

embodiments, N is at least five and five of the N biomarker proteins are INHBC, SHBG, ACY1, COL1A1, and RTN4R. In some embodiments, N is at least 16, and wherein 16 of the N biomarker proteins are ACY1, COL1A1, RTN4R, CRLF1, CBX7, KIN, SERPINA11, PELI2, TFF3, FABP12, INHBC, SHBG, FAM20B, COL15A1, MARCKSL1, and HTRA1.

[0015] In some embodiments, each biomarker capture reagent is an antibody or an aptamer. In some embodiments, each biomarker capture reagent is an aptamer. In some embodiments, at least one aptamer is a slow off-rate aptamer.

[0016] In some embodiments, each slow off-rate aptamer binds to a biomarker protein with an off rate ($t_{1/2}$) of ≥ 30 minutes, ≥ 60 minutes, ≥ 90 minutes, ≥ 120 minutes, ≥ 150 minutes, ≥ 180 minutes, ≥ 210 minutes, or ≥ 240 minutes.

[0017] In any of the embodiments described herein, the sample may be a blood sample. In any of the embodiments described herein, the sample may be selected from a serum sample and a plasma sample. In some embodiments, the sample is a plasma sample.

[0018] In any of the embodiments described herein, each of the N biomarker proteins is different from the other N biomarker proteins. In any of the embodiments described herein, at least one slow off-rate aptamer may comprise at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least 10 nucleotides with modifications. In some embodiments, the modifications are hydrophobic modifications. In some embodiments, the modifications are hydrophobic base modifications. In some embodiments, one or more of the modifications may be selected from the modifications shown in Figure 2.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] FIG. 1A-C show the AUC, sensitivity, and specificity of models comprising N biomarker proteins, as described in Example 4. FIG. 1A shows results for models in which one biomarker protein is added to the model at a time, as described in Example 4. FIG. 1B shows results for models in which one biomarker protein is removed from the model at a time, as described in Example 4. FIG. 1C shows results for models that comprise randomly selected N biomarkers from Table 1, as described in Example 4.

[0020] FIG. 2 shows certain nucleobase modifications that can be used in an aptamer.

[0021] FIG. 3 shows a nonlimiting exemplary computer system for use with various computer-implemented methods described herein.

[0022] FIG. 4 shows glucose tolerance status as determined by the method described in Example 2 (y-axis) relative to the classification determined by a traditional OGTT (x-axis).

DETAILED DESCRIPTION

[0023] While the invention will be described in conjunction with certain representative embodiments, it will be understood that the invention is defined by the claims, and is not limited to those embodiments.

[0024] One skilled in the art will recognize many methods and materials similar or equivalent to those described herein may be used in the practice of the present invention. The present invention is in no way limited to the methods and materials described.

[0025] Unless defined otherwise, technical and scientific terms used herein have the meaning 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 of the invention, certain methods, devices, and materials are described herein.

[0026] 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.

[0027] As used in this application, including the appended claims, the singular forms “a,” “an,” and “the” include the plural, unless the context clearly dictates otherwise, and may be 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.

[0028] 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 may include other elements not expressly listed.

[0029] The present application includes biomarkers, methods, devices, reagents, systems, and kits for determining whether a subject has or likely has impaired glucose tolerance, or has or is likely to develop pre-diabetes, and/or diabetes. In some embodiments, biomarkers, methods, devices, reagents, systems, and kits are provided for determining whether a subject with impaired glucose tolerance has or is likely to develop pre-diabetes or diabetes.

[0030] As used herein, the term “CRLF1:CLCF1 complex” is used to refer to CRLF1 and/or CLCF1 and/or the complex of CRLF1 and CLCF1. Thus, if a method comprises detecting the biomarker “CRLF1:CLCF1 complex,” the method may comprise detecting CRLF1, CLCF1, both CRLF1 and CLCF1, and/or the complex of CRLF1 and CLCF1. A

biomarker capture reagent that specifically binds CRLF1:CLCF1 complex may bind CRLF1 and/or CLCF1 and/or both CRLF1 and CLCF1 and/or the complex of CRLF1 and CLCF1.

[0031] In some embodiments, one or more biomarkers are provided for use either alone or in various combinations to determine whether a subject has or likely has impaired glucose tolerance or has or is likely to develop pre-diabetes and/or diabetes. As described in detail below, exemplary embodiments include the biomarkers provided in Table 1.

[0032] The biomarkers were identified using a multiplex aptamer-based assay. Table 1 lists biomarkers that are useful for distinguishing samples obtained from glucose tolerant individuals from samples obtained from individuals with impaired glucose tolerance.

[0033] The terms “sensitivity” and “specificity” are used herein with respect to the ability to correctly classify an individual, based on one or more biomarker levels detected in a biological sample, as having the disease or not having the disease. In some embodiments, the terms “sensitivity” and “specificity” may be used herein with respect to the ability to correctly classify an individual, based on one or more biomarker levels detected in a biological sample, as having impaired glucose tolerance or having normal glucose tolerance. In such embodiments, “sensitivity” indicates the performance of the biomarker(s) with respect to correctly classifying individuals with impaired glucose tolerance. “Specificity” indicates the performance of the biomarker(s) with respect to correctly classifying individuals who do not have impaired glucose tolerance. For example, 85% specificity and 90% sensitivity for a panel of biomarkers used to test a set of control samples (such as samples from healthy individuals or subjects known not to have impaired glucose tolerance) and test samples (such as samples from individuals with impaired glucose tolerance) indicates that 85% of the control samples were correctly classified as control samples by the panel, and 90% of the test samples were correctly classified as test samples by the panel.

[0034] In some embodiments, the terms “sensitivity” and “specificity” may be used herein with respect to the ability to correctly classify an individual, based on one or more biomarker levels detected in a sample, as having impaired glucose tolerance, which generally is an indication of pre-diabetes, and in some case may indicate diabetes. “Sensitivity” indicates the performance of the biomarker(s) with respect to correctly classifying individuals with impaired glucose tolerance, which generally is an indication of pre-diabetes, and in some case may indicate diabetes. “Specificity” indicates the performance of the biomarker(s) with respect to correctly classifying individuals who do not have impaired glucose tolerance, which generally is an indication of pre-diabetes, and in some case may indicate diabetes. For example, 85% specificity and 90% sensitivity for a panel of biomarkers used to test a set of samples from individuals with normal glucose tolerance would result in 85% of individuals correctly

classified. Similarly, in a set of samples from individuals with impaired glucose tolerance would correctly classify 90% of individuals.

[0035] In some embodiments, overall performance of a panel of one or more biomarkers is represented by the area-under-the-curve (AUC) value. The AUC value is derived from a receiver operating characteristic (ROC) curve. 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. 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., normal individuals and individuals with diabetes, or individuals with impaired glucose tolerance and individuals with probable diabetes). 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.

Typically, the feature data across the entire population 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.

[0036] As used herein, “obese” with reference to a subject refers to a subject with a BMI of 30 or greater.

[0037] “Biological sample” and “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, urine, saliva, peritoneal washings, ascites, cystic fluid, glandular fluid, lymph fluid, bronchial aspirate,

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). In some embodiments, 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, 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.

[0038] As used herein, “typical diet” means an individual’s daily eating habits. An individual’s typical diet may be the same, similar to, or different from any other individual’s typical diet. A typical diet excludes dietary changes, such as fasting longer than usual, eating more or less than usual, or any dietary change made to accommodate a medical test.

[0039] Further, in some embodiments, a biological sample may 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 may be treated as described herein for a sample from a single individual, and, for example, if a poor prognosis is established in the pooled sample, then each individual biological sample can be re-tested to determine which individual(s) have impaired glucose tolerance and/or have or are likely to develop pre-diabetes or diabetes.

[0040] “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” refers to 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 type or species of molecule or multi-molecular structure. Exemplary target molecules include proteins, polypeptides, nucleic acids, carbohydrates, lipids, polysaccharides, glycoproteins, hormones, receptors, antigens, antibodies, affybodies, 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. In some embodiments, a target molecule is a protein, in which case the target molecule may be referred to as a “target protein.”

[0041] As used herein, a “capture agent” or “capture reagent” refers to a molecule that is capable of binding specifically to a biomarker protein. A “biomarker protein capture reagent” refers to a molecule that is capable of binding specifically to a biomarker protein. Nonlimiting exemplary capture reagents include aptamers, antibodies, adnectins, ankyrins, other antibody mimetics and other protein scaffolds, autoantibodies, chimeras, small molecules, nucleic acids, lectins, ligand-binding receptors, imprinted polymers, avimers, peptidomimetics, hormone receptors, cytokine receptors, synthetic receptors, and modifications and fragments of any of the aforementioned capture reagents. In some embodiments, a capture reagent is selected from an aptamer and an antibody.

[0042] The term “antibody” refers to full-length antibodies of any species and fragments and derivatives of such antibodies, including Fab fragments, F(ab')₂ fragments, single chain antibodies, Fv fragments, and single chain Fv fragments. The term “antibody” also refers to synthetically-derived antibodies, such as phage display-derived antibodies and fragments, affybodies, nanobodies, etc.

[0043] 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. In some embodiments, a biomarker is a target protein.

[0044] As used herein, “biomarker level” and “level” 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 “level” depends on the specific design and components of the particular analytical method employed to detect the biomarker.

[0045] A “control level” of a target molecule refers to the level of the target molecule in the same sample type from an individual that does not have the disease or condition, or from an individual that is not suspected of having the disease or condition. A “control level” of a target molecule need not be determined each time the present methods are carried out, and may be a previously determined level that is used as a reference or threshold to determine whether the level in a particular sample is higher or lower than a normal level. In some embodiments, a control level in a method described herein is the level that has been observed in one or more subjects with normal glucose tolerance. In some embodiments, a control level in a method described herein is the level that has been observed in one or more subjects with impaired glucose tolerance, but not diabetes. In some embodiments, a control level in a method described herein is the average or mean level, optionally plus or minus a statistical variation, that has been observed in a plurality of normal subjects, or subjects with impaired glucose tolerance, but not diabetes.

