

(51) International Patent Classification:
C12Q 1/68 (2006.01)(21) International Application Number:
PCT/US2012/041385(22) International Filing Date:
7 June 2012 (07.06.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/494,831 8 June 2011 (08.06.2011) US(71) Applicant (for all designated States except US): **ALT-HEADX INCORPORATED** [US/US]; 3550 Dunhill Street, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **RHODES, Daniel, R.** [US/US]; 5528 Gallery Park Drive, Ann Arbor, MI 48103 (US). **MONFORTE, Joseph, A.** [US/US]; 1 Willow Lane, Kensington, CA 94707 (US).(74) Agent: **ELAMRANI, Samir**; Wilson Sonsini Goddard & Rosati, 650 Page Mill Road, Palo Alto, CA 94304 (US).

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

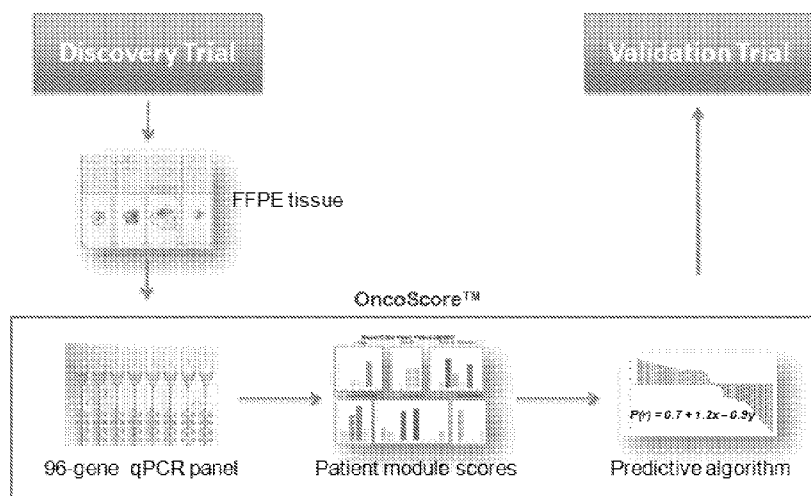
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE,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

Published:

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

(54) Title: DISEASE CLASSIFICATION MODULES

FIGURE 5



(57) Abstract: The embodiments herein provide methods and modules for predicting one or more outcomes of a disease in a patient, as well as predicting the effectiveness of clinical intervention. The methods and modules herein provide for personalized medicine which includes genomics- based methods and modules for making clinical predictions. Methods and models which can make more than one clinical prediction from a single diagnostic method are described herein, which can save costs in development of products and testing methods.



DISEASE CLASSIFICATION MODULES

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/494,831, filed 8 June 2011, which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] The generalized clinical paradigm involved making treatment decisions for a patient based on “standards of care” as determined by the results of studies performed on populations of other patients. Some diagnostic information from the patient may be included, but the assumption is that on average the patient will respond like other patients. This paradigm fails to take into account genetic variability of individuals within a population. As a result, the patient for which predictions are desired is not exactly like any of the patients in the populations of previous studies, and is only compared to an idealized average thereof. As a result, the outcome for the particular patient may vary from the ideal, potentially leading to ineffective treatment, adverse reactions to treatment and the like. The current generalized clinical paradigm requires that the study be designed to predict a limited number of clinical outcomes. This is expensive since clinical studies are costly and the development of associated diagnostic products is risky.

SUMMARY OF THE INVENTION

[0003] The embodiments herein provide methods and modules for predicting one or more outcomes of a disease in a patient, as well as predicting the effectiveness of clinical intervention. These predictions can save lives, improve health and avoid the cost and harm done by unnecessary treatments. The methods and modules herein provide for personalized medicine which includes genomics-based methods and modules for making clinical predictions. Methods and models which can make more than one clinical prediction from a single diagnostic method are described herein, which can save costs in development of products and testing methods.

[0004] In one aspect, described herein is a method for classifying a disease, the method comprising: (a) providing a plurality of disease classification Modules, wherein the Modules comprise a list of genetic markers selected by (i) forming a pairwise cluster associations network of genetic markers among a plurality of reference samples and (ii) reducing the cluster association network to a discrete set of Modules defined by highly interconnected clusters; (b) interrogating a biological sample obtained from a patient for presence of one or more genetic markers of one or more of the Modules; and (c) identifying the patient’s disease as belonging to

a Module having genetic markers consistent with the genetic markers interrogated in the biological sample.

[0005] In some embodiments, the genetic markers comprise genes having a similar expression profile across the plurality of reference samples, wherein the biological sample is interrogated for the expression of the genes, and wherein the patient's disease is identified as belonging to a Module based on the pattern of gene expression being consistent with the expression of said one or more gene members of the Module.

[0006] In some embodiments, a cluster of genetic markers is highly interconnected if the Pearson correlation is at least 0.5.

[0007] In some embodiments, the method further comprises predicting drug responsiveness based on identifying the patient's disease as belonging to a Module.

[0008] In some embodiments, the method further comprises predicting the risk of disease recurrence based on identifying the patient's disease as belonging to a Module.

[0009] In some embodiments, the method further comprises predicting the risk of metastasis based on identifying the patient's disease as belonging to a Module.

[0010] In some embodiments, the method further comprises selecting a therapeutic drug for the patient based on identifying the patient's disease as belonging to a Module.

[0011] In some embodiments, each genetic marker is selected for inclusion in a Module based on the marker being present in as little as 1% of the reference samples.

[0012] In some embodiments, each genetic marker is selected for inclusion in a Module based on the marker being present in as little as 5% of the reference samples.

[0013] In some embodiments, the number of reference samples is at least 100.

[0014] In some embodiments, the number of reference samples is at least 1000.

[0015] In some embodiments, the number of reference samples is at least 5000.

[0016] In some embodiments, the plurality of reference samples are selected from a plurality of cohorts.

[0017] In some embodiments, the number of cohorts is at least 3 and wherein the number of reference samples per cohort is at least 50.

[0018] In some embodiments, the reference samples are obtained from at least 5 cohorts.

[0019] In some embodiments, the reference samples are obtained from at least 10 cohorts.

[0020] In some embodiments, the reference samples are obtained from at least 25 cohorts.

[0021] In some embodiments, the reference samples are obtained from at least 50 cohorts.

[0022] In some embodiments, the reference samples are obtained from at least 100 cohorts.

[0023] In some embodiments, the Modules define independent biological functions.

[0024] In some embodiments, the independent biological functions define independent co-expression sets.

[0025] In one aspect, described herein is a diagnostic assay prepared based on one or more disease classification Modules, wherein the Modules comprise a list of genetic markers selected by (i) forming a pairwise cluster associations network of genetic markers among a plurality of reference samples and (ii) reducing the cluster association network to a discrete set of Modules defined by highly interconnected clusters.

[0026] In some embodiments, the assay provides diagnosis in connection with more than one disease state.

[0027] In one aspect, described herein is a disease classification Module comprising a list of genetic markers selected by (i) forming a pairwise cluster associations network of genetic markers among a plurality of reference samples and (ii) reducing the cluster association network to a discrete set of Modules defined by highly interconnected clusters.

[0028] In some embodiments, a cluster of genetic markers is highly interconnected if the Pearson correlation is at least 0.5.

[0029] One embodiment discloses a disease specific module wherein the module comprises a plurality of genetic markers selected first based on their similar expression profile across a plurality of reference samples.

[0030] One embodiment provides a module wherein the genetic markers selected first based on their similar expression profiles are named based on phenotypic information.

[0031] One embodiment provides a module wherein the plurality of reference samples are selected from a plurality of cohorts.

[0032] One embodiment provides a module wherein a genetic marker is selected based on the marker being present in as little as 1% of the reference samples.

[0033] One embodiment provides a module wherein a genetic marker is selected based on the marker being present in as little as 5% of the reference samples.

[0034] One embodiment provides a module wherein the number of reference samples is at least 100.

[0035] One embodiment provides a module wherein the number of reference samples is at least 1000.

[0036] One embodiment provides a module wherein the number of reference samples is at least 5000.

[0037] One embodiment provides a module wherein the reference samples are obtained from at least 5 cohorts.

[0038] One embodiment provides a module wherein the reference samples are obtained from at least 10 cohorts.

[0039] One embodiment provides a module wherein the reference samples are obtained from at least 25 cohorts.

[0040] One embodiment provides a module wherein the reference samples are obtained from at least 50 cohorts.

[0041] One embodiment provides a module wherein the reference samples are obtained from at least 100 cohorts.

[0042] One embodiment provides a module wherein the number of cohorts is at least 3 and wherein the number of reference samples per cohort is at least 50.

[0043] One embodiment provides plurality of modules wherein the modules define independent biological functions.

[0044] One embodiment provides plurality of modules wherein the independent biological functions define independent co-expression sets.

[0045] One embodiment provides a method for predicting a clinical outcome of a test patient, the method comprising comparing the expression profile of biomarkers associated with one or more modules.

[0046] One embodiment provides a method wherein the prediction is selected from the group consisting of drug sensitivity, drug insensitivity, recurrence and metastasis.

[0047] One embodiment provides a method wherein more than one prediction is made.

[0048] One embodiment provides diagnostic assay prepared based on one or more modules.

[0049] One embodiment provides diagnostic assay wherein the assay provides diagnosis in connection with more than one disease state.

BRIEF DESCRIPTION OF THE DRAWINGS

[0050] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0051] Figure 1 shows an example of detection rate for breast cancer samples.

[0052] Figure 2 shows an example average Cq for a number of endogenous controls across breast cancer samples.

[0053] Figure 3 shows an example of Cq values for various samples.

[0054] Figure 4 shows an example of Cq values for various samples.

[0055] Figure 5 shows an example of the methods described herein.

[0056] Figure 6 shows OncoScoresTM for a first exemplary patient.

[0057] Figure 7 shows OncoScoresTM for a second exemplary patient.

[0058] Figure 8 shows four module genes demonstrated expected gene-gene correlations and the genes also displayed correlations with expected clinical characteristics.

[0059] Figure 9 shows an example of Module score clinical trends.

[0060] Figure 10 shows an example of Module-based molecular stratification.

[0061] Figure 11 shows an example of long-term clinical follow-up for 287 patients.

[0062] Figure 12 shows an example of overall survival.

[0063] Figure 13 shows an example of neoadjuvant chemotherapy response.

[0064] Figure 14 shows an example of targeted therapy sensitivity.

INCORPORATION BY REFERENCE

[0065] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

PCT Patent Publication No. WO 2011/068839 A1, filed December 1, 2010, entitled “CLASSIFICATION OF CANCERS” is hereby incorporated by reference in its entirety.

DETAILED DESCRIPTION OF THE INVENTION

[0066] In contrast to the current generalized clinical paradigm, the present disclosure involves classifying diseases into one or more modules first based on genomic expression information. While other genomic methods are known, previous methods generally i) choose a cohort of patients by one or more phenotypic criteria, ii) interrogate the patients for one or more genomic markers and iii) correlate the presence of the marker with one or more clinical outcomes. The present disclosure recognizes that characterization of patients based on phenotype is susceptible to biases of the study and to biases of the particular clinicians performing the study.

Classification based on phenotype also restricts the breadth of genomic information from which improved predictions can be made. Disclosed herein are methods in which patient populations are not restricted based on phenotype.

[0067] Secondly, the present method differs from other genomic methods such as “gene expression profiling” by classifying disease samples into a plurality of modules. Gene expression profiling can stratify diseases such as cancer into molecular subtypes. The general approach has been to perform two-dimensional hierarchical clustering, identify sets of samples that cluster together (i.e., molecular subtypes), and then describe the sets of genes that best

correlate with the sets of samples. There are two main disadvantages to this approach. First, molecular subtypes are subject to study- specific biases such as sampling bias, tissue collection bias (e.g., stromal contamination), technology bias, tissue processing bias, and a host of others. As a result, a molecular subtype defined in one study may not be representative of molecular subtypes in general. Second, most analyses to date have assumed that every disease sample must fall into one molecular subtype, limiting practical associations of prognosis and treatment to a single molecular subtype that might not fully define an individual's disease. Provided herein is a new multi-dimensional approach of classifying diseases such as cancers.

Definitions

[0068] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. For the purposes of the present invention, the following terms are defined below.

[0069] The articles “a” and “an” are used herein to refer to one or to more than one (i.e. at least one) of the grammatical object of the article. By way of example, “an element” means one or more element.

[0070] Throughout this specification, unless the context requires otherwise, the words “comprise,” “comprises,” and “comprising” will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements.

[0071] A “patient” shall mean an individual person having a disease. A patient may be a training patient, test patient, or both training patient and test patient. A patient may be a member of one or more cohorts.

[0072] A “training patient” is known to have a disease, and also is known to have certain outcomes including but not limited to, resistances to certain drugs, sensitivities to certain drugs, metastasis or recurrence events, and the like. A training patient may be a member of one or more cohorts.

[0073] A “test patient” is known to, or suspected to have a disease. A test patient may be compared to “training patients” using the methods of the present invention to predict certain outcomes including but not limited to, resistances to certain drugs, sensitivities to certain drugs, metastasis or recurrence events, and the like.

