(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 December 2011 (22.12.2011)

PCT

(10) International Publication Number
WO 2011/160118 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2011/041080

(22) International Filing Date:
20 June 2011 (20.06.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/356,516 18 June 2010 (18.06.2010) US

(71) Applicant *(for all designated States except US)*: **MED BIOGENE INC.** [CA/CA]; 2386 East Mall, Gerald McGavin Building, #300, Vancouver, British Columbia V6T 1Z3 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SAAD, Fadia** [CA/CA]; 2072 Allenby Street, Victoria, british Columbia V8R 3C1 (CA). **SCHISLER, Jonathan, C.** [US/US]; 121 Hatchet Creek Ct., Morrisville, North Carolina 27560-9514 (US). **PARKER, Joel** [US/US]; 1113 Churchwood Drive, Apex, North Carolina 27502 (US). **BUERKI, Christine** [CH/CA]; 395 14th Ave W, #401, Vancouver, british Columbia BC V5Y 1X3 (CA).

(74) Agents: **TUSCAN, Michael** et al.; Cooley LLP, 777 6th Street, N.W., Suite 1100, Washington, District of Columbia 20001 (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, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: PROGNOSTIC AND PREDICTIVE GENE SIGNATURE FOR NON-SMALL CELL LUNG CANCER AND ADJUVANT CHEMOTHERAPY

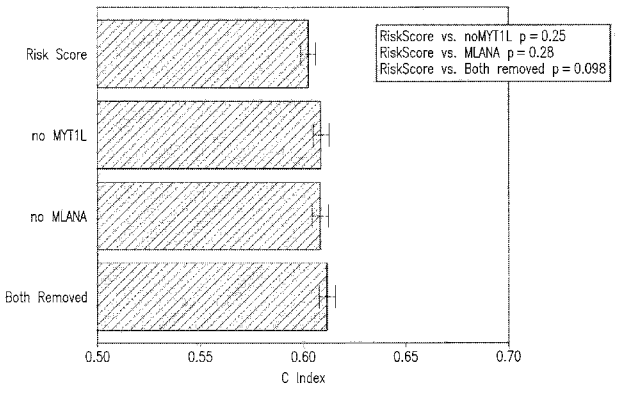


FIG. 7

(57) **Abstract**: The application provides methods of prognosing and classifying lung cancer patients into poor survival groups or good survival groups and for determining the benefit of adjuvant chemotherapy by way of a multigene signature. The application also includes kits and computer products for use in the methods of the application.

WO 2011/160118 A2

PROGNOSTIC AND PREDICTIVE GENE SIGNATURE FOR NON-SMALL CELL LUNG CANCER AND ADJUVANT CHEMOTHERAPY

Priority

[001] This application claims the benefit of U.S. Application No. 61/356,516, filed June 18, 2010, which is hereby incorporated by reference in its entirety.

Field

[002] The application relates to compositions and methods for prognosing and classifying non-small cell lung cancer and for determining the benefit of adjuvant chemotherapy.

Background of the invention

[003] In North America, lung cancer is the leading cancer in males and the leading cause of cancer deaths in both males and females[1]. Non-small cell lung cancer (NSCLC) represents 80% of all lung cancers and has an overall 5-year survival rate of only 16% [1]. Tumor stage is the primary determinant for treatment selection for NSCLC patients. Recent clinical trials have led to the adoption of adjuvant cisplatin-based chemotherapy in early stage NSCLC patients (Stages IB-IIIA). The 5-year survival advantage conferred by adjuvant chemotherapy in recent trials are 4% in the International Adjuvant Lung Trial (IALT) involving 1,867 stage I-III patients[2], 15% in the National Cancer Institute of Canada Clinical Trials Group (NCIC CTG) BR.10 Trial involving 483 stage IB-II patients[3], and 9% in the Adjuvant Navelbine International Trialist Association (ANITA) trial involving 840 stage IB-IIIA patients[4]. Pre-planned stratification analysis in the later two trials showed no significant survival benefit for stage IB patients[3, 4]. This was also demonstrated in the Cancer and Leukemia Group (CALGB) Trial 9633 that tested the benefit of chemotherapy on 344 stage IB patients receiving carboplatin and paclitaxel or observation[5]. Although initially presented in 2004 as a positive trial, recent survival analyses show no significant survival advantage with chemotherapy for either disease-free survival (HR=0.80, p=0.065) or overall survival (HR=0.83, p=0.12)[5]. In an attempt to draw an overall conclusion regarding the effectiveness of adjuvant cisplatin-based chemotherapy, the Lung Adjuvant Cisplatin Evaluation (LACE) meta-analysis[6] was conducted which synthesized information from the 5 largest published,

cisplatin-based trials that did not administer concurrent thoracic radiation [Adjuvant Lung Project Italy (ALPI)[7], Big Lung Trial (BLT)[8], IALT[2], BR.10[3], and ANITA[9]]. The study found a 5.3% absolute survival advantage at 5-year (HR=0.89, 95%CI 0.82-0.96, p=0.004). However, stratified analysis by stage showed that the stage IB patients did not benefit significantly from cisplatin treatment (HR=0.92, 95%CI 0.78-1.10). Moreover, a detriment for chemotherapy was suggested in stage IA patients (HR=1.41, 95%CI 0.96-2.09) [6]. Therefore, the current standard of treatment for patients with stage I NSCLC remains surgical resection alone. However, 30 to 40 percent of these stage I patients are expected to relapse after the initial surgery[10, 11], indicating that a subgroup of these patients might benefit from adjuvant chemotherapy.

[004] The lack of consistent prognostic molecular markers for early stage NSCLC patients led to attempts to identify novel gene expression signatures using genome wide microarray platforms. Such multi-gene signatures might be stronger than individual genes to predict poor prognosis and poor prognostic patients could potentially benefit from adjuvant therapies. Previous microarray studies have identified prognostic signatures that demonstrated minimal overlaps in the gene sets.[12-20] While only one of the early studies involved secondary signature validation in independent datasets[12], all recently reported signatures were tested for validation[13-16, 20]. Nevertheless, lack of direct overlaps between signatures remains. One of the potential confounding factors is that signatures were derived from patients operated at single institutions, which may introduce biases.

Summary of the invention

[005] As discussed in the Background section, certain patients suffering from NSCLC benefit from adjuvant chemotherapy. Attempts to identify systematically patient subpopulations in which adjuvant therapy would lead to increased survival or improve patient prognosis have generally failed. Efforts to assemble prognostic molecular markers have yielded various non-overlapping gene sets but have fallen short of establishing a gene signature with a minimal set of genes that is predictive regardless of the form of NSCLC (eg. adenocarcinoma or squamous cell carcinoma) or stage, and serves as a reliable classifier for adjuvant therapy benefit.

[006] As will be discussed in more detail below, a set of fifteen genes were previously identified by microarray analysis whose expression level is useful in the

prognosis of survival outcome and diagnosis of adjuvant therapy benefit. The prognostic and diagnostic value of the 15-gene set was verified by validation against independent data sets. In migrating this signature to a qPCR-based platform, it was discovered that a fewer number of genes can provide essentially the same predictive value, including a 13-gene signature. In various embodiments of the invention, the present disclosure provides methods and kits useful for obtaining and utilizing expression information for the fourteen or fewer genes to obtain prognostic and diagnostic information for a patient with NSCLC. In some embodiments, the invention provides methods and kits useful for obtaining and utilizing expression information of at least 5 of the 14 genes.

[007] In another aspect of the disclosure, it is shown that the genes maintain their predictive value when moved from microarray detection platform to quantitative PCR, and are applicable to both fresh frozen tissue and formalin-fixed paraffin-embedded (FFPE) tissue samples.

[008] The methods of the present disclosure generally involve obtaining from a patient relative expression data, at the DNA, messenger RNA (mRNA), or protein level, for each of the genes and micro RNAs (miRNAs) regulating those genes, included in the set. In some embodiments, the present disclosure involves processing the data and comparing the resulting information to one or more reference values. Relative expression levels are expression data normalized according to techniques known to those skilled in the art. Expression data may be normalized with respect to one or more genes with invariant expression, such as "housekeeping" genes. In some embodiments, expression data may be processed using standard techniques, such as transformation to a z-score, and/or software tools, such as RMAexpress v0.3.

[009] In one aspect, the invention provides a method for preparing a gene expression profile indicative of response to adjuvant chemotherapy for NSCLC. The method comprises determining the level of expression of at least five genes from Table 4A. Table 4A discloses 13 genes shown herein to maintain the predictive capacity of a 15 gene signature identified by a microarray detection format. The gene expression profile may be prepared from a fresh frozen tumor specimen or a FFPE specimen, and may be determined by quantitative PCR, or other amplification detection platform, which as shown herein is sufficient for maintaining the predictive

capacity. In various embodiments, the gene expression profile does not include expression levels for MLANA and/or MYT1L, and may be normalized based on the expression level of one or more additional genes, such as one or more of BAT1, TBP, PP1A, and GUSB. Exemplary target sequences, primer sequences, and probe sequences for preparing expression profiles are further described herein.

[010] In another aspect, the invention provides a method for predicting the benefit of adjuvant chemotherapy for a patient having non-small cell lung cancer. The method comprises determining a gene expression profile that includes the level of expression of from 5 to 14 genes each indicative of survival in a NSCLC population, with the genes being listed in Table 3. Preferably, at least 5 genes (e.g., 5, 6, 7, 8, 9, 10, 11, 12, or 13 genes) are listed in Table 4A. The gene expression profile is prepared from expression data obtained from fresh frozen or FFPE tumor tissue samples using a quantitative PCR detection platform. The profile is then classified to predict whether the patient will benefit from adjuvant chemotherapy, as described herein. For example, the profile may be classified by analyzing the gene expression levels of the 5 to 14 genes in connection with a classifier algorithm. The classifier algorithm may classify samples into a high risk group that is likely to benefit from adjuvant chemotherapy, or a low risk group where adjuvant chemotherapy is less likely to benefit the patient.

[011] In a third aspect, a multi-gene signature is provided for prognosing or classifying patients with lung cancer (e.g., NSCLC). In some embodiments, a five to fourteen-gene signature is provided, comprising reference values for each of the five to fourteen different genes based on relative expression data for each gene from a historical data set with a known outcome, such as good or poor survival, and/or known treatment, such as adjuvant chemotherapy. In one embodiment, four reference values are provided for each of the five to thirteen genes listed in Table 4A. In one embodiment, the reference values for each of the five to thirteen genes are principal component values, such as those set forth in Table 10 for example.

[012] In some embodiments, a fourteen-gene signature comprises reference values for each of fourteen different genes based on relative expression data for each gene from a historical data set with a known outcome and/or known treatment. In some embodiments, reference values are provided for one gene in addition to

those listed in Table 4A, and the gene is selected from those listed in Table 3. In some embodiments, a single reference value for each gene is provided.

[013] In one aspect, relative expression data from a patient are combined with the gene-specific reference values on a gene-by-gene basis for each of the five to fourteen genes to generate a test value which allows prognosis or therapy recommendation. In some embodiments, relative expression data are subjected to an algorithm that yields a single test value, or combined score, which is then compared to a control value obtained from the historical expression data for a patient or pool of patients. In some embodiments, the control value is a numerical threshold for predicting outcomes, for example good and poor outcome, or making therapy recommendations, for example adjuvant therapy in addition to surgical resection or surgical resection alone. In some embodiments, a test value or combined score greater than the control value is predictive, for example, of high risk (poor outcome) or benefit from adjuvant therapy, whereas a combined score falling below the control value is predictive, for example, of low risk (good outcome) or lack of benefit from adjuvant therapy.

[014] In one embodiment, the combined score is calculated from relative expression data multiplied by reference values, determined from historical data, for each gene. Accordingly, the combined score may be calculated using the algorithm of Formula I below:

$$\text{Combined score} = 0.557 \times PC1 + 0.328 \times PC2 + 0.43 \times PC3 + 0.335 \times PC4$$

Where PC1 is the sum of the relative expression level for each gene in a multi-gene signature multiplied by a first principal component for each gene in the multi-gene signature, PC2 is the sum of the relative expression level for each gene multiplied by a second principal component for each gene, PC3 is the sum of the relative expression level for each gene multiplied by a third principal component for each gene, and PC4 is the sum of the relative expression level for each gene multiplied by a fourth principal component for each gene. In some embodiments, the combined score is referred to as a risk score. A risk score for a subject can be calculated by applying Formula I to relative expression data from a test sample obtained from the subject.