[0046] As used herein, “individual” and “subject” are used interchangeably to refer 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 (such as impaired glucose tolerance) is not detectable by conventional diagnostic methods.

[0047] “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 impaired glucose tolerance includes distinguishing individuals who have impaired glucose tolerance from individuals who have normal glucose tolerance. The diagnosis of pre-diabetes or diabetes includes distinguishing individuals who have diabetes from individuals who have impaired glucose tolerance, but not probable diabetes, and from individuals with normal glucose tolerance.

[0048] “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.

[0049] “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” glucose tolerance can include, for example, any of the following: prognosing the future course of glucose tolerance in an individual; predicting whether impaired glucose tolerance will progress to pre-diabetes or diabetes; predicting whether a particular stage of pre-diabetes or diabetes will progress to a higher stage of pre-diabetes or diabetes; etc.

[0050] As used herein, “detecting” or “determining” with respect to a biomarker level includes the use of both the instrument used to observe and record a signal corresponding to a biomarker level and the material/s required to generate that signal. In various embodiments, the level 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.

[0051] As used herein, a “subject with impaired glucose tolerance” refers to a subject that has been diagnosed with impaired glucose tolerance. In some embodiments, impaired glucose tolerance is suspected during a routine checkup, monitoring of metabolic syndrome and obesity, or monitoring for possible side effects of drugs.

[0052] As used herein, a “subject with pre-diabetes” or a “subject with diabetes” refer to a subject that has been diagnosed with pre-diabetes or diabetes. In some embodiments, diagnosing the pre-diabetes or diabetes comprises a method described above for impaired glucose tolerance.

[0053] As used herein, a “subject at risk of developing” a condition refers to a subject with one or more risk factors or comorbidities of the condition. In some embodiments, the

condition is diabetes. Risk factors associated with developing diabetes include, but are not limited to, being 45 years or older, being male, being overweight or having a BMI of about 25 kg/m² or higher, a family history of diabetes, being physically active less than 3 time per week, race (e.g., African American, Hispanic/Latino American, American Indian or Alaskan Native), a history of having gestational diabetes and/or polycystic syndrome.

[0054] As used herein, “likely” means a probability higher than 0.50.

[0055] “Solid support” refers herein to any substrate having a surface to which molecules may be attached, directly or indirectly, through either covalent or non-covalent bonds. A “solid support” can have a variety of physical formats, which can include, for example, a membrane; a chip (e.g., a protein chip); a slide (e.g., a glass slide or coverslip); a column; a hollow, solid, semi-solid, pore- or cavity- containing particle, such as, for example, a bead; a gel; a fiber, including a fiber optic material; a matrix; and a sample receptacle. Exemplary sample receptacles include sample wells, tubes, capillaries, vials, and any other vessel, groove or indentation capable of holding a sample. A sample receptacle can be contained on a multi-sample platform, such as a microtiter plate, slide, microfluidics device, and the like. A support can be composed of a natural or synthetic material, an organic or inorganic material. The composition of the solid support on which capture reagents are attached generally depends on the method of attachment (e.g., covalent attachment). Other exemplary receptacles include microdroplets and microfluidic controlled or bulk oil/aqueous emulsions within which assays and related manipulations can occur. Suitable solid supports include, for example, plastics, resins, polysaccharides, silica or silica-based materials, functionalized glass, modified silicon, carbon, metals, inorganic glasses, membranes, nylon, natural fibers (such as, for example, silk, wool and cotton), polymers, and the like. The material composing the solid support can include reactive groups such as, for example, carboxy, amino, or hydroxyl groups, which are used for attachment of the capture reagents. Polymeric solid supports can include, e.g., polystyrene, polyethylene glycol tetraphthalate, polyvinyl acetate, polyvinyl chloride, polyvinyl pyrrolidone, polyacrylonitrile, polymethyl methacrylate, polytetrafluoroethylene, butyl rubber, styrenebutadiene rubber, natural rubber, polyethylene, polypropylene, (poly)tetrafluoroethylene, (poly)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

[0056] In various exemplary embodiments, methods are provided for determining whether a subject has or likely has impaired glucose tolerance, has pre-diabetes, and/or is likely to develop pre-diabetes or diabetes. In various embodiments, a method is provided for determining whether a subject has impaired glucose tolerance and/or is likely to develop pre-diabetes or diabetes, comprising obtaining a sample from the subject, forming a biomarker panel having N biomarker proteins, and detecting the level of each of the N biomarker proteins in the sample, wherein N is at least 3, and wherein at least one of the N biomarker proteins is selected from ACY1, COL1A1, RTN4R, CRLF1:CLCF1 complex, CBX7, KIN, SERPINA11, PELI2, TFF3, FABP12, GAD1, SVEP1, SOCS7, F9, STC1, MYOC, WFDC11, CALB1, CCL16, SMC02, CCL23, OSTM1, RNASE10, ITIH1, ZNF134, CFAP45, and SFTPD.

[0057] In various embodiments, each of the N biomarkers is selected from Table 1.

Table 1: Biomarkers for use in Embodiments Described Herein

Biomarker	UniProt ID (Protein)	Full Name
INHBC	P55103	Inhibin beta C chain
SHBG	P04278	Sex hormone-binding globulin
ACY1	Q03154	Aminoacylase-1
COL1A1	P02452	Collagen alpha-1(I) chain
RTN4R	Q9BZR6	Reticulon-4 receptor
CRLF1:CLCF1 complex	O75462:Q9UBD9	Cytokine receptor-like factor 1:Cardiotrophin-like cytokine factor 1 Complex
CBX7	O95931	Chromobox protein homolog 7
FAM20B	O75063	Glycosaminoglycan xylosylkinase
COL15A1	P39059	Collagen alpha-1(XV) chain
KIN	O60870	DNA/RNA-binding protein KIN17
SERPINA11	Q86U17	Serpin A11
PELI2	Q9HAT8	E3 ubiquitin-protein ligase pellino homolog 2
TFF3	Q07654	Trefoil factor 3
MARCKSL1	P49006	MARCKS-related protein
HTRA1	Q92743	Serine protease HTRA1
FABP12	A6NFH5	Fatty acid-binding protein 12
GAD1	Q99259	Glutamate decarboxylase 1
CHAD	O15335	Chondroadherin
CPM	P14384	Carboxypeptidase M
SVEP1	Q4LDE5	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
SOCS7	O14512	Suppressor of cytokine signaling 7
F9	P00740	Coagulation factor IX
DLK1	P80370	Protein delta homolog 1
HERC1	Q15751	Probable E3 ubiquitin-protein ligase HERC1

IL20RB	Q6UXL0	Interleukin-20 receptor subunit beta
STC1	P52823	Stanniocalcin-1
MYOC	Q99972	Myocilin
WFDC11	Q8NEX6	Protein WFDC11
MAP2K4	P45985	Dual specificity mitogen-activated protein kinase 4
CALB1	P05937	Calbindin
CCL16	O15467	C-C motif chemokine 16
SMCO2	A6NFE2	Single-pass membrane and coiled-coil domain-containing protein 2
GPX2	P18283	Glutathione peroxidase 2
CCL23	P55773	C-C motif chemokine 23
FGFR4	P22455	Fibroblast growth factor receptor 4
OSTM1	Q86WC4	Osteopetrosis-associated transmembrane protein 1
RNASE10	Q5GAN6	Inactive ribonuclease-like protein 10
ITIH1	P19827	Inter-alpha-trypsin inhibitor heavy chain H1
ZNF134	P52741	Zinc finger protein 134
CFAP45	Q9UL16	Cilia- and flagella-associated protein 45
SFTPD	P35247	Pulmonary surfactant-associated protein D

[0058] In some embodiments, the biomarkers are present at different levels in individuals with impaired glucose tolerance compared to individuals with normal glucose tolerance.

[0059] Detection of the differential levels of a biomarker in an individual can be used, for example, to permit the determination of whether an individual has or likely has impaired glucose tolerance, or whether an individual with impaired glucose tolerance is pre-diabetic or is likely to develop pre-diabetes. In some embodiments, any of the biomarkers described herein may be used to monitor individuals for development of impaired glucose tolerance, or to monitor individuals with impaired glucose tolerance for development of pre-diabetes or diabetes.

[0060] As an example of the manner in which any of the biomarkers described herein can be used to determine whether a subject has or likely has impaired glucose tolerance, levels of one or more of the described biomarkers in an individual who has not been diagnosed with impaired glucose tolerance, but has one or more impaired glucose tolerance risk factors or comorbidities, may indicate that the individual has developed impaired glucose tolerance at an earlier stage than would be determined using a different test. By detecting impaired glucose tolerance at an earlier stage, medical intervention may be more effective. Such medical intervention may include, but is not limited to, weight loss and blood sugar control. In some embodiments, therapeutic agents may be used, such as insulin or metformin.

[0061] Similarly, as a further example of the manner in which the biomarkers described herein can be used to determine whether a subject that has impaired glucose tolerance is developing pre-diabetes or diabetes, levels of one or more of the described biomarkers in an individual with impaired glucose tolerance may indicate that the individual is developing pre-diabetes or diabetes. By detecting pre-diabetes or diabetes at an earlier stage, medical intervention may be more effective. Such medical intervention may include, but is not limited to, weight loss and blood sugar control. In some embodiments, therapeutic agents may be used, such as insulin or metformin.

[0062] In addition, in some embodiments, a differential expression level of one or more of the biomarkers in an individual over time may be indicative of the individual's response to a particular therapeutic regimen. In some embodiments, changes in expression of one or more of the biomarkers during follow-up monitoring may indicate that a particular therapy is effective or may suggest that the therapeutic regimen should be altered in some way, such as by more aggressively controlling blood sugar, more aggressively pursuing weight loss, etc. In some embodiments, a constant expression level of one or more of the biomarkers in an individual over time may be indicative that an individual's impaired glucose tolerance is not worsening, or is not developing into pre-diabetes or diabetes.

