



US 20180267053A1

(19) **United States**

(12) **Patent Application Publication**
Urdea et al.

(10) **Pub. No.: US 2018/0267053 A1**

(43) **Pub. Date: Sep. 20, 2018**

(54) **SYSTEMS AND METHODS FOR
DEVELOPING DIAGNOSTIC TESTS BASED
ON BIOMARKER INFORMATION FROM
LEGACY CLINICAL SAMPLE SETS**

(71) Applicant: **True Health IP LLC**, Frisco, TX (US)

(72) Inventors: **Michael S. Urdea**, Alamo, CA (US);
Michael P. McKenna, Branford, CT
(US)

(21) Appl. No.: **15/918,489**

(22) Filed: **Mar. 12, 2018**

Related U.S. Application Data

(60) Continuation of application No. 15/007,786, filed on Jan. 27, 2016, now abandoned, which is a continuation of application No. 13/746,216, filed on Jan. 21, 2013, now abandoned, which is a division of application No. 13/561,913, filed on Jul. 30, 2012, now Pat. No. 8,357,497, which is a division of application No. 12/300,019, filed on Jun. 29, 2009, now Pat. No. 8,232,065, filed as application No. PCT/US07/11196 on May 8, 2007.

(60) Provisional application No. 60/798,867, filed on May 8, 2006.

Publication Classification

(51) **Int. Cl.**
G01N 33/68 (2006.01)
G06F 19/18 (2006.01)
C12Q 1/6883 (2006.01)
G01N 33/53 (2006.01)

(52) **U.S. Cl.**
CPC *G01N 33/6842* (2013.01); *G06F 19/18*
(2013.01); *G01N 2800/60* (2013.01); *G01N*
2800/56 (2013.01); *G01N 2570/00* (2013.01);
G01N 33/5302 (2013.01); *C12Q 1/6883*
(2013.01)

(57) **ABSTRACT**

Disclosed are systems and methods for developing diagnostic tests (e.g., detection, screening, monitoring, and prognostic tests) based on biomarker information from legacy clinical sample sets, for which only small sample volumes (e.g., about 0.05 to about 1.0 mL or less per sample) are typically available. For example, biomarkers (e.g., about 10, 50, 100, 150, 200, 300, or more) may be detected in the clinical samples through the use of single molecule detection and each biomarker may be detected in an assay that includes about 1 or less of a legacy clinical sample.