[0074] The term “biological sample” as used herein refers to a sample that may be obtained and assayed for gene expression. The biological sample can include a biological fluid (e.g., blood, cerebrospinal fluid, urine, plasma), tissue biopsy, and the like. In some embodiments, the sample

is a tissue sample, for example, tumor tissue, and may be fresh, frozen, or archival paraffin embedded tissue.

[0075] The term “gene expression” as used herein refers to the production of a gene product from a gene. A gene product can include, for example, RNA or protein. Gene expression can be measured directly or indirectly using known methods and those disclosed herein. Gene expression, as measured in a biological sample from a subject having cancer, can be modulated as compared to a control.

[0076] The term “coexpression,” as used herein refers to the relation of a gene's expression with the expression of one or more other genes. Genes that are coexpressed have a pattern of expression that is constant relative to one another, and each can be overexpressed, underexpressed, or remain the same relative to a control.

[0077] A “genetic marker” shall mean a molecular-based feature characteristic of, or correlated with a phenotype. Genetic markers can be mutations, gene deletions or duplications, polymorphisms or expression levels for example. Genetic markers can be detected by interrogating the DNA, RNA or proteins of a sample.

[0078] The term “module” as used herein refers to a specific list of coexpressed genes (gene members) useful for classifying a specific disease. For example, the gene members along with gene identification numbers (GenInfo Identifiers) for each gene member obtained from Genbank (<http://www.ncbi.nlm.nih.gov/Genbank/>) of certain cancer modules are disclosed in Tables 1-161 of PCT Patent Publication No. WO 2011/068839 A1, the tables of which are herein incorporated by reference and retain the same Table number reference in the present application. That is, Table 1 of PCT Patent Publication No. WO 2011/068839 A1 is Table 1 of the present application.

[0079] The term “gene expression profile” as used herein refers to an expression pattern of two or more gene members of a particular module. A gene expression profile can include the expression pattern of two or more (e.g., 2, 3, 4, 5, 10, 15, or more) gene members corresponding to the module of interest.

[0080] The term “primer” is defined as an oligonucleotide that, when paired with a strand of DNA, is capable of initiating synthesis of a primer extension product in the presence of a suitable polymerizing agent.

[0081] “Probe” refers to a molecule that binds to a specific sequence or sub-sequence or other moiety of another molecule. Unless otherwise indicated, the term “probe” typically refers to a polynucleotide probe that binds to another polynucleotide, often called the “target polynucleotide”, through complementary base pairing.

[0082] A drug that that shows “activity” against a disease identified as belonging to a particular module means that when the drug is administered to a patient having a disease belonging to the module, the disease exhibits a statistically significant reduction in proliferation or size, or an alteration in any other measurement of a type generally accepted as indicative of disease responsiveness.

[0083] “Drug sensitivity” in some embodiments means an outcome whereby the disease responds favorably (i.e. is cured or results in decreased adverse symptoms) to a drug. In other embodiments, the term “drug sensitivity” has an alternate meaning known to one of skill in the art.

[0084] “Drug insensitivity” in some embodiments means an outcome whereby the disease does not respond favorably (i.e. is not cured or does not result in decreased adverse symptoms) to a drug. In other embodiments, the term “drug insensitivity” has an alternate meaning known to one of skill in the art.

[0085] “Recurrence” in some embodiments means the return of a disease after an interval of time, or an increase in the severity of a disease or its symptoms after an interval of time. In other embodiments, the term “Recurrence” has an alternate meaning known to one of skill in the art.

[0086] “Metastasis” in some embodiments means the spread of a disease from one organ or part to another non-adjacent organ or part. In other embodiments, the term “Metastasis” has an alternate meaning known to one of skill in the art.

[0087] “Biological function” in some embodiments means that functions are associated with modules. In some embodiments “biological function” means that a function is associated with a particular module. In some embodiments, if a gene is a part of a module, it is not also part of another module. Therefore, independent biological functions in some embodiments means that biological functions are represented within independent modules. That is, a particular biological function is represented in a module that is independent from another module in which another biological function is represented. In other embodiments, the term “Biological function” has an alternate meaning known to one of skill in the art.

Gene expression profiles

[0088] A gene expression profile obtained from a patient's biological sample that is consistent with a pattern of gene expression of two or more members of a Module identifies the patient's disease as belonging to that particular Module. In an embodiment, a gene expression profile includes the expression pattern of, for example, 2-3, 2-5, 3-5, 5-8, 6-8, 8-10, 1-10, 1-15, 1-20, 5-10, 5-15, 5-20, 10-15, 10-20, or 15-20 gene members of a selected Module.

[0089] For example, to identify a patient's disease as belonging to a Module, a patient's sample can be interrogated for expression of one or more gene selected from gene members belonging to one or more of the presently identified Cancer Modules (Tables 1-161). In some embodiments, the interrogated genes include gene members from some or all of the Cancer Modules of a given cancer type, for example, breast, colon, brain, and the like. In some embodiments, the interrogated genes will be selected from more than one related cancer type, for example, breast and ovary.

[0090] The genes to be interrogated can be chosen from the members of a particular Module by any appropriate means. Considerations in determining which genes, which modules, and how many genes or modules to interrogate can include the availability of probes or primers for a particular gene member, the number of genes easily tested using the method of choice (e.g., microarray or PCR), the type of disease, the purpose of the analysis, and the like. For example, gene members ranked highest in each module can be selected; gene members associated with a particular metabolic pathway can be selected; or gene members can be randomly selected for interrogation. Modules can be selected for association with a known diagnosis, prognosis, drug response, and the like.

Examples of gene expression profiles

[0091] Examples of gene members of Modules to be interrogated include two or more of any of the gene members listed in the Cancer Module Tables 1-161. For example, for Bladder Cancer Module 1, two or more of any of the gene members listed in Table 1 can be interrogated. For Bladder Cancer Module 2, two or more of any of the gene members listed in Table 2 can be interrogated. Table 162 provides non-limiting examples of gene members that can be interrogated for identifying a patient's cancer as belonging to a specific Cancer Module.

[0092] The gene expression profile can indicate that a patient's cancer belongs to more than one Cancer Module, as described more fully in the Examples below. In some embodiments, interrogation of one gene member may be sufficient to identify the patient's disease as belonging to that module.

Analysis of gene expression profiles

[0093] The classification of a patient's disease as belonging to a Module is determined by interrogating a biological sample for the expression of one or more gene member of the Module and identifying the patient's disease as belonging to the Module if the module has a pattern of gene expression consistent with the pattern of expression of the one or more gene member interrogated in the biological sample.

[0094] Gene expression can be analyzed in a variety of platforms, assays, and methods, and is generally analyzed by amplification and/or detection of mRNA extracted from the subject's biological sample, or by detection of gene expression products such as, for example, cDNA and protein, or by analysis of genomic DNA. A subject's sample can be interrogated for the expression of a gene member of a Module using various known techniques as described below.

[0095] In some embodiments, a tissue other than the subject's diseased tissue can be interrogated for expression of one or more gene member of a Module. For example, a lymph node, blood, serum, or urine can be interrogated for expression of a gene member in a cancer. In such embodiments, presence of a nucleic acid, protein, or cell originating from the cancer can be interrogated in the selected tissue or fluid.

[0096] Prior to interrogation, the biological sample can be processed. Such processing can affect the way the analysis is performed on the sample. For example, formalin-fixed paraffin embedded samples are generally analyzed using different techniques than are used with fresh or frozen samples. Such differences will be apparent to those skilled in the art.

[0097] The sample is interrogated for the expression of at least one gene member of a Module using at least one analytic technique. An analytic technique can directly measure gene expression, for example, by RNA analysis or protein analysis, or can measure gene expression indirectly, for example by genomic analysis. A review of methods for quantitative detection of gene expression can be found, for example, in Nygaard, et al., 2009, Front Biosci, 14:552-69.

[0098] A gene expression profile can be identified from analysis of a sample, and can be compared to one or more Module to identify a patient's disease as belonging to one or more Module.

RNA analysis

[0099] In some embodiments, a biological sample containing RNA originating from a patient's disease can be interrogated using known methods. Biological samples can be purified as appropriate for the analytic technique to be used. Purification techniques can include, for example, laser-capture microdissection to isolate cancer cells from non-cancer tissue. Tumor cells in blood or other biological fluid can be isolated using purification techniques such as antibody-mediated purification, for example, fluorescence activated cell sorting, centrifugation- or gravity-based cell sorting, magnetic activated cell sorting, and the like.

[00100] RNA can be used directly as extracted and purified for analytic techniques such as Northern blotting, which can be used to measure relative RNA expression. Methods for isolating mRNA from a tissue sample for further analysis are described, for example, in Ausubel et al., 2002, Short Protocols in Molecular Biology, 4:4-1 - 4-29. Methods for isolating mRNA from

paraffin embedded tissues in particular are discussed, for example, in Penland, et al., 2007, Laboratory Investigation, 87:383-391. RNA isolation kits are commercially available, including, for example, Paraffin Block RNA Isolations Kits available from Ambion, Inc. (Austin, TX).

[00101] In some embodiments, RNA is subjected to gel electrophoresis and detected using probes labeled with a tag that may be radioactive. RNA levels in the sample can be compared with RNA levels from a reference sample, such as a normal control tissue, and the like, to determine relative expression levels of the selected gene member(s).

[00102] For some analytic techniques, RNA is processed to produce complementary DNA (cDNA) or complementary RNA (cRNA). A reverse transcriptase (RT) enzyme, in conjunction with appropriate primers (e.g., oligo(dT), T7-oligo(dT) primers, or random primers), can be used to reverse transcribe RNA into cDNA. Single stranded or double stranded cDNA can be used directly in PCR-based assays, such as non-quantitative PCR, quantitative PCR, and/or quantitative real time PCR. Quantitative PCR using cDNA can be used to analyze expression levels of the RNA, and PCR products from cDNA can be sequenced for mutation analysis. Additionally, cDNA can be used in microarray analysis to measure expression levels of the mRNA.

[00103] Complementary DNA can be further processed into cRNA. Typically, cRNA is produced from cDNA that incorporates a primer containing an RNA polymerase promoter, such as T7-oligo(dT) primer. An RNA polymerase that recognizes the promoter can be used to in vitro transcribe cRNA, resulting in linearly amplified cRNA from the cDNA template. Complementary RNA can be used in assays such as microarrays (e.g., Affymetrix GeneChips®) and the like, to analyze gene expression levels.

[00104] In some embodiments, cancer cells are left intact and RNA is analyzed using in situ analytic techniques. For example, RNA can be reverse transcribed in vitro, and subsequently analyzed using immunohistochemistry PCR, (Fassunkc, et al., Laboratory Investigation, 84:1520-5 (2004)) and the like.

[00105] Methods for analyzing expression RNA include, for example, Northern blotting (Brown, 2001, Curr Protoc Immunol, Chapter 10:10.12; Parker & Barnes, 1999, Methods in Molecular Biology 106:247-283), reverse transcriptase polymerase chain reaction (RT-PCR) (Nygaard, et al., 2009, Front Biosci, 14:552-69; Weis et al., 1992, Trends in Genetics, 8:263-64, massively parallel signature sequencing (MPSS) (Kutlu, 2009, BMC Med Genomics., 2:3; Brenner, 2000, Nature Biotechnol, 18:1021), Serial Analysis of Gene Expression (SAGE) (Boon, 2009, PLoS ONE, 4:e5134; Velculescu, 1995, Science, 270:368-9, 371), RNA-mediated annealing, selection, and ligation (RASL) assay (Yeakley, 2002, Nat Biotechnol, 20:353-8), a

cDNA mediated annealing, selection, extension, and ligation (DASL) assay (Abramovitz, 2008, *Biotechniques*, 44:417-423; Fan, 2004, *Genome Research*, 14:878-85), microarray techniques (Ravo, et al., 2008, *Lab Invest*, 88:430-40; Schena, 1996, *Proc Nat. Acad Sci USA*, 93:106-149), including Incyte's microarray technology, Affymetrix's GenChip technology; high throughput sequencing techniques developed by 454 Life Sciences, Inc. (Branford, CT) (Margulies, 2005, *Nature*, 437:376-80), and the like.

RT-PCR

[00106] RT-PCR methods useful for determining gene expression in a sample are described, for example, in Sambrook, 2001, *Molecular Cloning: A Laboratory Manual*. For example, clinical samples such as tumor biopsy tissue or archived paraffin embedded and/or frozen samples provide RNA templates for genetic analysis. General methods of performing PCR are described, for example, in Ausubel, et al., 2002, *Short Protocols in Molecular Biology*; Mullis and Faloona, 1987, *Methods Enzymol*, 155:335). Primers for performing RT-PCR can be obtained commercially, or can be designed using commercially available software (e.g., Scientific Software, Primer Designer 1).

DASL

[00107] In a DASL assay, total RNA is converted to cDNA using biotinylated primers, and the biotinylated DNA is attached to a streptavidin solid support, followed by the annealing of assay oligonucleotides to their target sequences in the cDNA. A pair of oligonucleotides is annealed to a given target site, with three to ten target sites per gene. The upstream annealed oligonucleotides are extended and ligated to corresponding nucleotides downstream to create a PCR template that is then amplified with universal PCR primers. The PCR products, having been labeled by incorporation of a labeled primer, are hybridized to capture sequences on the solid support array, and the fluorescence intensity is then measured for each bead.