[0063] In addition to testing biomarker levels as a stand-alone diagnostic test, biomarker levels can also be done in conjunction with determination of single nucleotide polymorphisms (SNPs) or other genetic lesions or variability that are indicative of increased risk of susceptibility of disease.

[0064] In addition to testing biomarker levels as a stand-alone diagnostic test, biomarker levels can also be done in conjunction with other impaired glucose tolerance screening methods. In some instances, methods using the biomarkers described herein may facilitate the medical and economic justification for implementing more aggressive treatments for impaired glucose tolerance or pre-diabetes or diabetes, more frequent follow-up screening, etc. The biomarkers may also be used to begin treatment in individuals at risk of developing impaired glucose tolerance, but who have not been diagnosed with impaired glucose tolerance, if the diagnostic test indicates they are likely to develop the disease.

[0065] In addition to testing biomarker levels in conjunction with other impaired glucose tolerance diagnostic methods, information regarding the biomarkers can also be evaluated in conjunction with other types of data, particularly data that indicates an individual's risk for impaired glucose tolerance. 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.

Detection and Determination of Biomarkers and Biomarker Levels

[0066] A biomarker level for the biomarkers described herein can be detected using any of a variety of known analytical methods. In one embodiment, a biomarker level is detected using a capture reagent. 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, adnectins, ankyrins, other antibody mimetics and other protein scaffolds, autoantibodies, chimeras, small molecules, F(ab')₂ fragments, single chain antibody fragments, Fv fragments, single chain Fv fragments, nucleic acids, lectins, ligand-binding receptors, affybodies, nanobodies, imprinted polymers, avimers, peptidomimetics, hormone receptors, cytokine receptors, and synthetic receptors, and modifications and fragments of these.

[0067] In some embodiments, a biomarker level is detected using a biomarker/capture reagent complex.

[0068] In some embodiments, the biomarker level 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.

[0069] In some embodiments, the biomarker level is detected directly from the biomarker in a biological sample.

[0070] In some embodiments, biomarkers are detected using a multiplexed format that allows for the simultaneous detection of two or more biomarkers in a biological sample. In some embodiments of the multiplexed format, capture reagents are immobilized, directly or indirectly, covalently or non-covalently, in discrete locations on a solid support. In some embodiments, 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 some embodiments, 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 analyze one or more of multiple biomarkers to be detected in a biological sample.

[0071] In one or more of the foregoing embodiments, a fluorescent tag can be used to label a component of the biomarker/capture reagent complex to enable the detection of the biomarker level. 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 level. 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.

[0072] In some embodiments, 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 some embodiments, the dye molecule includes a first type and a second type of dye molecule, such as, e.g., two different AlexaFluor molecules. In some embodiments, the dye molecule includes a first type and a second type of dye molecule, and the two dye molecules have different emission spectra.

[0073] 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.

[0074] In one or more embodiments, a chemiluminescence tag can optionally be used to label a component of the biomarker/capture complex to enable the detection of a biomarker level. Suitable chemiluminescent materials include any of oxalyl chloride, Rodamin 6G, $\text{Ru}(\text{bipy})_3^{2+}$, TMAE (tetrakis(dimethylamino)ethylene), Pyrogallol (1,2,3-trihydroxybenzene), Lucigenin, peroxyoxalates, Aryl oxalates, Acridinium esters, dioxetanes, and others.

[0075] In some embodiments, the detection method includes an enzyme/substrate combination that generates a detectable signal that corresponds to the biomarker level. 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.

[0076] In some embodiments, the detection method can be a combination of fluorescence, chemiluminescence, radionuclide or enzyme/substrate combinations that generate a measurable signal. In some embodiments, multimodal signaling could have unique and advantageous characteristics in biomarker assay formats.

[0077] In some embodiments, the biomarker levels for the biomarkers described herein can be detected using any 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 discussed below.

Determination of Biomarker Levels using Aptamer-Based Assays

[0078] 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. Patent No. 5,475,096 entitled “Nucleic Acid Ligands”; see also, e.g., U.S. Patent No. 6,242,246, U.S. Patent No. 6,458,543, and U.S. Patent 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 level corresponding to a biomarker.

[0079] 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.

[0080] 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.

[0081] 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.

[0082] 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. Patent 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. Patent No. 5,705,337 entitled “Systematic Evolution of Nucleic Acid Ligands by Exponential Enrichment: Chemi-SELEX.”

[0083] 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. Patent 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. Patent 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 No.

2009/0098549, entitled “SELEX and PHOTOSELEX”, which describes nucleic acid libraries having expanded physical and chemical properties and their use in SELEX and photoSELEX.

[0084] SELEX can also be used to identify aptamers that have desirable off-rate characteristics. *See* U.S. Publication No. US 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. Nonlimiting exemplary modified nucleotides include, for example, the modified pyrimidines shown in Figure 2. In some embodiments, an aptamer comprises at least one nucleotide with a modification, such as a base modification. In some embodiments, an aptamer comprises at least one nucleotide with a hydrophobic modification, such as a hydrophobic base modification, allowing for hydrophobic contacts with a target protein. Such hydrophobic contacts, in some embodiments, contribute to greater affinity and/or slower off-rate binding by the aptamer. Nonlimiting exemplary nucleotides with hydrophobic modifications are shown in Figure 2. In some embodiments, an aptamer comprises at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least 10 nucleotides with hydrophobic modifications, where each hydrophobic modification may be the same or different from the others. In some embodiments, at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least 10 hydrophobic modifications in an aptamer may be independently selected from the hydrophobic modifications shown in Figure 2.

[0085] In some embodiments, a slow off-rate aptamer (including an aptamers comprising at least one nucleotide with a hydrophobic modification) has an off-rate ($t_{1/2}$) of ≥ 30 minutes, ≥ 60 minutes, ≥ 90 minutes, ≥ 120 minutes, ≥ 150 minutes, ≥ 180 minutes, ≥ 210 minutes, or ≥ 240 minutes.

[0086] In some embodiments, an 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. Patent No. 6,544,776 entitled “Nucleic Acid Ligand Diagnostic Biochip”. These photoreactive aptamers are also referred to as photoaptamers. *See, e.g.,* U.S. Patent No. 5,763,177, U.S. Patent No. 6,001,577, and U.S. Patent 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. Patent 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 level corresponding to a biomarker in the test sample.

[0087] In some 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 affect 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.

[0088] 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. Publication No. 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.

[0089] 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.

[0090] 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. In some embodiments of the methods described herein, the molecular capture reagents comprise an aptamer or an antibody or the like and the specific target may be a biomarker shown in Table 1.

[0091] In some embodiments, 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.

[0092] An exemplary solution-based aptamer assay that can be used to detect a biomarker level 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.

[0093] A nonlimiting exemplary method of detecting biomarkers in a biological sample using aptamers is described in Example 3. See also Kraemer et al., *PLoS One* 6(10): e26332.

Determination of Biomarker Levels using Immunoassays

[0094] Immunoassay methods are based on the reaction of an antibody to its corresponding target or analyte, such as a biomarker protein, 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 and fragments thereof 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.

[0095] 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 level corresponding to the target in the unknown sample is established.

[0096] 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 (I^{125}) or fluorescence. Additional techniques include, for example, agglutination, nephelometry, turbidimetry, Western blot, immunoprecipitation, immunocytochemistry, immunohistochemistry, flow cytometry, Luminex assay, and others (see *ImmunoAssay: A Practical Guide*, edited by Brian Law, published by Taylor & Francis, Ltd., 2005 edition).

[0097] 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.

[0098] Methods of detecting and/or for 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.

[0099] 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 386 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 Levels using Gene Expression Profiling

[00100] Measuring mRNA in a biological sample may, in some embodiments, be used as a surrogate for detection of the level of the corresponding protein in the biological sample. Thus, in some embodiments, a biomarker or biomarker panel described herein can be detected by detecting the appropriate RNA.

[00101] In some embodiments, 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.

Detection of Biomarkers Using *In Vivo* Molecular Imaging Technologies

[00102] In some embodiments, a biomarker described herein may be used in molecular imaging tests. For example, an imaging agent can be coupled to a capture reagent, which can be used to detect the biomarker *in vivo*.

[00103] *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 biomarker *in vivo*.

[00104] The use of *in vivo* molecular imaging technologies is expanding due to various advances in technology. These advances include the development of new contrast agents or labels, such as radiolabels and/or fluorescent labels, which can provide strong signals within the body; and the development of powerful new imaging technology, which can detect and analyze these signals from outside the body, with sufficient sensitivity and accuracy to provide useful information. The contrast agent can be visualized in an appropriate imaging system, thereby providing an image of the portion or portions of the body in which the contrast agent is located. The contrast agent may be bound to or associated with a capture reagent, such as an aptamer or an antibody, for example, and/or with a peptide or protein, or an oligonucleotide (for example, for the detection of gene expression), or a complex containing any of these with one or more macromolecules and/or other particulate forms.

[00105] 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.

[00106] 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.

[00107] 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.

[00108] 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.

[00109] 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. 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 herein may be appropriately labeled and injected into an individual to detect the biomarker *in vivo*. The label used will be selected in accordance with the imaging modality to be used, as previously described. 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.

[00110] 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.

[00111] 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.

[00112] For a review of other techniques, see N. Blow, Nature Methods, 6, 465-469, 2009.

Determination of Biomarkers using Histology/Cytology Methods

[00113] In some embodiments, the biomarkers described herein may be detected in a variety of tissue samples using histological or cytological methods. 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, and sputum, can be used for cytology. Any of the biomarkers identified herein can be used to stain a specimen as an indication of disease.

[00114] In some embodiments, one or more capture reagent/s specific to the corresponding biomarker/s are used in a cytological evaluation of a 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.

[00115] In some embodiments, one or more capture reagent/s specific to the corresponding biomarkers are used in a histological evaluation of a 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.

[00116] 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.

[00117] 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.

[00118] 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 aptamers for the staining of the prepared cells.

Determination of Biomarker Levels using Mass Spectrometry Methods

[00119] A variety of configurations of mass spectrometers can be used to detect biomarker levels. 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)).