[015] In some embodiments, PC1 is the sum of the relative expression level for each gene provided in Table 4A multiplied by a first principal component for each gene, respectively, as set forth in Table 10; PC2 is the sum of the relative expression level for each gene provided in Table 4A multiplied by a second principal component for each gene, respectively, as set forth in Table 10; PC3 is the sum of the relative expression level for each gene provided in Table 4A multiplied by a third principal component for each gene, respectively, as set forth in Table 10; and PC4 is the sum of the relative expression level for each gene provided in Table 4A multiplied by a fourth principal component for each gene, respectively, as set forth in Table 10.

[016] The present disclosure provides a gene signature that is prognostic for survival as well as predictive for benefit from adjuvant chemotherapy.

[017] Accordingly in one embodiment, the application provides a method of prognosing or classifying a subject with non-small cell lung cancer comprising the steps:

a. determining the expression of thirteen biomarkers in a test sample from the subject, wherein the biomarkers correspond to genes in Table 4A, and

b. comparing the expression of the thirteen biomarkers in the test sample with expression of the thirteen biomarkers in a control sample,

wherein a difference or a similarity in the expression of the thirteen biomarkers between the control and the test sample is used to prognose or classify the subject with NSCLC into a poor survival group or a good survival group.

[018] In an aspect, the application provides a method of predicting prognosis in a subject with non-small cell lung cancer comprising the steps:

a. obtaining a subject biomarker expression profile in a sample of the subject;

b. obtaining a biomarker reference expression profile associated with a prognosis, wherein the subject biomarker expression profile and the biomarker reference expression profile each have thirteen values, each

value representing the expression level of a biomarker, wherein each biomarker corresponds to one gene in Table 4A; and

c. selecting the biomarker reference expression profile most similar to the subject biomarker expression profile, to thereby predict a prognosis for the subject.

[019] In another aspect, the prognoses and classifying methods of the application can be used to select treatment. For example, the methods can be used to select or identify subjects who might benefit from adjuvant chemotherapy. Accordingly, in one embodiment, the application provides a method of selecting a therapy for a subject with NSCLC, comprising the steps:

a. classifying the subject with NSCLC into a poor survival group or a good survival group according to the method of the application; and

b. selecting adjuvant chemotherapy for the poor survival group or no adjuvant chemotherapy for the good survival group.

[020] In another embodiment, the application provides a method of selecting a therapy for a subject with NSCLC, comprising the steps:

a. determining the expression of thirteen biomarkers in a test sample from the subject, wherein the thirteen biomarkers correspond to the thirteen genes in Table 4A;

b. comparing the expression of the thirteen biomarkers in the test sample with the thirteen biomarkers in a control sample;

c. classifying the subject in a poor survival group or a good survival group, wherein a difference or a similarity in the expression of the thirteen biomarkers between the control sample and the test sample is used to classify the subject into a poor survival group or a good survival group;

d. selecting adjuvant chemotherapy if the subject is classified in the poor survival group and selecting no adjuvant chemotherapy if the subject is classified in the good survival group.

[021] Another aspect of the application provides compositions for use with the methods described herein.

[022] The application also provides for kits used to prognose or classify a subject with NSCLC into a good survival group or a poor survival group or for selecting therapy for a subject with NSCLC that includes detection agents that can detect the expression products of the biomarkers.

[023] In one aspect, the present disclosure provides kits useful for carrying out the diagnostic and prognostic tests described herein. The kits generally comprise reagents and compositions for obtaining relative expression data for from 5 to 14 genes from Table 3, and including at least 5 genes from Table 4 (e.g., 5, 6, 7, 8, 9, 10, 11, 12, or 13 genes from Table 4A). In certain embodiments, the kits comprise reagents and compositions for obtaining relative expression levels of the genes described in Table 4A, and optionally, an additional gene selected from among those listed in Table 3. As will be recognized by the skilled artisans, the contents of the kits will depend upon the means used to obtain the relative expression information.

[024] Kits may comprise a labeled compound or agent capable of detecting protein product(s) or nucleic acid sequence(s) in a sample and means for determining the amount of the protein or mRNA in the sample (e.g., an antibody which binds the protein or a fragment thereof, or an oligonucleotide probe which binds to DNA or mRNA encoding the protein). Kits can also include instructions for interpreting the results obtained using the kit.

[025] In some embodiments, the kits are oligonucleotide-based kits, which may comprise, for example: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a marker protein or (2) a pair of primers useful for amplifying a marker nucleic acid molecule. Kits may also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kits can further comprise components necessary for detecting the detectable label (e.g., an enzyme or a substrate). The kits can also contain a control sample or a series of control samples which can be assayed and compared to the test sample. Each component of a kit can be enclosed within an individual

container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[026] In some embodiments, the kits are antibody-based kits, which may comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a marker protein; and, optionally, (2) a second, different antibody which binds to either the protein or the first antibody and is conjugated to a detectable label.

[027] A further aspect provides computer implemented products, computer readable mediums and computer systems that are useful for the methods described herein.

[028] Other features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

Brief description of the drawings

[029] The invention will now be described in relation to the drawings in which:

[030] Figure 1(A,B) shows the derivation and testing of a prognostic signature;

[031] Figure 2(A-F) shows the survival outcome based on the 15-gene signature in training and test sets;

[032] Figure 3(A-H) shows a comparison of chemotherapy vs. observation in low and high risk patients with microarray data;

[033] Figure 4 shows a consort diagram for microarray study of BR. 10 patients;

[034] Figure 5 shows the effect of adjuvant chemotherapy in microarray profiled patients;

[035] Figure 6 shows the effect of microarray batch processing at 2 different times. The samples were profiled in 2 batches at 2 times (January 2004 and June

2005). Unsupervised clustering shows that the expression patterns of these two batches differed significantly with samples arrayed on January 2004 aggregated in cluster 1 (93%) and samples arrayed on June 2005 in cluster 2 (73%).

[036] Figure 7 shows the risk score for the 15-gene signature ("Risk Score"), a 14-gene signature that omits MYT1L ("No MYTL1"), a 14-gene signature that omits MLANA ("No MLANA"), and a 13-gene signature that omits both MYTL1 and MLANA ("Both Removed"). There was no significant difference between the the 15 gene signature risk score and either 14-gene signature, from which either MYT1L or MLANA was removed (p=0.25 and p=0.28, respectively). There was also no significant diffrence between risk scores generated by the 15-gene and the 13-gene signatures –when both MYT1L and MLANA were removed (p=0.098).

[037] Figure 8 shows the process for evaluating the gene signature for RT-qPCR and with different probe chemistries.

[038] Figure 9 shows exemplary probe selection for ATP1B1.

[039] Figure 10(A-E) shows RT-qPCR on FFPE specimens, and assay efficiency.

[040] Figure 11(A-E) shows assay and selection of best performing reference gene.

[041] Figure 12(A,B) shows the correlation between frozen and FFPE samples.

[042] Figure 13 shows the correlation between qPCR and microarray-based risk scores.

Detailed description of the invention

[043] The application in various embodiments relates to a set of 13 or 14 biomarkers, and/or subsets thereof, that form a gene signature, and provides methods, compositions, computer implemented products, detection agents and kits for preparing gene expression profiles, and for prognosing or classifying a subject with non-small cell lung cancer (NSCLC) and for determining the benefit of adjuvant chemotherapy.

[044] In these and other embodiments, the application relates to preparing gene expression profiles from fresh frozen or FFPE NSCLC tumor tissue samples by

quantitative RT-PCR (RT-qPCR). Such gene expression profiles comprise the level of expression of from 5 to 14 (e.g., 13) genes that are correlate with survival in NSCLC, and such genes are listed in Tables 3 and 4A. The gene expression profiles may further be normalized using expression data from one or more normalization genes. The gene expression profiles may be used for classifying samples to predict the benefit of adjuvant chemotherapy.

[045] The term "biomarker" as used herein refers to a gene that is differentially expressed in tumor tissue excised from individuals with non-small cell lung cancer (NSCLC) according to prognosis and is predictive of different survival outcomes and of the benefit of adjuvant chemotherapy. In some embodiments, a 13-gene signature comprises 13 biomarker genes listed in Table 4A, or a subset thereof. One optional additional biomarker for a 14-gene signature may be selected from the genes listed in Table 3.

[046] Accordingly, one aspect of the invention is a method of prognosing or classifying a subject with non-small cell lung cancer, comprising the steps:

a. determining the expression of thirteen biomarkers in a test sample from the subject, wherein the biomarkers correspond to genes in Table 4A, and

b. comparing the expression of the thirteen biomarkers in the test sample with expression of the thirteen biomarkers in a control sample,

wherein a difference or a similarity in the expression of the thirteen biomarkers between the control and the test sample is used to prognose or classify the subject with NSCLC into a poor survival group or a good survival group.

[047] In another aspect, the application provides a method of predicting prognosis in a subject with non-small cell lung cancer (NSCLC) comprising the steps:

a. obtaining a subject biomarker expression profile in a sample of the subject;

b. obtaining a biomarker reference expression profile associated with a prognosis, wherein the subject biomarker expression profile and the biomarker reference expression profile each have thirteen values, each value representing the

expression level of a biomarker, wherein each biomarker corresponds to a gene in Table 4; and

c. selecting the biomarker reference expression profile most similar to the subject biomarker expression profile, to thereby predict a prognosis for the subject.

[048] The term "reference expression profile" as used herein refers to the expression of the biomarkers or genes, such as those listed in Table 3 and Table 4A associated with a clinical outcome in a NSCLC patient. The reference expression profile comprises values (e.g., 5 to 14 values), each value representing the expression level of a biomarker, wherein each biomarker corresponds to one gene in Table 3 or Table 4A. The reference expression profile is identified using one or more samples comprising tumor tissue (NSCLC) wherein the expression is similar between related samples defining an outcome class or group such as poor survival or good survival and is different to unrelated samples defining a different outcome class such that the reference expression profile is associated with a particular clinical outcome. The reference expression profile is accordingly a reference profile of the expression of the genes in Table 3 or Table 4A (including the gene sets and subsets described herein), to which the subject expression levels of the corresponding genes in a patient sample (e.g., the gene expression profile) are compared in methods for determining or predicting clinical outcome.

[049] As used herein, the term “control” refers to a specific value or dataset that can be used to prognose or classify the value e.g expression level or reference expression profile obtained from the test sample associated with an outcome class. In one embodiment, a dataset may be obtained from samples from a group of subjects known to have NSCLC and good survival outcome or known to have NSCLC and have poor survival outcome or known to have NSCLC and have benefited from adjuvant chemotherapy or known to have NSCLC and not have benefited from adjuvant chemotherapy. The expression data of the biomarkers in the dataset can be used to create a “control value” that is used in testing samples from new patients. A control value is obtained from the historical expression data for a patient or pool of patients with a known outcome. In some embodiments, the control value is a numerical threshold for predicting outcomes, for example good and

poor outcome, or making therapy recommendations, for example adjuvant therapy in addition to surgical resection or surgical resection alone.

[050] In some embodiments, the "control" is a predetermined value for the set of biomarkers (e.g., set of 5 to 14 biomarkers) obtained from NSCLC patients whose biomarker expression values and survival times are known. Alternatively, the "control" is a predetermined reference profile for the set of biomarkers obtained from NSCLC patients whose survival times are known. Using values from known samples allows one to develop an algorithm for classifying new patient samples into good and poor survival groups as described in the Example.

[051] Accordingly, in one embodiment, the control is a sample from a subject known to have NSCLC and good survival outcome. In another embodiment, the control is a sample from a subject known to have NSCLC and poor survival outcome.