Single Molecule Detector

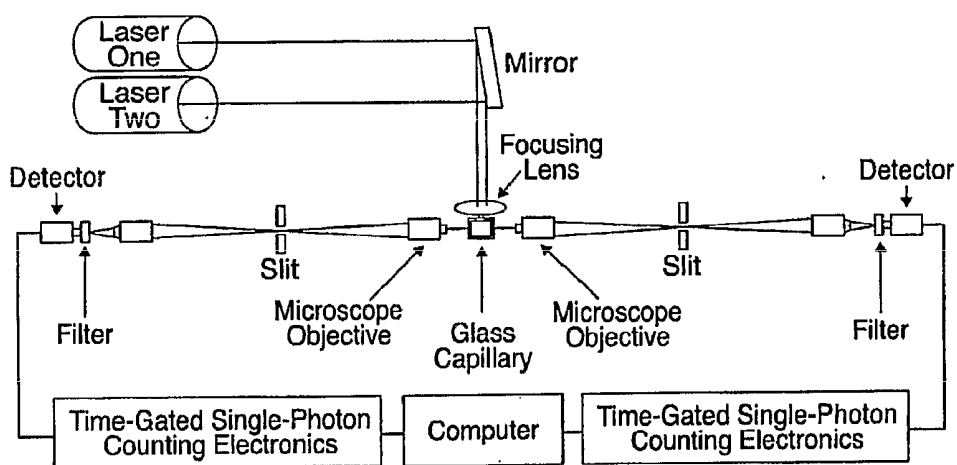


Fig. 1

Heart of the SMD Instrument

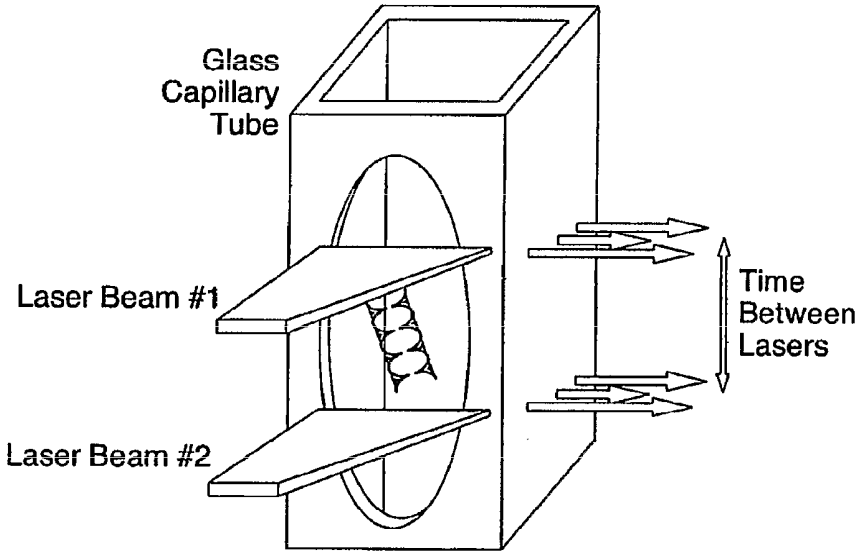


Fig. 2

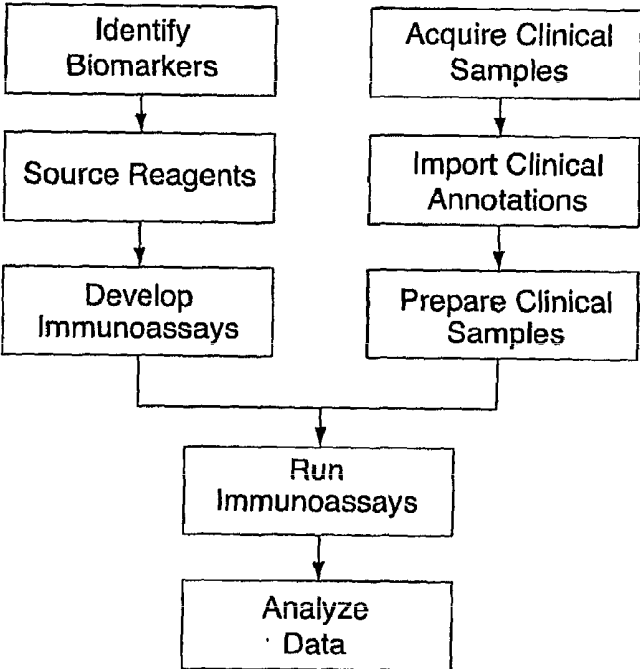


Fig. 3

IL-1RA Standard curve

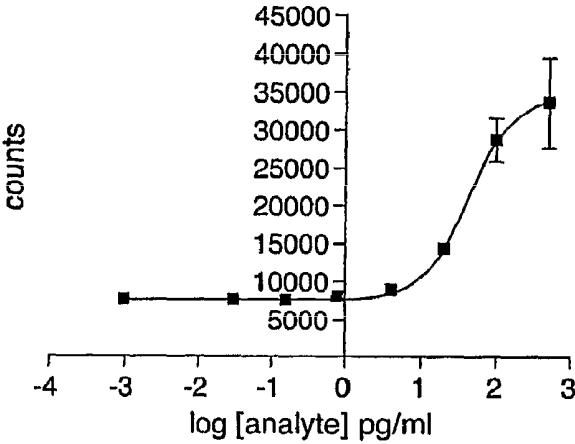


Fig. 4

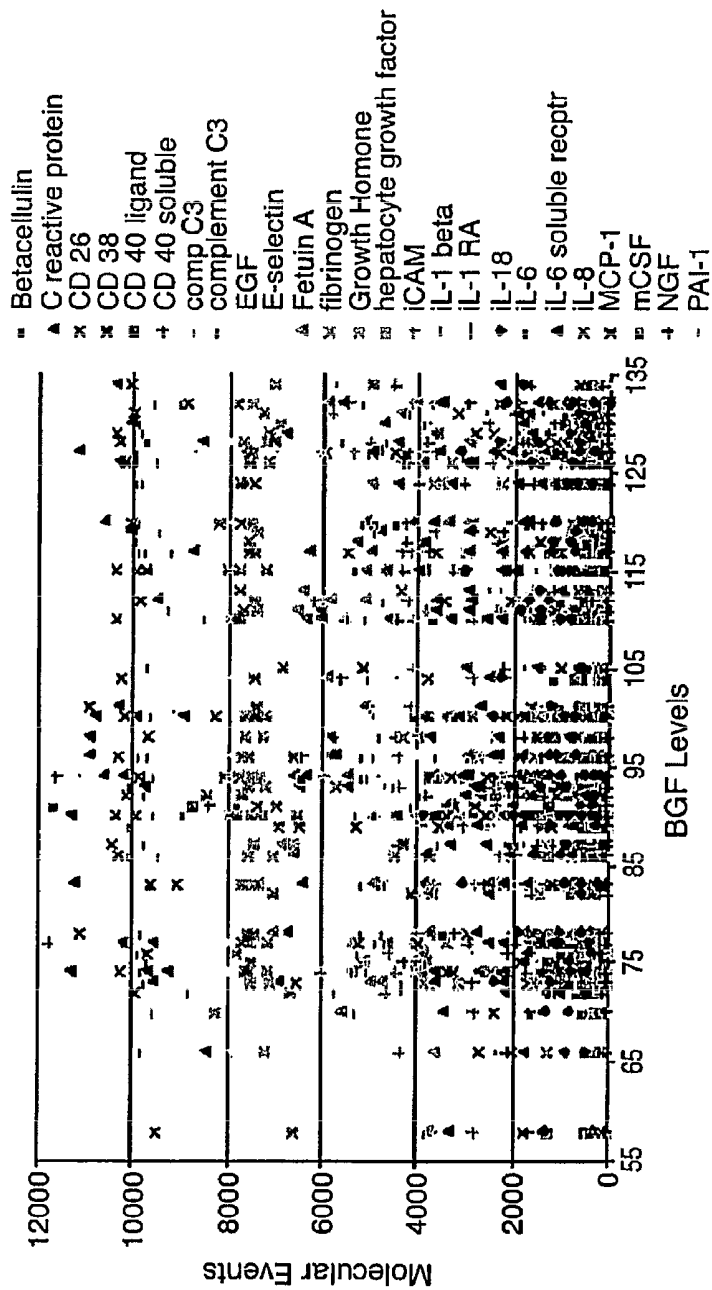


Fig. 5

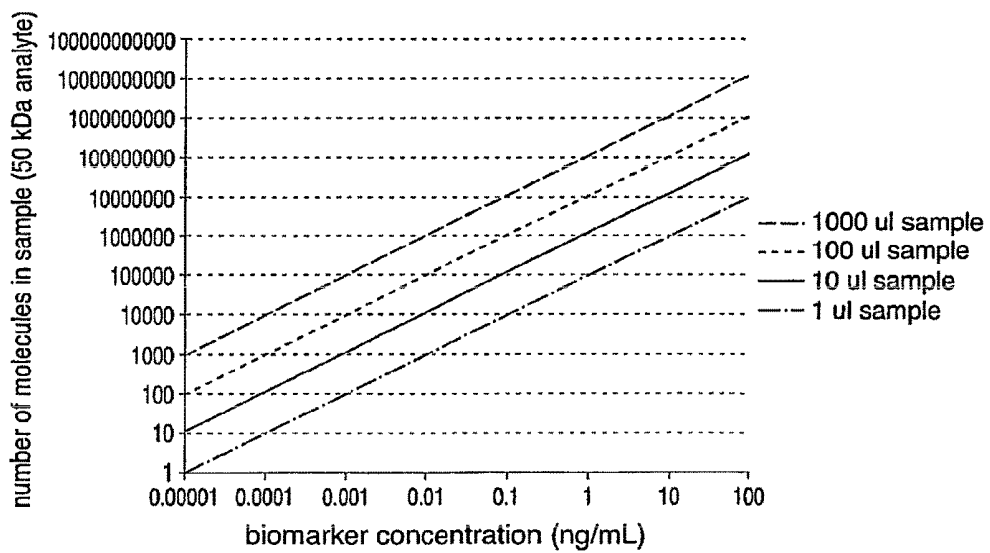
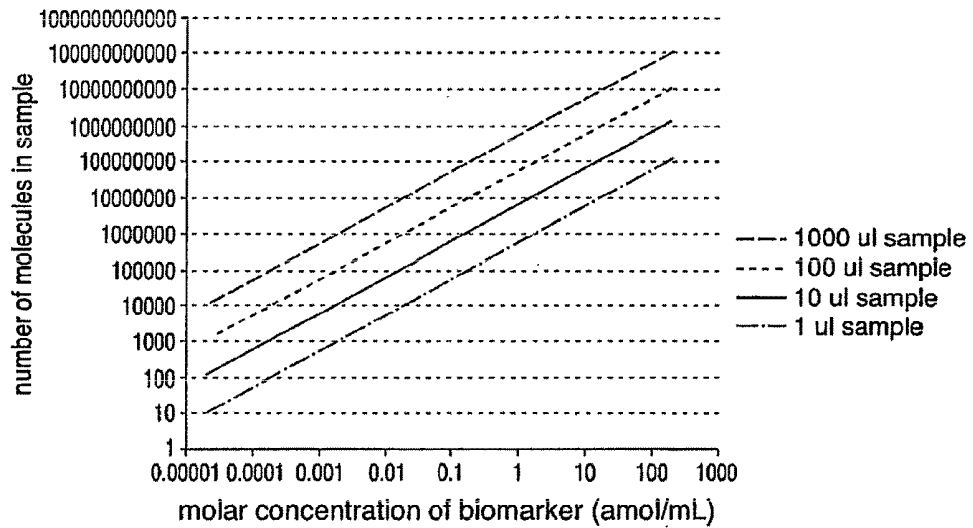


Fig. 6

Company	Product	Technology Description	Detection Limit (Molecules)	CV
<u>Ciphergen</u>	System 4000	SELDI Mass Spectrometry	6.0E+11	20%
<u>ForteBio</u>	Octet	Bio-layer interferometry	1.9E+08	15%
<u>Luminex</u>	100 IS	Coded bead and flow detection system	7.2E+05	10%
<u>Meso-scale</u>	Sector	Electro Chemiluminescence	1.3E+06	10%
<u>Singulex</u>	Erenna	Single molecule fluorescent detection	1.0E+05	10%
<u>US Genomics</u>	Trilogy 2020	Single molecule fluorescent detection	2.0E+04	10%
<u>Zeptosens</u>	ZeptoReader	Planar wave guide array detection	1.7E+06	15%

Fig. 7

**SYSTEMS AND METHODS FOR
DEVELOPING DIAGNOSTIC TESTS BASED
ON BIOMARKER INFORMATION FROM
LEGACY CLINICAL SAMPLE SETS**

PRIORITY CLAIM

[0001] This application is a continuation of U.S. patent application Ser. No. 15/007,786, filed on Jan. 27, 2016, which is a continuation of U.S. patent application Ser. No. 13/746,216, filed on Jan. 21, 2013, which is a divisional of U.S. patent application Ser. No. 13/561,913, filed Jul. 30, 2012 (now U.S. Pat. No. 8,357,497), which is a divisional of U.S. patent application Ser. No. 12/300,019, filed Jun. 29, 2009 (now U.S. Pat. No. 8,232,065), which is an U.S. 371 National Stage application of PCT Application No. PCT/US2007/011196, filed May 8, 2007, which claims priority to U.S. Provisional Patent Application No. 60/798,867, filed May 8, 2006, the entire contents of each which are incorporated herein by reference and relied upon.

FIELD OF THE INVENTION

[0002] Embodiments of the present invention relate to systems and methods for developing diagnostic tests and, more specifically, to systems and methods for developing diagnostic tests based on biomarker information from legacy clinical sample sets, for which only small sample sizes (e.g., about 0.05 to 1.0 mL per sample) are typically available. In a preferred embodiment, the biomarker information is detected in the clinical samples through the use of single molecule detection.

BACKGROUND OF THE INVENTION

[0003] Diagnostic tests have been provided for detecting, screening, monitoring, and/or predicting the future development of various health states (e.g., disease states) in a subject. Typically, the detecting, screening, monitoring, or prognosis is provided by a diagnostic test based, at least in part, on the level(s) of one or more biological markers ("biomarkers") in a clinical sample taken from the subject (e.g., the subject's blood), or the presence thereof. Such biomarkers are selected because the presence, absences, or levels of such biomarkers alone or in combination are indicative of the presence, stage, or future clinical course of the health state. Often times, but not necessarily, the diagnostic test may additionally be based on clinical information concerning the subject. Determining an appropriate diagnosis or prognosis for a subject can, for example, advantageously increase the subject's chances for survival and/or recovery.

[0004] Diagnostic tests must undergo a development stage during which the tests are formulated (and optionally tested/validated) using previously collected samples stored for future research and development needs. This process is prior to their use in diagnosing or predicting the development of disease in subjects in real time. The information used to formulate and validate the tests typically comes from clinical samples for a cohort of subjects for whom at least some biochemical and clinical data is known regarding the presence or absence of the health state under consideration. Thus, traditionally a party who is desirous of developing a diagnostic test for a given health state is required to commit significant resources to the collection of clinical samples (and optionally clinical information such as medical history)

from subjects who have, and/or lack, the health state, often at various stages. This data collection process can take many years, depending on the type of disease being considered and the party's relative access to suitable subjects.

[0005] Traditional approaches for developing diagnostic tests also require the clinical samples that are collected to have sufficiently large volumes, and such large samples cannot always be readily obtained. Specifically, traditional biomolecular detection approaches require large sample volumes in order to allow for the selection of a set of biomarkers that will be useful in the determination of a patient's health state. Of all the biomarkers that are evaluated (e.g., 1-3, 150-300 biomarkers, or 1000 or more), only those biomarkers that are determined to aid in the determination of the health state in a patient are included in the final diagnostic test. For example, according to one approach, single-biomarker multiple ELISAs used to measure the presence or level of 300 biomarkers typically require a serum or plasma sample size of about 30 mL of specimen per individual (i.e., 100 uL per assay times 300 biomarkers). The required sample volume becomes 90 mL of specimen per individual if the assays are done in triplicate. This is a very large volume and is very impractical. In addition, few studies have ever been conducted where so much clinical sample was collected. Multiplexing, which involves measuring multiple biomarkers in the same reaction vessel, can reduce the overall required sample volume by way of conservation but requires compatibility between all the assay components and typically compromises sensitivity through increased background effects. As a result, on an assay by assay basis, individual assays are typically 10 or more fold more sensitive than their counterpart within a multiplexed assay.

[0006] In view of the foregoing, it would be desirable to provide systems and methods for developing diagnostic tests in which access to suitable clinical samples is improved and which rely on smaller sample volumes.

SUMMARY OF THE INVENTION

[0007] The above and other objects and advantages of the present invention are provided in accordance with the principles of the present invention described herein. Embodiments of the present invention relate to systems and methods for developing diagnostic tests based on biomarker information from legacy clinical sample sets, for which only small sample volumes (e.g., about 0.05 to 1.0 mL per individual) are typically available. As used herein, a "legacy clinical sample set" is one or more clinical samples (e.g., 10 to 5000 samples or more) collected in the past (i.e., retrospective sample collections). The use of legacy clinical samples, as opposed to performing the process of collecting clinical samples prospectively, reduces the resources and time that must be committed to developing new diagnostic tests. Legacy clinical samples may be from, for example, one or more past studies that occurred over a span of 1 to 40 years or more, which studies may be accompanied by tens to thousands of clinical parameters, traditional laboratory measurements that are considered risk factors or that provide additive information to enable a better clinical decision to be made, and other previously measured information (e.g., clinical data such as the subject's age, weight, ethnicity, medical history, and/or other information). In most cases, the legacy clinical samples are serum or plasma samples that have been stored for years at -80 degrees Centigrade or -20

degrees Centigrade. In other examples, a legacy clinical sample can include, for example, blood cells, ascites fluid, interstitial fluid, bone marrow, sputum, urine, or other biological sample. Examples of such past studies, which are included for the purpose of illustration and not limitation, are listed below:

- [0008] 1. DPP (Diabetes Prevention Program)—An NIH sponsored trail that studied the impact of lifestyle modifications, metformin vs. placebo. This study had 2.8 years follow-up with diabetes outcomes.
- [0009] 2. IRAS (Insulin Resistance Atherosclerosis Study)—Studied the impact of insulin resistance on the development of cardiovascular disease.
- [0010] 3. ARIC (Atherosclerosis Risk in Communities Study)—This study includes CVD and cardiovascular outcomes.
- [0011] 4. Finnish Diabetes Prevention Study—studied the impact of lifestyle changes on the development of diabetes.
- [0012] 5. Israeli Diabetes Research Group (MELANY)—Studied the development of diabetes in healthy normal subjects from the Israeli military
- [0013] 6. HDDRISC (Heart Disease and Diabetes Risk Indicators in a Screened Cohort)—collection of diabetes and cardiovascular outcomes.
- [0014] 7. WSCOPS (West of Scotland Coronary Prevention Study)—studied the impact of pravastatin on reduction of LDL and reduction in myocardial events
- [0015] 8. ASCOT (Anglo-Scandinavian Cardiac Outcomes Trial)—studied the impact of different medicines for lowering blood pressure and cholesterol. CVD outcomes collected.
- [0016] 9. SOF (Study of Osteoporotic Fractures)—Study looks for predictors of fracture in women over 65 years of age
- [0017] 10. NORA (National Osteoporosis Risk Assessment)—Studied fracture outcomes in women with varying BMD levels.
- [0018] 11. Framingham Heart Study—Related to identifying the common factors or characteristics that contribute to CVD by following its development over a long period of time in a large group of participants who had not yet developed overt symptoms of CVD or suffered a heart attack or stroke.
- [0019] 12. CARDIA—(Coronary Artery Risk Development in Young Adults) A longitudinal study designed to trace the development of risk factors for coronary heart disease in a cohort of 18-30 year olds (1985) in four U.S. cities.
- [0020] 13. Reykjavik Study—A long-term prospective population-based cardiovascular study of 33-79 year olds with 4 to 20 year follow-up (1967-91), in Iceland.
- [0021] 14. Malmo Preventive Project—A prospective, population-based study of the effects of interventions on mortality and cardiovascular morbidity in 32-51 year olds (1974-1992) in Sweden
- [0022] 15. Heart Protection Study—A very large, prospective, double-blind, randomized, controlled trial investigating prolonged use (>5 years) of a statin and an antioxidant vitamin cocktail in individuals 40 to 80 years old in the United Kingdom who had an elevated risk for CHD.
- [0023] 16. 4S (Scandinavian Simvastatin Survival Study)—Large double-blind, randomized trial

designed to evaluate the effect of a statin on mortality and morbidity in patients with coronary heart disease (CHD).

- [0024] 17. DREAM (Diabetes Reduction Assessment with ramipril and rosiglitazone Medication) Study—A large, double-blind, randomized, placebo-controlled trial evaluating the effects of an ACE inhibitor and/or a thiazolidinedione on the development of diabetes, death, or regression to normoglycaemia in adults aged 30 years or more with impaired fasting glucose and/or impaired glucose tolerance, and no previous cardiovascular disease.
- [0025] 18. Physician's Health Study—a large cohort of apparently healthy male U.S. physicians aged 40 to 84 years in 1982, followed prospectively for an average of 60.2 months
- [0026] 19. WHI (Women's Health Initiative)—A very large, prospective study, involving both clinical trial and observational components, of women 50 to 79 years of age in the U.S., and is designed to examine the relationship between health, lifestyle, and risk factors for a variety of specific diseases, including CHD
- [0027] 20. WHS (Women's Health Study) A very large, double-blind, randomized, placebo-controlled trial to evaluate the effects of vitamin E and low-dose aspirin on cardiovascular disease and cancer in apparently healthy U.S. women, age 45 and older, which also included an observational extension
- [0028] 21. NHS (Nurses' Health Study)—A very large, prospective cohort study of nurses aged 30-55 (in 1976) designed to assess the long term effects of oral contraceptive use
- [0029] 22. NHS II (Nurses' Health Study II)—A very large, prospective cohort study of nurses aged 25-42 (in 1989) designed to assess the long term effects of oral contraceptives, diet and lifestyle risks.
- [0030] In an embodiment of the present invention, methods and systems are provided for developing a diagnostic test for determining a health state in a patient (e.g., a test for a predicting or diagnosing disease such as diabetes, osteoporosis, pre-osteoporosis, or any other disease), in which at least one biomarker is detected in at least one legacy clinical sample. For example, the biomarker may be detected in an immunoassay that includes about 1 uL or less of the legacy clinical sample. The detection may be performed by, for example, a single molecule detector. Typically, although not necessarily, developing a new diagnostic test comprises detecting multiple biomarkers from multiple clinical samples, including samples from subjects known have a given health state, or with respect to reference ranges from a known normal population. The detected biomarker(s) are then analyzed for an association with the health state. For example, a statistical analysis may be performed to determine whether the biomarker statistically correlates with the presence or absence of the health state, or alternatively correlates with the existing gold standard (whether biomarker, clinical parameter, or otherwise) used for defining the presence of the health state (for example, fasting glucose level for diabetes, blood pressure for hypertension as a health state, or coronary imaging scores or percentage occlusions/stenosis for coronary artery disease). Alternatively or additionally, the analysis may involve determining whether the inclusion of the biomarker in a formula or machine learning analysis increases an ability of a math-

ematical function resulting from the machine learning analysis to determine the health state in a patient.

[0031] In another embodiment, clinical parameters (e.g., age, weight, ethnicity, medical history, and/or other clinical information) that accompany the legacy clinical sample(s) may also be analyzed for an association with the health state.

[0032] In yet another embodiment, methods and systems are provided for developing a diagnostic test for determining a health state in a patient, in which a plurality of biomarkers (e.g., 10-300 biomarkers) are detected in a legacy clinical sample through the use of a corresponding plurality of immunoassays, where the total amount of the legacy clinical sample that is used across the plurality of immunoassays is less than about 1 mL (e.g., less than about 0.05 mL). Typically, multiple legacy clinical samples are analyzed in the same fashion, and the detected biomarkers are then analyzed for an association with the disease.

[0033] In another embodiment, a diagnostic test is used to screen or monitor a patient for a given health state. The test is developed using any of the methods disclosed herein for screening legacy clinical samples. For example, at least one biomarker indicative of the presence, absence, or likelihood of developing the health state and identified by the methods described herein is employed in the test and its presence, absence, or level is determined.

[0034] Other features and advantages of the invention will be apparent from the following detailed description and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] For a better understanding of the present invention, and not intending to limit the scope of the invention in any way, reference is made to the following description, taken in conjunction with the accompanying drawings, in which like reference characters refer to like parts throughout, and in which:

[0036] FIGS. 1 and 2 are illustrative diagrams of a single molecule detector in accordance with an embodiment of the present invention;

[0037] FIG. 3 is a flowchart of illustrative stages involved in developing a diagnostic test in accordance with an embodiment of the present invention;

[0038] FIG. 4 shows a typical result for a working standard curve used in the development of immunoassays in accordance with an embodiment of the present invention;

[0039] FIG. 5 shows illustrative single molecule detection data in accordance with an embodiment of the present invention;

[0040] FIG. 6 shows a table indicating the actual number of analyte molecules present in a sample across the ranges of various sample sizes and starting analyte molar concentrations; and

[0041] FIG. 7 shows, without intending any limitation, the detection limit of selected biomarker assay technologies that are commercially available, indicating their typical analytical reproducibility performance characteristic (coefficient of variation) at these starting analyte concentrations. There are many additional technologies being applied to improve the sensitivity of single molecule detection, including microscopic techniques (atomic force microscopy, magnetic resonance force microscopy, scanning electrochemical microscopy, scanning tunneling microscopy) and spectroscopic techniques (fluorescence correlation spectroscopy, evanescent wave induced fluorescence spectroscopy, scanning

near-field optical microscopy, scanning enhanced raman spectroscopy, surface plasma resonance).

DETAILED DESCRIPTION OF THE INVENTION

[0042] Embodiments of the present invention relate to systems and methods for developing diagnostic tests for diagnosing, and predicting the future development of, various health states (e.g., health states including disease-specific states as well as other non-disease specific states) in a subject. Examples of diseases are osteoporosis, pre-osteoporosis, diabetes, cancer, and any other disease. In one embodiment of the present invention, systems and methods are provided for developing diagnostic tests based on biomarker information from legacy clinical sample sets, for which only small sample sizes (e.g., about 0.05 to 1.0 mL or less) are typically available. In a preferred embodiment, the biomarker information is extracted from the clinical samples through the use of single molecule detection.

Definitions

[0043] “Biomarker” in the context of the present invention encompasses, without limitation, proteins, nucleic acids, and metabolites, together with their polymorphisms, isoforms, mutations, derivatives, variants, modifications, and precursors, including nucleic acids and pro-proteins, cleavage products, receptors (including soluble and transmembrane receptors), subunits, fragments, ligands, protein-ligand complexes, multimeric complexes, and degradation products, elements, related metabolites, and other analytes or sample-derived measures. Biomarkers can also include mutated proteins or mutated nucleic acids. Biomarkers also include any calculated indices created mathematically or combinations of any one or more of the foregoing measurements, including temporal trends and differences. The term “analyte” as used herein can mean any substance to be measured and can encompass electrolytes and elements, such as calcium.

[0044] “Clinical parameters” encompasses all non-sample or non-analyte markers of subject health status or other characteristics, such as, without limitation, age (AGE), ethnicity (RACE), gender (SEX), diastolic blood pressure (DBP) and systolic blood pressure (SBP), family history (FHx), height (HT), weight (WT), waist (Waist) and hip (Hip) circumference, body-mass index (BMI), past Gestational Diabetes Mellitus (GDM), resting heart rate, EMG, EEG, body temperature, and sleep states.

[0045] A “formula,” “algorithm,” or “model” is any mathematical equation, algorithmic, analytical or programmed process, or statistical technique that takes one or more continuous or categorical inputs (herein called “parameters”) and calculates an output value, sometimes referred to as an “index” or “index value”. Non-limiting examples of “formulas” include sums, ratios, and regression operators, such as coefficients or exponents, biomarker value transformations and normalizations (including, without limitation, those normalization schemes based on clinical parameters, such as gender, age, or ethnicity), rules and guidelines, statistical classification models, and neural networks trained on historical populations. Of particular use in combining markers are linear and non-linear equations and statistical classification analyses to determine the relationship between levels of the biomarkers detected in a subject sample and the

subject's risk of disease (for example). In panel and combination construction, of particular interest are structural and syntactic statistical classification algorithms, and methods of risk index construction, utilizing pattern recognition features, including established techniques such as cross correlation, Principal Components Analysis (PCA), factor rotation, Logistic Regression (LogReg), Linear Discriminant Analysis (LDA), Eigengene Linear Discriminant Analysis (ELDA), Support Vector Machines (SVM), Random Forest (RF), Recursive Partitioning Tree (RPART), as well as other related decision tree classification techniques, Shruken Centroids (SC), StepAIC, Kth-Nearest Neighbor, Boosting, Decision Trees, Neural Networks, Bayesian Networks, Support Vector Machines, and Hidden Markov Models, among others. Many of these techniques are useful either combined with a biomarker selection technique, such as forward selection, backwards selection, or stepwise selection, complete enumeration of all potential panels of a given size, genetic algorithms, or they may themselves include biomarker selection methodologies in their own technique. These may be coupled with information criteria, such as Akaike's Information Criterion (AIC) or Bayes Information Criterion (BIC), in order to quantify the tradeoff between additional biomarkers and model improvement, and to aid in minimizing overfit. The resulting predictive models may be validated in other studies, or cross-validated in the study they were originally trained in, using such techniques as Leave-One-Out (LOO) and 10-Fold cross-validation (10-Fold-CV).