[00120] Protein biomarkers and biomarker levels can be detected and measured by any of the following: electrospray ionization mass spectrometry (ESI-MS), ESI-MS/MS, ESI-MS/(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.

[00121] Sample preparation strategies are used to label and enrich samples before mass spectroscopic characterization of protein biomarkers and determination biomarker levels. Labeling methods include but are not limited to isobaric tag for relative and absolute quantitation (iTRAQ) and stable isotope labeling with amino acids in cell culture (SILAC). Capture reagents used to selectively enrich samples for candidate biomarker proteins prior to mass spectroscopic analysis include but are not limited to aptamers, antibodies, nucleic acid

probes, chimeras, small molecules, an F(ab')₂ fragment, a single chain antibody fragment, an Fv fragment, a single chain Fv fragment, a nucleic acid, a lectin, a ligand-binding receptor, affybodies, nanobodies, ankyrins, domain antibodies, alternative antibody scaffolds (e.g. diabodies etc) imprinted polymers, avimers, peptidomimetics, peptoids, peptide nucleic acids, threose nucleic acid, a hormone receptor, a cytokine receptor, and synthetic receptors, and modifications and fragments of these.

[00122] The foregoing assays enable the detection of biomarker levels that are useful in the methods described herein, where the methods comprise detecting, in a biological sample from an individual, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine biomarkers selected from the biomarkers in Table 1. In accordance with any of the methods described herein, biomarker levels 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

[00123] In some embodiments, a biomarker “signature” for a given diagnostic test contains a set of biomarkers, each biomarker having characteristic levels in the populations of interest. Characteristic levels, in some embodiments, may refer to the mean or average of the biomarker levels for the individuals in a particular group. In some embodiments, a diagnostic method described herein can be used to assign an unknown sample from an individual into one of two groups, either having impaired glucose tolerance or normal glucose tolerance. In some embodiments, a diagnostic method described herein can be used to assign an unknown sample from an individual into one of two groups, either normal or impaired glucose tolerance. In some embodiments, a diagnostic method described herein can be used to assign an unknown sample from an individual into one of three groups: normal glucose tolerance, impaired glucose tolerance without pre-diabetes or diabetes, and pre-diabetes or diabetes.

[00124] 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 levels. In some instances, classification methods are performed using supervised learning techniques in which 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.

[00125] Common approaches for developing diagnostic classifiers include decision trees; bagging + boosting + 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.

[00126] 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. Training a naïve Bayesian classifier is an example of such a supervised learning technique (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). Training of a naïve Bayesian classifier is described, e.g., in U.S. Publication Nos: 2012/0101002 and 2012/0077695.

[00127] Since typically there are many more potential biomarker levels 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 biomarkers used in developing the classifier, by assuming that the biomarker 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.

[00128] 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 biomarkers 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.

[00129] 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). The addition of subsequent biomarkers with good KS distances (>0.3 , for example) will, in general, improve the classification performance if the subsequently added biomarkers are independent of the first biomarker. Using the sensitivity plus specificity as a classifier score, many high scoring classifiers can be generated 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.)

[00130] Another way to depict classifier performance is through a receiver operating characteristic (ROC), or simply ROC curve or ROC plot. The ROC is a graphical plot of the sensitivity, or true positive rate, vs. false positive rate ($1 - \text{specificity}$ or $1 - \text{true negative rate}$), for a binary classifier system as its discrimination threshold is varied. The ROC can also be represented equivalently by plotting the fraction of true positives out of the positives (TPR = true positive rate) vs. the fraction of false positives out of the negatives (FPR = false positive rate). Also known as a Relative Operating Characteristic curve, because it is a comparison of two operating characteristics (TPR & FPR) as the criterion changes. The area under the ROC curve (AUC) is commonly used as a summary measure of diagnostic accuracy. It can take values from 0.0 to 1.0. The AUC has an important statistical property: the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance (Fawcett T, 2006. An introduction to ROC analysis. Pattern Recognition Letters .27: 861–874). This is equivalent to the Wilcoxon test of ranks (Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.).

[00131] Exemplary embodiments use any number of the biomarkers listed in Table 1, in various combinations to produce diagnostic tests for identifying individuals with impaired glucose tolerance. The biomarkers listed in Table 1 can be combined in many ways to produce

classifiers. In some embodiments, panels of biomarkers are comprised of different sets of biomarkers depending on a specific diagnostic performance criterion that is selected. For example, certain combinations of biomarkers may produce tests that are more sensitive (or more specific) than other combinations.

[00132] In some embodiments, once a panel is defined to include a particular set of biomarkers and a classifier is constructed from a set of training data, the diagnostic test parameters are complete. In some embodiments, a biological sample is 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.

[00133] In some embodiments, a sample is optionally diluted and run in a multiplexed aptamer assay, and data is assessed as follows. First, the data from the assay are optionally normalized and calibrated, and the resulting biomarker levels are used as input to a Bayes classification scheme. Second, the log-likelihood 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. In some embodiments, the individual log-likelihood risk factors computed for each biomarker level can be reported as well.

Kits

[00134] Any combination of the biomarkers described herein 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.

[00135] In some embodiments, 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, and optionally (b) one or more software or computer program products for predicting whether the individual from whom the biological sample was obtained has or likely has impaired glucose tolerance or has pre-diabetes, or is likely to develop pre-diabetes or diabetes.. 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.

[00136] In some embodiments, a kit comprises a solid support, capture reagents, and at least one signal generating material. 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.

[00137] The kits can also contain one or more reagents (e.g., solubilization buffers, detergents, washes, or buffers) for processing a 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.

[00138] In some embodiments, kits are provided for the analysis of impaired glucose tolerance, wherein the kits comprise PCR primers for one or more biomarkers described herein. In some embodiments, a kit may further include instructions for use and correlation of the biomarkers with impaired glucose tolerance and/or pre-diabetes or diabetes prognosis. In some embodiments, a kit may include a DNA array containing the complement of one or more of the biomarkers described herein, 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.

[00139] For example, a kit can comprise (a) reagents comprising at least one capture reagent for determining the level of one or more biomarkers in a sample, and optionally (b) one or more algorithms or computer programs for performing the steps of comparing the amount of each biomarker quantified in the sample to one or more predetermined cutoffs. In some embodiments, an algorithm or computer program assigns a score for each biomarker quantified based on said comparison and, in some embodiments, combines the assigned scores for each biomarker quantified to obtain a total score. Further, in some embodiments, an algorithm or computer program compares the total score with a predetermined score, and uses the comparison to determine whether the individual has impaired glucose tolerance. 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

[00140] Once a biomarker or biomarker panel is selected, a method for assessing whether a subject has or likely has impaired glucose tolerance, or has pre-diabetes, or is likely to develop pre-diabetes or diabetes may comprise the following: 1) collect or otherwise obtain a biological sample from the subject; 2) perform an analytical method to detect and measure the biomarkers in the panel in the biological sample; and 3) report the results of the biomarker levels. In some embodiments, the results of the biomarker levels are reported qualitatively rather than quantitatively, such as, for example, a proposed diagnosis (“impaired glucose tolerance,” or

“pre-diabetes,” or simply a positive / negative result where “positive” and “negative” are defined. In some embodiments, a method for assessing impaired glucose tolerance in an individual may comprise the following: 1) collect or otherwise obtain a biological sample; 2) perform an analytical method to detect and measure the biomarkers in the panel in the biological sample; 3) perform any data normalization or standardization; 4) calculate each biomarker level; and 5) report the results of the biomarker levels. In some embodiments, the biomarker levels are combined in some way and a single value for the combined biomarker levels is reported. In this approach, in some embodiments, the reported value may be a single number determined from the sum of all the biomarker calculations that is compared to a pre-set threshold value that is an indication of the presence or absence of a condition. 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 a condition.

[00141] 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 Figure 3. With reference to Figure 3, 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.

[00142] With respect to Figure 3, 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.

[00143] In one aspect, the system can comprise a database containing features of biomarkers characteristic of impaired glucose tolerance and/or pre-diabetes. 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.

[00144] In one aspect, the system further comprises one or more devices for providing input data to the one or more processors.

[00145] The system further comprises a memory for storing a data set of ranked data elements.

[00146] 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.

[00147] 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.

[00148] 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.

[00149] 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.

[00150] 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.

[00151] 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.

[00152] 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.

[00153] The methods and apparatus for analyzing 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.

[00154] The 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 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 disease status and/or diagnosis. Identifying impaired glucose tolerance, pre-diabetes, and/or probable diabetes 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.

[00155] 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.

[00156] 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.

[00157] In one aspect, a computer program product is provided for indicating whether an individual has impaired glucose tolerance, and/or whether an individual has or is likely to develop pre-diabetes, and/or is likely to develop diabetes. 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 levels that correspond to one or more of the biomarkers described herein, and code that executes a classification method that indicates the impaired glucose tolerance status of the individual as a function of the biomarker levels.

[00158] 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, programmable logic arrays (PLAs), or application-specific integrated circuits (ASICs).

[00159] 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.

Methods of Treatment

[00160] In some embodiments, following a determination that a subject has or likely has impaired glucose tolerance, or has pre-diabetes, or is likely to develop pre-diabetes or diabetes, the subject undergoes a therapeutic regimen to delay or prevent worsening of the disease. Nonlimiting exemplary therapeutic regimens for impaired glucose tolerance, pre-diabetes, and/or probable diabetes include weight loss and blood sugar control. In some embodiments, a subject is given a therapeutic agent, such as insulin or metformin.

[00161] In some embodiments, methods of monitoring impaired glucose tolerance are provided. In some embodiments, the present methods of determining whether a subject has impaired glucose tolerance are carried out at a time 0. In some embodiments, the method is carried out again at a time 1, and optionally, a time 2, and optionally, a time 3, etc., in order to monitor the progression of the impaired glucose tolerance in the subject. In some embodiments, different biomarkers are used at different time points, depending on the current state of the individual's disease and/or depending on the rate at which the disease is believed or predicted to progress.