[052] A person skilled in the art will appreciate that the comparison between the expression of the biomarkers in the test sample and the expression of the biomarkers in the control will depend on the control used. For example, if the control is from a subject known to have NSCLC and poor survival, and there is a difference in expression of the biomarkers between the control and test sample, then the subject can be prognosed or classified in a good survival group. If the control is from a subject known to have NSCLC and good survival, and there is a difference in expression of the biomarkers between the control and test sample, then the subject can be prognosed or classified in a poor survival group. For example, if the control is from a subject known to have NSCLC and good survival, and there is a similarity in expression of the biomarkers between the control and test sample, then the subject can be prognosed or classified in a good survival group. For example, if the control is from a subject known to have NSCLC and poor survival, and there is a similarity in expression of the biomarkers between the control and test sample, then the subject can be prognosed or classified in a poor survival group.

[053] As used herein, a “reference value” refers to a gene-specific coefficient derived from historical expression data. The multi-gene signatures of the present disclosure comprise gene-specific reference values. In some embodiments, the multi-gene signature comprises one reference value for each gene in the signature.

In some embodiments, the multi-gene signature comprises four reference values for each gene in the signature. In some embodiments, the reference values are the first four components derived from principal component analysis for each gene in the signature.

[054] The term "differentially expressed" or "differential expression" as used herein refers to a difference in the level of expression of the biomarkers that can be assayed by measuring the level of expression of the products of the biomarkers, such as the difference in level of messenger RNA transcript expressed or proteins expressed of the biomarkers. In a preferred embodiment, the difference is statistically significant. The term "difference in the level of expression" refers to an increase or decrease in the measurable expression level of a given biomarker as measured by the amount of messenger RNA transcript and/or the amount of protein in a sample as compared with the measurable expression level of a given biomarker in a control. In one embodiment, the differential expression can be compared using the ratio of the level of expression of a given biomarker or biomarkers as compared with the expression level of the given biomarker or biomarkers of a control, wherein the ratio is not equal to 1.0. For example, an RNA or protein is differentially expressed if the ratio of the level of expression in a first sample as compared with a second sample is greater than or less than 1.0. For example, a ratio of greater than 1, 1.2, 1.5, 1.7, 2, 3, 3, 5, 10, 15, 20 or more, or a ratio less than 1, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05, 0.001 or less. In another embodiment the differential expression is measured using p-value. For instance, when using p-value, a biomarker is identified as being differentially expressed as between a first sample and a second sample when the p-value is less than 0.1, preferably less than 0.05, more preferably less than 0.01, even more preferably less than 0.005, the most preferably less than 0.001.

[055] The term "similarity in expression" as used herein means that there is no or little difference in the level of expression of the biomarkers between the test sample and the control or reference profile. For example, similarity can refer to a fold difference compared to a control. In a preferred embodiment, there is no statistically significant difference in the level of expression of the biomarkers.

[056] The term "most similar" in the context of a reference profile refers to a reference profile that is associated with a clinical outcome that shows the greatest number of identities and/or degree of changes with the subject profile.

[057] The term "prognosis" as used herein refers to a clinical outcome group such as a poor survival group or a good survival group associated with a disease subtype which is reflected by a reference profile such as a biomarker reference expression profile or reflected by expression levels of the biomarkers disclosed herein. The prognosis provides an indication of disease progression and includes an indication of likelihood of death due to lung cancer. In one embodiment the clinical outcome class includes a good survival group and a poor survival group.

[058] The term "prognosing or classifying" as used herein means predicting or identifying the clinical outcome group that a subject belongs to according to the subject's similarity to a reference profile or biomarker expression level associated with the prognosis. For example, prognosing or classifying comprises a method or process of determining whether an individual with NSCLC has a good or poor survival outcome, or grouping an individual with NSCLC into a good survival group or a poor survival group.

[059] The term "good survival" as used herein refers to an increased chance of survival as compared to patients in the "poor survival" group. For example, the biomarkers of the application can prognose or classify patients into a "good survival group". These patients are at a lower risk of death after surgery.

[060] The term "poor survival" as used herein refers to an increased risk of death as compared to patients in the "good survival" group. For example, biomarkers or genes of the application can prognose or classify patients into a "poor survival group". These patients are at greater risk of death from surgery.

[061] Accordingly, in one embodiment, the biomarker reference expression profile comprises a poor survival group. In another embodiment, the biomarker reference expression profile comprises a good survival group.

[062] The term "subject" as used herein refers to any member of the animal kingdom, preferably a human being that has NSCLC or that is suspected of having NSCLC.

[063] NSCLC patients are classified into stages, which are conventionally used to determine therapy. Staging classification testing may include any or all of history, physical examination, routine laboratory evaluations, chest x-rays, and chest computed tomography scans or positron emission tomography scans with infusion of contrast materials. For example, stage I includes cancer in the lung, but has not spread to adjacent lymph nodes or outside the chest. Stage I is divided into two categories based on the size of the tumor (IA and IB). Stage II includes cancer located in the lung and proximal lymph nodes. Stage II is divided into 2 categories based on the size of tumor and nodal status (IIA and IIB). Stage III includes cancer located in the lung and the lymph nodes. Stage III is divided into 2 categories based on the size of tumor and nodal status (IIIA and IIIB). Stage IV includes cancer that has metastasized to distant locations. The term "early stage NSCLC" includes patients with Stage I to IIIA NSCLC. These patients are treated primarily by complete surgical resection.

[064] In an aspect, a multi-gene signature is prognostic of patient outcome and/or response to adjuvant chemotherapy. In some embodiments, a minimal signature for 5 to 13 genes is provided. In one embodiment, the signature comprises reference values for each of the 5 to 13 genes listed in Table 4A. In some embodiments, the 5 to 13-gene signature is associated with the early stages of NSCLC. Accordingly, in connection with any aspect or embodiment of the invention described herein, the subject may have stage I or stage II NSCLC. In some embodiments, a 5 to 13-gene signature is prognostic of patient outcome and/or response to adjuvant chemotherapy.

[065] In some embodiments, the multi-gene signature comprises four coefficients, or reference values, for each gene in the signature. In one embodiment, the four coefficients are the first four principal components derived from principal component analysis described in Example 1 below. In one embodiment, the 5 to 13-gene signature comprises the principal component values listed in Table 10 below.

[066] The term "test sample" as used herein refers to any cancer-affected fluid, cell or tissue sample from a subject which can be assayed for biomarker expression products and/or a reference expression profile, e.g. genes differentially expressed in subjects with NSCLC according to survival outcome. In connection

with any aspect or embodiment of the invention described herein, the test sample may be a fresh frozen or a formalin-fixed paraffin-embedded (FFPE) tumor tissue sample. As disclosed herein, such tumor specimens can be used to provide accurate gene expression profiles for the purpose of classifying samples. Further, using such samples, minimal gene sets of 5 to 14 genes may be employed. RNA may be isolated from tissues using techniques known in the art. RNA Methodologies, A laboratory guide for isolation and characterization, 2nd edition, 1998, Robert E. Farrell, Jr., Ed., Academic Press. For example, RNA may be isolated from frozen tissue samples by homogenization in guanidinium isothiocyanate and acid phenol-chloroform extraction. Commercial kits are available for isolating RNA, including for use with FFPE specimens.

[067] The phrase "determining the expression of biomarkers" as used herein refers to determining or quantifying RNA or proteins expressed by the biomarkers. The term "RNA" includes mRNA transcripts, and/or specific spliced variants of mRNA. The terms "RNA product of the biomarker," "biomarker RNA," or "target RNA" as used herein refers to RNA transcripts transcribed from the biomarkers and/or specific spliced variants. In the case of "protein", it refers to proteins translated from the RNA transcripts transcribed from the biomarkers. The term "protein product of the biomarker" or "biomarker protein" refers to proteins translated from RNA products of the biomarkers.

[068] A person skilled in the art will appreciate that a number of methods can be used to detect or quantify the level of RNA products of the biomarkers within a sample, including arrays, such as microarrays, RT-PCR (including quantitative RT-PCR), nuclease protection assays, multiplex assays including nanostring technology, and Northern blot analyses. Any analytical procedure capable of permitting specific and quantifiable (or semi-quantifiable) detection of the biomarkers may be used in the methods herein presented, such as the microarray and quantitative PCR, *e.g.* quantitative RT-PCR, methods set forth herein, and methods known to those skilled in the art.

[069] Accordingly, in one embodiment, the biomarker expression levels are determined using arrays, optionally microarrays, RT-PCR, optionally quantitative RT-PCR, nuclease protection assays or Northern blot analyses.

[070] In some embodiments, the biomarker expression levels are determined by using an array. cDNA microarrays consist of multiple (usually thousands) of different cDNA probes spotted (usually using a robotic spotting device) onto known locations on a solid support, such as a glass microscope slide. Microarrays for use in the methods described herein comprise a solid substrate onto which the probes are covalently or non-covalently attached. The cDNAs are typically obtained by PCR amplification of plasmid library inserts using primers complementary to the vector backbone portion of the plasmid or to the gene itself for genes where sequence is known. PCR products suitable for production of microarrays are typically between 0.5 and 2.5 kB in length. In a typical microarray experiment, RNA (either total RNA or poly A RNA) is isolated from cells or tissues of interest and is reverse transcribed to yield cDNA. Labeling is usually performed during reverse transcription by incorporating a labeled nucleotide in the reaction mixture. A microarray is then hybridized with labeled RNA, and relative expression levels calculated based on the relative concentrations of cDNA molecules that hybridized to the cDNAs represented on the microarray. Microarray analysis can be performed by commercially available equipment, following manufactuer's protocols, such as by using Affymetrix GeneChip technology, Agilent Technologies cDNA microarrays, Illumina Whole-Genome DASL array assays, or any other comparable microarray technology.

[071] In some embodiments, probes capable of hybridizing to one or more biomarker RNAs or cDNAs are attached to the substrate at a defined location ("addressable array"). Probes can be attached to the substrate in a wide variety of ways, as will be appreciated by those in the art. In some embodiments, the probes are synthesized first and subsequently attached to the substrate. In other embodiments, the probes are synthesized on the substrate. In some embodiments, probes are synthesized on the substrate surface using techniques such as photopolymerization and photolithography.

[072] In some embodiments, microarrays are utilized in a RNA-primed, Array-based Klenow Enzyme ("RAKE") assay. *See* Nelson, P.T. *et al.* (2004) *Nature Methods* 1(2):1-7; Nelson, P.T. *et al.* (2006) *RNA* 12(2):1-5, each of which is incorporated herein by reference in its entirety. In these embodiments, total RNA is isolated from a sample. Optionally, small RNAs can be further purified from the total RNA sample. The RNA sample is then hybridized to DNA probes immobilized at the

5'-end on an addressable array. The DNA probes comprise a base sequence that is complementary to a target RNA of interest, such as one or more biomarker RNAs capable of specifically hybridizing to a nucleic acid comprising a sequence that is identically present in one of the genes listed in Table 4A under standard hybridization conditions.

[073] In some embodiments, the addressable array comprises DNA probes for no more than the 13 genes listed in Table 4A. In some embodiments, the addressable array comprises DNA probes for each of the 13 genes listed in Table 4A and optionally, no more than one additional gene selected from those listed in Table 3.

[074] In some embodiments, quantitation of biomarker RNA expression levels requires assumptions to be made about the total RNA per cell and the extent of sample loss during sample preparation. In some embodiments, the addressable array comprises DNA probes for each of the 13 genes listed in Table 4A and, optionally, one, two, three, or four housekeeping genes (or "normalization genes").

[075] In some embodiments, expression data are pre-processed to correct for variations in sample preparation or other non-experimental variables affecting expression measurements. For example, background adjustment, quantile adjustment, and summarization may be performed on microarray data, using standard software programs such as RMAexpress v0.3, followed by centering of the data to the mean and scaling to the standard deviation.

[076] After the sample is hybridized to the array, it is exposed to exonuclease I to digest any unhybridized probes. The Klenow fragment of DNA polymerase I is then applied along with biotinylated dATP, allowing the hybridized biomarker RNAs to act as primers for the enzyme with the DNA probe as template. The slide is then washed and a streptavidin-conjugated fluorophore is applied to detect and quantitate the spots on the array containing hybridized and Klenow-extended biomarker RNAs from the sample.