[0046] "Frank Disease" in the context of the present invention, is a clearly manifest, unmistakable, evident, or symptomatic disease state that unequivocally meets the definition of the disease set forth by a professional medical organization, such as the World Health Organization.

[0047] "Health state" encompasses disease states (e.g., presence, absence, or risk of developing a disease and likely responses to therapies for the disease) as well as other states not necessarily related to a specific disease such as environmental exposure, nutritional status, neurological function, immune status, organ function, and blood chemistry. Generally, determining a health state in a patient/subject involves determining that the patient should be classified within a given one of a plurality of populations (e.g., healthy vs. unhealthy, in a 2-population example).

[0048] A "legacy subject" is a subject (defined below) for which one or more clinical samples is included in a legacy clinical sample set.

[0049] A "live subject" is a subject for whom a determination (e.g., diagnosis or prognosis of disease) is made by a diagnostic test that has been developed in accordance with the principles of the present invention.

[0050] A "legacy clinical sample" is a clinical sample for an individual from a legacy clinical sample set (which set may have multiple samples for multiple individuals), where the volume of the sample meets a sample requirement (defined below) and the biomarker information from the sample may be used to develop a diagnostic test in accordance with the principles of the present invention.

[0051] A "live clinical sample" is a clinical sample from which biomarker information is evaluated by a diagnostic test in order to provide a determination (e.g., diagnosis or prognosis) for a corresponding live subject.

[0052] "Measuring" or "measurement" means assessing the presence, absence, quantity or amount (which can be an

effective amount) of either a given substance within a clinical or subject-derived sample, including the derivation of qualitative or quantitative concentration levels of such substances, or otherwise evaluating the values or categorization of a subject's clinical parameters. Alternatively, the term "detecting" or "detection" may be used and is understood to cover all measuring or measurement as described herein.

[0053] "Risk" in the context of the present invention, relates to the probability that an event will occur over a specific time period (e.g., conversion to frank Diabetes) and can mean a subject's "absolute" risk or "relative" risk. Absolute risk can be measured with reference to either actual observation post-measurement for the relevant time cohort, or with reference to index values developed from statistically valid historical cohorts that have been followed for the relevant time period. Relative risk refers to the ratio of absolute risks of a subject compared either to the absolute risks of low risk cohorts or an average population risk, which can vary by how clinical risk factors are assessed. Odds ratios, the proportion of positive events to negative events for a given test result, are also commonly used (odds are according to the formula $p/(1-p)$ where p is the probability of event and $(1-p)$ is the probability of no event) to no-conversion. Alternative continuous measures which may be assessed in the context of the present invention include time to health state (e.g., disease) conversion and therapeutic conversion risk reduction ratios.

[0054] "Pre-Disease" in the context of the present invention refers to a state that is intermediate between that defined as the normal homeostatic and metabolic state and states seen in Frank Disease. Pre-disease states can include abnormalities of homeostatic regulation, abnormal physiological measurements, abnormal morphometric measurements, and/or states in which abnormal levels of clinical parameters or biomarkers are present at a specific time point. Abnormalities are measurement outside the normal range as defined by professional medical organizations, such as the World Health Organization. "Pre-Disease" states, in the context of the present invention, are states, in an individual or in a population, having a higher than normal expected rate of disease conversion to frank disease. When a continuous measure of Pre-Disease conversion risk is produced, having a "pre-disease condition" encompasses any expected annual rate of conversion above that seen in a normal reference or general unselected normal prevalence population.

[0055] "Risk evaluation," or "evaluation of risk" in the context of the present invention encompasses making a prediction of the probability, odds, or likelihood that an event or health state may occur, the rate of occurrence of the event or conversion from one health state to another (e.g., from a normoglycemic condition to a pre-diabetic condition or pre-Diabetes, or from a pre-diabetic condition to pre-Diabetes or Diabetes). Risk evaluation can also comprise prediction of future levels, scores or other indices of disease, either in absolute or relative terms in reference to a previously measured population. The methods of the present invention may be used to make continuous or categorical measurements of the risk of conversion between health states. Embodiments of the invention can also be used to discriminate between normal and pre-diseased subject cohorts. In other embodiments, the present invention may be used so as to discriminate pre-diseased from diseased, or diseased from normal. Such differing use may require dif-

ferent biomarker combinations in individual panel, mathematical algorithm(s), and/or cut-off points, but be subject to the same aforementioned measurements of accuracy for the intended use.

[0056] A “sample” in the context of the present invention is a biological sample isolated from a subject and can include, by way of example and not limitation, whole blood, serum, plasma, blood cells, endothelial cells, tissue biopsies, lymphatic fluid, ascites fluid, interstitial fluid (also known as “extracellular fluid” and encompasses the fluid found in spaces between cells, including, inter alia, gingival crevicular fluid), bone marrow, cerebrospinal fluid (CSF), saliva, mucous, sputum, sweat, urine, or any other secretion, excretion, or other bodily fluids.

[0057] A “sample requirement” is the volume of starting sample required by a given assay technology in order to achieve an acceptable level of performance (coefficient of variation).

[0058] A “subject” in the context of the present invention is preferably a mammal. The mammal can be a human, non-human primate, mouse, rat, dog, cat, horse, or cow, but are not limited to these examples. Mammals other than humans can be advantageously used as subjects that represent animal models of disease, pre-disease, or a pre-disease condition. A subject can be male or female. A subject can be one who has been previously diagnosed or identified as having a health state (e.g., disease, pre-disease, or a pre-disease condition), and optionally has already undergone, or is undergoing, a therapeutic intervention for the health state. Alternatively, a subject can also be one who has not been previously diagnosed as having a given health state. For example, a subject can be one who exhibits one or more risk factors for a disease, pre-disease, or a pre-disease condition, or a subject who does not exhibit disease risk factors, or a subject who is asymptomatic for a disease, pre-disease, or pre-disease conditions. A subject can also be one who is suffering from or at risk of developing disease, pre-disease, or a pre-disease condition.

[0059] “Traditional laboratory risk factors” correspond to biomarkers isolated or derived from subject samples and which are currently evaluated in the clinical laboratory and used in traditional global risk assessment algorithms (e.g., Stern, Framingham, Finland Diabetes Risk Score, ARIC Diabetes, and Archimedes). Traditional laboratory risk factors commonly tested from subject blood samples include, but are not limited to, total cholesterol (CHOL), LDL (LDL/LDL-C), HDL (HDL/HDL-C), VLDL (VLDL-C), triglycerides (TRIG), glucose (including, without limitation, the fasting plasma glucose (Glucose) and the oral glucose tolerance test (OGTT)) and HBA1c (HBA1C) levels.

INDICATIONS OF THE INVENTION

[0060] Embodiments of the present invention allow for the determining of a health state in a patient. For example, the risk of developing disease, pre-disease, or a pre-disease condition typically can be detected with a pre-determined level of predictability by measuring an “effective amount” of a biomarker in a test sample (e.g., a subject derived sample), and comparing the effective amounts to reference or index values, often utilizing mathematical algorithms or formulas in order to combine information from results of multiple individual biomarkers and from non-analyte clinical parameters into a single measurement or index. When appropriate, subjects identified as having an increased risk for a health

state can optionally be selected to receive treatment regimens, such as administration of prophylactic or therapeutic compounds, or implementation of exercise regimens or dietary supplements to prevent or delay the onset of, for example, disease, pre-disease, or a pre-disease condition or other adverse health conditions.

[0061] The amount of the biomarker can be measured in a test sample and compared to a normal control level, utilizing techniques such as reference limits, discrimination limits, or risk defining thresholds to define cutoff points and abnormal values for a health state. The normal control level means the level of one or more biomarkers or combined biomarker indices typically found in a subject not having the health state. Such normal control level and cutoff points may vary based on whether a biomarker is used alone or in a formula combining with other biomarkers into an index. Alternatively, the normal control level can be a database of biomarker patterns from previously tested subjects who did not convert to the health state over a clinically relevant time horizon.

[0062] The present invention may be used to make continuous or categorical measurements of the risk of conversion to an adverse health state (e.g., disease), thus diagnosing and defining the risk spectrum of a category of subjects defined as predisposed to the adverse health state. In the categorical scenario, the methods of the present invention can be used to discriminate between (for example) normal and pre-diseased subject cohorts. In other embodiments, the present invention may be used so as to discriminate pre-disease from disease, or diseased from normal. Other non-disease specific health states can also be determined. Such differing use may require different biomarker combinations in individual panel, mathematical algorithm, and/or cut-off points, but be subject to the same aforementioned measurements of accuracy for the intended use.

[0063] Identifying patients that are predisposed to adverse health states (e.g., pre-disease states) enables the selection and initiation of various therapeutic interventions or treatment regimens in order to delay, reduce or prevent those patients’ conversion to the adverse health states (e.g., disease). Levels of a specific amount of biomarker also may allow for the course of treatment of the health state (e.g., disease, pre-disease, or a pre-disease condition) to be monitored. For example, in this method, a biological sample can be provided from a subject undergoing treatment regimens, e.g., drug treatments, for a disease. Such treatment regimens can include, but are not limited to, exercise regimens, dietary supplementation, weight loss, surgical intervention, device implantation, and treatment with therapeutics or prophylactics used in subjects diagnosed or identified with various health states. If desired, biological samples are obtained from the subject at various time points before, during, or after treatment.

[0064] The present invention can also be used to screen patient or subject populations in any number of settings. For example, a health maintenance organization, public health entity or school health program can screen a group of subjects to identify those requiring interventions, as described above, or for the collection of epidemiological data. Insurance companies (e.g., health, life or disability) may screen applicants in the process of determining coverage or pricing, or existing clients for possible intervention. Data collected in such population screens, particularly when tied to any clinical progression to conditions like disease,

pre-disease, or a pre-disease condition, will be of value in the operations of, for example, health maintenance organizations, public health programs and insurance companies. Such data arrays or collections can be stored in machine-readable media and used in any number of health-related data management systems to provide improved healthcare services, cost effective healthcare, improved insurance operation, etc. See, for example, U.S. Patent Application No.; U.S. Patent Application No. 2002/0038227; U.S. Patent Application No. US 2004/0122296; U.S. Patent Application No. US 2004/0122297; and U.S. Pat. No. 5,018,067, which are hereby incorporated by reference herein in their entireties. Such systems can access the data directly from internal data storage or remotely from one or more data storage sites. Thus, in a health-related data management system, wherein risk of developing a diabetic condition for a subject or a population comprises analyzing disease risk factors, the present invention provides an improvement comprising use of a data array encompassing the biomarker measurements as defined herein and/or the resulting evaluation of risk from those biomarker measurements.