[00108] Complete custom designed DASL assay panels for up to 1536 genes comprising 1-3 probe groups per gene are available commercially from Illumina, Inc. (San Diego, CA), as well as a standard DASL human cancer panel comprising a set of probe groups targeting 502 genes that have been associated with cancer in the literature.

MassARRAY

[00109] The MassARRAY system is used to isolate and reverse transcribe RNA to cDNA. The cDNA is amplified, dephosphorylated, extended with primers, and then placed onto a chip array for analysis via MALDI-TOF mass spectrometry. Hardware and software for carrying out MassARRAY analysis is commercially available from Sequenom, Inc. (San Diego, CA).

SAGE

[00110] In SAGE, multiple sequence tags of about 10-14 base pairs, each corresponding to a unique position within an RNA transcript are linked together to form extended molecules for sequencing and identifying the sequence of multiple tags simultaneously. A transcript's expression pattern can be quantitated by determining the abundance of a given tag and identifying the gene corresponding to that tag. Kits for performing SAGE as well as software for analyzing SAGE data are commercially available, including, for example, the I-SAGE Kit (Invitrogen, Carlsbad, CA). SAGE data can be used to search, for example, the SAGEmap database at www.ncbi.nlm.nih.gov/SAGE.

Protein and polypeptide analysis

[00111] In some embodiments, a biological sample containing a protein or polypeptide originating from a patient's cancer can be interrogated. Such a sample can comprise cancer cells, or can contain a protein or polypeptide substantially free of cancer cells, such as protein or polypeptide that is secreted from a cancer cell or is released during cancer cell necrosis. Protein and/or polypeptide expression levels can be used to infer gene expression levels since mRNA levels are generally well correlated with protein expression levels (Guo, et al., 2008, *Acta Biochim Biophys Sin*, 40:426-36).

[00112] Tumor cells can remain unpurified or can be purified using known methods. Depending on analysis method(s) used, more or less purification may be desired.

[00113] In some embodiments, protein/polypeptide levels can be determined using Western blotting with antibodies specific to a protein/polypeptide gene product of a gene member of a Module, for example those modules listed in Tables 1-161. Similarly, other antibody-based assays, such as enzyme-linked immunosorbent assays (ELISAs) or protein arrays (see, for example, Joos and Bachman, 2009, *Front Biosci*, 14:4376-85), can utilize antibodies to measure protein/polypeptide levels.

[00114] In some embodiments, a protein or polypeptide can be detected using a molecule other than an antibody. For example, a protein receptor can be used to detect the presence of its cognate ligand or vice versa. Other methods for detecting polypeptides include mass spectrometry.

[00115] A chosen method of protein/polypeptide analysis used can depend on the source of the protein/polypeptide. For example, more sensitive methods would be desirable for measuring the level of proteins or polypeptides that are dilute in the biological sample, for example proteins secreted by a cancer into blood. Conversely, the analysis method does not need to be as sensitive if the source of the protein is concentrated, such as protein extracted from a cancer cell sample.

[00116] The method of protein/polypeptide analysis used can also depend on the number of proteins and/or polypeptides that are to be measured. For example, Western blotting can be used if only a few proteins/polypeptides are to be measured, while a protein array would be useful for detecting many proteins and/or polypeptides.

[00117] The results of protein expression level analysis can be compared to protein expression levels in a control to infer relative gene expression levels and identify a gene expression pattern. In some embodiments, it is possible to infer a gene expression pattern based on absolute protein expression levels. The gene expression pattern can then be matched to a pattern of expression of gene members of a disclosed Module, such as those in Tables 1-161.

Genomic analysis

[00118] In some embodiments, a biological sample containing DNA originating from a patient's disease can be interrogated. DNA analysis results can be used to infer gene expression levels as described below.

[00119] Biological samples can be purified as appropriate for the analytic technique being used. For example, purified cancer cells can be appropriate for analyzing acetylation or methylation status of cancer DNA, whereas cancer cell purification can be less important when analyzing for the presence of a DNA sequence mutation.

[00120] DNA is extracted from the biological sample or purified cancer cells using known techniques. A DNA sample can be subjected to one or more types of analysis, including DNA sequence analysis, SNP (single nucleotide polymorphism) analysis, gene copy number analysis, nucleic acid insertions and/or deletions, viral insertions, and acetylation/methylation status. Other appropriate types of analyses, and techniques for performing such analyses, are known. For example, DNA can be amplified using polymerase chain reaction (PCR) and used in various protocols such as SNP analysis (see, for example, Kwok, 2002, "Single Nucleotide Polymorphisms: Methods and Protocols." In: *Methods in Molecular Biology*, Vol. 212. Walker (ed.). Humana Press).

[00121] Gene copy number variance can be analyzed using a variety of known methods, assays, and platforms. A review of methods for detecting and analyzing copy number variations (CNV) can be found, for example, in Lee, et al., 2008, *Cytogenet Genome Res*, 123:333-42. Specific methods for performing CNV analysis include, for example, qt-PCR (Wu, et al., 2007, *J Immunol*, 179:3012-25), DNA microarrays based upon fluorescent in situ hybridization (FISH) or SNP arrays (Redon, et al., 2006, *Nature*, 444:444-54), sequencing methodologies such as those reviewed in Hall, 2007, *J Exp Biol*, 209:1518-25), RFLP/Southern blot analysis (Yang, et al., 2007, *Am J Hum Genet*, 80:1037-54), ligation detection reaction (LDR) (Seo, et al., 2007,

BMC Genet, 8:81), invader assays (Pielberg, et al., 2003, Genome Res, 13:2171-7), and pyrosequencing (Soderback, et al., 2005, Clin Chem, 51:522-31).

[00122] In some embodiments, cancer cells are left intact and DNA is analyzed using in situ techniques, such as fluorescent in situ hybridization in conjunction with fluorescence microscopy, fluorescence activated cell sorting, or image scanning flow cytometry (Basiji et al., 2007, Clin Lab Med, 27(3):653).

[00123] The results of genomic analysis can be compared to a control to identify differences between the control DNA sequence, gene copy number, and nucleic acid acetylation/methylation status of the sample DNA, as appropriate for the analysis method(s) used. Any differences can be identified and used to infer gene expression.

[00124] For example, increases in DNA acetylation of a gene would be expected to be associated with an increase in expression of that gene. Conversely, increases in DNA methylation of a gene would be expected to be associated with a decrease in expression of that gene.

Kits and Tools

[00125] Representative tools applying the newly identified associations between genetic expression profiles and Modules include assay systems for microarray, protein array, ELISA, hybridization, amplification, PCR, DASL, SAGE, and the like systems, as well as kits, chips, cards, multi-well assay plates, probes and primers, and the like, adapted and designed to measure the expression of one or more gene members of one or more Module, such as those selected from Tables 1-161.

[00126] In some embodiments, panels of nucleic acid probes and/or primers can be designed to amplify and/or detect the presence of the gene members of one or more Module, such as those selected from Tables 1-161. Such probes include isolated genes, including reference and mutated genes, or portions of such genes, isolated mRNA, cDNA derived from these, amplified nucleic acids, and the like useful, for example in microarray and hybridization platforms. Such primers include nucleic acids flanking a desired amplicon and useful in amplification reactions such as PCR to amplify a desired gene or portion of a gene for detection and quantifying gene expression.

[00127] In some embodiments, panels of binding molecules can be produced to bind, detect, and/or quantify gene expression products of one or more gene member of one or more Module, such as proteins or peptides. Such binding molecules can include antibodies, ligands, ligand receptors, small molecules, and the like.

[00128] An assay substrate such as a hybridization plate, chip, card, and the like, is adapted and designed to include primer pairs and/or probes that amplify and/or identify and/or sequence and thereby measure gene expression in a sample obtained from a subject.

[00129] Kits include reagents and tools useful to measure gene expression and include, for example, nucleic acid probes and/or primers designed to amplify, detect, and/or quantify in a sample one or more gene member from one or more Module, such as those selected from Tables 1-161.

Methods for Disease Classification

[00130] It can be easily appreciated that the present disclosure is not limited to cancer, but is applicable to diseases in general including diabetes, Parkinson's disease, heart disease and such. Cancers are but one embodiment

Clustering genes by co-expression

[00131] In one embodiment, genomic data may be measured from a population of training patients using methods disclosed herein or generally known to those skilled in the art. Analytical results may also be combined from a plurality of previous studies, or data may be combined from both measurement and previous studies. In one example, gene expression data from cancer patients may be found in the Oncomine database at <http://www.oncomine.org>.

[00132] Gene expression data from various sources may be normalized. In one example, normalization may be performed as described in Rhodes et al., Neoplasia, 2007 Feb;9(2):166-80.

[00133] Average linkage hierarchical clustering may be performed on the normalized data. In one embodiment, the Pearson correlation may be used as the distance metric.

[00134] Discrete gene expression clusters may then be created. In one embodiment, the clusters with the most features having a minimum Pearson correlation of 0.5 and a minimum of 10 features were identified (Rhodes, Neoplasia, 2007).

[00135] Pair-wise association analysis may then be performed on each pair of clusters. In one embodiment, this was performed by counting the number of overlapping genes, computing an odds ratio and calculating a p-value based on Fisher's exact test. Significant associations were defined as those with at least 3 genes overlapping, an odds ratio > 10, and a p-value < 1E-6.

Meta-analysis of Gene expression in Cancer

[00136] In one embodiment, analytical results of gene expression in cancer patients was obtained from the Oncomine database at <http://www.oncomine.org> and was processed and normalized as described in Rhodes et al., Neoplasia, 2007 Feb;9(2):166-80. Datasets from the 15 most represented cancer types were analyzed. Average linkage hierarchical clustering, using the

Pearson correlation as the distance metric, was performed on each dataset. Up to 10,000 features (but not more than 50% of all features) with the largest standard deviations were included in the analysis. To reduce the hierarchical clustering results to discrete gene expression clusters, the clusters with the most features having a minimum Pearson correlation of 0.5 and a minimum of 10 features were identified (Rhodes, Neoplasia, 2007). Pair-wise association analysis was performed on each pair of clusters, counting the number of overlapping genes, computing an odds ratio and calculating a p-value based on Fisher's exact test. Significant associations were defined as those with at least 3 genes overlapping, an odds ratio > 10 , and a p-value $< 1E-6$.

Creation of the modules

[00137] The method further provides for creation of modules from the gene clusters. In one example, a network representation was used to visualize the pairwise cluster associations and identify modules of highly interconnected clusters. In one embodiment, the network representation is Cytoscape.

[00138] The cluster association network may then be reduced to a discrete set of modules. In one embodiment, edges without at least two supporting indirect associations were removed and nodes and edges that linked two otherwise mostly unlinked sets of interlinked clusters were removed.

[00139] Each module is defined as a list of interlinked clusters.

[00140] Following creation of modules, the method provides for scoring the genes within a module. In one embodiment, representative genes were ranked for each module based on the number of clusters in which they were a member.

Identification of Cancer Modules

[00141] In one embodiment, a network representation (Cytoscape) was used to visualize the pairwise cluster associations and identify modules of highly interconnected clusters. To reduce the cluster association network to a discrete set of modules, edges without at least two supporting indirect associations were removed and nodes and edges that linked two otherwise mostly unlinked sets of interlinked clusters were removed. Each Cancer Module was defined as a list of interlinked clusters.

[00142] Representative genes were ranked for each module based on the number of clusters in which they were a member. Identified Cancer Modules for 15 cancer types are shown in Tables 1-161.

Numbers of patients and cohorts

[00143] The disclosed methods may be performed with large number of patients, large number of cohorts, or large numbers of patients and large numbers of cohorts. Increased numbers of

patients and/or cohorts may lead to improved predictions. For large numbers of patients and/or cohorts, genetic marker patterns as rare as 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9% or 10% of the population can be detected.

[00144] In some embodiment, the number of cohorts is at least one of: at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, and at least 100. In another embodiment, the number of cohorts is at least one of: from 2 to 20, from 2 to 50, from 2 to 100, from 10 to 20, from 10 to 50, from 10 to 100, from 25 to 50, from 25 to 100, and from 50 to 75.

[00145] In some embodiments the number of patients is at least one of: at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 150, at least 200, at least 250, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1200, at least 1400, at least 1600, at least 1800, at least 2000, at least 3000, at least 4000, and at least 5000. In another embodiment, the number of patients is at least one of: from 100 to 200, from 100 to 500, from 100 to 1000, from 100 to 5000, from 500 to 1000, from 500 to 3000, from 500 to 5000, from 1000 to 5000, and from 3000 to 5000.

Adding some biological names to the modules

[00146] While the embodiments disclosed herein provide gene selection based on first analyzing expression of the genes without regard to phenotypic or other biological aspects, phenotype may be considered in a second step.

[00147] In some embodiments, biological or phenotypic names are given after the modules have been created based on genomic data. For example, when choosing the subset of genes to be queried, it may be beneficial to choose genes with independent biological functions.

Methods for predicting efficacy of drug treatment

[00148] The newly identified Modules allow for identifying a population of patients responsive to a selected drug. Generally, methods for predicting a population of patients responsive to a drug involve identifying the patient's disease as belonging to one or more Module, such as those of Tables 1-161 and determining in a training group of patients, each patient's response to the administered drug as well as one or more Module to which the patient's disease belongs (such as Tables 1-161). Responsiveness of test patients having diseases identified as belonging to one or more Module can then be predicted according to the

demonstrated response of the training group patient's having diseases belonging to that one or more Module.