[077] In some embodiments, the RNA sample is reverse transcribed using a biotin/poly-dA random octamer primer. The RNA template is digested and the biotin-containing cDNA is hybridized to an addressable microarray with bound probes that permit specific detection of biomarker RNAs. In typical embodiments, the microarray

includes at least one probe comprising at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, even at least 20, 21, 22, 23, or 24 contiguous nucleotides identically present in each of the genes listed in Table 4A. After hybridization of the cDNA to the microarray, the microarray is exposed to a streptavidin-bound detectable marker, such as a fluorescent dye, and the bound cDNA is detected. *See* Liu C.G. *et al.* (2008) Methods 44:22-30, which is incorporated herein by reference in its entirety.

[078] In one embodiment, the array is a U133A chip from Affymetrix. In another embodiment, a plurality of nucleic acid probes that are complementary or hybridizable to an expression product of the genes listed in Table 4A are used on the array. In a particular embodiment, the probe target sequences are listed in Table 9. In some embodiments, the probe target sequences are selected from SEQ ID NO: 4, 11-15, 22, 26, 35, 78, 130, 133, and 169. In one embodiment, thirteen probes are used, each probe hybridizable to a different target sequence selected from SEQ ID NO: 4, 11-15, 22, 26, 35, 78, 130, 133, and 169. In some embodiments, a plurality of nucleic acid probes that are complementary or hybridizable to an expression product of some or all the genes listed in Table 3 are used on the array. In some embodiments, the probe target sequences are selected from those listed in Table 11. In some embodiments, the probe target sequences are selected from SEQ ID NO:1-172.

[079] The term "nucleic acid" includes DNA and RNA and can be either double stranded or single stranded.

[080] The term "hybridize" or "hybridizable" refers to the sequence specific non-covalent binding interaction with a complementary nucleic acid. In a preferred embodiment, the hybridization is under high stringency conditions. Appropriate stringency conditions which promote hybridization are known to those skilled in the art, or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1 6.3.6. For example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C may be employed.

[081] The term "probe" as used herein refers to a nucleic acid sequence that will hybridize to a nucleic acid target sequence. In one example, the probe

hybridizes to an RNA product of the biomarker or a nucleic acid sequence complementary thereof. The length of probe depends on the hybridization conditions and the sequences of the probe and nucleic acid target sequence. In one embodiment, the probe is at least 8, 10, 15, 20, 25, 50, 75, 100, 150, 200, 250, 400, 500 or more nucleotides in length.

[082] In some embodiments, compositions are provided that comprise at least one biomarker or target RNA-specific probe. The term "target RNA-specific probe" encompasses probes that have a region of contiguous nucleotides having a sequence that is either (*i*) identically present in one of the genes listed in Tables 3 or 4A, or (ii) complementary to the sequence of a region of contiguous nucleotides found in one of the genes listed in Tables 3 or 4A, where "region" can comprise the full length sequence of any one of the genes listed in Tables 3 or 4A, a complementary sequence of the full length sequence of any one of the genes listed in Tables 3 or 4A, or a subsequence thereof.

[083] In some embodiments, target RNA-specific probes consist of deoxyribonucleotides. In other embodiments, target RNA-specific probes consist of both deoxyribonucleotides and nucleotide analogs. In some embodiments, biomarker RNA-specific probes comprise at least one nucleotide analog which increases the hybridization binding energy. In some embodiments, a target RNA-specific probe in the compositions described herein binds to one biomarker RNA in the sample.

[084] In some embodiments, more than one probe specific for a single biomarker RNA is present in the compositions, the probes capable of binding to overlapping or spatially separated regions of the biomarker RNA.

[085] It will be understood that in some embodiments in which the compositions described herein are designed to hybridize to cDNAs reverse transcribed from biomarker RNAs, the composition comprises at least one target RNA-specific probe comprising a sequence that is identically present in a biomarker RNA (or a subsequence thereof).

[086] In some embodiments, a biomarker RNA is capable of specifically hybridizing to at least one probe comprising a base sequence that is identically present in one of of the genes listed in Table 4A. In some embodiments, a

biomarker RNA is capable of specifically hybridizing to at least one nucleic acid probe comprising a sequence that is identically present in one of of the genes listed in Table 3. In some embodiments, a target RNA is capable of specifically hybridizing to at least one nucleic acid probe, and comprises a sequence that is identical to a sequence selected from SEQ ID NO:1-172, or a sequence listed in Table 11. In some embodiments, a target RNA is capable of specifically hybridizing to at least one nucleic acid probe, and comprises a sequence that is identical to a sequence listed in Table 9. In some embodiments, a target RNA is capable of specifically hybridizing to at least one nucleic acid probe, and comprises a sequence that is identical to a sequence selected from SEQ ID NO: 4, 11-15, 22, 26, 35, 78, 130, 133, and 169. In some embodiments, a biomarker RNA is capable of specifically hybridizing to at least one probe comprising a base sequence that is identically present in one of the genes listed in Table 4A.

[087] In some embodiments, the composition comprises a plurality of target or biomarker RNA-specific probes each comprising a region of contiguous nucleotides comprising a base sequence that is identically present in one or more of the genes listed in Table 4A, or in a subsequence thereof. In some embodiments, the composition comprises a plurality of target or biomarker RNA-specific probes each comprising a region of contiguous nucleotides comprising a base sequence that is complementary to a sequence listed in Table 9. In some embodiments, the composition comprises a plurality of target RNA-specific probes each comprising a region of contiguous nucleotides comprising a base sequence that is complementary to a sequence selected from SEQ ID NO: 4, 11-15, 22, 26, 35, 78, 130, 133, and 169.

[088] As used herein, the terms "complementary" or "partially complementary" to a biomarker or target RNA (or target region thereof), and the percentage of "complementarity" of the probe sequence to that of the biomarker RNA sequence is the percentage "identity" to the reverse complement of the sequence of the biomarker RNA. In determining the degree of "complementarity" between probes used in the compositions described herein (or regions thereof) and a biomarker RNA, such as those disclosed herein, the degree of "complementarity" is expressed as the percentage identity between the sequence of the probe (or region thereof) and the reverse complement of the sequence of the biomarker RNA that best aligns

therewith. The percentage is calculated by counting the number of aligned bases that are identical as between the 2 sequences, dividing by the total number of contiguous nucleotides in the probe, and multiplying by 100.

[089] In some embodiments, the microarray comprises probes comprising a region with a base sequence that is fully complementary to a target region of a biomarker RNA. In other embodiments, the microarray comprises probes comprising a region with a base sequence that comprises one or more base mismatches when compared to the sequence of the best-aligned target region of a biomarker RNA.

[090] As noted above, a "region" of a probe or biomarker RNA, as used herein, may comprise or consist of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or more contiguous nucleotides from a particular gene or a complementary sequence thereof. In some embodiments, the region is of the same length as the probe or the biomarker RNA. In other embodiments, the region is shorter than the length of the probe or the biomarker RNA.

[091] In some embodiments, the microarray comprises thirteen probes each comprising a region of at least 10 contiguous nucleotides, such as at least 11 contiguous nucleotides, such as at least 13 contiguous nucleotides, such as at least 14 contiguous nucleotides, such as at least 15 contiguous nucleotides, such as at least 16 contiguous nucleotides, such as at least 17 contiguous nucleotides, such as at least 18 contiguous nucleotides, such as at least 19 contiguous nucleotides, such as at least 20 contiguous nucleotides, such as at least 21 contiguous nucleotides, such as at least 22 contiguous nucleotides, such as at least 23 contiguous nucleotides, such as at least 24 contiguous nucleotides, such as at least 25 contiguous nucleotides with a base sequence that is identically present in one of the genes listed in Table 4.

[092] In some embodiments, the microarray component comprises thirteen probes each comprising a region with a base sequence that is identically present in each of the genes listed in Table 4A. In some embodiments, the microarray comprises fourteen probes, each of which comprises a region with a base sequence that is identically present in each of the genes listed in Table 4A and, optionally, one of the genes listed in Table 3.

[093] In another embodiment, the biomarker expression levels are determined by using quantitative RT-PCR. The first step is the isolation of mRNA from a target sample. The starting material is typically total RNA isolated from human tumors or tumor cell lines. General methods for mRNA extraction are well known in the art and are disclosed in standard textbooks of molecular biology, including Ausubel et al., Current Protocols of Molecular Biology, John Wiley and Sons (1997). Methods for RNA extraction from paraffin embedded tissues are disclosed, for example, in Rupp and Locker, Lab Invest. 56:A67 (1987), and De Andres et al., BioTechniques 18:42044 (1995). In particular, RNA isolation can be performed using purification kit, buffer set and protease from commercial manufacturers, such as Qiagen, according to the manufacturer's instructions. For example, total RNA from cells in culture can be isolated using Qiagen RNeasy mini-columns. Numerous RNA isolation kits are commercially available.

[094] In some embodiments, the primers used for quantitative RT-PCR comprise a forward and reverse primer for each of 5 to 13 genes listed in Table 4A. In one embodiment, the primers used for quantitative RT-PCR of the 5 to 13 genes are listed in Table 7 and/or Table 16. In one embodiment, primers comprising sequences identical to the sequences of SEQ ID NO: 173-198 and 203-206 are used for quantitative RT-PCR, wherein primers with sequences identifical to SEQ ID NO:173-185, 203 and 204 are forward primers and primers with sequences identifical to SEQ ID NO:186-198, 205 and 206 are reverse primers.

[095] In some embodiments the analytical method used for detecting at least one biomarker RNA in the methods set forth herein includes real-time quantitative RT-PCR. *See* Chen, C. *et al.* (2005) *Nucl. Acids Res.* 33:e179, which is incorporated herein by reference in its entirety. Although PCR can use a variety of thermostable DNA-dependent DNA polymerases, it typically employs the Taq DNA polymerase, which has a 5'-3' nuclease activity but lacks a 3'-5' proofreading endonuclease activity. In some embodiments, RT-PCR is done using a TaqMan® assay sold by Applied Biosystems, Inc. In a first step, total RNA is isolated from the sample. In some embodiments, the assay can be used to analyze about 10 ng of total RNA input sample, such as about 9 ng of input sample, such as about 8 ng of input sample, such as about 7 ng of input sample, such as about 6 ng of input

sample, such as about 5 ng of input sample, such as about 4 ng of input sample, such as about 3 ng of input sample, such as about 2 ng of input sample, and even as little as about 1 ng of input sample containing RNA. In some embodiments, RT-PCR is done using a probe based on the Locked Nucleic Acid technology sold as Universal Probe Library (UPL) by Hoffman Laroche.

[096] The TaqMan® assay utilizes a stem-loop primer that is specifically complementary to the 3'-end of a biomarker RNA. The step of hybridizing the stem-loop primer to the biomarker RNA is followed by reverse transcription of the biomarker RNA template, resulting in extension of the 3' end of the primer. The result of the reverse transcription step is a chimeric (DNA) amplicon with the step-loop primer sequence at the 5' end of the amplicon and the cDNA of the biomarker RNA at the 3' end. Quantitation of the biomarker RNA is achieved by RT-PCR using a universal reverse primer comprising a sequence that is complementary to a sequence at the 5' end of all stem-loop biomarker RNA primers, a biomarker RNA-specific forward primer, and a biomarker RNA sequence-specific TaqMan® probe.

[097] The assay uses fluorescence resonance energy transfer ("FRET") to detect and quantitate the synthesized PCR product. Typically, the TaqMan® probe comprises a fluorescent dye molecule coupled to the 5'-end and a quencher molecule coupled to the 3'-end, such that the dye and the quencher are in close proximity, allowing the quencher to suppress the fluorescence signal of the dye via FRET. When the polymerase replicates the chimeric amplicon template to which the TaqMan® probe is bound, the 5'-nuclease of the polymerase cleaves the probe, decoupling the dye and the quencher so that FRET is abolished and a fluorescence signal is generated. Fluorescence increases with each RT-PCR cycle proportionally to the amount of probe that is cleaved.

[098] Exemplary probes for use with the five to thirteen genes, in connection with preparing gene expression profiles and classifying samples, are disclosed in Table 16. Exemplary probes for additional genes as may be desired, such as one or more additional genes from Table 3, are commercially available or may be prepared according to conventional methods.