[0065] A machine-readable storage medium can comprise a data storage material encoded with machine readable data or data arrays which, when using a machine programmed with instructions for using said data, is capable of use for a variety of purposes, such as, without limitation, subject information relating to health state risk factors over time or in response to drug therapies, drug discovery, and the like. Measurements of effective amounts of the biomarkers of the invention and/or the resulting evaluation of risk from those biomarkers can be implemented in computer programs executing on programmable computers, comprising, inter alia, a processor, a data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one output device. Program code can be applied to input data to perform the functions described above and generate output information. The output information can be applied to one or more output devices, according to methods known in the art. The computer may be, for example, a personal computer, microcomputer, or workstation of conventional design.

[0066] Each program can be implemented in a high level procedural or object oriented programming language to communicate with a computer system. However, the programs can be implemented in assembly or machine language, if desired. The language can be a compiled or interpreted language. Each such computer program can be stored on a storage media or device (e.g., ROM or magnetic diskette or others as defined elsewhere in this disclosure) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The health-related data management system of the invention may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform various functions described herein. Levels of a specific amount of one or more biomarkers can then be determined and compared to a reference value, e.g. a control subject or population whose state is known or an index value or baseline value. The reference sample or index value or baseline value may be taken or derived from one or more

subjects who have been exposed to the treatment, or may be taken or derived from one or more subjects who are at low risk of developing a health state (e.g., disease, pre-disease, or a pre-disease condition), or may be taken or derived from subjects who have shown improvements in risk factors (such as clinical parameters or traditional laboratory risk factors as defined herein) as a result of exposure to treatment. Alternatively, the reference sample or index value or baseline value may be taken or derived from one or more subjects who have not been exposed to the treatment. For example, samples may be collected from subjects who have received initial treatment for disease, pre-disease, or a pre-disease condition and subsequent treatment for disease, pre-disease, or a pre-disease condition to monitor the progress of the treatment. A reference value can also comprise a value derived from risk prediction algorithms or computed indices from population studies such as those disclosed herein.

[0067] The biomarkers of the present invention can thus be used to generate a reference biomarker profile of those subjects who do not have a health state (e.g., impaired glucose tolerance in the case of Diabetes), and would not be expected to develop the health state. The biomarkers disclosed herein can also be used to generate a subject biomarker profile taken from subjects who have a health state such as disease, pre-disease, or a pre-disease condition. The subject biomarker profiles can be compared to a reference biomarker profile to diagnose or identify subjects at risk for developing the health state, to monitor the progression of the health state (e.g., disease), as well as the rate of progression of the health state, and to monitor the effectiveness of any treatments for the health state. The reference and subject biomarker profiles of the present invention can be contained in a machine-readable medium, such as but not limited to, digital and analog media like those readable by a VCR, CD-ROM, DVD-ROM, USB flash media, among others. Such machine-readable media can also contain additional test results, such as, without limitation, measurements of clinical parameters and traditional laboratory risk factors. Alternatively or additionally, the machine-readable media can also comprise subject information such as medical history and any relevant family history. The machine-readable media can also contain information relating to other disease-risk algorithms and computed indices such as those described herein.

[0068] A diagnostic test that is developed in accordance with the principles of the present invention can be used to make a determination for a live subject (e.g., a diagnosis or prognosis) based, at least in part, on the presence or level(s) of one or more biomarkers present in a live clinical sample from the live subject. The levels are determined, as is understood to those of ordinary skill in the art, within the sensitivity and specificity parameters of the test format selected (e.g., a biomarker is "absent" if its level is below the test's limit of detection or some other cut-off value). For example, one such diagnostic test may involve comparing the subject's biomarker level(s) to a reference value. As another example, the diagnostic test may involve evaluating the live subject's biomarker level(s) (and optionally other information for the subject such as, for example, age, weight, ethnicity, medical history, and/or other clinical information) with a formula or model that produces a diagnostic or prognostic score for the live subject.

[0069] A diagnostic test for a given health state may be developed, at least in part, through the use of a legacy

clinical sample set. The legacy clinical sample set may include samples for a cohort of legacy subjects, for whom at least some data is known regarding the presence or absence of the health state. For example, a diagnostic test may be developed based on samples for legacy subjects who are known to have a given disease. Alternatively or additionally, the diagnostic test may be developed based on clinical samples for legacy subjects who are known to lack the disease or other health state.

[0070] Theoretically, an almost limitless number of biomarkers are available for selection within the process of developing a diagnostic test. However, only a subset of all available biomarkers (e.g., between 10 and 300) are typically selected per disease area, which subset of biomarkers may be identified by physicians and/or other sources of information (e.g., medical journals) with expertise in the disease area. Biomarkers may also be derived from de novo research using “open” proteomics profiling technologies such as mass spectrometry, LC-LC mass spectrometry, 2-D gel electrophoresis, protein arrays, western blots, reverse western tissue blots, etc.

[0071] In an embodiment of the present invention, systems and methods are provided for developing a diagnostic test, according to which (i) a set of one or more legacy clinical samples is received (e.g., 50 to 5000 legacy samples), (ii) the levels of a selected subset of biomarkers are measured from the sample(s), and (iii) the biomarker levels (and optionally clinical parameters) are analyzed for an association with the health state under consideration. This analyzing may involve, for example, using statistical analysis to determine whether a particular one or more biomarkers (and optionally particular level(s) of those biomarkers and/or clinical parameters) is correlated statistically with the presence, absence, or risk of developing the health state (e.g., progression to disease states of different severities), and/or to select one or more therapies or to monitor therapy response/efficacy. In some embodiments, a biomarker panel can be constructed and a formula derived specifically to enhance performance for use also in subjects undergoing therapeutic interventions, or a separate panel and formula may alternatively be used solely in such patient populations. An aspect of the invention is the use of specific known characteristics of biomarkers and their changes in such subjects for such panel construction and formula derivation. Such modifications may enhance the performance of various indications noted above in prevention of adverse health states, and diagnosis, therapy, monitoring, and prognosis of a health state. The

biomarkers may vary under therapeutic intervention for the health state, whether lifestyle (e.g. diet and exercise), surgical (e.g., bariatric surgery) or pharmaceutical (e.g., one of the various classes of drugs mentioned herein or known to modify common risk factors or risk of disease) intervention. The biomarkers may also vary based on environmental exposure, nutritional status, neurological function, immune status, organ function, and/or blood chemistry. Alternatively or additionally, the analyzing of the biomarker may involve determining whether the inclusion of particular biomarker(s) in a formula or machine learning analysis (e.g., support vector or neural network analysis) increases the relative ability of a mathematical function resulting from the analysis to diagnose or predict the health state in a subject. Generally, machine learning is a form of artificial intelligence whereby information learned from a computer-assisted analysis of data can be used to generate a function that describes dependencies in data. This computer-assisted, machine learning analysis may be performed by any suitable software, hardware, or combination thereof (a “machine learning tool”). Suitable examples of machine learning tools will be apparent to those of ordinary skill in the art and therefore will not be described in detail.

[0072] A key feature of embodiments of the invention is the ability to profile tens, hundreds or even thousands of biomarkers in a single small legacy sample. It will be apparent that the invention thus allows the profiling of several classes of biomarkers, and the testing of multiple members of each class, in order to gain insight into the biological mechanisms of a health state and the interaction of such biomarkers. In the preferred embodiment, this encompasses two or more biomarker members per class, more preferably five or more, and most preferably ten or more. As will be appreciated by one skilled in the art, such classes include, without limitation, cytokines and chemokines, such as chemoattractants and inflammatory molecules such as acute phase reactants, signaling molecules, adhesion molecules, biomarkers of immunity (including subclasses, such as those related to individual immune cell lines such as macrophages, T-cells, neutrophils, eosinophils, etc), biomarkers of angiogenesis and endothelial function, and biomarkers of glucose and lipid metabolism and energy storage. Several of these classes overlap, in particular with respect to the cytokine, chemokine, and growth factor members of each. Selected representative examples of such classes and their members are given in the table below, without limiting the foregoing in any way.

Examples of Classes of Molecules	Examples of Genes and Molecules in the Class
Acute Phase Reactants	SAA1, CRP, IL1, IL6, IL8, TNFA, FTL, A2M, MBL, SAP
Angiogenesis & Endothelial Function	VEGF, CD36, ANG1, ANG2/ANGPT2, ENG, FGF2, PDGF
Cell Adhesion	ICAM, DPP4/CD26, CD38, SELE, SELP, CD62L, VCAM, ITGA1, ITGA2, ITGA4, ITGAL, ITGAX, ITGB1, ITGB2, ITGB3
Cell Proliferation & Death	AKT1, CASP2, CASP8, CASP9, IGF, TNF/TNFA, TNFR1, TNFSF10, TNFSF11, CDK2, FAS, FASLG
Chemokines	CCL1, MCP-1/CCL2, CCL3, CCL4, CCL5, CCL6, MCP-3/CCL7, MCP-2/CCL8, CCL9, CCL11, CCL12, MCP-4/CCL13, CCL19, CCL21, CCL24, CCL26, CCL27, CXCL1, CXCL2, IP10/CXCL10, IL8, CX3CL1
Cytokines	IL1B, IL1RN, IL2, IL3, IL4, IL5, IL6, IL8, IL10, IL12, IL12B, IL13, IL18, BTC, TGFA, TGFB, TNF, CSF1, CSF2, CSF3, IFNG

-continued

Examples of Classes of Molecules	Examples of Genes and Molecules in the Class
Coagulation	C2, C3, C4, C5, C9, C1, F2, F12, PROC, PROS1, SERPING1, FGA, VWF, D-dimer
Growth Factors & Hormones	EGF, GH1, NGFB, ADIPOQ, IGF, CSF1, CSF2, CSF3, PDGF, EPO, FGF2, GDF8, GDF9, GH1, IGF1, TGFB1, TPO, EFG, HGF, FGF, IGF, BMP1, BMP2, BMP3, BMP7
Inflammation	CSF1, CSF2, CSF3, IFNG, CD40LG, CD40, C3, C5A, TNF, IL1, IL8, SELP,
Lipid Metabolism	Lipoprotein(a), LEP, ADIPOQ, AGRP, NPY
Energy Homeostasis	INS, glucose, HBA1c, C-peptide, IGF-1, AKT2
Proteolysis	MMP2, MMP9, SERPINA1, heparin, SERPIND1, PAI-1/SERPINE1, TIMP1, TIMP2, CASP3

[0073] Another key aspect of the invention is, in a preferred embodiment, utilizing a single molecule detector, with the ability to range multiple orders of concentration magnitude by using the stochastic and quantum nature of single molecule detection. In particular, biomarkers within the plasma proteome, including many of those cited above, are known to span many orders of magnitude in their molar concentration, as seen in the literature. Without limitation of the foregoing, a review of such concentrations cited from literature for cardiovascular and cancer related plasma proteins is described in Anderson, "Candidate-Based Proteomics in the Search for Biomarkers of Cardiovascular Disease", *J Physiol* 563.1 pp 23-60 (2005), and Anderson, "A List of Candidate Cancer Biomarkers for Targeted Proteomics", *Biomarker Insights* 2: 1-48 (2006), which are hereby incorporated by reference herein in their entireties. As shown in the table below and in FIG. 6, this range of concentrations rapidly approaches single molecule requirements, particularly when combined with the smaller volume samples commonly available in legacy clinical sample sets.