Other Methods

[00162] In some embodiments, the biomarkers and methods described herein are used to determine a medical insurance premium and/or a life insurance premium. In some embodiments, the results of the methods described herein are used to determine a medical insurance premium and/or a life insurance premium. In some such instances, an organization that provides medical insurance or life insurance requests or otherwise obtains information concerning a subject's impaired glucose tolerance or pre-diabetes or likelihood of developing pre-diabetes or diabetes status and uses that information to determine an appropriate medical insurance or life insurance premium for the subject. In some embodiments, the test is requested by, and paid for by, the organization that provides medical insurance or life insurance.

[00163] In some embodiments, the biomarkers and methods described herein are used to predict and/or manage the utilization of medical resources. In some such embodiments, the methods are not carried out for the purpose of such prediction, but the information obtained from the method is used in such a prediction and/or management of the utilization of medical resources. For example, a testing facility or hospital may assemble information from the present

methods for many subjects in order to predict and/or manage the utilization of medical resources at a particular facility or in a particular geographic area.

EXAMPLES

[00164] 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. 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. Multiplex Aptamer Assay and Statistical Approaches for Biomarker Identification

[00165] A multiplex aptamer assay was used to analyze test samples and control samples to identify biomarkers predictive of impaired glucose tolerance. The multiplexed analysis used in this experiment included aptamers to detect approximately 5,000 proteins in blood from small sample volumes (~65 μ l of serum or plasma), with low limits of detection (1 pM median), ~7 logs of dynamic range, and ~5% median coefficient of variation. The multiplex aptamer assay is described, generally, e.g., in Gold et al. (2010) *Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery*. PLoS ONE 5(12): e15004; and U.S. Publication Nos: 2012/0101002 and 2012/0077695.

Impaired Glucose Tolerance Classifier

[00166] A panel of forty-one biomarkers chosen by stability selection are shown in Table 1 and were supplied to a random forest algorithm to generate a model. The model applied to the biomarkers herein is a classification model, and more specifically an elastic net logistic regression model.

[00167] The β_{hat} values for each biomarker are shown in Table 2 below. The β_{hat} values show the relative change in biomarker level in a test sample, relative to a control sample, that indicates that the test sample was obtained from a subject that has impaired glucose tolerance.

Table 2: β_{hat} Values of Certain Biomarkers

Biomarker	β_{hat}
ACY1	-0.3312025
CALB1	0.09414296
CBX7	-0.2496155
CCL16	-0.0916518
CCL23	-0.0823261
CFAP45	0.0371442

CHAD	0.19457556
COL15A1	0.243721
COL1A1	0.30089257
CPM	-0.1903201
CRLF1:CLCF1 complex	0.25463585
DLK1	-0.1543516
F9	-0.1638971
FABP12	-0.1992025
FAM20B	0.24420748
FGFR4	0.07566835
GAD1	-0.1973748
GPX2	-0.0828715
HERC1	-0.1413612
HTRA1	-0.2101563
IL20RB	0.13331569
INHBC	-0.3886077
ITIH1	0.04455428
KIN	0.23433508
MAP2K4	0.09578616
MARCKSL1	0.21269464
MYOC	0.12263523
OSTM1	0.07294053
PELI2	0.22836152
RNASE10	-0.0558454
RTN4R	-0.2672434
SERPINA11	0.22982337
SFTPD	0.00588235
SHBG	0.38576004
SMCO2	-0.0838454
SOCS7	-0.1756156
STC1	-0.1320602
SVEP1	-0.1857574
TFF3	-0.2216243
WFDC11	0.1132431
ZNF134	-0.0436936

Development Cohort and Model Development

[00168] The development cohort was from [EGC1] a population-based study in over 12,000 male and female participants (aged 29 – 64 years) from the UK. The objective of the study was to identify the genetic and lifestyle risk factors that lead to diabetes, obesity, and related health conditions in a general population. Participants with clinically diagnosed diabetes, clinically diagnosed psychotic illness, terminal illness, pregnancy, or who were unable to walk unaided were excluded from this study. The cohort included sample measurements from the participants using a multiplex assay described herein from Phase 1 (baseline) and Phase 2 (~6-year post-baseline follow-up visit) samples taken at four study enrollment sites.

[00169] Sample-handling protocols changed between the two measurements. Due to this change, Phase 1 showed a wider distribution by site in processing times than Phase 2. During the Phase 1 (baseline) visit, priority was given to conducting the study participant testing (OGTT, DEXA, treadmill, etc.) in an efficient manner, leading to a range in time for blood sample processing from immediately after to a few hours after sample collection. For the Phase 2 visit, additional staff were provided, and all samples were processed promptly at the time of collection. Thus, the second timepoint (Phase 2) has more consistent sample processing across study sites than the first timepoint (Phase 1). To account for the protocol change, and to increase model robustness across longitudinal data, the datasets used for model development, verification, and validation included measurements from Phase 1 and Phase 2. These data included 7,116 Phase 1 samples (from participants that only attended Phase 1) and 5,003 Phase 2 samples (from participants that attended Phase 1 and 2). Only one measurement from each participant with a valid 2h-OGTT plasma glucose measurement was used to maintain the assumption of independence across samples.

[00170] In this data set, the prevalence of individuals with impaired glucose tolerance, as measured by a standard OGTT, was 6.5%, as shown in Table 3. Estimates of impaired glucose tolerance, or prediabetes, across the greater UK during the timeframe of this study were 10.5% (See <https://www.diabetes.co.uk/pre-diabetes.html>). In the United States, current estimates are higher, with a predicted prevalence of 33.9% in adults 18 and older (See <https://www.niddk.nih.gov/health-information/health-statistics/diabetes-statistics>). In a higher prevalence population such as the US, the PPV (positive predictive value) for this test result is likely to improve, and the NPV (negative predictive value) is likely to drop slightly compared to test performance in the low prevalence population. There is a slight risk of underprediction in the US prevalence population which is acceptable in the current target market of self-pay, concierge medicine.

[00171] The cohort data set was split 70%/15%/15% into training, verification and validation data sets, respectively. Training and verification data set demographics, as measured by a standard OGTT, are listed in Table 3.

Table 3

Dataset	Covariate	Measure	Total	Impaired Glucose Tolerance	Normal Glucose Tolerance
Training	Sample Size	N	8,224	532 (6.5%)	7,692 (93.5%)
	Age	Mean (SD)	48.1 (7.5)	51.0 (6.9)	47.9 (7.5)
		Median	48.0	52.0	48.0

		Range	(29-64)	(30-63)	(29-64)
	Sex	Male	3,774 (45.9%)	285 (3.5%)	3,489 (42.4%)
		Female	4,395 (53.4%)	244 (3.0%)	4,151 (50.5%)
Verification	Sample Size	N	1,762	114 (6.5%)	1,648 (93.5%)
	Age	Mean (SD)	48.2 (7.6)	51.5 (6.3)	48.0 (7.6)
		Median	48.0	52.5	48
		Range	(30-64)	(37-62)	(30-64)
	Sex	Male	785 (44.6%)	53 (3.0%)	732 (41.5%)
		Female	963 (54.7%)	61 (3.5%)	902 (51.2%)

[00172] When developing predictive models using machine learning techniques, multiple data sets should be used to identify the model that has the best predictive capabilities. To this end, the following strategy for splitting the data was used. The data was split into three sets: a training set (used for identifying top models through cross-validation), a verification set (a second training set that allows us to tune the parameters of the top models), and a validation test set (a hold-out set that is only used to assess the final model and is not used for model development). Splitting the data three ways is the most ideal and requires a large sample size and thus is not as commonly employed nor is it considered required for model development. This approach mitigates issues with overfitting when performing feature selection and parameter estimation.

[00173] The cohort included sample measurements using a multiplex assay as described herein at two timepoints, with a change in sample-handling protocols between the two measurements, therefore, the datasets used for model development, verification, and validation included measurements from both timepoints. The cohort selected for this analysis included 7,116 samples from individuals who only had samples at Phase 1, plus only the Phase 2 data from the 5,003 individuals with data at both timepoints.

[00174] To ensure quality of the data, four pre-processing steps were performed before the data were analyzed:

1. Normalization via ANML: Adaptive normalization by maximum likelihood (ANML) was used to correct for dilution specific sample and assay biases including pipetting errors, changes

in reagent concentration, assay timing, and other sources of systematic variability. Scale factors were calculated by maximizing the probability a sample's measurement came from the reference distribution (control sample set)). Analytes that exceeded a Z-score of $|2|$ relative to the reference distribution were dropped from these computations to mitigate the bias of sample-handling artifacts or other large proteomic changes.

2. Data quality control (QC): This step checked sample handling and normalization issues. Sample data was first normalized to remove hybridization variation within a run. This was followed by median normalization across calibrator samples to remove other assay biases within the run. Overall scaling was then performed on a per plate basis to remove overall intensity differences between runs. Calibration was then performed to remove assay differences between runs. Finally, median normalization to a reference was performed on the QC, buffer and individual samples.

3. Pre-analytics: In this step, the relationship between clinical variables and normalization scale factors was investigated to ensure that there is minimal correlation between the two.

[00175] 4. Missing data: There were no missing data points that required subject or sample removal before model development. The model was developed to be dichotomous: a patient has normal glucose tolerance (corresponding to an OGTT glucose measurement of <7.8 mmol/L) or impaired glucose tolerance (corresponding to an OGTT glucose measurement of ≥ 7.8 mmol/L). The response variable was calculated within applied modeling BI.

[00176] After data quality control and pre-analytics, model development was completed in two steps, listed as follows:

1. Proof of Concept (POC): A univariate and machine learning analyses designed to understand if there is any evidence of signal for the endpoint of interest.

2. Refinement: Modeler-directed analysis that confirms and expands models produced in POC; addresses any other additional concerns about the data and model.