[099] In some embodiments, quantitation of the results of RT-PCR assays is done by constructing a standard curve from a nucleic acid of known concentration

and then extrapolating quantitative information for biomarker RNAs of unknown concentration. In some embodiments, the nucleic acid used for generating a standard curve is an RNA of known concentration. In some embodiments, the nucleic acid used for generating a standard curve is a purified double-stranded plasmid DNA or a single-stranded DNA generated *in vitro*.

[0100] In some embodiments, where the amplification efficiencies of the biomarker nucleic acids and the endogenous reference are approximately equal, quantitation is accomplished by the comparative C_t (cycle threshold, *e.g.*, the number of PCR cycles required for the fluorescence signal to rise above background) method. C_t values are inversely proportional to the amount of nucleic acid target in a sample. In some embodiments, C_t values of the target RNA of interest can be compared with a control or calibrator, such as RNA from normal tissue. In some embodiments, the C_t values of the calibrator and the target RNA samples of interest are normalized to an appropriate endogenous housekeeping gene (see above).

[0101] In addition to the TaqMan® assays, other RT-PCR chemistries useful for detecting and quantitating PCR products in the methods presented herein include, but are not limited to, UPL probes, Molecular Beacons, Scorpion probes and SYBR Green detection.

[0102] In some embodiments, Molecular Beacons can be used to detect and quantitate PCR products. Like TaqMan® probes, Molecular Beacons use FRET to detect and quantitate a PCR product via a probe comprising a fluorescent dye and a quencher attached at the ends of the probe. Unlike TaqMan® probes, Molecular Beacons remain intact during the PCR cycles. Molecular Beacon probes form a stem-loop structure when free in solution, thereby allowing the dye and quencher to be in close enough proximity to cause fluorescence quenching. When the Molecular Beacon hybridizes to a target, the stem-loop structure is abolished so that the dye and the quencher become separated in space and the dye fluoresces. Molecular Beacons are available, *e.g.*, from Gene Link™ (*see* http://www.genelink.com/newsite/products/mbintro.asp).

[0103] In some embodiments, Scorpion probes can be used as both sequence-specific primers and for PCR product detection and quantitation. Like

Molecular Beacons, Scorpion probes form a stem-loop structure when not hybridized to a target nucleic acid. However, unlike Molecular Beacons, a Scorpion probe achieves both sequence-specific priming and PCR product detection. A fluorescent dye molecule is attached to the 5'-end of the Scorpion probe, and a quencher is attached to the 3'-end. The 3' portion of the probe is complementary to the extension product of the PCR primer, and this complementary portion is linked to the 5'-end of the probe by a non-amplifiable moiety. After the Scorpion primer is extended, the target-specific sequence of the probe binds to its complement within the extended amplicon, thus opening up the stem-loop structure and allowing the dye on the 5'-end to fluoresce and generate a signal. Scorpion probes are available from, *e.g,* Premier Biosoft International (*see* http://www.premierbiosoft.com/tech_notes/Scorpion.html).

[0104] In some embodiments, RT-PCR detection is performed specifically to detect and quantify the expression of a single biomarker RNA. The biomarker RNA, in typical embodiments, is selected from a biomarker RNA capable of specifically hybridizing to a nucleic acid comprising a sequence that is identically present in one of the genes set forth in Table 4A. In some embodiments, the biomarker RNA specifically hybridizes to a nucleic acid comprising a sequence that is identically present in at least one of the genes in Table 3.

[0105] In various other embodiments, RT-PCR detection is utilized to detect, in a single multiplex reaction, each of 5 to 14 (e.g., 13) biomarker RNAs. The biomarker RNAs, in some embodiments, are capable of specifically hybridizing to a nucleic acid comprising a sequence that is identically present in one of the thirteen genes listed in Table 4A.

[0106] In some multiplex embodiments, a plurality of probes, such as TaqMan probes, each specific for a different RNA target, is used. In typical embodiments, each target RNA-specific probe is spectrally distinguishable from the other probes used in the same multiplex reaction.

[0107] In some embodiments, quantitation of RT-PCR products is accomplished using a dye that binds to double-stranded DNA products, such as SYBR Green. In some embodiments, the assay is the QuantiTect SYBR Green PCR assay from Qiagen. In this assay, total RNA is first isolated from a sample. Total

RNA is subsequently poly-adenylated at the 3'-end and reverse transcribed using a universal primer with poly-dT at the 5'-end. In some embodiments, a single reverse transcription reaction is sufficient to assay multiple biomarker RNAs. RT-PCR is then accomplished using biomarker RNA-specific primers and an miScript Universal Primer, which comprises a poly-dT sequence at the 5'-end. SYBR Green dye binds non-specifically to double-stranded DNA and upon excitation, emits light. In some embodiments, buffer conditions that promote highly-specific annealing of primers to the PCR template (*e.g.,* available in the QuantiTect SYBR Green PCR Kit from Qiagen) can be used to avoid the formation of non-specific DNA duplexes and primer dimers that will bind SYBR Green and negatively affect quantitation. Thus, as PCR product accumulates, the signal from SYBR green increases, allowing quantitation of specific products.

[0108] RT-PCR is performed using any RT-PCR instrumentation available in the art. Typically, instrumentation used in real-time RT-PCR data collection and analysis comprises a thermal cycler, optics for fluorescence excitation and emission collection, and optionally a computer and data acquisition and analysis software.

[0109] In some embodiments, the method of detectably quantifying one or more biomarker RNAs includes the steps of: (a) isolating total RNA; (b) reverse transcribing a biomarker RNA to produce a cDNA that is complementary to the biomarker RNA; (c) amplifying the cDNA from step (b); and (d) detecting the amount of a biomarker RNA with RT-PCR.

[0110] As described above, in some embodiments, the RT-PCR detection is performed using a FRET probe, which includes, but is not limited to, a TaqMan® probe, a Molecular beacon probe and a Scorpion probe. In some embodiments, the RT-PCR detection and quantification is performed with a TaqMan® probe, i.e., a linear probe that typically has a fluorescent dye covalently bound at one end of the DNA and a quencher molecule covalently bound at the other end of the DNA. The FRET probe comprises a base sequence that is complementary to a region of the cDNA such that, when the FRET probe is hybridized to the cDNA, the dye fluorescence is quenched, and when the probe is digested during amplification of the cDNA, the dye is released from the probe and produces a fluorescence signal. In

such embodiments, the amount of biomarker RNA in the sample is proportional to the amount of fluorescence measured during cDNA amplification.

[0111] The TaqMan® probe typically comprises a region of contiguous nucleotides comprising a base sequence that is complementary to a region of a biomarker RNA or its complementary cDNA that is reverse transcribed from the biomarker RNA template (i.e., the sequence of the probe region is complementary to or identically present in the biomarker RNA to be detected) such that the probe is specifically hybridizable to the resulting PCR amplicon. In some embodiments, the probe comprises a region of at least 6 contiguous nucleotides having a base sequence that is fully complementary to or identically present in a region of a cDNA that has been reverse transcribed from a biomarker RNA template, such as comprising a region of at least 8 contiguous nucleotides, or comprising a region of at least 10 contiguous nucleotides, or comprising a region of at least 12 contiguous nucleotides, or comprising a region of at least 14 contiguous nucleotides, or even comprising a region of at least 16 contiguous nucleotides having a base sequence that is complementary to or identically present in a region of a cDNA reverse transcribed from a biomarker RNA to be detected.

[0112] Preferably, the region of the cDNA that has a sequence that is complementary to the TaqMan® probe sequence is at or near the center of the cDNA molecule. In some embodiments, there are independently at least 2 nucleotides, such as at least 3 nucleotides, such as at least 4 nucleotides, such as at least 5 nucleotides of the cDNA at the 5'-end and at the 3'-end of the region of complementarity.

[0113] In typical embodiments, all biomarker RNAs are detected in a single multiplex reaction. In these embodiments, each TaqMan® probe that is targeted to a unique cDNA is spectrally distinguishable when released from the probe. Thus, each biomarker RNA is detected by a unique fluorescence signal.

[0114] In some embodiments, expression levels may be represented by gene transcript numbers per nanogram of cDNA. To control for variability in cDNA quantity, integrity and the overall transcriptional efficiency of individual primers, RT-PCR data can be subjected to standardization and normalization against one or

more housekeeping genes as has been previously described. *See e.g.*, Rubie et al., Mol. Cell. Probes 19(2):101-9 (2005).

[0115] Appropriate genes for normalization in the methods described herein include those as to which the quantity of the product does not vary between between different cell types, cell lines or under different growth and sample preparation conditions. In some embodiments, endogenous housekeeping genes useful as normalization controls in the methods described herein include, but are not limited to, ACTB, BAT1, B2M, EDS, IP08, TBP, PP1A, GUSB, U6 snRNA, RNU44, RNU 48, and U47. In typical embodiments, at least one endogenous housekeeping gene for use in normalizing the measured quantity of RNA is selected from ACTB, BAT1, B2M, EDS, TBP, U6 snRNA, U6 snRNA, RNU44, RNU 48, and U47. For example, the methods and kits of the invention may employ 2, 3, or 4 normalization genes selected from IP08, BAT1, TBP, PP1A, and GUSB. In certain embodiments, the tissue sample is frozen, and the normalization genes include 2, 3, or 4 of IP08, BAT1, TBP, and PP1A. In other embodiments, the sample is an FFPE sample, and the normalization genes include 2, 3, or 4 of BAT1, TBP, PP1A, and GUSB. In some embodiments, one housekeeping gene is used for normalization. In some embodiments, two, three, four or more housekeeping genes are used for normalization.

[0116] In some embodiments, labels that can be used on the FRET probes include colorimetric and fluorescent labels such as Alexa Fluor dyes, BODIPY dyes, such as BODIPY FL; Cascade Blue; Cascade Yellow; coumarin and its derivatives, such as 7-amino-4-methylcoumarin, aminocoumarin and hydroxycoumarin; cyanine dyes, such as Cy3 and Cy5; eosins and erythrosins; fluorescein and its derivatives, such as fluorescein isothiocyanate; macrocyclic chelates of lanthanide ions, such as Quantum DyeTM; Marina Blue; Oregon Green; rhodamine dyes, such as rhodamine red, tetramethylrhodamine and rhodamine 6G; Texas Red; fluorescent energy transfer dyes, such as thiazole orange-ethidium heterodimer; and, TOTAB.

[0117] Specific examples of dyes include, but are not limited to, those identified above and the following: Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500. Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor

633, Alexa Fluor 647, Alexa Fluor 660, Alexa Fluor 680, Alexa Fluor 700, and, Alexa Fluor 750; amine-reactive BODIPY dyes, such as BODIPY 493/503, BODIPY 530/550, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY 630/650, BODIPY 650/655, BODIPY FL, BODIPY R6G, BODIPY TMR, and, BODIPY-TR; Cy3, Cy5, 6-FAM, Fluorescein Isothiocyanate, HEX, 6-JOE, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, REG, Rhodamine Green, Rhodamine Red, Renographin, ROX, SYPRO, TAMRA, 2', 4',5',7'-Tetrabromosulfonefluorescein, and TET.

[0118] Specific examples of fluorescently labeled ribonucleotides useful in the preparation of RT-PCR probes for use in some embodiments of the methods described herein are available from Molecular Probes (Invitrogen), and these include, Alexa Fluor 488-5-UTP, Fluorescein-12-UTP, BODIPY FL-14-UTP, BODIPY TMR-14-UTP, Tetramethylrhodamine-6-UTP, Alexa Fluor 546-14-UTP, Texas Red-5-UTP, and BODIPY TR-14-UTP. Other fluorescent ribonucleotides are available from Amersham Biosciences (GE Healthcare), such as Cy3-UTP and Cy5-UTP.