Concentration of 50 kDa molecule	pg/ml L	amol/ml L	Molecules/mL
50 mg/mL	50,000,000,000	1,000,000,000	6.02×10^{17}
10 mg/mL	10,000,000,000	200,000,000	1.02×10^{17}
1 mg/mL	1,000,000,000	20,000,000	1.02×10^{16}
100 ug/mL	100,000,000	2,000,000	1.02×10^{14}
10 ug/mL	10,000,000	200,000	1.02×10^{14}
1 ug/mL	1,000,000	20,000	1.02×10^{13}
100 ng/mL	100,000	2,000	1.02×10^{12}
10 ng/mL	10,000	200	1.02×10^{11}
1 ng/mL	1,000	20	1.02×10^{10}
100 pg/mL	100	2	1.02×10^9
10 pg/mL	10	0.2	1.02×10^8
1 pg/mL	1	0.02	1.02×10^7

[0074] Concentration ranges of common biomarkers within the plasma proteome, indicating the disagreement of biomarker discovery technology such as mass spec across sample sets in the literature are also shown in Anderson et al., "The Human Plasma Proteome: History, Character, and Diagnostic Prospects", *Molecular & Cellular Proteomics* 1.11, pp. 845-867 (2002) and Anderson et al., "The Human Plasma Proteome: A Nonredundant List Developed By Combination of Four Separate Sources", *Molecular & Cellular Proteomics* 3.4, pp. 311-326 (2004), which are hereby incorporated by reference herein in their entireties. Such disagreement further demonstrates the different detection

system needs inherent when encountering broad concentration ranges, which may occur both across many analytes and across many differing health states. FIG. 5 demonstrates the practice of the invention across multiple orders of magnitude in concentration, and across representative biomarkers of each of the aforementioned classes.

PRACTICE OF THE INVENTION

[0075] In a preferred embodiment, the biomarker levels are measured from the clinical sample(s) through the use of a single molecule detector. Suitable single molecule detection equipment is described in U.S. Patent Application Publication Nos. 2004/0166514 A1, 2005/0164205 A1, and 2006/0003333 A1, the disclosures of which are hereby incorporated by reference herein in their entireties. Other examples of single molecule detectors that can be used in accordance with preferred embodiments of the present invention are described in U.S. Patent Application Publication No. 2005/0221408, PCT Publication No. WO 2005/089524, and Richard Brown et al., "Review of Techniques for Single Molecule Detection in Biological Applications, National Physical Laboratory Report, 2001, the disclosures of which are hereby incorporated by reference herein in their entireties. Generally, a single molecule detector operates under the principle that the ultimate, and desired, detection of biomarker information occurs at the level of individual molecules, interactions between molecules, and molecular complexes. Such individual molecules, molecular interactions, and/or molecular complexes can be detected by flow cytometry, single molecule electrophoresis, ion-channel switch membrane biosensor, or other single-molecule analytical instrumentation. Single molecule information can be cumulated over multiple molecular events, providing dynamic quantification of biomarker levels within a clinical sample, allowing the sparing use of very small samples. Data acquisition of such events may be halted when a sufficient number of events are received within a given sample volume to reliably quantitate (e.g. reliably here meaning with a coefficient of variation of 20% or less) a given biomarker's concentration using a presumed Poisson or binomial probability distribution function, as known by one skilled in the art. Such dynamic quantitation of very small sample volumes is a key aspect of the invention as practiced using single molecule detectors.

[0076] Accordingly, embodiments of the present invention contemplate the specific application of single molecule detection to the development of diagnostic tests based on legacy clinical sample sets. Namely, it has been determined

by the present inventors that single molecule detection can detect the presence of biomarker or levels thereof with a suitable sensitivity using only about 1 μL or less of sample per single-biomarker immunoassay (for example). Any suitable analyte recognition unit (e.g., antibodies, aptamers, molecular imprints, probes, primers etc. which have differentially greater affinity for a biomarker of interest) and signal detection technique can be used with a single molecule detection reader in accordance with the present invention. Additionally, it will be understood that the present invention is not limited to the use of immunoassays. Thus, for example, to develop a diagnostic test based on an initial subset of (for example) 300 biomarkers, the use of single molecule detection allows requires a sample size of only about 0.3 mL (i.e., 1 μL per assay*300 biomarkers), or about 0.9 mL if the assays are done in triplicate. The assay may use a 96-well, 384-well format or any other suitable assay configuration. Any multiplexing within the assay will only further reduce the required sample size. The present inventors have applied this knowledge to the discovery that diagnostic tests can be developed based on legacy clinical samples which, as described, are typically available in sizes of 0.05 to 1.0 mL or less. Additional details regarding an illustrative single molecule detection system are provided below.

[0077] In some embodiments, the single molecule detection system can rely on single-molecule fluorescence. Thus, in such embodiments, no polymerases, enzymes or proteins, or any amplification processes are necessary so sample preparation times and complexity are minimal. In other embodiments, the single molecule detection may utilize labeled antibodies. Such labels for individual antibody (or other suitable biomarker recognition units) may themselves be constructed of a plurality of individual fluorescent molecules, further amplifying the signal derived from each single complex multi-fold, and further reducing the detection technique requirements for single molecule detection (such multiplexing of fluorophores may be achieved using beads, dendrimers, polysaccharides and other natural and synthetic polymers, amongst other techniques well described in the art). In one embodiment, the basic detection apparatus may comprise one or two lasers (or a single laser source split into two beams), focusing light-collection optics, one or two single photon detectors, and detection electronics under computer control. FIGS. 1 and 2 are illustrative diagrams of a single molecule detector in accordance with an illustrative, but non-limiting, embodiment of the present invention. A sample compartment is also included and may comprise two reservoirs that hold the solution being analyzed. The reservoirs can be connected by tubing to a glass capillary cell.

[0078] The system also may include a glass capillary flow cell. For example, two laser beams (5 μm in diameter) are optically focused about 100 μm apart and perpendicular to the length of the sample-filled capillary tube. The lasers generally are operated at particular wavelengths depending upon the nature of the detection probe to be excited. An interrogation volume of the detection system may be determined by the diameter of the laser beam and by the segment of the laser beam selected by the optics that direct light to the detectors. The interrogation volume is preferably set such that, with an appropriate sample concentration, single molecules (such as single biomarker-recognition unit hybrids, single nucleic acid probes or single probe-target hybrids) are

present in the interrogation volume during each time interval over which observations are made. Another embodiment of an apparatus for use in accordance with the present invention uses the same capillary flow cell and detection system, but only uses a single laser beam and detector.

[0079] With the above-described instrument configuration (5 μm laser beam) approximately 0.25% of the fluorescent molecules in the solution pass through the laser beams and are typically detectable. This percentage can be increased by configuring each laser beam such that it forms a narrow band perpendicular to the length of the capillary. Such an arrangement can raise the percentage of detectable molecules to approximately 5% of the molecules in the solution. Other configurations illuminating larger areas of the capillary have been calculated to enable detection of up to (for example) 50% of the fluorescent molecules present in a sample. The device has the capability of detecting single molecules in real time, allowing the detection of a fixed number of counts independent of time, and enabling dynamic quantification and concentration range finding during the course of the initial detection period. This feature allows faster readouts of samples as setting a count threshold (for example, at 1000 molecular events or such other effective level, giving a statistically valid quantitation of a biomarker within a sample) is often much faster than a fixed time point (1 minute). For higher biomarker concentrations, preparatory sample dilution may nonetheless be required in order to avoid reaching the count threshold too rapidly in such single molecule detector configurations.

[0080] FIG. 3 is a flowchart of illustrative, exemplary stages involved in developing a diagnostic test in accordance with some of the embodiments of the present invention, including: identification of biomarker candidates, sourcing of reagents, assay development, procurement of clinical samples, interrogation of clinical samples with biomarker assays, and analysis of the data to identify predictive markers and incorporate the results into predictive tests. These illustrative stages are described in greater detail below.

[0081] Identify biomarkers: Biomarkers may be identified by way of a comprehensive search through scientific and patent literature, supplemented with expert review. Based on an understanding of biological mechanisms associated with progression in a given disease area, standard search terms are developed to generate disease-specific databases containing typically thousands of journal articles and hundreds of patents. Canonical pathways, homology, and linkage studies are alternative means of identifying putative biomarkers for a given disease state, as are cell line and animal experiments utilizing mRNA expression under response to stimuli, active agents (drugs, siRNAs, etc.), or in disease-specific organisms (knock-outs, nude mice, ApoE deficient mice, etc.) as are well known to those versed in the art of biomarker discovery. Analytical techniques on larger sample volumes, or pooled sample volumes, may also be used as in Granger, et al. Discovery of Proteins Related to Coronary Artery Disease Using Industrial-Scale Proteomics Analysis of Pooled Plasma, American Heart Journal v152 (3) September 2006, which is hereby incorporated by reference herein in its entirety. Each article and patent is read to identify candidates which are organized in a spreadsheet. For each biomarker, standardized nomenclature derived from human genome databases is applied to eliminate redundancy and enter standardized annotations.

[0082] A score for evidence level is assigned to prioritize the potential value of each biomarker based on experimental data. The evidence level may be combined with protein cellular expression localization to create an overall prioritized list of biomarkers for each disease. At the end of this process, the list of candidates is typically 150-400 biomarkers, but may be more or less. Illustrative lists of biomarkers for use in developing diagnostic tests for diabetes and osteoporosis are described in U.S. Provisional Patent Application Nos. 60/725,462, filed Oct. 11, 2005, 60/771,077, filed Feb. 6, 2006, Ser. No. 11/546,874, filed Oct. 11, 2006, Ser. No. 11/703,400, filed Feb. 6, 2007, and U.S. application Ser. No. 11/788,260, filed Apr. 18, 2007, titled "Diabetes-Associated Markers and Methods of Use Thereof" and bearing attorney docket no. 24748-502 CIP, which are all hereby incorporated by reference herein in their entireties.