[00177] Only the training data was used in the POC step. Univariate analyses using t-tests, KS test, Mann-Whitney, and Wald tests on the logistic regression coefficient were performed to determine if there was a statistically significant association between any analytes and OGTT status. For each univariate test, multiple testing was corrected for with the false discovery rate (FDR), calculated using the Benjamini-Hochberg procedure (Hochberg, et. al), and the Bonferroni-corrected p-value. Preliminary elastic net logistic regression models were also created to assess if minimal performance metrics were met. The models were developed using 5 repeats of 10-fold cross-validation. Because of the class imbalance (only 6.5% of the training data were labeled "glucose impaired"), down sampling within cross-validation was employed. Initial model performance criteria were met, providing sufficient evidence to

move the test into model refinement.

[00178] Models developed in refinement used the cohort training and verification data sets. Initial models were calculated using 5 repeats of 10-fold cross-validation on the training data, using down sampling within cross-validation to accommodate the class imbalance in the endpoint. Top models selected using the accuracy of the model, where $\text{Accuracy} = (\text{True positive} + \text{true negative})/n$. This metric was used because it represents a balance between sensitivity and specificity (unlike the AUC, which can be very high even for models with low sensitivity or low specificity.) Accuracy was used during refinement as an analysis tool but the acceptance criteria for the model remained the combination of AUC, sensitivity, and specificity. These top models were then further refined using the verification data and the various model hardening tools.

[00179] The main approach used for model building included repeated rounds of elastic-net penalization models where features were filtered repeatedly until model performance did not improve.

[00180] The algorithm for the repeated elastic-net feature reduction modeling was as follows:

1. Build an elastic-net model that filters features based on rank.
2. From the top model created in Step 1, retain all features that have estimates not equal to zero.
3. Build a new elastic model that includes all features retained in Step 2.
4. Repeat Steps 2 and 3 at least ten times, and then continue until the accuracy no longer increases.

[00181] This process was performed multiple times, with features filtered in a variety of ways, including: by univariate rank (including the top 100, 200, or 500 features); removal of features statistically significantly associated ($\text{FDR} \leq 0.01$) with fasting status; removal of features associated with interference testing failure; and removal of features based on variability in external data sets used for modeling hardening.

[00182] This final model was also assessed for robustness by examining different synthetic data sets, checking impacts of imputation on minimum and maximum values, and assessment of sample handling effects on model predictions.

[00183] The predictive performance of the best model was then examined at both Phase 1 and Phase 2 to ensure no significant differences were observed between the two time points. Additionally, the residuals of the models were examined for correlation with age, sex, or visit.

[00184] Data QC showed that 219 samples failed row-check, meaning at least one of the hybridization or three median scale factors were outside the 0.4 to 2.5 range, indicating technical issues (e.g., clogs) with that particular sample that would not be fixed by running the sample

again. Additionally, there were 18 outlier samples with at least 5% of measurements more than 6 MADs from the median signal, and two samples with large normalization scale factors. These 239 samples (1.4%) were removed from further analyses. Finally, only analytes that passed target confirmation specificity testing were used for analyses.

[00185] A PCA plot of PC1 vs PC2 also showed a possible non-linear relationship between the two principal components, however, these did not appear to be significantly different based on the collection time point (Phase 1 vs. Phase 2) so were not of huge concern. The comparison of the two time points was important because the sample collection protocol changed between Phase 1 and Phase 2, so it was important to establish that the protocol change was not the largest source of variation in assay signal.

[00186] Pre-analytics did not show evidence of strong relationships between any of the endpoints and the normalization scale factors at the first or second time points.

[00187] The POC results showed a number of analytes significant at different FDR levels. Those numbers and percentages are shown in Table 4 for the univariate t-tests.

Table 4

False Discovery Rate (FDR)	Number of analytes (%) less than or equal to FDR level
0.10	1,895
0.05	1,548
0.01	1,122

[00188] The best performing model was an elastic net logistic regression model with an AUC of 0.856 and a sensitivity/specificity of 0.78/0.77, which exceeded the feasibility criterion of an AUC, sensitivity, and specificity ≥ 0.70 .

[00189] Elastic net logistic regression models were used in refinement, because this type of model showed the most success in the POC stage and it reduces the burden placed on the model transfer process to software. The performance of the model on the training and verification data is shown in Table 5. The 95% bootstrapped intervals are shown in parentheses in Table 5.

Table 5

Dataset	AUC	Sensitivity	Specificity	Accuracy
Training	0.794 (0.778,0.811)	0.825 (0.791, 0.859)	0.764 (0.755, 0.772)	0.768 (0.758, 0.777)
Verification	0.760 (0.726, 0.797)	0.772 (0.705, 0.839)	0.749 (0.728, 0.770)	0.750 (0.729, 0.770)

[00190] The effects of imputation on out-of-range values was examined in two stages.

1. Using the training data, the minimum and maximum acceptable RFU values for each analyte were calculated as follows:
 - a. Using the training data, generate a bootstrapped data set (of the same sample size as the training data).
 - b. Split the data into k-folds, where $k = \max \{10, \text{sample size}/10\}$.
 - c. Within each fold and for each aptamer, calculate the minimum and maximum value in the data set and store them.
 - d. Repeat steps A-C 100 times.
 - e. Across the $k*100$ repeats, for each aptamer, calculate the standard deviation.
 - f. The aptamer-specific minimum/maximum RFU values are the minimum/maximum training data RFU values $\pm 2SD$ (where the SD is from step E).
2. The following steps were performed using the minimum and maximum values calculated in step 1 and the verification data:
 - a. Among the $n*p$ observations in the verification data set (where n is the sample size of the verification data and p denotes the number of aptamers in the model), randomly sample 1.5% of the RFU values to be out-of-range. (The 1.5% is based on empirical observations on previous assay performance and RFU measurements.) These row- and column-specific values are tagged.
 - b. Replace the tagged values in the verification data with the maximum and minimum from 1(F). Calculate the predictive accuracy (as shown in the equation in Section 5.4) of the final model using this data set.
 - c. Replace the tagged values in the verification data with 0 (which is akin to removal of the aptamer from the model). Calculate the predictive accuracy (using the same method as step B) of the final model using this data set.
 - d. Repeat steps A-E 100 times. Summarize the predictive accuracy metrics. The imputation method that provides the highest predictive accuracy metrics is the one that is recommended for production.

[00191] Out of range aptamers should be Winsorized, as the resulting predictive metrics were the same (see Table 6) and this process is most commonly used.

Table 6: Imputation results for out-of-range aptamers

Imputation Method	Accuracy
Original verification data	0.75
Winsorized verification data	0.75

Analyte removal in verification data	0.75
--------------------------------------	------

Example 2: Model Validation

[00192] The final model as assessed on the 15% hold-out validation (test) set. This data was stored in a separate folder in the test repository and was only examined for the purposes of making the validation data demographics table (see Table 7).

[00193] The model predictions are probabilities that denote the probability the individual has impaired glucose tolerance. The cut-off threshold is 0.5, with individuals who have probabilities higher than the cut-off classified as “impaired glucose tolerance.” Values closer to “1” denote subjects with greatest likelihood of impaired glucose tolerance.

[00194] In the validation phase, predicted probabilities and their associated classifiers were calculated on the validation data. The AUC, sensitivity, and specificity each should be and, in fact, were greater than or equal to 0.70 (see Table 8).

[00195] The validation dataset consisted of the final 15% of the cohort dataset with 1,761 patients in total. The demographics of this dataset (Table 7) were qualitatively the same as the training and verification sets (Table 3).

Table 7

Dataset	Covariate	Measure	Total	Impaired Glucose Tolerance	Normal Glucose Tolerance
Validation	Sample Size	N	1,761	107 (6.1%)	1,654 (93.9%)
	Age	Mean (SD)	47.8 (7.6)	50.3 (7.6)	47.6 (7.6)
		Median	48.0	52.0	48.0
		Range	(29-64)	(34-63)	(29-64)
	Sex	Male	836 (47.5%)	61 (3.5%)	775 (44.0%)
		Female	918 (52.1%)	45 (2.6%)	873 (9.6%)

[00196] The AUC was calculated by using the final forty-one biomarker panel model to classify the final 15% holdout cohort validation data set. This data was not used in either POC or Refinement. The final model passed validation, with an AUC, sensitivity, and specificity greater than 0.70. The predictive metric results for the validation data set is shown in Table 8 below. The 95% bootstrapped intervals are shown in parentheses in Table 8.

Table 8

Dataset	AUC	Sensitivity	Specificity	Accuracy
Validation	0.764 (0.723, 0.806)	0.794 (0.723, 0.869)	0.734 (0.713, 0.754)	0.738 (0.716, 0.758)

[00197] Figure 4 shows boxplots of the probabilities of impaired glucose tolerance predicted by the forty-one biomarker panel model, stratified by diagnosis based on values from a standard oral glucose tolerance test (normal, impaired, and probable diabetes) on the validation data set. The separation between the three groups is fairly strong, indicating a robust model.

[00198] The model met the validation criteria of an AUC/Sensitivity/Specificity $\geq$ 0.7/0.7/0.7 (which is based, in part, on the 10-year incident diabetes predictive value of 2-hour OGTT plasma glucose levels). The final model has an AUC of 0.764, with a sensitivity/specificity of 0.794/0.734 on the hold-out validation set. The conclusion of this report is that the test has met the clinical acceptance criteria and can move into production.

Example 3: Exemplary Biomarker Detection Using Aptamers

[00199] An exemplary method of detecting one or more biomarkers in a sample is described, e.g., in Kraemer et al., *PLoS One* 6(10): e26332, and is described below. Three different methods of quantification: microarray-based hybridization, a Luminex bead-based method, and qPCR, are described.