[0119] Examples of fluorescently labeled deoxyribonucleotides useful in the preparation of RT-PCR probes for use in the methods described herein include Dinitrophenyl (DNP)-1'-dUTP, Cascade Blue-7-dUTP, Alexa Fluor 488-5-dUTP, Fluorescein-12-dUTP, Oregon Green 488-5-dUTP, BODIPY FL-14-dUTP, Rhodamine Green-5-dUTP, Alexa Fluor 532-5-dUTP, BODIPY TMR-14-dUTP, Tetramethylrhodamine-6-dUTP, Alexa Fluor 546-14-dUTP, Alexa Fluor 568-5-dUTP, Texas Red-12-dUTP, Texas Red-5-dUTP, BODIPY TR-14-dUTP, Alexa Fluor 594-5-dUTP, BODIPY 630/650-14-dUTP, BODIPY 650/665-14-dUTP; Alexa Fluor 488-7-OBEA-dCTP, Alexa Fluor 546-16-OBEA-dCTP, Alexa Fluor 594-7-OBEA-dCTP, Alexa Fluor 647-12-OBEA-dCTP. Fluorescently labeled nucleotides are commercially available and can be purchased from, e.g., Invitrogen.

[0120] In some embodiments, dyes and other moieties, such as quenchers, are introduced into nucleic acids used in the methods described herein, such as FRET probes, via modified nucleotides. A "modified nucleotide" refers to a nucleotide that has been chemically modified, but still functions as a nucleotide. In some embodiments, the modified nucleotide has a chemical moiety, such as a dye or quencher, covalently attached, and can be introduced into an oligonucleotide, for

example, by way of solid phase synthesis of the oligonucleotide. In other embodiments, the modified nucleotide includes one or more reactive groups that can react with a dye or quencher before, during, or after incorporation of the modified nucleotide into the nucleic acid. In specific embodiments, the modified nucleotide is an amine-modified nucleotide, *i.e.,* a nucleotide that has been modified to have a reactive amine group. In some embodiments, the modified nucleotide comprises a modified base moiety, such as uridine, adenosine, guanosine, and/or cytosine. In specific embodiments, the amine-modified nucleotide is selected from 5-(3-aminoallyl)-UTP; 8-[(4-amino)butyl]-amino-ATP and 8-[(6-amino)butyl]-amino-ATP; N6-(4-amino)butyl-ATP, N6-(6-amino)butyl-ATP, N4-[2,2-oxy-bis-(ethylamine)]-CTP; N6-(6-Amino)hexyl-ATP; 8-[(6-Amino)hexyl]-amino-ATP; 5-propargylamino-CTP, 5-propargylamino-UTP. In some embodiments, nucleotides with different nucleobase moieties are similarly modified, for example, 5-(3-aminoallyl)-GTP instead of 5-(3-aminoallyl)-UTP. Many amine modified nucleotides are commercially available from, e.g., Applied Biosystems, Sigma, Jena Bioscience and TriLink.

[0121] In some embodiments, the methods of detecting at least one biomarker RNA described herein employ one or more modified oligonucleotides, such as oligonucleotides comprising one or more affinity-enhancing nucleotides. Modified oligonucleotides useful in the methods described herein include primers for reverse transcription, PCR amplification primers, and probes. In some embodiments, the incorporation of affinity-enhancing nucleotides increases the binding affinity and specificity of an oligonucleotide for its target nucleic acid as compared to oligonucleotides that contain only deoxyribonucleotides, and allows for the use of shorter oligonucleotides or for shorter regions of complementarity between the oligonucleotide and the target nucleic acid.

[0122] In some embodiments, affinity-enhancing nucleotides include nucleotides comprising one or more base modifications, sugar modifications and/or backbone modifications.

[0123] In some embodiments, modified bases for use in affinity-enhancing nucleotides include 5-methylcytosine, isocytosine, pseudoisocytosine, 5-bromouracil, 5-propynyluracil, 6-aminopurine, 2-aminopurine, inosine, diaminopurine, 2-chloro-6-aminopurine, xanthine and hypoxanthine.

[0124] In some embodiments, affinity-enhancing modifications include nucleotides having modified sugars such as 2'-substituted sugars, such as 2'-O-alkyl-ribose sugars, 2'-amino-deoxyribose sugars, 2'-fluoro- deoxyribose sugars, 2'-fluoro-arabinose sugars, and 2'-O-methoxyethyl-ribose (2'MOE) sugars. In some embodiments, modified sugars are arabinose sugars, or d-arabino-hexitol sugars.

[0125] In some embodiments, affinity-enhancing modifications include backbone modifications such as the use of peptide nucleic acids (*e.g.,* an oligomer including nucleobases linked together by an amino acid backbone). Other backbone modifications include phosphorothioate linkages, phosphodiester modified nucleic acids, combinations of phosphodiester and phosphorothioate nucleic acid, methylphosphonate, alkylphosphonates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters, methylphosphorothioate, phosphorodithioate, p-ethoxy, and combinations thereof.

[0126] In some embodiments, the oligomer includes at least one affinity-enhancing nucleotide that has a modified base, at least nucleotide (which may be the same nucleotide) that has a modified sugar, and at least one internucleotide linkage that is non-naturally occurring.

[0127] In some embodiments, the affinity-enhancing nucleotide contains a locked nucleic acid ("LNA") sugar, which is a bicyclic sugar. In some embodiments, an oligonucleotide for use in the methods described herein comprises one or more nucleotides having an LNA sugar. In some embodiments, the oligonucleotide contains one or more regions consisting of nucleotides with LNA sugars. In other embodiments, the oligonucleotide contains nucleotides with LNA sugars interspersed with deoxyribonucleotides. *See, e.g.*, Frieden, M. *et al.* (2008) Curr. Pharm. Des. 14(11):1138-1142.

[0128] The term "primer" as used herein refers to a nucleic acid sequence, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a nucleic acid strand is induced (e.g. in the presence of nucleotides and an

inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon factors, including temperature, sequences of the primer and the methods used. A primer typically contains 15-25 or more nucleotides, although it can contain less. The factors involved in determining the appropriate length of primer are readily known to one of ordinary skill in the art. In one embodiment, primer sets for 5 to 13 genes whose expression levels are determined, are selected from Table 7 and/or Table 16.

[0129] In addition, a person skilled in the art will appreciate that a number of methods can be used to determine the amount of a protein product of the biomarker of the invention, including immunoassays such as Western blots, ELISA, and immunoprecipitation followed by SDS-PAGE and immunocytochemistry.

[0130] Accordingly, in another embodiment, an antibody is used to detect the polypeptide products of the biomarkers listed in Table 4A. In another embodiment, the sample comprises a tissue sample. In a further embodiment, the tissue sample is suitable for immunohistochemistry.

[0131] The term "antibody" as used herein is intended to include monoclonal antibodies, polyclonal antibodies, and chimeric antibodies. The antibody may be from recombinant sources and/or produced in transgenic animals. The term "antibody fragment" as used herein is intended to include Fab, Fab', F(ab')2, scFv, dsFv, ds-scFv, dimers, minibodies, diabodies, and multimers thereof and bispecific antibody fragments. Antibodies can be fragmented using conventional techniques. For example, F(ab')2 fragments can be generated by treating the antibody with pepsin. The resulting F(ab')2 fragment can be treated to reduce disulfide bridges to produce Fab' fragments. Papain digestion can lead to the formation of Fab fragments. Fab, Fab' and F(ab')2, scFv, dsFv, ds-scFv, dimers, minibodies, diabodies, bispecific antibody fragments and other fragments can also be synthesized by recombinant techniques.

[0132] Conventional techniques of molecular biology, microbiology and recombinant DNA techniques are within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, 1989,

Molecular Cloning: A Laboratory Manual, Second Edition; Oligonucleotide Synthesis (M.J. Gait, ed., 1984); Nucleic Acid Hybridization (B.D. Harnes & S.J. Higgins, eds., 1984); A Practical Guide to Molecular Cloning (B. Perbal, 1984); and a series, Methods in Enzymology (Academic Press, Inc.); Short Protocols In Molecular Biology, (Ausubel *et al.*, ed., 1995).

[0133] For example, antibodies having specificity for a specific protein, such as the protein product of a biomarker, may be prepared by conventional methods. A mammal, (e.g. a mouse, hamster, or rabbit) can be immunized with an immunogenic form of the peptide which elicits an antibody response in the mammal. Techniques for conferring immunogenicity on a peptide include conjugation to carriers or other techniques well known in the art. For example, the peptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassay procedures can be used with the immunogen as antigen to assess the levels of antibodies. Following immunization, antisera can be obtained and, if desired, polyclonal antibodies isolated from the sera.

[0134] To produce monoclonal antibodies, antibody producing cells (lymphocytes) can be harvested from an immunized animal and fused with myeloma cells by standard somatic cell fusion procedures thus immortalizing these cells and yielding hybridoma cells. Such techniques are well known in the art, (e.g. the hybridoma technique originally developed by Kohler and Milstein (Nature 256:495-497 (1975)) as well as other techniques such as the human B-cell hybridoma technique (Kozbor *et al.*, Immunol. Today 4:72 (1983)), the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.*, Methods Enzymol, 121:140-67 (1986)), and screening of combinatorial antibody libraries (Huse *et al.*, Science 246:1275 (1989)). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with the peptide and the monoclonal antibodies can be isolated.

[0135] In some embodiments, recombinant antibodies are provided that specifically bind protein products of the genes listed in Table 4, and optionally the expression product(s) of one gene selected from among those listed in Table 3. Recombinant antibodies include, but are not limited to, chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, single-

chain antibodies and multi-specific antibodies. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine monoclonal antibody (mAb) and a human immunoglobulin constant region. (See, e.g., Cabilly et al., U.S. Pat. No. 4,816,567; and Boss et al., U.S. Pat. No. 4,816,397, which are incorporated herein by reference in their entirety). Single-chain antibodies have an antigen binding site and consist of single polypeptides. They can be produced by techniques known in the art, for example using methods described in Ladner et al, U.S. Pat. No. 4,946,778 (which is incorporated herein by reference in its entirety); Bird et al., (1988) Science 242:423-426; Whitlow et al., (1991) Methods in Enzymology 2:1-9; Whitlow et al., (1991) Methods in Enzymology 2:97-105; and Huston et al., (1991) Methods in Enzymology Molecular Design and Modeling: Concepts and Applications 203:46-88. Multi-specific antibodies are antibody molecules having at least two antigen-binding sites that specifically bind different antigens. Such molecules can be produced by techniques known in the art, for example using methods described in Segal, U.S. Pat. No. 4,676,980 (the disclosure of which is incorporated herein by reference in its entirety); Holliger et al., (1993) Proc. Natl. Acad. Sci. USA 90:6444-6448; Whitlow et al., (1994) Protein Eng 7:1017-1026 and U.S. Pat. No. 6,121,424.

[0136] Monoclonal antibodies directed against any of the expression products of the genes listed in Table 4A and can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with the polypeptide(s) of interest. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Pat. No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) Bio/Technology 9:1370-1372; Hay et al. (1992) Hum. Antibod. Hybridomas 3:81-85; Huse et al. (1989) Science 246:1275-1281; Griffiths et al. (1993) EMBO J 12:725-734.

[0137] Humanized antibodies are antibody molecules from non-human species having one or more complementarity determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. (See, e.g., Queen, U.S. Pat. No. 5,585,089, which is incorporated herein by reference in its entirety.) Humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Pat. No. 4,816,567; European Patent Application 125,023; Better et al. (1988) Science 240:1041-1043; Liu et al. (1987) Proc. Natl. Acad. Sci. USA 84:3439-3443; Liu et al. (1987) J. Immunol. 139:3521-3526; Sun et al. (1987) Proc. Natl. Acad. Sci. USA 84:214-218; Nishimura et al. (1987) Cancer Res. 47:999-1005; Wood et al. (1985) Nature 314:446-449; and Shaw et al. (1988) J. Natl. Cancer Inst. 80:1553-1559); Morrison (1985) Science 229:1202-1207; Oi et al. (1986) Bio/Techniques 4:214; U.S. Pat. No. 5,225,539; Jones et al. (1986) Nature 321:552-525; Verhoeyan et al. (1988) Science 239:1534; and Beidler et al. (1988) J. Immunol. 141:4053-4060.