[00149] For example, methods for predicting patient responsiveness can involve interrogating biological samples obtained from cancer patients in a training group for expression of one or more gene member of one or more Cancer Module, identifying each training patient's cancer as belonging to a Cancer Module having a pattern of gene expression consistent with the expression of the one or more interrogated gene, identifying response of each training patient's cancer to a test drug as responsive or non-responsive to the drug, identifying one or more Cancer Module as consistent with training patient responsiveness or non-responsiveness to the drug, and predicting responsiveness or non-responsiveness of test patients with a cancer identified as belonging to the one or more identified Cancer Module. Thus, a patient's likelihood of responsiveness or non-responsiveness can be correlated with the level of responsiveness or non-responsiveness associated with a Cancer Module assigned to the patient's cancer.

Methods for selecting a therapeutic drug

[00150] Gene expression modules are also useful in selecting a therapeutic drug for individual patients. Generally, methods for selecting a therapeutic drug involve interrogating a biological sample obtained from a patient for expression of one or more gene member of one or more Module, identifying the patient's disease as belonging to a Module having a pattern of gene expression consistent with the expression of the one or more interrogated gene, and selecting a drug having demonstrated activity for diseases belonging to the identified one or more Module and/or not selecting a drug that does not demonstrate activity for diseases belonging to the identified one or more Module.

Methods for predicting metastasis or recurrence

[00151] Modules are also useful in identifying a patient population at risk for metastasis and/or recurrence. Generally, methods for identifying a patient population at risk for metastasis and/or recurrence include determining in a training group of patients, demonstration of metastasis/no metastasis and/or recurrence/nonrecurring and also identifying each training patient's disease as belonging to one or more Module, such as those of Tables 1-161. Risk of metastasis and/or recurrence of test patients having diseases identified as belonging to the one or more Module can then be predicted according to the demonstrated metastasis/no metastasis and/or recurrence/no recurrence of the training group patients having diseases belonging to the identified Module.

[00152] For example, methods for predicting cancer metastasis and/or recurrence can involve interrogating biological samples obtained from cancer patients in a training group for expression of one or more gene member of one or more Cancer Module, identifying each training patient's

cancer as belonging to a Cancer Module having a pattern of gene expression consistent with the expression of the one or more interrogated gene, identifying the incidence of metastasis and/or recurrence of each training patient's cancer, identifying one or more Cancer Module as consistent with training patient incidence of metastasis and/or recurrence, and predicting risk of metastasis and/or recurrence in test patients with a cancer identified as belonging to the one or more identified Cancer Module. Thus, a patient's risk of metastasis and/or recurrence can be correlated with the risk of metastasis and/or recurrence associated with a Cancer Module assigned to the patient's cancer.

Methods of Treatment

[00153] Gene expression modules are also useful in selecting a method of treatment for individual patients. Generally, methods for selecting a method of treatment involve interrogating a biological sample obtained from a patient for expression of one or more gene member of one or more Module, identifying the patient's disease as belonging to a Module having a pattern of gene expression consistent with the expression of the one or more interrogated gene, administering a drug having demonstrated activity for diseases belonging to the identified one or more Module, and not administering a drug demonstrating a lack of activity for diseases belonging to the identified one or more Modules.

[00154] Methods for determining an appropriate means and dosage of administration of a drug for a particular patient's disease can be determined generally or can be identified as consistent with effective treatment modalities for the identified Module.

Making multiple predictions from the modules

[00155] Current genomics technologies including microarrays and next-generation sequencing platforms are suitable for routine clinical testing, however in the current paradigm, every assay is custom designed for a single therapy. This model is prohibitive for clinical development due to the time, cost and risk involved in their development. One embodiment discloses a single assay with the ability to predict therapeutic responses for a wide variety of therapies and combinations.

Other Methods

[00156] As described herein, the disclosed Modules are useful for identifying patient populations having a unique pattern of gene expression that can be correlated with disease prognosis, therapy, resistance, and the like.

Table 163 – Genes for the 96 well breast cancer assay

assay	gene	type	module	Priority Rank
Hs01070859_g1	DDX3X	Module	Breast-1	
Hs00200318_m1	PICALM	Module	Breast-1	
Hs00922166_m1	VAMP3	Module	Breast-1	
Hs00270129_m1	FOXA1	Module	Breast-11	
Hs00922328_m1	GATA3	Module	Breast-11	
Hs00225445_m1	MLPH	Module	Breast-11	
Hs00175403_m1	CTSS	Module	Breast-12	
Hs00160164_m1	PLEK	Module	Breast-12	
Hs01004159_m1	SRGN	Module	Breast-12	
Hs00175619_m1	PCSK1	Module	Breast-15	
Hs00938962_m1	SNAP25	Module	Breast-15	
Hs00919233_m1	DERL1	Module	Breast-16	
Hs01091048_m1	MRPL13	Module	Breast-16	
Hs00206458_m1	TTC35	Module	Breast-16	
Hs00193386_m1	GPD1	Module	Breast-18	
Hs00160173_m1	PLIN	Module	Breast-18	
Hs00198830_m1	RBP4	Module	Breast-18	
Hs00182073_m1	MX1	Module	Breast-19	
Hs00242943_m1	OAS1	Module	Breast-19	
Hs00942643_m1	OAS2	Module	Breast-19	
Hs00996789_g1	CCNA2	Module	Breast-2	
Hs01554513_g1	MAD2L1	Module	Breast-2	
Hs00964100_g1	UBE2C	Module	Breast-2	
Hs00375382_m1	HEATR6	Module	Breast-20	

assay	gene	type	module	Priority Rank
Hs00186230_m1	PPM1D	Module	Breast-20	
Hs00177357_m1	RPS6KB1	Module	Breast-20	
Hs00273612_m1	CPSF1	Module	Breast-21	
Hs00226123_m1	FBXL6	Module	Breast-21	
Hs00171565_m1	ZNF7	Module	Breast-21	
Hs00233566_m1	CD79A	Module	Breast-22	
Hs00174811_m1	POU2AF1	Module	Breast-22	
Hs00171292_m1	TNFRSF17	Module	Breast-22	
Hs01028956_m1	COL1A2	Module	Breast-23	
Hs00943809_m1	COL3A1	Module	Breast-23	
Hs00893923_m1	COL5A2	Module	Breast-23	
Hs00610257_g1	DUSP1	Module	Breast-24	
Hs00152928_m1	EGR1	Module	Breast-24	
Hs01119267_g1	FOS	Module	Breast-24	
Hs00260553_m1	C17orf37	Module	Breast-25	
Hs01001595_m1	ERBB2	Module	Breast-25	
Hs00918009_g1	GRB7	Module	Breast-25	
Hs01566350_m1	INTS4	Module	Breast-3	
Hs00372778_m1	NARS2	Module	Breast-3	
Hs00213155_m1	RSF1	Module	Breast-3	
Hs00234981_m1	CCL14	Module	Breast-6	
Hs01011079_s1	DARC	Module	Breast-6	
Hs00412974_m1	MFAP4	Module	Breast-6	
Hs00217757_m1	BRF2	Module	Breast-7	

assay	gene	type	module	Priority Rank
Hs00200360_m1	ERLIN2	Module	Breast-7	
Hs00951195_m1	RAB11FIP1	Module	Breast-7	
Hs00204411_m1	GABRP	Module	Breast-8	
Hs03928965_s1	SFRP1	Module	Breast-8	
Hs00366918_m1	SOX10	Module	Breast-8	
Hs00233476_m1	BMP7	Amp		
Hs00765553_m1	CCND1	Amp		
Hs00364847_m1	CDK4	Amp		
Hs01076091_m1	EGFR	Amp		
Hs99999003_m1	MYC	Amp		
Hs00919915_m1	ZNF217	Amp		
Hs01001598_g1	ERBB2	Breast Gene		
Hs01046816_m1	ESR1	Breast Gene		
Hs01046818_m1	ESR1	Breast Gene		
Hs01556702_m1	PGR	Breast Gene		
PTEN	PTEN	Breast Gene		
Hs00258937_m1	AGTR1	COPA		
Hs00153988_m1	ALOX15B	COPA		96
Hs00218464_m1	BEX1	COPA		91
Hs00154977_m1	EN1	COPA		86
Hs00361426_m1	FABP7	COPA		
Hs00181331_m1	GRIA2	COPA		
Hs00365956_m1	HOXA9	COPA		
Hs01699178_g1	KRT6A	COPA		84

assay	gene	type	module	Priority Rank
Hs01084628_m1	LBP	COPA		
Hs00938423_g1	MED24	COPA		87
Hs00964352_gH	PSMB3	COPA		94
Hs00161045_m1	PVALB	COPA		90
Hs00200740_m1	SCRG1	COPA		85
Hs00186798_m1	SLC4A4	COPA		89
Hs01370702_m1	TDRD12	COPA		95
Hs01380839_m1	TDRD9	COPA		92
Hs00610327_m1	TSPAN8	COPA		88
Hs00357333_g1	ACTB	Housekeeping		
Hs00855401_g1	ATP5E	Housekeeping		
Hs00829989_gH	GPX1	Housekeeping		
Hs99999908_m1	GUSB	Housekeeping		
Hs03043885_g1	RPL13A	Housekeeping		
Hs00420895_gH	RPLP0	Housekeeping		
Hs00200497_m1	PROSC	Module/Amp		93
Hs00608289_m1	ALK	Target		
Hs01582073_m1	AURKA	Target		
Hs00240792_m1	FGFR2	Target		
Hs01106910_g1	FGFR4	Target		
Hs00181385_m1	IGF1R	Target		
Hs00922194_g1	KIT	Target		
Hs01565583_m1	MET	Target		
Hs00183486_m1	PDGFRA	Target		

EXAMPLES

EXAMPLE 1 – Breast cancer modules and uses thereof

[00157] Breast cancer is a highly heterogeneous disease as evidenced by comprehensive genetic studies which have revealed multiple subtypes using gene expression profiling and cell lineage classifier analyses. Previous studies have characterized different subtypes including normal breast-like, luminal epithelial A, luminal epithelial B, Her 2 over-expression and basal type carcinoma. However, the genetic variation within breast cancer is far more diverse than these core subtypes, and it is necessary to fully characterize this diversity in order to move beyond simple prognosis and to specifically predict drug sensitivity.

[00158] In accordance with one embodiment, a review of global gene expression and SNP-based cytogenetic data of more than 5,000 breast cancer patients is performed using the Oncomine™ database. The analysis is able to characterize approximately 30 different genetic variations that are shared by 1% or more of the breast cancer population. These core, independent variables reflect diverse elements of the disease at a molecular level including cell lineage, dysregulated core biological functions, factors of cell growth, and the tumor microenvironment. Further genetic subtypes are characterized within the various large and focal genomic amplifications, such as Her2 and Myc, as well as focal expression events present subpopulations of patients. In aggregate these genetic variables represent all of the major genetic factors that present within breast cancer.

[00159] Currently biomarker/diagnostic approaches have tended to be over-tailored to specific clinical questions and therefore have lacked broad applicability, with every diagnostic test requiring a custom gene set and tailored signature and in some cases, requiring separate validated assays using multiple technologies and consequent splitting of clinical samples. To overcome these limitations, disclosed herein is a single, 96-gene qRT-PCR test for rapid breast cancer companion diagnostics development using FFPE tumor tissue. All 30 of the core variables or “modules” are represented by this test which reports on both gene expression and chromosomal amplification events. It is shown that this single test, with its multiple modules, can report on standard histopathological parameters, such as ER, PR and Her2, and reproduce existing prognostic and predictive genomic signatures.

[00160] A co-expression meta-analysis on 5,339 breast cancer samples from Oncomine™ identified highly co-expressed sets of genes (modules) across multiple breast cancer microarray datasets, with each module consisting on average of 450 genes (range: 11–962). These modules represented expected subclasses (e.g., basal, luminal A, luminal B), as well as additional subclasses (e.g., immune response, proliferation). The 96 gene panel which collectively

predicted to identify each of the modules were selected by testing 384 candidate genes in FFPE–RT-PCR. The approach was tested on 65 FFPE samples with known histological parameters such as ER, PR, and HER2. The patterns of module expression in retrospective studies was matched in expected ways with prognosis and drug response.

EXAMPLE 2 - Using Quantitative RT-PCR to Identify Cancers Belonging to a Cancer Module

[00161] A tumor biopsy is obtained from a patient. Messenger RNA is purified from the biopsy using a Dynabeads® Oligo(dT)25 mRNA purification kit (Invitrogen, Carlsbad, CA), according to the manufacturer's protocol. Briefly, tumor cells are lysed by grinding the sample in liquid nitrogen to form crude lysate. The lysate is added to washed Dynabeads® Oligo(dT)25 beads and allowed to incubate at room temperature to allow the annealing of poly-A mRNA to the beads. The beads are recovered with the bound mRNA using a magnet, and other cell components are washed away. The mRNA is eluted from the beads for use in RT-PCR.