Reagents

[00200] HEPES, NaCl, KCl, EDTA, EGTA, MgCl₂ and Tween-20 may be purchased, e.g., from Fisher Biosciences. Dextran sulfate sodium salt (DxSO₄), nominally 8000 molecular weight, may be purchased, e.g., from AIC and is dialyzed against deionized water for at least 20 hours with one exchange. KOD EX DNA polymerase may be purchased, e.g., from VWR. Tetramethylammonium chloride and CAPSO may be purchased, e.g., from Sigma-Aldrich and streptavidin-phycoerythrin (SAPE) may be purchased, e.g., from Moss Inc. 4-(2-Aminoethyl)-benzenesulfonylfluoride hydrochloride (AEBSF) may be purchased, e.g., from Gold Biotechnology. Streptavidin-coated 96-well plates may be purchased, e.g., from Thermo Scientific (Pierce Streptavidin Coated Plates HBC, clear, 96-well, product number 15500 or 15501). NHS-PEO4-biotin may be purchased, e.g., from Thermo Scientific (EZ-Link NHS-PEO4-Biotin, product number 21329), dissolved in anhydrous DMSO, and may be stored frozen in single-use aliquots. IL-8, MIP-4, Lipocalin-2, RANTES, MMP-7, and MMP-9 may be purchased, e.g., from R&D Systems. Resistin and MCP-1 may be purchased, e.g., from PeproTech, and tPA may be purchased, e.g., from VWR.

Nucleic acids

[00201] Conventional (including amine- and biotin-substituted) oligodeoxynucleotides may be purchased, e.g., from Integrated DNA Technologies (IDT). Z-Block is a single-stranded oligodeoxynucleotide of sequence 5'-(AC-BnBn)7-AC-3', where Bn indicates a benzyl-substituted deoxyuridine residue. Z-block may be synthesized using conventional phosphoramidite chemistry. Aptamer capture reagents may also be synthesized by conventional phosphoramidite chemistry, and may be purified, for example, on a 21.5×75 mm PRP-3 column, operating at 80°C on a Waters Autopurification 2767 system (or Waters 600 series semi-automated system), using, for example, a timberline TL-600 or TL-150 heater and a gradient of triethylammonium bicarbonate (TEAB) / ACN to elute product. Detection is performed at 260 nm and fractions are collected across the main peak prior to pooling best fractions.

Buffers

[00202] Buffer SB18 is composed of 40 mM HEPES, 101 mM NaCl, 5 mM KCl, 5 mM MgCl₂, and 0.05% (v/v) Tween 20 adjusted to pH 7.5 with NaOH. Buffer SB17 is SB18 supplemented with 1 mM trisodium EDTA. Buffer PB1 is composed of 10 mM HEPES, 101 mM NaCl, 5 mM KCl, 5 mM MgCl₂, 1 mM trisodium EDTA and 0.05% (v/v) Tween-20 adjusted to pH 7.5 with NaOH. CAPSO elution buffer consists of 100 mM CAPSO pH 10.0 and 1 M NaCl. Neutralization buffer contains of 500 mM HEPES, 500 mM HCl, and 0.05% (v/v) Tween-20. Agilent Hybridization Buffer is a proprietary formulation that is supplied as part of a kit (Oligo aCGH/ChIP-on-chip Hybridization Kit). Agilent Wash Buffer 1 is a proprietary formulation (Oligo aCGH/ChIP-on-chip Wash Buffer 1, Agilent). Agilent Wash Buffer 2 is a proprietary formulation (Oligo aCGH/ChIP-on-chip Wash Buffer 2, Agilent). TMAC hybridization solution consists of 4.5 M tetramethylammonium chloride, 6 mM trisodium EDTA, 75 mM Tris-HCl (pH 8.0), and 0.15% (v/v) Sarkosyl. KOD buffer (10-fold concentrated) consists of 1200 mM Tris-HCl, 15 mM MgSO₄, 100 mM KCl, 60 mM (NH₄)₂SO₄, 1% v/v Triton-X 100 and 1 mg/mL BSA.

Sample preparation

[00203] Serum (stored at -80 °C in 100 µL aliquots) is thawed in a 25 °C water bath for 10 minutes, then stored on ice prior to sample dilution. Samples are mixed by gentle vortexing for 8 seconds. A 6% serum sample solution is prepared by dilution into 0.94× SB17 supplemented with 0.6 mM MgCl₂, 1 mM trisodium EGTA, 0.8 mM AEBSF, and 2 µM Z-Block. A portion of the 6% serum stock solution is diluted 10-fold in SB17 to create a 0.6% serum stock. 6% and 0.6% stocks are used, in some embodiments, to detect high- and low-abundance analytes, respectively.

Capture reagent (aptamer) and streptavidin plate preparation

[00204] Aptamers are grouped into 2 mixes according to the relative abundance of their cognate analytes (or biomarkers). Stock concentrations are 4 nM for each aptamer, and the final concentration of each aptamer is 0.5 nM. Aptamer stock mixes are diluted 4-fold in SB17 buffer, heated to 95°C for 5 min and cooled to 37°C over a 15 minute period prior to use. This denaturation-renaturation cycle is intended to normalize aptamer conformer distributions and thus ensure reproducible aptamer activity in spite of variable histories. Streptavidin plates are washed twice with 150 µL buffer PB1 prior to use.

Incubation and plate capture

[00205] Heat-cooled 2× Aptamer mixes (55 µL) are combined with an equal volume of 6% or 0.6% serum dilutions, producing incubation mixes containing 3% and 0.3% serum. The plates are sealed with a Silicone Sealing Mat (AxyMat Silicone sealing mat, VWR) and incubated for 1.5 h at 37°C. Incubation mixes are then transferred to the wells of a washed 96-well streptavidin plate and further incubated on an Eppendorf Thermomixer set at 37°C, with shaking at 800 rpm, for two hours.

Manual Assay

[00206] Unless otherwise specified, liquid is removed by dumping, followed by two taps onto layered paper towels. Wash volumes are 150 µL and all shaking incubations are done on an Eppendorf Thermomixer set at 25°C, 800 rpm. Incubation mixes are removed by pipetting, and plates are washed twice for 1 minute with buffer PB1 supplemented with 1 mM dextran sulfate and 500 µM biotin, then 4 times for 15 seconds with buffer PB1. A freshly made solution of 1 mM NHS-PEO4-biotin in buffer PB1 (150 µL/well) is added, and plates are incubated for 5 minutes with shaking. The NHS-biotin solution is removed, and plates washed 3 times with buffer PB1 supplemented with 20 mM glycine, and 3 times with buffer PB1. Eighty-five µL of buffer PB1 supplemented with 1 mM D_xSO₄ is then added to each well, and plates are irradiated under a BlackRay UV lamp (nominal wavelength 365 nm) at a distance of 5 cm for 20 minutes with shaking. Samples are transferred to a fresh, washed streptavidin-coated plate, or an unused well of the existing washed streptavidin plate, combining high and low sample dilution mixtures into a single well. Samples are incubated at room temperature with shaking for 10 minutes. Unadsorbed material is removed and the plates washed 8 times for 15 seconds each with buffer PB1 supplemented with 30% glycerol. Plates are then washed once with buffer PB1. Aptamers are eluted for 5 minutes at room temperature with 100 µL CAPSO elution buffer. 90 µL of the eluate is transferred to a 96-well HybAid plate and 10 µL neutralization buffer is added.

Semi-Automated Assay

[00207] Streptavidin plates bearing adsorbed incubation mixes are placed on the deck of a BioTek EL406 plate washer, which is programmed to perform the following steps: unadsorbed material is removed by aspiration, and wells are washed 4 times with 300 μ L of buffer PB1 supplemented with 1 mM dextran sulfate and 500 μ M biotin. Wells are then washed 3 times with 300 μ L buffer PB1. One hundred fifty μ L of a freshly prepared (from a 100 mM stock in DMSO) solution of 1 mM NHS-PEO4-biotin in buffer PB1 is added. Plates are incubated for 5 minutes with shaking. Liquid is aspirated, and wells are washed 8 times with 300 μ L buffer PB1 supplemented with 10 mM glycine. One hundred μ L of buffer PB1 supplemented with 1 mM dextran sulfate are added. After these automated steps, plates are removed from the plate washer and placed on a thermoshaker mounted under a UV light source (BlackRay, nominal wavelength 365 nm) at a distance of 5 cm for 20 minutes. The thermoshaker is set at 800 rpm and 25°C. After 20 minutes irradiation, samples are manually transferred to a fresh, washed streptavidin plate (or to an unused well of the existing washed plate). High-abundance (3% serum+3% aptamer mix) and low-abundance reaction mixes (0.3% serum+0.3% aptamer mix) are combined into a single well at this point. This “Catch-2” plate is placed on the deck of BioTek EL406 plate washer, which is programmed to perform the following steps: the plate is incubated for 10 minutes with shaking. Liquid is aspirated, and wells are washed 21 times with 300 μ L buffer PB1 supplemented with 30% glycerol. Wells are washed 5 times with 300 μ L buffer PB1, and the final wash is aspirated. One hundred μ L CAPSO elution buffer are added, and aptamers are eluted for 5 minutes with shaking. Following these automated steps, the plate is then removed from the deck of the plate washer, and 90 μ L aliquots of the samples are transferred manually to the wells of a HybAid 96-well plate that contains 10 μ L neutralization buffer.

Hybridization to custom Agilent 8×15k microarrays

[00208] 24 μ L of the neutralized eluate is transferred to a new 96-well plate and 6 μ L of 10× Agilent Block (Oligo aCGH/ChIP-on-chip Hybridization Kit, Large Volume, Agilent 5188–5380), containing a set of hybridization controls composed of 10 Cy3 aptamers is added to each well. Thirty μ L 2× Agilent Hybridization buffer is added to each sample and mixed. Forty μ L of the resulting hybridization solution is manually pipetted into each “well” of the hybridization gasket slide (Hybridization Gasket Slide, 8-microarray per slide format, Agilent). Custom Agilent microarray slides, bearing 10 probes per array complementary to 40 nucleotide random region of each aptamer with a 20× dT linker, are placed onto the gasket slides according to the manufacturers' protocol. The assembly (Hybridization Chamber Kit – SureHyb-enabled, Agilent) is clamped and incubated for 19 hours at 60°C while rotating at 20 rpm.