[0138] In some embodiments, humanized antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of a polypeptide corresponding to a protein product. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995) Int. Rev. Immunol. 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Pat. Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies

such as Abgenix, Inc. (Fremont, Calif.), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

[0139] Antibodies may be isolated after production (e.g., from the blood or serum of the subject) or synthesis and further purified by well-known techniques. For example, IgG antibodies can be purified using protein A chromatography. Antibodies specific for a protein can be selected or (e.g., partially purified) or purified by, e.g., affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) expression product may be produced, and covalently or non-covalently coupled to a solid support such as, for example, a chromatography column. The column can then be used to affinity purify antibodies specific for the protein products of the genes listed in Tables 3 and 4A from a sample containing antibodies directed against a large number of different epitopes, thereby generating a substantially purified antibody composition, i.e., one that is substantially free of contaminating antibodies. By a substantially purified antibody composition it is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than those of the protein products of the genes listed in Tables 3 and 4A, and preferably at most 20%, yet more preferably at most 10%, and most preferably at most 5% (by dry weight) of the sample is contaminating antibodies. A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein.

[0140] In some embodiments, substantially purified antibodies may specifically bind to a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a cytoplasmic domain or cytoplasmic membrane of a protein product of one of the genes listed in Table 4A. In an embodiment, substantially purified antibodies specifically bind to a secreted sequence or an extracellular domain of the amino acid sequences of a protein product of one of the genes listed in Tables 3 and 4A.

[0141] In some embodiments, antibodies directed against a protein product of one of the genes listed in Tables 3 and 4A can be used to detect the protein products or fragment thereof (e.g., in a cellular lysate or cell supernatant) in order to evaluate the level and pattern of expression of the protein. Detection can be facilitated by the use of an antibody derivative, which comprises an antibody coupled

to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

[0142] A variety of techniques can be employed to measure expression levels of each of the thirteen genes in Table 4 as well as the genes in Table 3, given a sample that contains protein products that bind to a given antibody. Examples of such formats include, but are not limited to, enzyme immunoassay (EIA), radioimmunoassay (RIA), Western blot analysis and enzyme linked immunoabsorbant assay (ELISA). A skilled artisan can readily adapt known protein/antibody detection methods for use in determining protein expression levels of the thirteen products of the genes listed in Tables 4 and optionally additional products of one gene selected from those listed in Table 3.

[0143] In one embodiment, antibodies, or antibody fragments or derivatives, can be used in methods such as Western blots or immunofluorescence techniques to detect the expressed proteins. In some embodiments, either the antibodies or proteins are immobilized on a solid support. Suitable solid phase supports or carriers include any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite.

[0144] One skilled in the art will know many other suitable carriers for binding antibody or antigen, and will be able to adapt such support for use with the present disclosure. The support can then be washed with suitable buffers followed by treatment with the detectably labeled antibody. The solid phase support can then be

washed with the buffer a second time to remove unbound antibody. The amount of bound label on the solid support can then be detected by conventional means.

[0145] Immunohistochemistry methods are also suitable for detecting the expression levels of the prognostic markers. In some embodiments, antibodies or antisera, including polyclonal antisera, and monoclonal antibodies specific for each marker may be used to detect expression. The antibodies can be detected by direct labeling of the antibodies themselves, for example, with radioactive labels, fluorescent labels, hapten labels such as, biotin, or an enzyme such as horse radish peroxidase or alkaline phosphatase. Alternatively, unlabeled primary antibody is used in conjunction with a labeled secondary antibody, comprising antisera, polyclonal antisera or a monoclonal antibody specific for the primary antibody. Immunohistochemistry protocols and kits are well known in the art and are commercially available.

[0146] Immunological methods for detecting and measuring complex formation as a measure of protein expression using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), fluorescence-activated cell sorting (FACS) and antibody arrays. Such immunoassays typically involve the measurement of complex formation between the protein and its specific antibody. These assays and their quantitation against purified, labeled standards are well known in the art (Ausubel, supra, unit 10.1-10.6). A two-site, monoclonal-based immunoassay utilizing antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may be employed (Pound (1998) Immunochemical Protocols, Humana Press, Totowa N.J.).

[0147] Numerous labels are available which can be generally grouped into the following categories:

(a) Radioisotopes, such as ^{36}S, ^{14}C, ^{125}I, ^{3}H, and ^{131}I. The antibody variant can be labeled with the radioisotope using the techniques described in Current Protocols in Immunology, vol 1-2, Coligen et al., Ed., Wiley-Interscience, New York, Pubs. (1991) for example and radioactivity can be measured using scintillation counting.

(b) Fluorescent labels such as rare earth chelates (europium chelates) or fluorescein and its derivatives, rhodamine and its derivatives, dansyl, Lissamine, phycoerythrin and Texas Red are available. The fluorescent labels can be conjugated to the antibody variant using the techniques disclosed in Current Protocols in Immunology, supra, for example. Fluorescence can be quantified using a fluorimeter.

(c) Various enzyme-substrate labels are available and U.S. Pat. Nos. 4,275,149, 4,318,980 provides a review of some of these. The enzyme generally catalyzes a chemical alteration of the chromogenic substrate which can be measured using various techniques. For example, the enzyme may catalyze a color change in a substrate, which can be measured spectrophotometrically. Alternatively, the enzyme may alter the fluorescence or chemiluminescence of the substrate. Techniques for quantifying a change in fluorescence are described above. The chemiluminescent substrate becomes electronically excited by a chemical reaction and may then emit light which can be measured (using a chemiluminometer, for example) or donates energy to a fluorescent acceptor. Examples of enzymatic labels include luciferases (e.g., firefly luciferase and bacterial luciferase; U.S. Pat. No. 4,737,456), luciferin, 2,3-dihydrophthalazinediones, malate dehydrogenase, urease, peroxidase such as horseradish peroxidase (HRPO), alkaline phosphatase, .beta.-galactosidase, glucoamylase, lysozyme, saccharide oxidases (e.g., glucose oxidase, galactose oxidase, and glucose-6-phosphate dehydrogenase), heterocyclic oxidases (such as uricase and xanthine oxidase), lactoperoxidase, microperoxidase, and the like. Techniques for conjugating enzymes to antibodies are described in O'Sullivan et al., Methods for the Preparation of Enzyme-Antibody Conjugates for Use in Enzyme Immunoassay, in Methods in Enzyme. (Ed. J. Langone & H. Van Vunakis), Academic press, New York, 73: 147-166 (1981).

[0148] In some embodiments, a detection label is indirectly conjugated with the antibody. The skilled artisan will be aware of various techniques for achieving this. For example, the antibody can be conjugated with biotin and any of the three broad categories of labels mentioned above can be conjugated with avidin, or vice versa. Biotin binds selectively to avidin and thus, the label can be conjugated with

the antibody in this indirect manner. Alternatively, to achieve indirect conjugation of the label with the antibody, the antibody is conjugated with a small hapten (e.g. digoxin) and one of the different types of labels mentioned above is conjugated with an anti-hapten antibody (e.g. anti-digoxin antibody). In some embodiments, the antibody need not be labeled, and the presence thereof can be detected using a labeled antibody, which binds to the antibody.

[0149] The gene signatures described herein can be used to select treatment for NCSLC patients. As explained herein, the biomarkers can classify patients with NSCLC into a poor survival group or a good survival group and into groups that might benefit from adjuvant chemotherapy or not.

[0150] Accordingly, in one embodiment, the application provides a method of selecting a therapy for a subject with NSCLC, comprising the steps:

(a) classifying the subject with NSCLC into a poor survival group or a good survival group according to the methods described herein; and

(b) selecting adjuvant chemotherapy for the subject classified as being in the poor survival group or no adjuvant chemotherapy for the subject classified as being in the good survival group.

[0151] In another embodiment, the application provides a method of selecting a therapy for a subject with NSCLC, comprising the steps:

(a) determining the expression of from 5 to thirteen biomarkers in a test sample from the subject, wherein the five to thirteen biomarkers correspond to the genes in Table 4;

(b) comparing the expression of the five to thirteen biomarkers in the test sample with the five to thirteen biomarkers in a control sample;

(c) classifying the subject in a poor survival group or a good survival group, wherein a difference or a similarity in the expression of the five to thirteen biomarkers between the control sample and the test sample is used to classify the subject into a poor survival group or a good survival group; and

(d) selecting adjuvant chemotherapy if the subject is classified in the poor survival group and selecting no adjuvant chemotherapy if the subject is classified in the good survival group.

[0152] The term "adjuvant chemotherapy" as used herein means treatment of cancer with chemotherapeutic agents after surgery where all detectable disease has been removed, but where there still remains a risk of small amounts of remaining cancer. Typical chemotherapeutic agents include cisplatin, carboplatin, vinorelbine, gemcitabine, doccetaxel, paclitaxel and navelbine.

[0153] In another aspect, the application provides compositions useful in detecting changes in the expression levels of the 13 genes listed in Table 4A. Accordingly in one embodiment, the application provides a composition comprising a plurality of isolated nucleic acid sequences wherein each isolated nucleic acid sequence hybridizes to:

(a) a RNA product of one of the 13 genes listed in Table 4; and/or

(b) a nucleic acid complementary to a),

wherein the composition is used to measure the level of RNA expression of the 13 genes. In a particular embodiment, the plurality of isolated nucleic acid sequences comprise isolated nucleic acids hybridizable to the 13 probe target sequences as set out in Table 9. In one embodiment, the plurality of isolated nucleic acid sequences comprise isolated nucleic acids hybridizable to SEQ ID NO 4, 11-15, 22, 26, 35, 78, 130, 133, and 169.

[0154] In another embodiment, the application provides a composition comprising 13 forward and 13 reverse primers for amplifying a region of each gene listed in Table 4A. In a particular embodiment, the 26 primers are as set out in Table 7. An additional primer set is described in Table 16. Any combination of the primer sets listed in Table 7 and Table 16 can be used to amplify the genes listed in Table 4A. In one embodiment, the 26 primers each comprise a sequence that is identical to the sequence of one of SEQ ID NO: 173-198 and 203-206. In certain embodiments, L1CAM cDNA is amplified using the forward and revserse primers disclosed in Table 16. In this or another embodiment, MDM2 cDNA is amplified using forward and reverse primers disclosed in Table 16.

[0155] In a further aspect, the application also provides an array that is useful in detecting the expression levels of the 13 genes set out in Table 4A (or a subset thereof including at least 5 genes from Table 4A). Accordingly, in one embodiment, the application provides an array comprising for each gene shown in Table 4A (or

the subset) one or more nucleic acid probes complementary and hybridizable to an expression product of the gene. In a particular embodiment, the array comprises the nucleic acid probes hybridizable to the probe target sequences listed in Table 9. In one embodiment, the array comprises the nucleic acid probes hybridizable to sequences identical to each of SEQ ID NO: 4, 11-15, 22, 26, 35, 78, 130, 133, and 169.

[0156] In yet another aspect, the application also provides for kits used to prognose or classify a subject with NSCLC into a good survival group or a poor survival group or to select a therapy for a subject with NSCLC that includes detection agents that can detect the expression products of the biomarkers. Accordingly, in one embodiment, the application provides a kit to prognose or classify a subject with early stage NSCLC comprising detection agents that can detect the expression products of from 5 to 14 biomarkers, wherein the 5 to 14 biomarkers comprise from 5 to 13 genes in Table 4A. The set of up to 14 biomarkers may further comprise genes from Table 3.

[0157] In one embodiment, the application provides a kit to select a therapy for a subject with NSCLC, comprising detection agents that can detect the expression products of from 5 to 14 biomarkers, wherein the 5 to 14 biomarkers comprise 5 to 13 genes in Table 4A, and optionally genes from Table 3.

[0158] For example, the kit may comprise a primer set for amplifying a target sequence in each of from 5 to 14 genes from Table 3, at least 5 of which are from Table 4 (e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13 genes from Table 4A). Exemplary target sequences are shown in Table 11. In certain embodiments, the primer set contains primer pairs (forward and reverse primers) for amplifying 5 to 13 genes from Table 4A. In certain embodiments, the kit further comprises a primer set for amplifying at least one normalization gene, such as one or more normalization genes described herein. Additionally, the kit may comprise at least one probe for detecting each target sequence, including in connection with the detection platforms described herein (e.g., TaqMan™).