[00162] The purified mRNA is reverse transcribed using a RETROscript® cDNA kit (Ambion®, Austin, TX), according to the manufacturer's protocol for two-step RT-PCR. Briefly, 20-200 ng mRNA is mixed with random decamer primers and denatured at 85° C. The primers are then allowed to anneal to the mRNA template on ice. A dNTP mixture, MMTV-RT, an RNase inhibitor and RT buffer are added, and the mixture is incubated at 42-44° C to allow reverse transcription. The reverse transcriptase is deactivated by a brief incubation at 92° C. The cDNA is used in PCR immediately, or can be stored at -20° C for later PCR analysis.

[00163] The cDNA is analyzed using the LightCycler®, thermocycler (Roche Diagnostics Corporation, Indianapolis, IN). PCR is performed using the LightCycler® Multiplex DNA Master HybProbe kit (Roche), according to the manufacturer's protocol, to assay for up to four gene targets at once. A segregation panel containing Cancer Modules is chosen based on the cancer type (e.g., breast cancer, bladder cancer, colon cancer, etc.). HybProbe probes designed for any or each of the selected target sequences in each of the Cancer Modules for the chosen segregation panel are used. For example, if the cancer is breast cancer, probes representing at least one gene from each of Breast Cancer Modules 1-25 of Tables 22-46 would be used. Three to five genes from each of the modules for a specific cancer type can be used. Genes from fewer than all of the Cancer Modules of a given cancer type can be used. Genes from Cancer Modules of multiple cancer types can be used, for example, breast and ovary Cancer Modules. At least one set of HybProbe probes can be used to detect a reference gene, such as actin, that can be used to normalize cDNA content. The cDNA is mixed with master mix (containing Taq DNA polymerase, reaction buffer, MgCl₂, and dNTP mix with dUTP instead of dTTP), additional

MgCl₂, if necessary, up to 4 sets of HybProbes (including at least one set targeted to a normalizing gene), and nuclease free water. The mixture is put into LightCycler® capillaries and placed in the LightCycler® thermocycler. PCR is run using a cycle appropriate for the probes used. Typically, the samples are subjected to a denaturing step at 95° C for 5-10 minutes. The samples are then denatured (95° C, for 10 seconds), annealed (temperature depending on the probes used, for 5-15 seconds), and the primers extended (72° C, for a length of time dependent on the expected length of the PCR products), for several cycles as needed for the signal strength of each of the probes to plateau. The samples are then slowly heated to denature the probes from the PCR product to determine melting temperature, which can then be used to determine purity of the PCR product.

[00164] The number of cycles required for the signal of each probe set in the tumor samples to reach a set threshold can be compared to the number of cycles required in a control sample to reach the same threshold in order to determine the relative expression level of the initial mRNA. The threshold point of the normalizing gene can be used to normalize general mRNA quantity between the tumor sample and the control sample and allow accurate comparison of gene expression levels between the two samples. The relative expression for each measured gene can be used to identify the patient's cancer as belonging to one or more of the Cancer Modules set forth in Tables 1-161.

[00165] The patient's prognosis, including response to therapy, recurrence, and/or metastasis, can be predicted based on the Cancer Module identified, for example, using RT-PCR and the expression profiles of the Cancer Modules set forth in Tables 1-161. Prognosis is based on the prognosis, recurrence, metastasis, or drug response demonstrated by cancers belonging to the same Cancer Module(s) of the patient's cancer as determined by comparing the patient's cancer gene expression pattern with that of the disclosed Cancer Modules.

EXAMPLE 3 - Using Microarray to Identify Cancers Belonging to a Cancer Module

[00166] Purified total RNA is obtained from a cancer patient's tumor biopsy using standard procedures. The RNA is further prepared for Affymetrix® GeneChip® analysis using the GeneChip® 3' IVT Express Kit (Affymetrix®, Santa Clara, CA), according to the manufacturer's protocol. Briefly, 50-500 ng total RNA is mixed with diluted polyA RNA controls, and RNase-free water to a total volume of 5 The mixture is combined with a first strand cDNA synthesis master mix containing RT enzyme and buffer and incubated at 42° C for two hours to produce first strand cDNA. Second strand cDNA is produced by combining the first-strand cDNA with a second strand master mix containing DNA polymerase and buffer, and incubating at 16° C for an hour, followed by an incubation period of 10 minutes at 65° C. The

cDNA is then in vitro transcribed by adding an IVT master mix containing an IVT enzyme mix, biotin label, and buffer, and incubating at 40° C for 4 to 16 hours to produce biotin-labeled cRNA (i.e., aRNA). The cRNA is purified, washed, and eluted using the magnetic beads, wash buffer, and elution solution provided in the kit. The labeled cRNA is fragmented into fragments having 35-200 nucleotides using the fragmentation buffer provided in the kit.

[00167] Fragmented and labeled cRNA is prepared for hybridization to a GeneChip® Human Genome Focus Array containing all the genes from each Cancer Module for a given cancer type, according to the manufacturer's instructions. Briefly, the fragmented cRNA is mixed with a hybridization cocktail containing hybridization controls, control oligonucleotide B2, DMSO and buffer, and incubated at 99° C for 5 minutes, and then at 45° C for 5 minutes, while the array is prewet and incubated with prehybridization mix at 45° C. The prehybridization mix is then replaced with the hybridization cocktail containing the labeled, fragmented cRNA, and incubated for 16 hours at 45° C.

[00168] Following hybridization, the array is washed, stained, and scanned using the Affymetrix® Hybridization, Wash, and Stain kit, according to the manufacturer's protocol. Briefly, the array is washed twice using the provided wash buffers, stained with a first stain cocktail, washed, stained with a second stain cocktail, stained again with the first stain cocktail, washed, and then filled with a buffer. The array is then scanned using an Agilent GeneArray® Scanner or a GeneChip® Scanner 3000. The raw scanning data can be normalized according to the controls included at various steps during processing, and relative and absolute gene expression can be determined from the scanned array. The relative expression for each measured gene can be used to identify the patient's cancer as belonging to one or more of the Cancer Modules set forth in Tables 1-161.

[00169] The patient's prognosis, including response to therapy, recurrence, and/or metastasis, can be predicted based on the Cancer Module identified, for example, using microarray analysis and the expression profiles of the Cancer Modules set forth in Tables 1-161. Prognosis is based on the prognosis demonstrated by cancers belonging to the same Cancer Module(s) as that of the patient's cancer as determined by comparing the patient's cancer gene expression pattern with that of the disclosed Cancer Modules.

EXAMPLE 4 - Use of Cancer Modules to Identify a Tumor Signature

[00170] Genes in a Cancer Module (Tables 1-161) can be categorized by gene ontology and relation to cell signaling pathways. Gene ontology information is identified for a gene using data available from the GO Consortium (<http://www.geneontology.org/>) and other known methods, such as the online search tool for gene ontology, AmiGO (<http://amigo.geneontology.org/cgi->

bin/amigo/go.cgi), and the like. For example, a search using AmiGO for COL1A2 (found in Bladder Cancer Module 6) indicates that COL1A2 is categorized as relating to the ontologies shown below in Table 163.

[00171] Gene ontology information for the genes in a selected Cancer Module can be analyzed for patterns, either manually, or using computer software designed to identify patterns in gene ontology data. Patterns can include a high frequency of genes with related ontologies or association with a cell signaling pathway. For instance, Bladder Cancer Module 2 contains a high percentage of member genes having ontologies related to protein production.

[00172] As shown in Table 164, for example, 9 of 20 member genes of Bladder Cancer Module 2 are classified as having a molecular function in Gene Ontology Accession No. go:000373 structural constituent of ribosome, defined as the action of a molecule that contributes to the structural integrity of the ribosome. Sixteen of 20 member genes of Bladder Cancer Module 2 are classified as having a biological process in Gene Ontology Accession No. go:0006414 translational elongation, defined as the successive addition of amino acid residues to a nascent polypeptide chain during protein biosynthesis. Fifteen of 20 member genes of Bladder Cancer Module 2 are classified in the cell composition Gene Ontology Accession No. go:0005829 cytosol, defined as the part of the cytoplasm that does not contain organelles but does contain other particulate matter, such as protein complexes.

[00173] Table 164 does not include all ontological classifications for each gene.

[00174] Gene Ontology patterns identified within a Cancer Module can be used to define a cancer signature. For example, Bladder Cancer Module 2 can be identified as having a protein biosynthesis signature based on the ontologies of the member genes that define Bladder Cancer Module 2 (Table 164).

[00175] Tumor signatures are useful for predicting cancer sensitivity to certain classes of drugs. For example, cancers classified in Bladder Cancer Module 2 can be predicted to be sensitive to translation inhibitors such as tedanolides and related molecules. Sensitivity of cancers belonging to a Cancer Module with a known cancer signature to a certain drug class can be confirmed using in vitro and in vivo experiments. In some embodiments, cancer sensitivity can be confirmed using retroactive studies on cancer samples from patients treated with known classes of drugs.

[00176] Figure 1 shows an embodiment of the invention. Figure 2 shows an embodiment of the invention. Figure 3 shows an example of Cq values for various samples. Figure 4 shows an example of Cq values for various samples. Figure 5 shows an example of the methods described herein. Figure 6 shows OncoScoresTM for a first exemplary patient. Figure 7 shows

OncoScores™ for a second exemplary patient. Figure 8 shows four module genes demonstrated expected gene-gene correlations and the genes also displayed correlations with expected clinical characteristics. Figure 9 shows an example of Module score clinical trends. Figure 10 shows an example of Module-based molecular stratification. Figure 11 shows an example of long-term clinical follow-up for 287 patients. Figure 12 shows an example of overall survival. Figure 13 shows an example of neoadjuvant chemotherapy response. Figure 14 shows an example of targeted therapy sensitivity.

[00177] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

[00178] Aspects of the systems and methods described herein may be implemented as functionality programmed into any of a variety of circuitry, including programmable logic devices (PLDs), such as field programmable gate arrays (FPGAs), programmable array logic (PAL) devices, electrically programmable logic and memory devices and standard cell-based devices, as well as application specific integrated circuits (ASICs). Some other possibilities for implementing aspects of the systems and methods include: microcontrollers with memory, embedded microprocessors, firmware, software, etc. Furthermore, aspects of the systems and methods may be embodied in microprocessors having software-based circuit emulation, discrete logic (sequential and combinatorial), custom devices, fuzzy (neural network) logic, quantum devices, and hybrids of any of the above device types. Of course the underlying device technologies may be provided in a variety of component types, e.g., metal-oxide semiconductor field-effect transistor (MOSFET) technologies like complementary metal-oxide semiconductor (CMOS), bipolar technologies like emitter-coupled logic (ECL), polymer technologies (e.g., silicon-conjugated polymer and metal-conjugated polymer-metal structures), mixed analog and digital, etc.

[00179] It should be noted that the various functions or processes disclosed herein may be described as data and/or instructions embodied in various computer-readable media, in terms of their behavioral, register transfer, logic component, transistor, layout geometries, and/or other characteristics. Computer-readable media in which such formatted data and/or instructions may

be embodied include, but are not limited to, non-volatile storage media in various forms (e.g., optical, magnetic or semiconductor storage media) and carrier waves that may be used to transfer such formatted data and/or instructions through wireless, optical, or wired signaling media or any combination thereof. Examples of transfers of such formatted data and/or instructions by carrier waves include, but are not limited to, transfers (uploads, downloads, email, etc.) over the Internet and/or other computer networks via one or more data transfer protocols (e.g., HTTP, FTP, SMTP, etc.). When received within a computer system via one or more computer-readable media, such data and/or instruction-based expressions of components and/or processes under the systems and methods may be processed by a processing entity (e.g., one or more processors) within the computer system in conjunction with execution of one or more other computer programs.

[00180] Unless the context clearly requires otherwise, throughout the description and the claims, the words 'comprise,' 'comprising,' and the like are to be construed in an inclusive sense as opposed to an exclusive or exhaustive sense; that is to say, in a sense of 'including, but not limited to.' Words using the singular or plural number also include the plural or singular number respectively. Additionally, the words 'herein,' 'hereunder,' 'above,' 'below,' and words of similar import refer to this application as a whole and not to any particular portions of this application. When the word 'or' is used in reference to a list of two or more items, that word covers all of the following interpretations of the word: any of the items in the list, all of the items in the list and any combination of the items in the list.

[00181] The above description of illustrated embodiments of the systems and methods is not intended to be exhaustive or to limit the systems and methods to the precise form disclosed. While specific embodiments of, and examples for, the systems and methods are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the systems and methods, as those skilled in the relevant art will recognize. The teachings of the systems and methods provided herein can be applied to other processing systems and methods, not only for the systems and methods described above.

[00182] The elements and acts of the various embodiments described above can be combined to provide further embodiments. These and other changes can be made to the systems and methods in light of the above detailed description.

[00183] In general, the terms used should not be construed to limit the systems and methods to the specific embodiments disclosed in the specification and the claims, but should be construed to include all processing systems that operate under the claims. Accordingly, the systems and methods are not limited by the disclosure.