Post Hybridization Washing

[00209] Approximately 400 mL Agilent Wash Buffer 1 is placed into each of two separate glass staining dishes. Slides (no more than two at a time) are disassembled and separated while submerged in Wash Buffer 1, then transferred to a slide rack in a second staining dish also containing Wash Buffer 1. Slides are incubated for an additional 5 minutes in Wash Buffer 1 with stirring. Slides are transferred to Wash Buffer 2 pre-equilibrated to 37°C and incubated for 5 minutes with stirring. Slides are transferred to a fourth staining dish containing acetonitrile, and incubated for 5 minutes with stirring.

Microarray Imaging

[00210] Microarray slides are imaged with an Agilent G2565CA Microarray Scanner System, using the Cy3-channel at 5 µm resolution at 100% PMT setting, and the XRD option enabled at 0.05. The resulting TIFF images are processed using Agilent feature extraction software version 10.5.1.1 with the GE1_105_Dec08 protocol.

Luminex probe design

[00211] Probes immobilized to beads have 40 deoxynucleotides complementary to the 3' end of the 40 nucleotide random region of the target aptamer. The aptamer complementary region is coupled to Luminex Microspheres through a hexaethyleneglycol (HEG) linker bearing a 5' amino terminus. Biotinylated detection deoxyoligonucleotides comprise 17–21 deoxynucleotides complementary to the 5' primer region of target aptamers. Biotin moieties are appended to the 3' ends of detection oligos.

Coupling of probes to Luminex Microspheres

[00212] Probes are coupled to Luminex Microplex Microspheres essentially per the manufacturer's instructions, but with the following modifications: amino-terminal oligonucleotide amounts are 0.08 nmol per 2.5×10^6 microspheres, and the second EDC addition is 5 µL at 10 mg/mL. Coupling reactions are performed in an Eppendorf ThermoShaker set at 25°C and 600 rpm.

Microsphere hybridization

[00213] Microsphere stock solutions (about 40000 microspheres/µL) are vortexed and sonicated in a Health Sonics ultrasonic cleaner (Model: T1.9C) for 60 seconds to suspend the microspheres. Suspended microspheres are diluted to 2000 microspheres per reaction in $1.5 \times$ TMAC hybridization solutions and mixed by vortexing and sonication. Thirty-three µL per reaction of the bead mixture are transferred into a 96-well HybAid plate. Seven µL of 15 nM biotinylated detection oligonucleotide stock in $1 \times$ TE buffer are added to each reaction and mixed. Ten µL of neutralized assay sample are added and the plate is sealed with a silicon cap mat seal. The plate is first incubated at 96°C for 5 minutes and incubated at 50°C without

agitation overnight in a conventional hybridization oven. A filter plate (Dura pore, Millipore part number MSBVN1250, 1.2 μm pore size) is prewetted with 75 μL 1 $\times$ TMAC hybridization solution supplemented with 0.5% (w/v) BSA. The entire sample volume from the hybridization reaction is transferred to the filter plate. The hybridization plate is rinsed with 75 μL 1 $\times$ TMAC hybridization solution containing 0.5% BSA and any remaining material is transferred to the filter plate. Samples are filtered under slow vacuum, with 150 μL buffer evacuated over about 8 seconds. The filter plate is washed once with 75 μL 1 $\times$ TMAC hybridization solution containing 0.5% BSA and the microspheres in the filter plate are resuspended in 75 μL 1 $\times$ TMAC hybridization solution containing 0.5% BSA. The filter plate is protected from light and incubated on an Eppendorf Thermalmixer R for 5 minutes at 1000 rpm. The filter plate is then washed once with 75 μL 1 $\times$ TMAC hybridization solution containing 0.5% BSA. 75 μL of 10 $\mu\text{g/mL}$ streptavidin phycoerythrin (SAPE-100, MOSS, Inc.) in 1 $\times$ TMAC hybridization solution is added to each reaction and incubated on Eppendorf Thermalmixer R at 25°C at 1000 rpm for 60 minutes. The filter plate is washed twice with 75 μL 1 $\times$ TMAC hybridization solution containing 0.5% BSA and the microspheres in the filter plate are resuspended in 75 μL 1 $\times$ TMAC hybridization solution containing 0.5% BSA. The filter plate is then incubated protected from light on an Eppendorf Thermalmixer R for 5 minutes, 1000 rpm. The filter plate is then washed once with 75 μL 1 $\times$ TMAC hybridization solution containing 0.5% BSA. Microspheres are resuspended in 75 μL 1 $\times$ TMAC hybridization solution supplemented with 0.5% BSA, and analyzed on a Luminex 100 instrument running Xponent 3.0 software. At least 100 microspheres are counted per bead type, under high PMT calibration and a doublet discriminator setting of 7500 to 18000.

qPCR read-out

[00214] Standard curves for qPCR are prepared in water ranging from 108 to 102 copies with 10-fold dilutions and a no-template control. Neutralized assay samples are diluted 40-fold into diH₂O. The qPCR master mix is prepared at 2 $\times$ final concentration (2 $\times$ KOD buffer, 400 μM dNTP mix, 400 nM forward and reverse primer mix, 2 $\times$ SYBR Green I and 0.5 U KOD EX). Ten μL of 2 $\times$ qPCR master mix is added to 10 μL of diluted assay sample. qPCR is run on a BioRad MyIQ iCycler with 2 minutes at 96°C followed by 40 cycles of 96°C for 5 seconds and 72°C for 30 seconds.

Example 4. Analysis of Biomarker Panel Model

[00215] Models comprising various combinations of the 41 biomarkers listed in Tables 1 and 2 were analyzed to determine the AUC, sensitivity, and specificity of the panels comprising the various biomarker combinations. Table 9 below shows the model results when the biomarker proteins of the panel were added one by one or were removed one by one in the order shown in

Table 9. Thus, the first line of Table 9 shows the model results when the panel included only INHBC (for the “add one by one” results) or included all of the forty-one listed biomarkers listed except for INHBC (for the “remove one by one” results). The results in Table 9 show that once the panel included at least the first five biomarker proteins, specifically, INHBC, SHBG, ACY1, COL1A1, and RTN4R, the AUC improved less significantly with the addition of each subsequent biomarker protein. The sensitivity and specificity values continued improving with the addition of the remaining biomarker proteins. The results in Table 9 also show that the performance of model panels missing the first ten biomarker proteins, specifically, INHBC, SHBG, ACY1, COL1A1, RTN4R, CRLF1, CBX7, FAM20B COL15A and KIN dropped significantly, as measured by AUC, sensitivity, and specificity. The results shown in Table 9 are also shown in Figures 1A and 1B in graph form.

Table 9: Add One by One and Remove One by One Model Results

Biomarker	Add One by One			Remove One by One		
	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
INHBC	0.500	1.000	0.000	0.731	0.702	0.761
SHBG	0.711	0.746	0.677	0.737	0.702	0.772
ACY1	0.739	0.772	0.707	0.720	0.693	0.746
COL1A1	0.758	0.807	0.710	0.723	0.684	0.762
RTN4R	0.749	0.772	0.726	0.706	0.640	0.772
CRLF1:CLCF1 complex	0.743	0.772	0.714	0.723	0.711	0.735
CBX7	0.748	0.789	0.706	0.709	0.684	0.734
FAM20B	0.766	0.816	0.716	0.707	0.667	0.748
COL15A1	0.775	0.816	0.735	0.724	0.719	0.729
KIN	0.768	0.807	0.728	0.721	0.693	0.748
SERPINA11	0.757	0.789	0.725	0.710	0.658	0.763
PELI2	0.752	0.781	0.723	0.700	0.658	0.742
TFF3	0.753	0.781	0.725	0.711	0.667	0.755
MARCKSL1	0.753	0.781	0.725	0.712	0.684	0.740
HTRA1	0.771	0.798	0.744	0.682	0.640	0.725
FABP12	0.766	0.798	0.734	0.679	0.640	0.718
GAD1	0.763	0.789	0.737	0.688	0.658	0.718
CHAD	0.780	0.807	0.754	0.689	0.684	0.694
CPM	0.753	0.754	0.752	0.631	0.579	0.684
SVEP1	0.760	0.763	0.757	0.641	0.579	0.704
SOCS7	0.786	0.816	0.755	0.634	0.579	0.689
F9	0.770	0.781	0.758	0.625	0.588	0.661
DLK1	0.765	0.772	0.757	0.619	0.588	0.650
HERC1	0.779	0.807	0.752	0.628	0.596	0.660
IL20RB	0.772	0.789	0.754	0.637	0.623	0.652
STC1	0.778	0.798	0.758	0.640	0.640	0.640
MYOC	0.746	0.737	0.754	0.653	0.702	0.604
WFDC11	0.765	0.781	0.749	0.629	0.640	0.618
MAP2K4	0.756	0.763	0.749	0.636	0.667	0.606

CALB1	0.750	0.728	0.772	0.643	0.684	0.602
CCL16	0.761	0.763	0.760	0.538	0.544	0.532
SMCO2	0.759	0.763	0.755	0.559	0.561	0.557
GPX2	0.765	0.754	0.775	0.556	0.553	0.560
CCL23	0.759	0.754	0.763	0.539	0.526	0.552
FGFR4	0.757	0.763	0.750	0.493	0.474	0.541
OSTM1	0.756	0.754	0.758	0.542	0.535	0.548
RNASE10	0.757	0.763	0.752	0.532	0.518	0.547
ITIH1	0.747	0.728	0.766	0.520	0.474	0.487
ZNF134	0.752	0.737	0.768	0.514	0.439	0.533
CFAP45	0.756	0.746	0.766	0.515	1.000	0.000
SFTPD	0.765	0.763	0.766	NA	NA	NA

[00216] An analysis of the AUC, sensitivity, and specificity of the model was also performed in which the panel comprised N biomarkers from Tables 1 and 2 in which the biomarkers were added one by one in random order. The results are shown in FIG. 1C. The results show that the performance of the model improves with the addition of more biomarkers to the panel regardless of which biomarkers selected from Tables 1 and 2 are added. Further, 10 random biomarker proteins will generally result in an AUC of 0.7.

[00217] 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. 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. 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 specifically listed biomarkers can be specifically excluded either as an individual biomarker or as a biomarker from any panel.