[0083] Source Reagents: Table 1 below shows a large and diverse array of vendors that may be used to source immunoreagents as a starting point for assay development. Using the prioritized list of markers, a search for capture antibodies, detection antibodies, and analytes may be performed that can be used to configure a working sandwich immunassay.

[0084] For example, in one disease area, diabetes, 156 of 208 biomarkers were successfully sourced. Depending on the specific disease area, it is anticipated that anywhere from 50 to 80% of the biomarkers on any list are available from commercial sources. The reagents are ordered and received into inventory.

TABLE 1

Immunoreagent Vendors
Company
Abazyme
AbCam
AbGent
AbKem
Abnova
Absea Biotechnology
Academy Biomed
Accurate Chemical and Scientific Corporation
Acris
Advanced Immunochemical, Inc.
Advanced Targeting Systems
Affibody
Affiniti Research Products Limited
Affinity Biologicals
Affinity Bioreagents
Alexis Biochemicals
Alomone Labs
Alpha Diagnostic Intl.
AlphaGenix
American Diagnostica Inc.
American Qualex
American Research Products
American Type Culture Collection
Anaspec
ANAWA Trading SA
Ancell
AngioBio
Angio-Proteomie
Aniara
Anogen
Antibodies Incorporated
AntibodyBen
AntibodyShop
Apotech
APTEC Diagnostics

TABLE 1-continued

Immunoreagent Vendors
Company
Araclon Biotech
Assay Designs
Athens Research and Technology
Austral Biologicals
Aves Labs
Aviva Antibody
Axxora
Babraham Technix
Bachem
Beckman Coulter, Inc.
Bender Medsystems
Bethyl Laboratories
Bio Research Canada
BioCore
BioCytex
Biodesign International
Biogenesis
BioGenex
BioLegend
Biomarket
Biomeda Corporation
Biomedical Technologies
BIOMOL International
BioProcessing
Biosense Laboratories
BioSeptra
Biosonda
BioSource International
BiosPacific
Biostride
Biotrend
Biovendor Laboratory Medicine
Biovet
BMA Biomedical
Boston Biochem
Brendan Scientific
Calbiochem
Caltag
Cambio
CanAg Diagnostics
Capralogics
Capricorn Products
Cayman Chemical Company
Cedarlane Laboratories
Cell Marque
Cell Sciences
Cell Signaling Technology
Cemines
Chang Bioscience
Chemicon International
Chemokine
Clonogene
Clontech
Cortex Biochem
Covance Research Products
Cytolab
Cytopulse
Cytoshop
CytoStore
DAKO
Deltabiolabs
Development Studies Hybridoma Bank
Diaclone
Diagnostic BioSystems
Diagnostic Systems Laboratory
Diasorin
Diatec
Dolfin
Dutch Diagnostics
East Coast Biologics
eBioscience
Echelon Research Laboratorie
ECM Biosciences

TABLE 1-continued

Immunoreagent Vendors
Company
EnCor Biotechnology
Endocrine Technologies
Enzo Biochem
Epitomics
Euroclone
Euro-Diagnostica
Eurogentec
Everest Biotech
Exalpha
EXBIO Praha
EY Laboratories
FabGennix Int.
Fitzgerald Industries International
Fortron Bio Science
Fusion Antibodies
FutureImmune Immunologic Technical and Consulting Services
Gallus Immunotech
G-Biosciences
GEMAC
Genesis Biotech
G-Biosciences
Genex
Genhot Laboratories
Genway Biotech
GloboZymes
Good Biotech
Green Mountain Antibodies
Groovy Blue Genes Biotech
Haematologic Technologies
Hampton Research
Histoline Laboratoires
HyCult Biotechnology
HyTest
IBL
IBT
IDS
Imgenex
IMMCO Diagnostics
Immunodetect
Immunodiagnostik
ImmunoGlobe Antikoerpertechnik
ImmunoKontakt
ImmunologicalsDirect
Immunology Consultants Laboratory
Immunometrics
Immuno-Precise Services
Immunostar
Immunostep
ImmunoTools
Immunovision
Immuquest
Biogenex
Innova Biosciences
Innovation Automation
Innovex
Insight Biotechnology
International Enzymes
Invitek
Invitrogen
IQ Products
Isconova
ISL (Immune Systems Ltd)
Jackson ImmunoResearch Laboratory
KCH Scientific
Kirkegaard & Perry Laboratorie
KMI Diagnostics
Koma Biotech
Kordia Laboratory Supplies
Lab Vision Corporation
LabFrontier Life Science Institute
LAE Biotechnology Company
Lampire Biological Laborator
Lee Laboratories

TABLE 1-continued

Immunoreagent Vendors
Company
Leinco Technologies
Lifescreeen
Linco Research
Maine Biotechnology Services
MBL International
Mediclone
Medix Biochemica
MedSystems Diagnostics GmbH
MicroPharm Ltd.
MilleGen
MitoSciences
MoBiTec
ModiQuest
Molecular Innovations
Molecular Probes
MP Biomedicals
Mubio Products
NatuTec
Neoclone
Neuromics
New England Biolabs
Nordic Immunological Laboratories
Norrin Laboratories
Novocastra
Novus Biologicals
OEM Concepts, Inc.
Oncogene Research Products
Open Biosystems
Orbigen
Oxford Biotechnology
Pacific Immunology
Pall Corporation
Panvera
PBL Biomedical Laboratories
Peprotech, Inc.
PerkinElmer Life Sciences
Perseus Proteomics
Pharmingen
Phoenix Pharmaceuticals
PickCell Laboratories
Pierce Chemical Company
PlasmaLab International, Inc.
Polymun Scientific
Polysciences, Inc.
PRF&L
Pro-Chem
Progen
Promab Biotechnologies
Promega Corporation
ProSci
Proteogenix
Protos Immunoresearch
QED Biosciences, Inc.
Quidel Corporation
R&D Systems
Randox
Repligen
Research Diagnostics
Roboscreen
Rockland Immunochemicals
Rose Biotech
Santa Cruz Biotechnology
SCIpac
Scottish Agricultural Science
ScyTek Laboratories
Seikagaku America
Seramon
Serological Corporation
Serotec
SigmaAldrich
Signature Immunologics
Signet Laboratories
Silver Lake Research

TABLE 1-continued

Immunoreagent Vendors
Company
Southern Biotechnology Associates
SPI-BIO
Statens Serum Institut
StemCell Technologies
Sterogene Bioseparations
Strategic Biosolutions
Stressgen
Structure Probe, Inc. (SPI)
SWant
Synaptic Systems GmbH
SynthOrg Biochemicals, Ltd.
Technopharm
Terra Nova Biotechnology
Tetra Link International
The Biotech Source
TiterMax
Transmissible Spongiform Encephalopathy Research Center
Trevigen
Trillium Diagnostics
Triple Point Biologics
Tulip Biolabs
Union Stem Cell & Gene Engineering Company
Upstate Biotechnology
US Biological
Vector Laboratories
Ventana Medical Systems, Inc
Vision BioSystems
Wako Pure Chemical Industrie
WolvoBiotech Company
Zeptomatrix

[0085] Develop Immunoassays: Immunoassays are preferably developed in three steps, Prototyping, Validation, and Kit Release.

[0086] Prototyping: Prototyping may be done using standard ELISA formats if the two antibodies used in the assay are from different host species. Using standard conditions, anti-host secondary antibodies conjugated with horse radish peroxidase are evaluated in a standard curve. If a good standard curve is detected, the assay proceeds to the next step. Assays that have same host antibodies go directly to the next step (i.e., mouse monoclonal sandwich assays).

[0087] Validation: Validation of a working assay may be performed using single molecule detection technology. The detection antibody is first conjugated to fluorescent molecules, typically Alexa 647. The conjugations use standard NHS ester chemistry, for example, according to the manufacturer. Once the antibody is labeled, the assay is tested in a sandwich assay format using standard conditions. Each assay well is solubilized in a denaturing buffer, and the material read on the single molecule detection platform.

[0088] FIG. 2 shows a typical result for a working standard curve. Once a working standard curve is demonstrated, the assay may be applied to 24 serum samples (for example) to determine the normal distribution of the target analyte across clinical samples. The amount of serum required to measure the biomarker within the linear dynamic range of the assay is determined, and the assay proceeds to kit release. In the present example, based on 39 validated assays, 0.004 microliters are used per well on average.

[0089] Kit Release: Each component of the kit including manufacturer, catalog numbers, lot numbers, stock and working concentrations, standard curve, and serum requirements may be compiled into a standard operating procedures

for each biomarker assay. This kit may then be released for use in testing clinical research samples.

[0090] Acquiring Clinical Samples: Depending on the specification of the diagnostic test being developed, the clinical samples preferably have (for example) clinical annotations that track progression of disease, and preferably also include measurements of underlying mechanisms or disease phenotypes, and/or have disease outcomes using longitudinal samples over time. Relationships with the investigators may then be developed, and a contractual agreement is put into place. For each clinical study, the typical volumes range from 0.1 to 1 mL.

[0091] Import Clinical Annotations: Samples arrive frozen on dry ice, and each sample is stored at -80 C. Each sample typically has tens to hundreds of clinical annotations associated with it. The clinical annotations associated with each sample set may be brought into a standardized nomenclature prior to import. All of the clinical annotations associated with each sample are then imported into a relational database.

[0092] Prepare Clinical Samples: The frozen aliquots are thawed and aliquotted for use in the laboratory. Each clinical sample is thawed on ice, and aliquots are dispensed into barcoded tubes (daughter tubes). Each daughter tube is stored at -80 C until it is needed for immunoassays. The daughter tubes are then arrayed into sample plates. Each barcoded daughter tube to be assayed is arrayed into bar-coded 96 or 384 well plates (sample plates). This daughter tube to sample plate well mapping is tracked by the relational database.

[0093] Run Immunoassays: Each sample plate is now prepared for immunoassays. In one example, 384 well barcoded assay plates may be dedicated to one biomarker per plate. Typically, 4-12 assay plates are derived from each sample plate dependent on the amount of serum required for each assay. The sample plate goes through a series of dilutions to ensure that the clinical samples are at an appropriate dilution for each immunoassay. The clinical samples are then deposited into the assay plate wells in triplicate for each marker. Again, tracking of each sample plate well to assay plate well is tracked in the relational database. The assays may then be processed using standard immunoassay procedures, and the assay plate is read on a single molecule detection instrument. Each run contains data for a single biomarker across multiple clinical samples, typically around one hundred. The resulting data files may then be imported back into the relational database, where standard curves can be calculated and the concentration values for each biomarker for each sample can be calculated. FIG. 3 shows an example of single molecule detection data across 92 samples for 25 biomarkers.

[0094] Analyze Data: The quantitative biomarker data can now be correlated to the clinical annotations associated with each sample. Any number of statistical formula or machine learning approaches on single or multiple markers can be used to identify disease states or risk for disease or biomarker patterns that have commercial potential to diagnose or prognose disease state (for example).