CLAIMS

What is claimed is:

1. A method for classifying a disease, the method comprising:
 - a. providing a plurality of disease classification Modules, wherein the Modules comprise a list of genetic markers selected by (i) forming a pairwise cluster associations network of genetic markers among a plurality of reference samples and (ii) reducing the cluster association network to a discrete set of Modules defined by highly interconnected clusters;
 - b. interrogating a biological sample obtained from a patient for presence of one or more genetic markers of one or more of the Modules; and
 - c. identifying the patient's disease as belonging to a Module having genetic markers consistent with the genetic markers interrogated in the biological sample.
2. The method of Claim 1, wherein the genetic markers comprise genes having a similar expression profile across the plurality of reference samples, wherein the biological sample is interrogated for the expression of the genes, and wherein the patient's disease is identified as belonging to a Module based on the pattern of gene expression being consistent with the expression of said one or more gene members of the Module.
3. The method of Claim 1, wherein a cluster of genetic markers is highly interconnected if the Pearson correlation is at least 0.5.
4. The method of Claim 1, further comprising predicting drug responsiveness based on identifying the patient's disease as belonging to a Module.
5. The method of Claim 1, further comprising predicting the risk of disease recurrence based on identifying the patient's disease as belonging to a Module.
6. The method of Claim 1, further comprising predicting the risk of metastasis based on identifying the patient's disease as belonging to a Module.

7. The method of Claim 1, further comprising selecting a therapeutic drug for the patient based on identifying the patient's disease as belonging to a Module.
8. The method of Claim 1, wherein each genetic marker is selected for inclusion in a Module based on the marker being present in as little as 1% of the reference samples.
9. The method of Claim 1, wherein each genetic marker is selected for inclusion in a Module based on the marker being present in as little as 5% of the reference samples.
10. The method of Claim 1, wherein the number of reference samples is at least 100.
11. The method of Claim 1, wherein the number of reference samples is at least 1000.
12. The method of Claim 1, wherein the number of reference samples is at least 5000.
13. The method of Claim 1, wherein the plurality of reference samples are selected from a plurality of cohorts.
14. The method of Claim 13, wherein the number of cohorts is at least 3 and wherein the number of reference samples per cohort is at least 50.
15. The method of Claim 1, wherein the reference samples are obtained from at least 5 cohorts.
16. The method of Claim 1, wherein the reference samples are obtained from at least 10 cohorts.
17. The method of Claim 1, wherein the reference samples are obtained from at least 25 cohorts.
18. The method of Claim 1, wherein the reference samples are obtained from at least 50 cohorts.
19. The method of Claim 1, wherein the reference samples are obtained from at least 100 cohorts.
20. The method of Claim 1, wherein the Modules define independent biological functions.

21. The method of Claim 20, wherein the independent biological functions define independent co-expression sets.
22. A diagnostic assay prepared based on one or more disease classification Modules, wherein the Modules comprise a list of genetic markers selected by (i) forming a pairwise cluster associations network of genetic markers among a plurality of reference samples and (ii) reducing the cluster association network to a discrete set of Modules defined by highly interconnected clusters.
23. The assay of Claim 22 wherein the assay provides diagnosis in connection with more than one disease state.
24. A disease classification Module comprising a list of genetic markers selected by (i) forming a pairwise cluster associations network of genetic markers among a plurality of reference samples and (ii) reducing the cluster association network to a discrete set of Modules defined by highly interconnected clusters.
25. The module of Claim 24, wherein a cluster of genetic markers is highly interconnected if the Pearson correlation is at least 0.5.
26. A disease specific module wherein the module comprises a plurality of genetic markers selected first based on similarity of their expression profiles across a plurality of reference samples.
27. The module of claim 26, wherein the genetic markers selected first based on similarity of their expression profiles are named based on phenotypic information.
28. The module of claim 26, wherein the plurality of reference samples are selected from a plurality of cohorts.
29. The module of Claim 26, wherein each genetic marker is selected based on the marker being present in as little as 1% of the reference samples.

30. The module of Claim 26, wherein each genetic marker is selected based on the marker being present in as little as 5% of the reference samples.
31. The module of Claim 26, wherein the number of reference samples is at least 100.
32. The module of Claim 26, wherein the number of reference samples is at least 1000.
33. The module of Claim 26, wherein the number of reference samples is at least 5000.
34. The module of Claim 26, wherein the reference samples are obtained from at least 5 cohorts.
35. The module of Claim 26, wherein the reference samples are obtained from at least 10 cohorts.
36. The module of Claim 26, wherein the reference samples are obtained from at least 25 cohorts.
37. The module of Claim 26, wherein the reference samples are obtained from at least 50 cohorts.
38. The module of Claim 26, wherein the reference samples are obtained from at least 100 cohorts.
39. The module of Claim 26, wherein the number of cohorts is at least 3 and wherein the number of reference samples per cohort is at least 50.
40. A plurality of modules according to claim 26, wherein the modules define independent biological functions.
41. The modules of claim 40, wherein the independent biological functions define independent co-expression sets.
42. A method for predicting a clinical outcome of a test patient, the method comprising comparing the expression profile of biomarkers associated with one or more modules according to claim 24.

- 43. The method of claim 42, wherein the prediction is selected from the group consisting of drug sensitivity, drug insensitivity, recurrence and metastasis.
- 44. The method of claim 43, wherein more than one prediction is made.
- 45. A diagnostic assay prepared based on one or more modules according to claim 24.
- 46. The assay of claim 45 wherein the assay provides diagnosis in connection with more than one disease state.

FIGURE 1

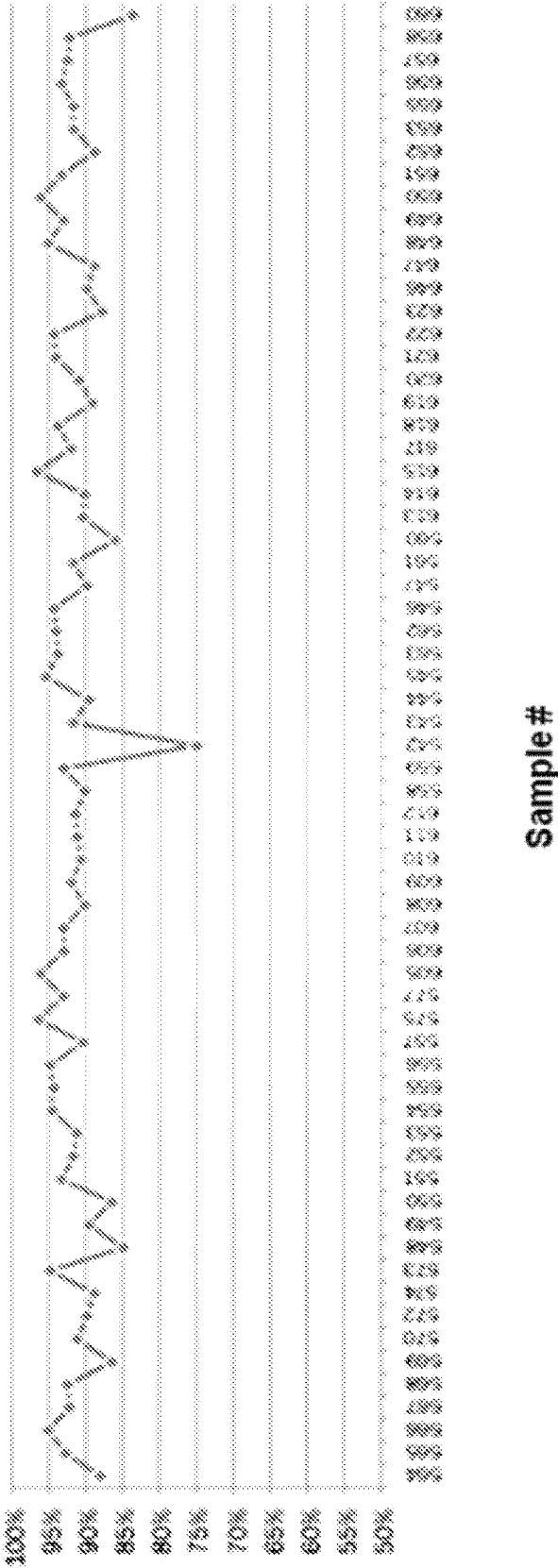


FIGURE 2

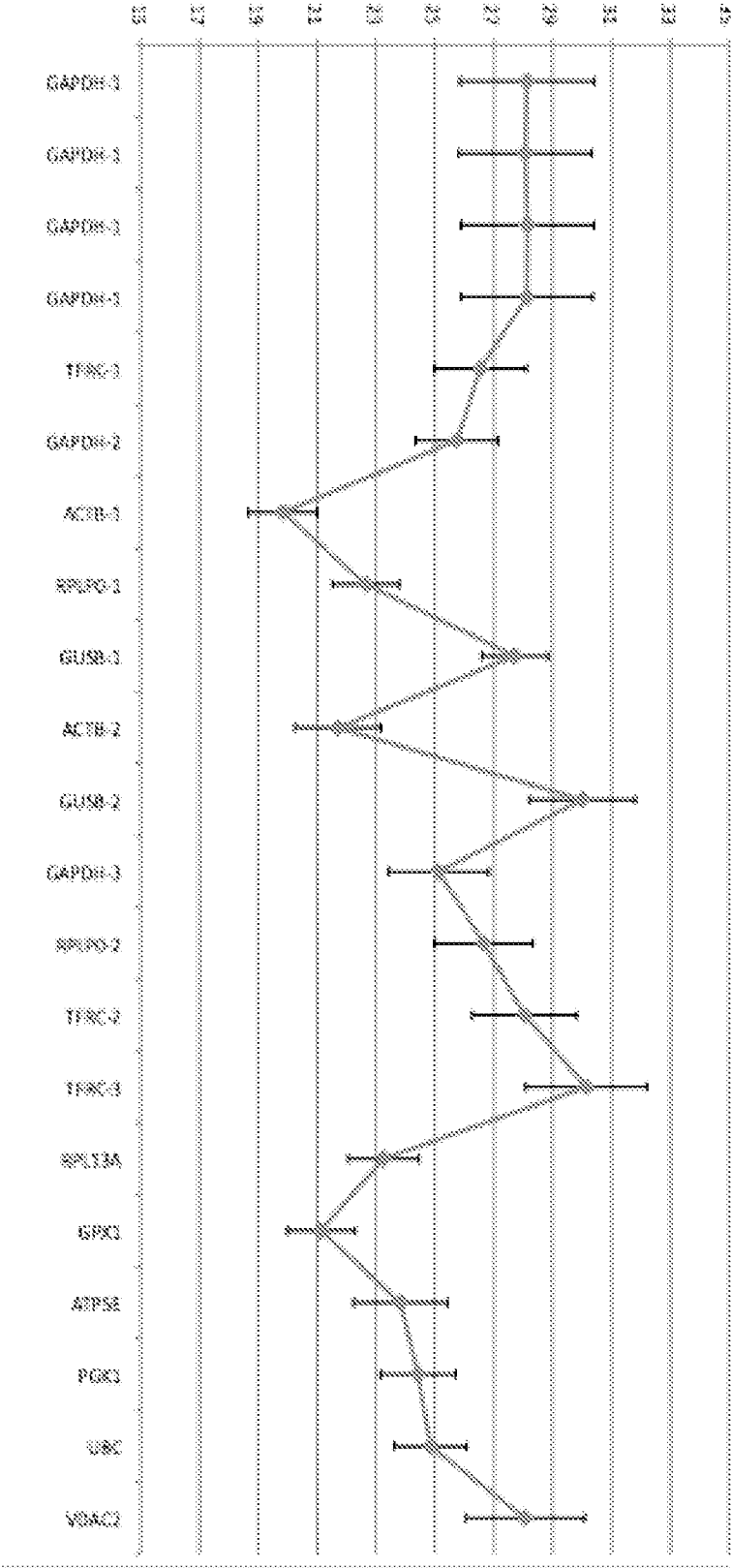


FIGURE 3

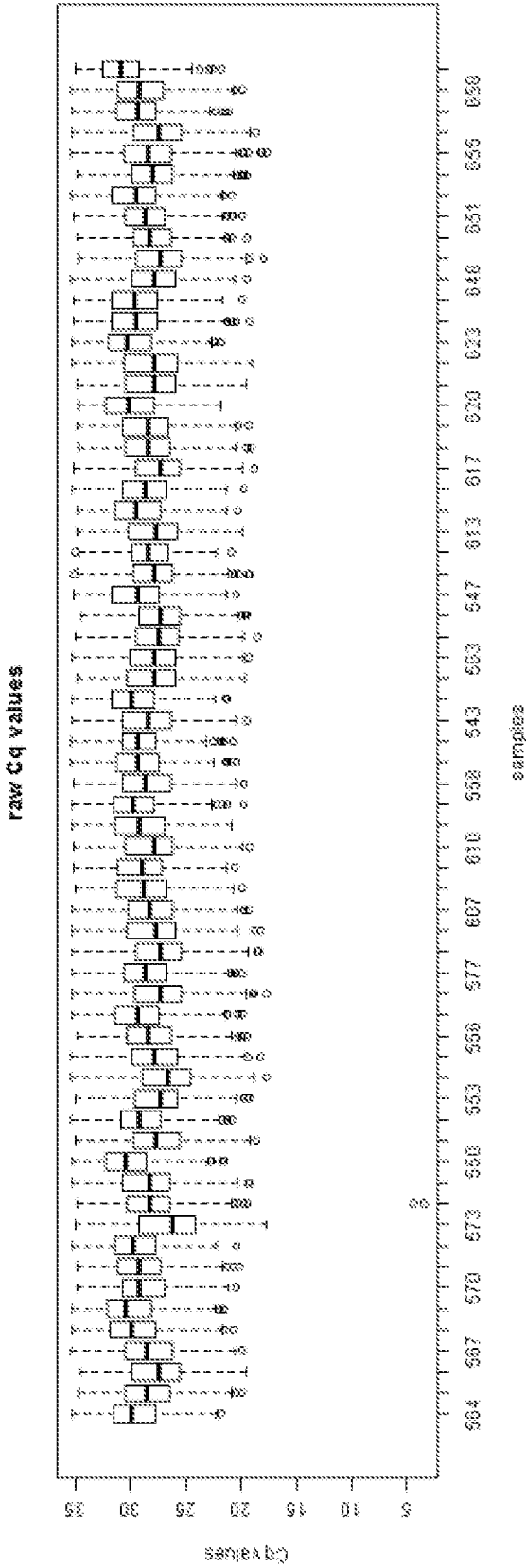


FIGURE 4

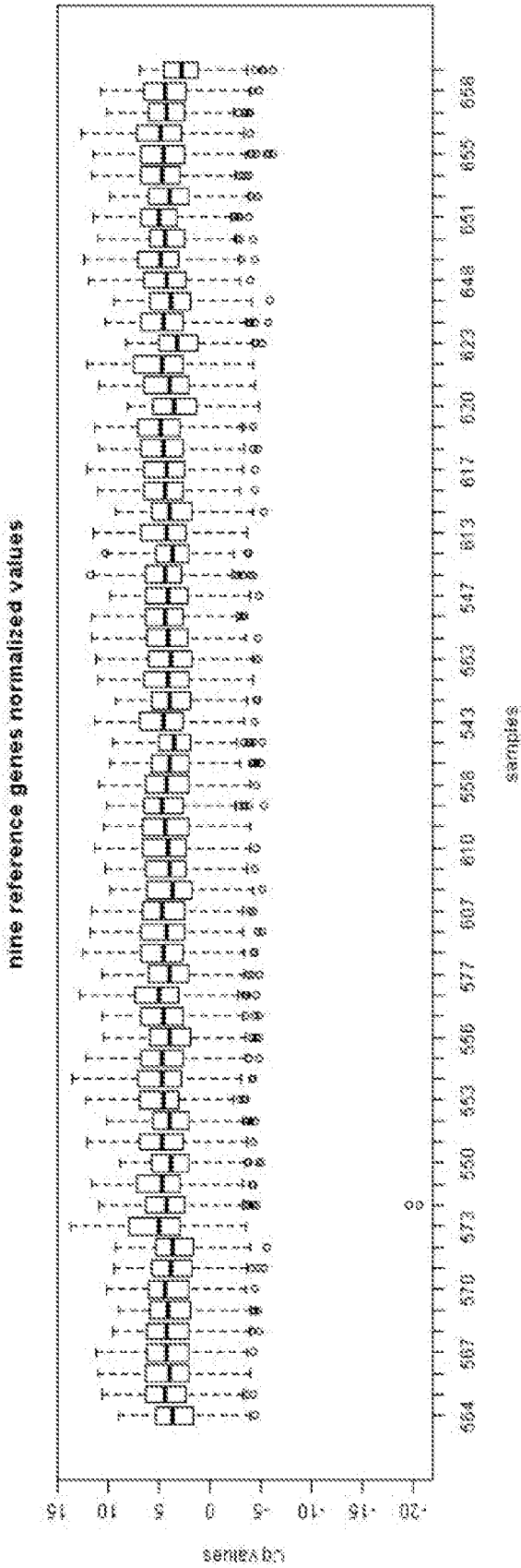


FIGURE 5

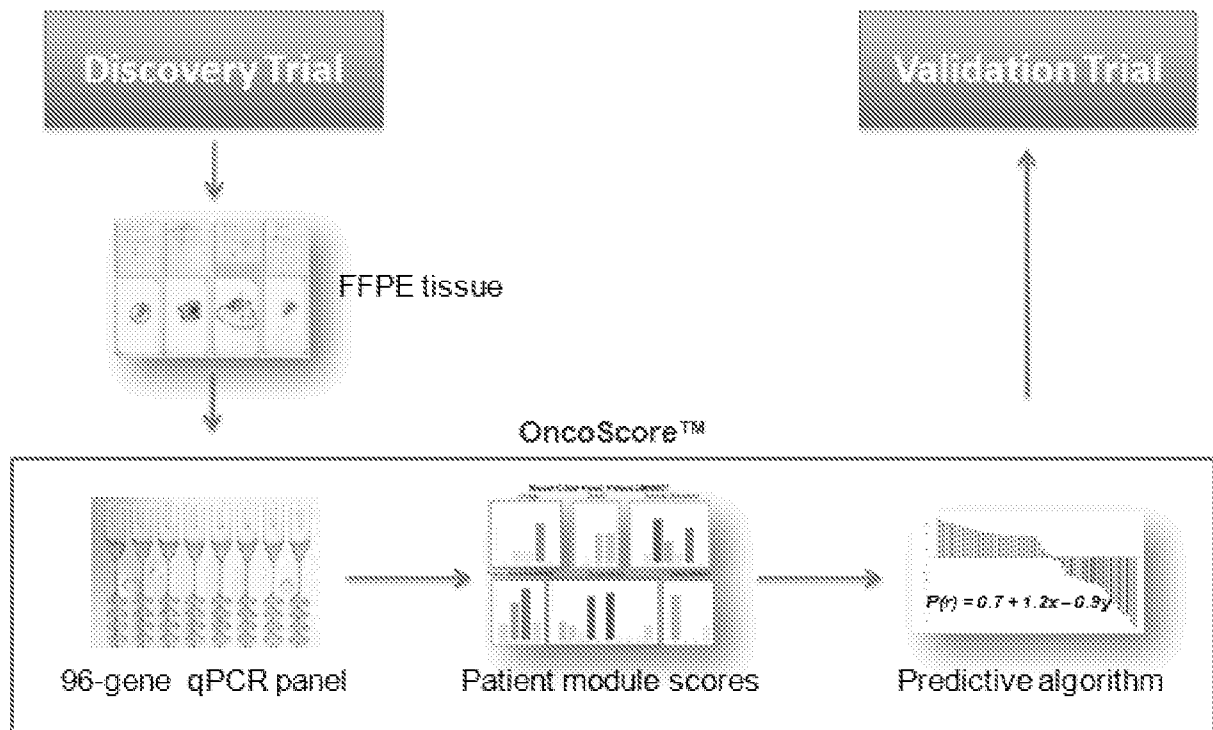


FIGURE 6

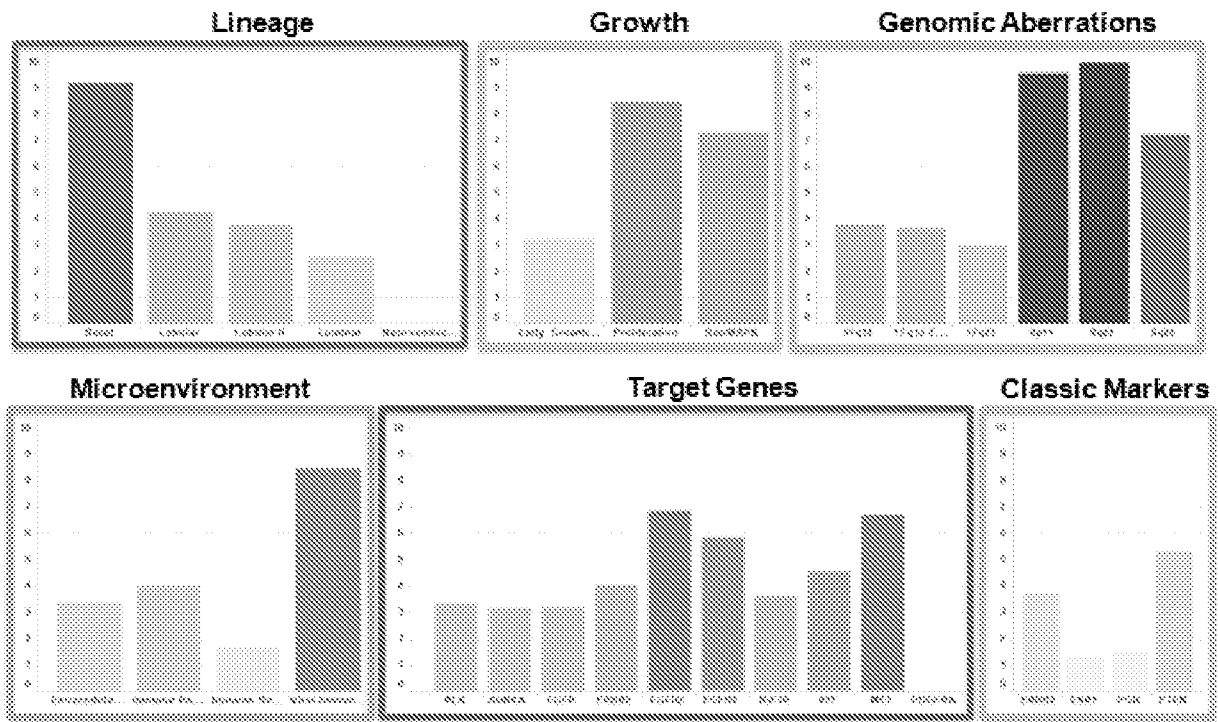


FIGURE 7

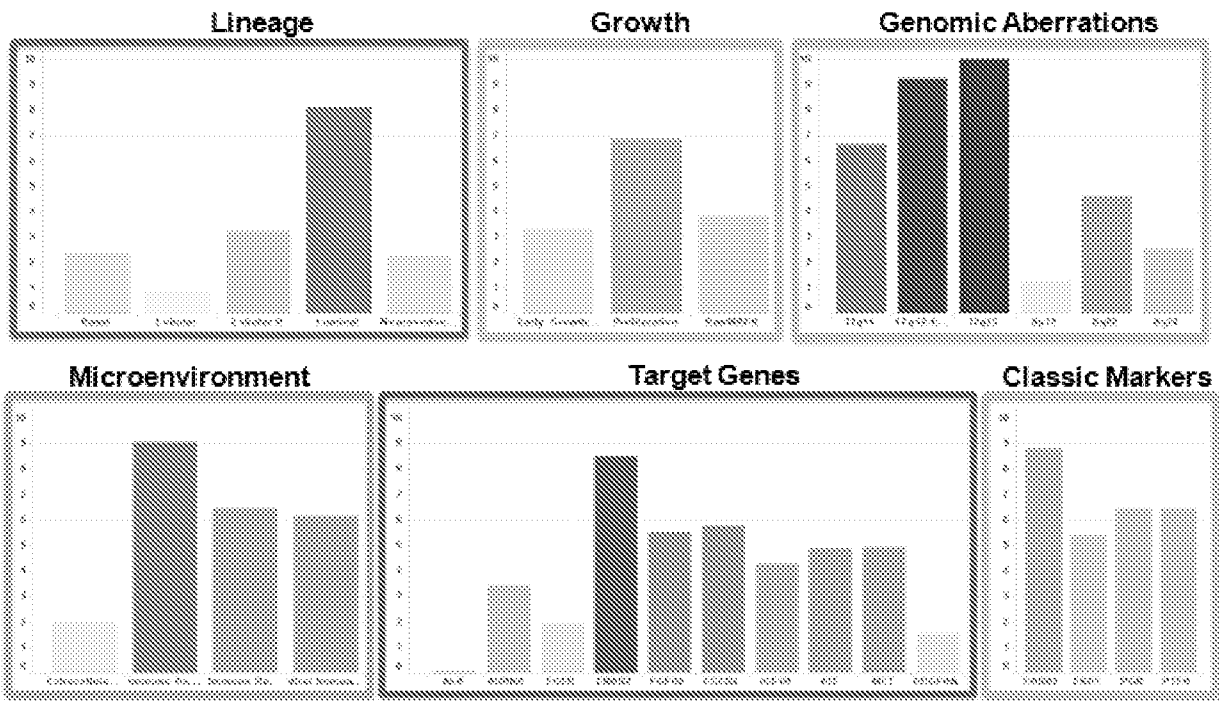


FIGURE 8

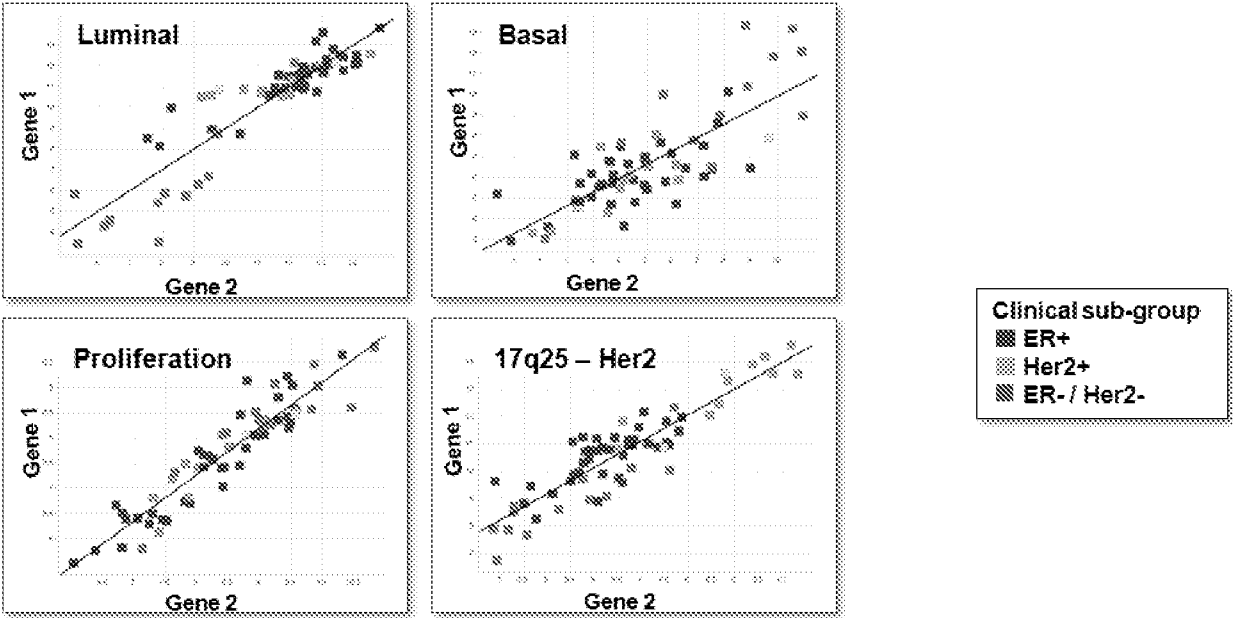


FIGURE 9

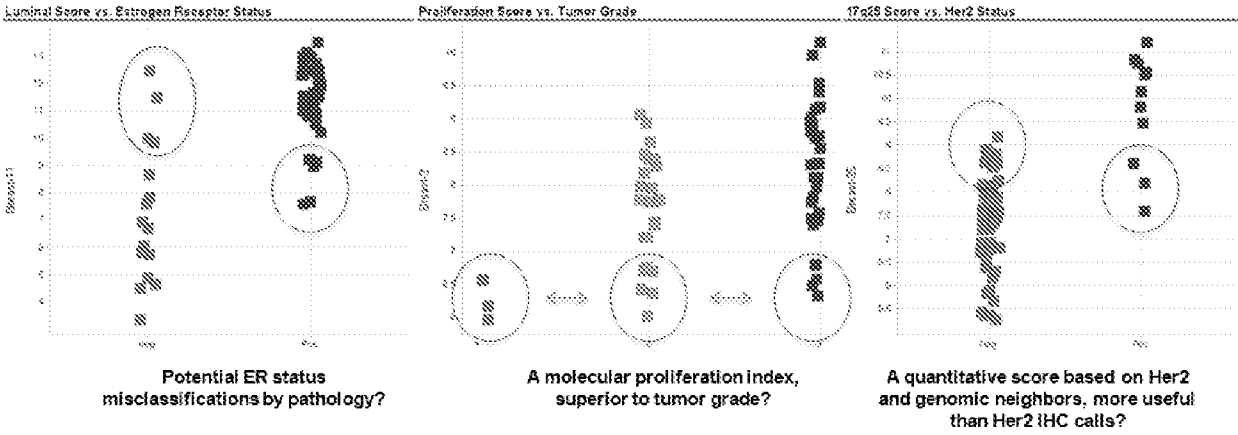


FIGURE 10

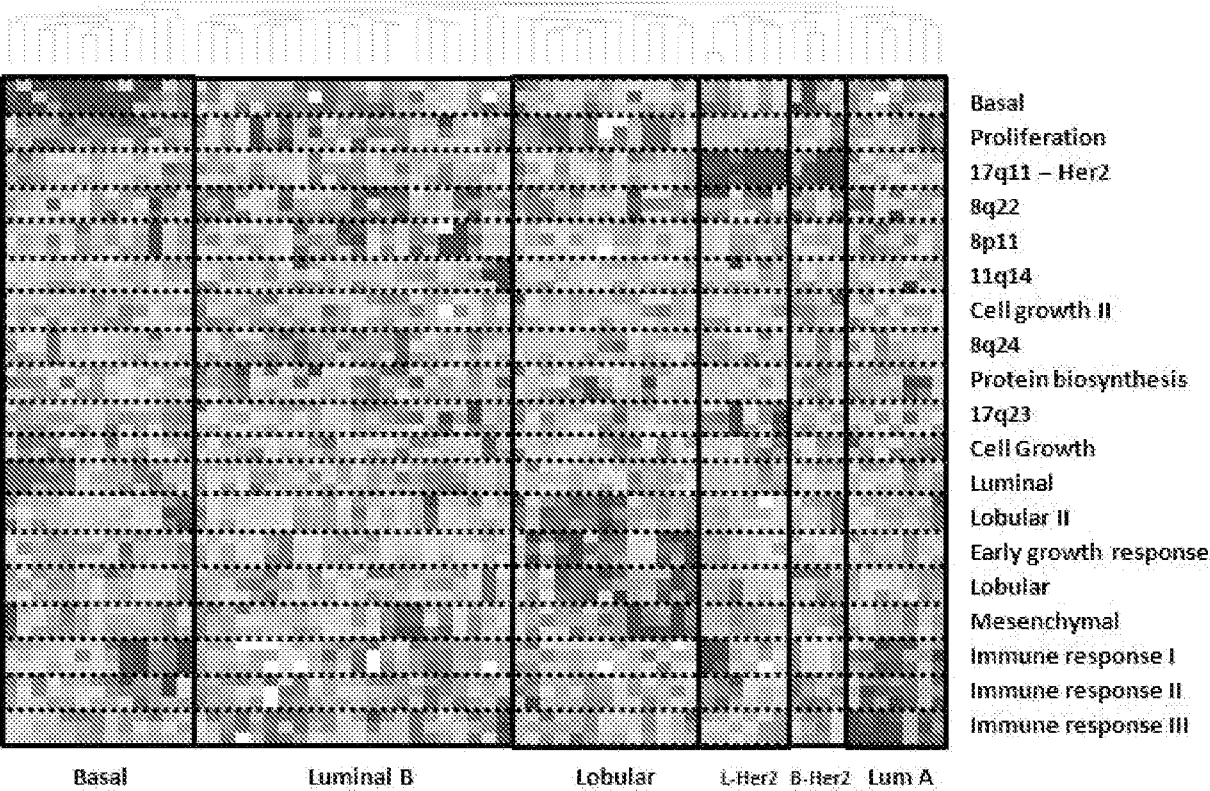


FIGURE 11

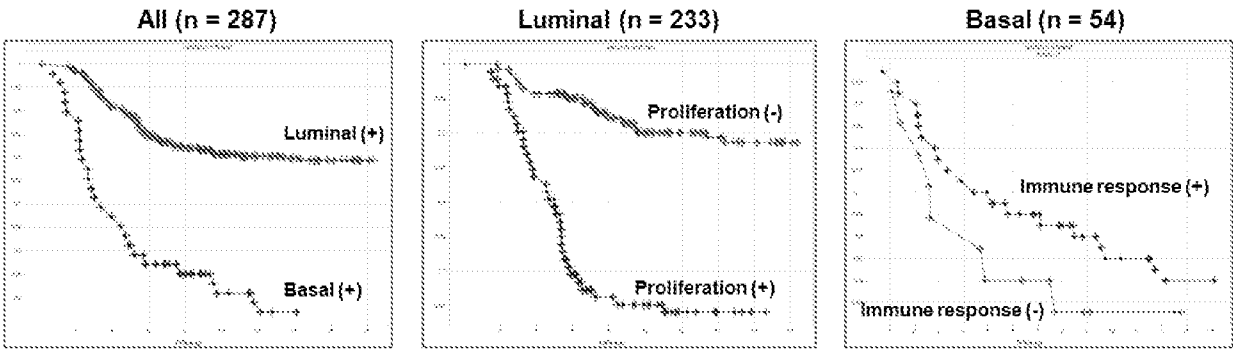


FIGURE 12

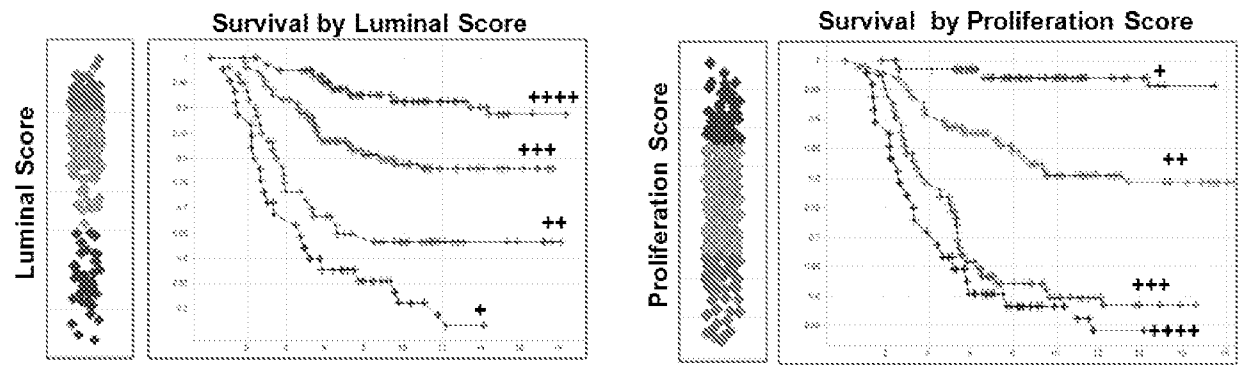


FIGURE 13

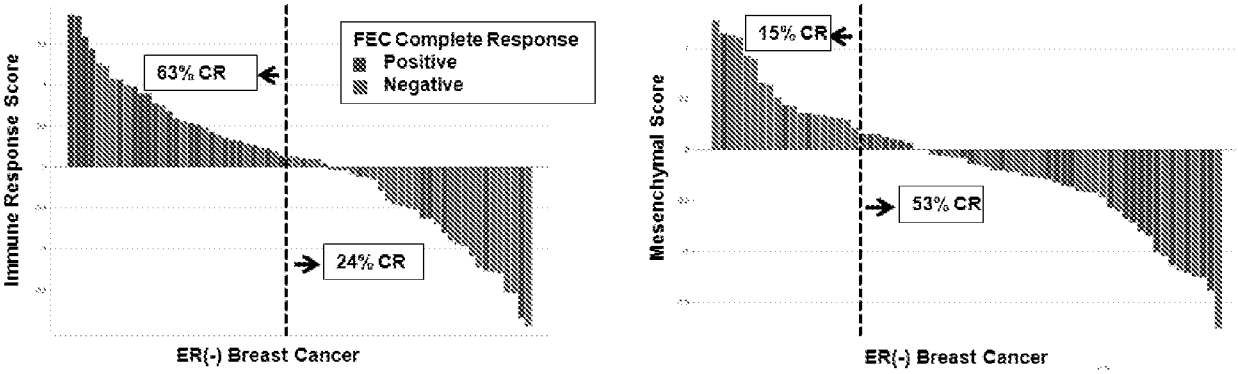
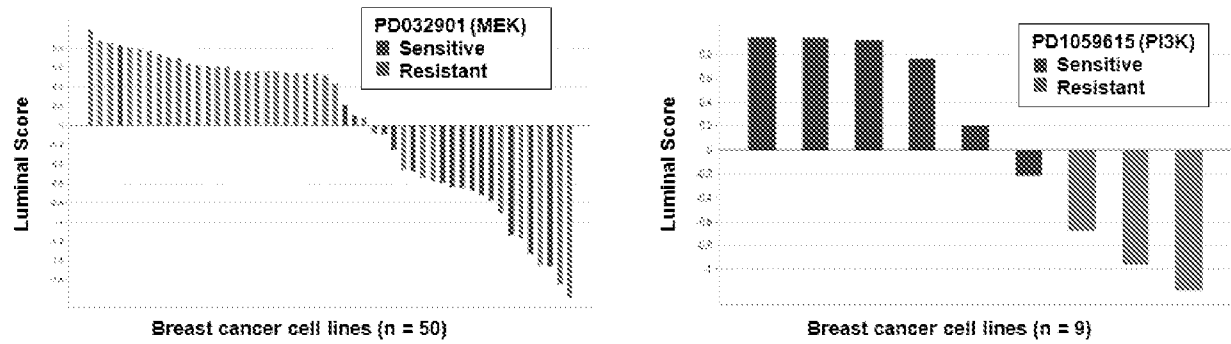


FIGURE 14



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2012/041385

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68 (2012.01)

USPC - 435/6.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - C12Q 1/68, 70; C40B 30/02; G01N 33/48; G06F 19/00 (2012.01)

USPC - 435/6, 7.1, 91.1; 506/8; 702/20

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase, Google Patents, Google, PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2010/0285980 A1 (SHAK et al) 11 November 2010 (11.11.2010) entire document	1-46
Y	US 2008/0305965 A1 (MOORHOUSE et al) 11 December 2008 (11.12.2008) entire document	1-46
Y	US 2010/0233691 A1 (BANKAITIS-DAVIS et al) 16 September 2010 (16.09.2010) entire document	13-19, 28, 34-39
A	US 7,939,261 B2 (BAKER et al) 10 May 2011 (10.05.2011) entire document	1-46
A	US 2001/0036631 A1 (MCGRATH et al) 01 November 2001 (01.11.2001) entire document	1-46

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

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

20 July 2012

Date of mailing of the international search report

09 AUG 2012

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774