[0095] The following is an illustrative example of a Standard Operating Procedure (SOP) for use in developing diagnostic tests in accordance with an embodiment of the present invention.

Assay Analyte: C-Reactive Protein

Components:

[0096]

Component	Vendor	Catalog Number	Lot Number
C-Reactive Protein	US Biologicals	C7907-26A	L5042910
Capture Antibody	US Biologicals	C7907-09	L4030562
Detection Antibody	US Biologicals	C7907-10	L2121306M

[0097] 1. Plate Coating: Coat and Block immunoassay plates for analyte capture

[0098] 1.1 Materials

[0099] 1.1.1. NUNC Maxisorp 384 well plates, Cat. No. 460518

[0100] 1.1.2. NUNC Acetate Sealers, Cat. No. 235306

[0101] 1.1.3. Coating buffer

[0102] 1.1.3.1. 0.05 M carbonate, pH 9.6

[0103] 1.1.3.2. Store at 4° C. for up to 2 months

[0104] 1.1.4. Capture Antibody

[0105] 1.1.5. Wash buffer A

[0106] 1.1.5.1. PBS with 0.1% TWEEN 20

[0107] 1.1.5.2. Store at room temperature for up to 2 months

[0108] 1.1.6. Blocking buffer

[0109] 1.1.6.1. 1% BSA, 5% sucrose, 0.05% NaN₃ in PBS

[0110] 1.1.6.2. Store at 4° C. for up to 1 month

[0111] 1.1.7. Microplate washer

[0112] 1.2. Procedure

[0113] 1.2.1. Dilute capture antibody to 1 microgram/mL in coating buffer. (Prepare immediately before use)

[0114] 1.2.2. Add 20 microliters of diluted capture antibody per well

[0115] 1.2.3. Seal and shake for 2 minutes on plate shaker

[0116] 1.2.4. Centrifuge 1000rpm 2 min, 25° C.

[0117] 1.2.5. Incubate overnight at room temperature (no shaking)

[0118] 1.2.6. Wash 3× with 100 microliters wash buffer A

[0119] 1.2.7. Add 30 microliters blocking buffer per well

[0120] 1.2.8. Seal and shake for 2 minutes on Jitterbug setting 7

[0121] 1.2.9. Centrifuge 1000rpm 2 min, 25° C.

[0122] 1.2.10. Incubate at least 2 hour at room temperature (no shaking)

[0123] 1.2.11. Dump plate and blot upside down (no wash)

[0124] 1.2.12. Air dry the blocked plates (uncovered) at least 5 hours at room temperature

[0125] 1.2.13. Cover the dry plates with acetate sealer

[0126] 1.2.14. Store at 4° C. for up to one month

[0127] 2. Single Molecule Detection Assay: Add clinical samples to coated plates and quantify

[0128] 2.1. Materials

[0129] 2.1.1. Coated, blocked NUNC Maxisorp 384 well plate

[0130] 2.1.2. NUNC Acetate Sealers, Cat. No. 235306

[0131] 2.1.3. Assay buffer

[0132] 2.1.3.1. BS* with 1% BSA, 0.1% TRITON X-100

[0133] 2.1.3.2. Store 4° C. for up to 1 month

[0134] 2.1.4. Standard Calibrator diluent

[0135] 2.1.4.1. Assay buffer+additional 5% BSA,

[0136] 2.1.4.2. Enough volume for standard curve, including 0 pg/ml.

[0137] 2.1.4.3. Make fresh for use.

[0138] 2.1.5. Standard Molecule Control

[0139] 2.1.6. Detection Antibody: A647 labeled antibody

[0140] 2.1.7. Assay Wash buffer B

[0141] 2.1.7.1. BS* with 0.02% TRITON X-100 and 0.001% BSA

[0142] 2.1.7.2. 500 ml per assay plate

[0143] 2.1.7.3. Store at 4° C. for up to 1 month

[0144] 2.1.8. Elution Buffer

[0145] 2.1.8.1. 4 M urea, 1× BS with 0.02% TRITON X-100 and 0.001% BSA

[0146] 2.1.8.2. Approx 8 ml per assay plate

[0147] 2.1.9. Microplate shaker (Jitterbug), set at "7"

[0148] 2.1.10. Microplate washer

[0149] 2.1.11. Centrifuge

[0150] 2.2. Procedure

[0151] 2.3. Record

[0152] 2.3.1. Plate assay plate number, kit lot number, and sample plates used

[0153] 2.4. Standard Curve

[0154] 2.4.1. Dilute control to 100 ng/ml in calibrator

[0155] 2.4.2. Prepare ½ serial dilutions from 100 ng/ml to 0.01 pg/ml in calibrator diluent

[0156] 2.5. Sample Dilution

[0157] 2.5.1. Dilute samples 1:400 in assay buffer

[0158] 2.6. Capture and Detection

[0159] 2.6.1. Add 20 microliters/well of standards

[0160] 2.6.2. Add 20 microliters/well diluted unknowns

[0161] 2.6.3. Seal w/ acetate sealing tape. Shake for 2 minutes on plate shaker

[0162] 2.6.4. Incubate overnight at room temperature

[0163] 2.6.5. Dilute Detection antibody labeled A647 antibody to 50 ng/ml in assay buffer.

[0164] 2.6.6. Aspirate. Wash 5× with 100 ul wash buffer B

[0165] 2.6.7. Blot upside down

[0166] 2.6.8. Add 20 microliters/well diluted detection antibody

[0167] 2.6.9. Seal w/ acetate sealing tape. Shake for 2 minutes on plate shaker

[0168] 2.6.10. Incubate 2 hours at room temperature

[0169] 2.6.11. Aspirate. Wash 5× with 100 ul wash buffer B

[0170] 2.6.12. Blot upside down

[0171] 2.6.13. Add 20 microliters/well elution buffer.

[0172] 2.6.14. Seal w/ acetate sealing tape. Shake for 2 minutes on plate shaker.

[0173] 2.6.15. Incubate ½ hour at 25° C.

[0174] 2.6.16. Centrifuge on 1000 rpm for 2 min, 25° C.

[0175] 2.7. Analyze on Single Molecule Detection instrument

Other Embodiments

[0176] It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

1. A single detection means of evaluating the health state of a human subject, comprising obtaining a measurement of a least one clinical biomarker from at least one live clinical sample isolated from said human subject, and inputting said measurement(s) into a model that calculates an output value correlated to said health state, the improvement comprising using as said model an algorithm that was developed by measurement of multiple development biomarkers comprising said clinical biomarker(s) from at least one legacy clinical sample set annotated for said health state, said measurement comprising use of legacy clinical samples having a sample volume of 1 ml or less, and analyzing said measurement of multiple biomarkers for an association with said health state.

2. The means of claim 1, wherein said measurement of multiple development biomarkers comprises measuring at least two biomarkers from said sample volume.

3. The means of claim 2, wherein said measurement of multiple development biomarkers comprises measuring at least 10 biomarkers from a sample volume less than about 0.5 milliliters.

4. The means of claim 2, wherein said measurement of multiple development biomarkers comprises measuring at least 20 biomarkers from said sample volume.

5. The means of claim 2, wherein said measurement of multiple development biomarkers comprises measuring at least 100 biomarkers from said sample volume.

6. The means of claim 2, wherein said measurement of multiple development biomarkers comprises measuring at least 200 biomarkers from said sample volume.

7. The means of claim 2, wherein said measurement of multiple development biomarkers comprises measuring at least 300 biomarkers from said sample volume.

8. The means of claim 1, wherein said measurement of multiple development biomarkers comprises, for each biomarker, measuring said biomarker in an assay of said legacy clinical sample, wherein said assay used about 1 microliter (μL) or less of said sample volume for each biomarker.

9. The means of claim 8, wherein at least 10 development biomarkers are measured per legacy clinical sample.

10. The means of claim 8, wherein said measurement of multiple development biomarkers uses single molecule detection to measure said multiple development biomarkers in said legacy clinical sample.

11. The means of claim 10, wherein said measurement of biomarkers in the legacy clinical sample by single molecule detection consists of dynamic quantitation.

12. The means of claim 1, wherein said health state is the presence or absence of a disease.

13. The means of claim 1, wherein said health state is the pre-disease or pre-disease condition.

14. The means of claim 1, wherein said health state is the risk of developing a disease.

15. The means of claim 12, wherein said absence of a disease is further defined to be a normal state or pre-disease state.

16. The means of claim 1, wherein said biomarkers comprise traditional laboratory risk factors.

17. The means of claim 1, wherein said live clinical sample isolated from said human subject is whole blood, serum, plasma, blood cells, endothelial, cells, tissue biopsies, lymphatic fluid, ascites fluid, interstitial fluid, bone marrow, cerebrospinal fluid, saliva, sputum, sweat, or urine.

18. The means of claim 17, wherein said live clinical sample is plasma or serum.

19. The means of claim 17, wherein said sample is from a human subject undergoing one or more treatment regimens.

20. The means of claim 19, wherein said treatment regimens are selected from a group consisting of therapeutics, prophylactics, exercise regimens, dietary supplementation, weight loss, surgical intervention, device implantation and exercise regimens.

* * * * *

专利名称(译)	基于来自遗留临床样本集的生物标记信息开发诊断测试的系统和方法		
公开(公告)号	US20180267053A1	公开(公告)日	2018-09-20
申请号	US15/918489	申请日	2018-03-12
[标]申请(专利权)人(译)	真健康有限责任公司		
申请(专利权)人(译)	真健康有限责任公司		
当前申请(专利权)人(译)	真健康有限责任公司		
[标]发明人	URDEA MICHAEL S MCKENNA MICHAEL P		
发明人	URDEA, MICHAEL S. MCKENNA, MICHAEL P.		
IPC分类号	G01N33/68 G06F19/18 C12Q1/6883 G01N33/53		
CPC分类号	G06F19/18 G01N2800/60 G01N2800/56 C12Q1/6883 G01N33/6842 G01N33/5302 G01N2570/00 G16B20/00		
优先权	12/300019 2009-06-29 US PCT/US2007/011196 2007-05-08 WO 60/798867 2006-05-08 US		
外部链接	Espacenet USPTO		

摘要(译)

公开了用于基于来自遗留临床样品组的生物标记物信息开发诊断测试 (例如, 检测, 筛选, 监测和预后测试) 的系统和方法, 其中仅小样品体积 (例如, 约0.05至约1.0mL或更低)。样品) 通常是可用的。例如, 可以通过使用单分子检测在临床样品中检测生物标志物 (例如, 约 10,50,100,150,200,300或更多), 并且可以在包括约1的分析中检测每种生物标志物。或更少的遗留临床样本。

Single Molecule Detector