[0159] The materials and methods of the present disclosure are ideally suited for preparation of kits produced in accordance with well known procedures. In some embodiments, kits comprise agents (like the polynucleotides and/or antibodies

described herein as non-limiting examples) for the detection of expression of the disclosed sequences, such as for example, SEQ ID NO: 4, 11-15, 22, 26, 35, 78, 130, 133, and 169, the target sequences listed in Table 9, or the target sequences listed in Table 11. Kits, may comprise containers, each with one or more of the various reagents (sometimes in concentrated form), for example, pre-fabricated microarrays, buffers, the appropriate nucleotide triphosphates (e.g., dATP, dCTP, dGTP and dTTP; or rATP, rCTP, rGTP and UTP), reverse transcriptase, DNA polymerase, RNA polymerase, and one or more primer complexes (e.g., appropriate length poly(T) or random primers linked to a promoter reactive with the RNA polymerase). A set of instructions will also typically be included.

[0160] In some embodiments, a kit may comprise a plurality of reagents, each of which is capable of binding specifically with a target nucleic acid or protein. Suitable reagents for binding with a target protein include antibodies, antibody derivatives, antibody fragments, and the like. Suitable reagents for binding with a target nucleic acid (e.g. a genomic DNA, an mRNA, a spliced mRNA, a cDNA, or the like) include complementary nucleic acids. For example, nucleic acid reagents may include oligonucleotides (labeled or non-labeled) fixed to a substrate, labeled oligonucleotides not bound with a substrate, pairs of PCR primers, molecular beacon probes, and the like.

[0161] In some embodiments, kits may comprise additional components useful for detecting gene expression levels. By way of example, kits may comprise fluids (e.g. SSC buffer) suitable for annealing complementary nucleic acids or for binding an antibody with a protein with which it specifically binds, one or more sample compartments, a material which provides instruction for detecting expression levels, and the like.

[0162] In some embodiments, kits for use in the RT-PCR methods described herein comprise one or more target RNA-specific FRET probes and one or more primers for reverse transcription of target RNAs or amplification of cDNA reverse transcribed therefrom.

[0163] In some embodiments, one or more of the primers is "linear". A "linear" primer refers to an oligonucleotide that is a single stranded molecule, and typically does not comprise a short region of, for example, at least 3, 4 or 5

contiguous nucleotides, which are complementary to another region within the same oligonucleotide such that the primer forms an internal duplex. In some embodiments, the primers for use in reverse transcription comprise a region of at least 4, such as at least 5, such as at least 6, such as at least 7 or more contiguous nucleotides at the 3'-end that has a base sequence that is complementary to region of at least 4, such as at least 5, such as at least 6, such as at least 7 or more contiguous nucleotides at the 5'-end of a target RNA.

[0164] In some embodiments, the kit further comprises one or more pairs of linear primers (a "forward primer" and a "reverse primer") for amplification of a cDNA reverse transcribed from a target RNA. Accordingly, in some embodiments, the forward primer comprises a region of at least 4, such as at least 5, such as at least 6, such as at least 7, such as at least 8, such as at least 9, such as at least 10 contiguous nucleotides having a base sequence that is complementary to the base sequence of a region of at least 4, such as at least 5, such as at least 6, such as at least 7, such as at least 8, such as at least 9, such as at least 10 contiguous nucleotides at the 5'-end of a target RNA. Furthermore, in some embodiments, the reverse primer comprises a region of at least 4, such as at least 5, such as at least 6, such as at least 7, such as at least 8, such as at least 9, such as at least 10 contiguous nucleotides having a base sequence that is complementary to the base sequence of a region of at least 4, such as at least 5, such as at least 6, such as at least 7, such as at least 8, such as at least 9, such as at least 10 contiguous nucleotides at the 3'-end of a target RNA.

[0165] In some embodiments, the kit comprises at least a first set of primers for amplification of a cDNA that is reverse transcribed from a target RNA capable of specifically hybridizing to a nucleic acid comprising a sequence identically present in one of the genes listed in Table 4A. In some embodiments, the kit comprises at least thirteen sets of primers, each of which is for amplification of a different target RNA capable of specifically hybridizing to a nucleic acid comprising a sequence identically present in a different gene listed in Table 4A. In one embodiment, the kit comprises thirteen forward and thirteen reverse primers described in Table 7, comprising sequences identical to SEQ ID NOs 173-198. In some embodiments, the kit comprises at least one set of primers that is capable of amplifying more than one cDNA reverse transcribed from a target RNA in a sample.

[0166] In some embodiments, probes and/or primers for use in the compositions described herein comprise deoxyribonucleotides. In some embodiments, probes and/or primers for use in the compositions described herein comprise deoxyribonucleotides and one or more nucleotide analogs, such as LNA analogs or other duplex-stabilizing nucleotide analogs described above. In some embodiments, probes and/or primers for use in the compositions described herein comprise all nucleotide analogs. In some embodiments, the probes and/or primers comprise one or more duplex-stabilizing nucleotide analogs, such as LNA analogs, in the region of complementarity.

[0167] In some embodiments, the compositions described herein also comprise probes, and in the case of RT-PCR, primers, that are specific to one or more housekeeping genes for use in normalizing the quantities of target RNAs. Such probes (and primers) include those that are specific for one or more products of housekeeping genes selected from ACTB, BAT1, B2M, EDS, IP08, TBP, PPA1, GUSB, U6 snRNA, RNU44, RNU 48, and U47.

[0168] In some embodiments, the kits for use in real time RT-PCR methods described herein further comprise reagents for use in the reverse transcription and amplification reactions. In some embodiments, the kits comprise enzymes such as reverse transcriptase, and a heat stable DNA polymerase, such as Taq polymerase. In some embodiments, the kits further comprise deoxyribonucleotide triphosphates (dNTP) for use in reverse transcription and amplification. In further embodiments, the kits comprise buffers optimized for specific hybridization of the probes and primers.

[0169] In some embodiments, kits are provided containing antibodies to each of the protein products of the genes listed in Table 4A, conjugated to a detectable substance, and instructions for use. Kits may comprise an antibody, an antibody derivative, or an antibody fragment, which binds specifically with a marker protein, or a fragment of the protein. Such kits may also comprise a plurality of antibodies, antibody derivatives, or antibody fragments wherein the plurality of such antibody agents binds specifically with a marker protein, or a fragment of the protein.

[0170] In some embodiments, kits may comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological

sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such kits can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail herein for nucleic acid arrays and similar methods have been developed for antibody arrays.

[0171] A person skilled in the art will appreciate that a number of detection agents can be used to determine the expression of the biomarkers. For example, to detect RNA products of the biomarkers, probes, primers, complementary nucleotide sequences or nucleotide sequences that hybridize to the RNA products can be used. To detect protein products of the biomarkers, ligands or antibodies that specifically bind to the protein products can be used.

[0172] Accordingly, in one embodiment, the detection agents are probes that hybridize to the 13 biomarkers. In a particular embodiment, the probe target sequences are as set out in Table 9. In one embodiment, the probe target sequences are identical to SEQ ID NO: 4, 11-15, 22, 26, 35, 78, 130, 133, and 169. In another embodiment, the detection agents are forward and reverse primers that amplify a region of each of the 13 genes listed in Table 4A. In a particular embodiment, the primers are as set out in Table 7 and Table 16. In one embodiment, the primers comprise one or more of the polynucleotide sequences (or one or more primer sets) of SEQ ID NO: 173-198 and 203-206.

[0173] A person skilled in the art will appreciate that the detection agents can be labeled.

[0174] The label is preferably capable of producing, either directly or indirectly, a detectable signal. For example, the label may be radio-opaque or a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, ^{123}I, ^{125}I, ^{131}I; a fluorescent (fluorophore) or chemiluminescent (chromophore) compound, such as fluorescein isothiocyanate, rhodamine or luciferin; an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase; an imaging agent; or a metal ion.

[0175] The kit can also include a control or reference standard and/or instructions for use thereof. In addition, the kit can include ancillary agents such as vessels for storing or transporting the detection agents and/or buffers or stabilizers.

[0176] In some aspects, a multi-gene signature is provided for prognosis or classifying patients with lung cancer. In some embodiments, a 5 to fourteen (e.g., thirteen)-gene signature is provided, comprising reference values for each of the genes based on relative expression data from a historical data set with a known outcome, such as good or poor survival, and/or known treatment, such as adjuvant chemotherapy. In one embodiment, four reference values are provided for each of the thirteen genes listed in Table 4A. In one embodiment, the reference values for each of the thirteen genes are principal component values set forth in Table 10.

[0177] In one aspect, relative expression data from a patient are combined with the gene-specific reference values on a gene-by-gene basis for each of the thirteen, and, optionally, additional genes, to generate a test value which allows prognosis or therapy recommendation. In some embodiments, relative expression data are subjected to an algorithm that yields a single test value, or combined score, which is then compared to a control value obtained from the historical expression data for a patient or pool of patients.

[0178] In some embodiments, the control value is a numerical threshold for predicting outcomes, for example good and poor outcome, or making therapy recommendations for a subject, for example adjuvant chemotherapy in addition to surgical resection or surgical resection alone. In some embodiments, a test value or combined score greater than the control value is predictive, for example, of a poor outcome or benefit from adjuvant chemotherapy, whereas a combined score falling below the control value is predictive, for example, of a good outcome or lack of benefit from adjuvant chemotherapy for a subject.

[0179] In some embodiments, a method for prognosing or classifying a subject with NSCLC comprises:

(a) measuring expression levels of from 5 to 14 (e.g., 13) biomarkers from Table 4 in a test sample,

(b) calculating a combined score or test value for the subject from the expression levels , and,

(c) comparing the combined score to a control value,

Wherein a combined score greater than the control value is used to classify a subject into a high risk or poor survival group and a combined score lower than the control value is used to classify a subject into a lower risk or good survival group.

[0180] In one embodiment, the combined score is calculated from relative expression data multiplied by reference values, determined from historical data, for each gene. Accordingly, the combined score may be calculated using Formula I below:

$$\text{Combined score} = 0.557 \times PC1 + 0.328 \times PC2 + 0.43 \times PC3 + 0.335 \times PC4$$

Where PC1 is the sum of the relative expression level for each gene in a multi-gene signature multiplied by a first principal component for each gene in the multi-gene signature, PC2 is the sum of the relative expression level for each gene multiplied by a second principal component for each gene, PC3 is the sum of the relative expression level for each gene multiplied by a third principal component for each gene, and PC4 is the sum of the relative expression level for each gene multiplied by a fourth principal component for each gene. In some embodiments, the combined score is referred to as a risk score. A risk score for a subject can be calculated by applying Formula I to relative expression data from a test sample obtained from the subject.

[0181] In some embodiments, PC1 is the sum of the relative expression level for each gene provided in Table 4A multiplied by a first principal component for each gene, respectively, as set forth in Table 10; PC2 is the sum of the relative expression level for each gene provided in Table 4A multiplied by a second principal component for each gene, respectively, as set forth in Table 10; PC3 is the sum of the relative expression level for each gene provided in Table 4A multiplied by a third principal component for each gene, respectively, as set forth in Table 10; and PC4 is the sum of the relative expression level for each gene provided in Table 4A multiplied by a fourth principal component for each gene, respectively, as set forth in Table 10.

[0182] In one embodiment, the control value is equal to -0.1. A subject with a risk score of more than -0.1 is classified as high risk (poor prognosis). A patient with a risk score of less than -0.1 is classified as lower risk (good prognosis). In some embodiments, adjuvant chemotherapy is recommended for a subject with a risk

score of more than -0.1 and not recommended for a subject with a risk score of less than -0.1.

[0183] In some embodiments, the invention involves classifying the sample into a high risk or low risk group as described herein. For example, samples may be classified on the basis of threshold values or based upon Mean and/or Median expression levels in high risk patients versus low-risk patients. Various classification schemes are known for classifying samples between two or more classes or groups, and these include, without limitation: Principal Components Analysis, Naïve Bayes, Support Vector Machines, Nearest Neighbors, Decision Trees, Logistic, Artificial Neural Networks, Penalized Logistic Regression, and Rule-based schemes. In addition, the predictions from multiple models can be combined to generate an overall prediction. For example, a "majority rules" prediction may be generated from the outputs of a Naïve Bayes model, a Support Vector Machine model, and a Nearest Neighbor model.

[0184] Thus, a classification algorithm or "class predictor" may be constructed to classify samples. The process for preparing a suitable class predictor is reviewed in R. Simon, Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data, *British Journal of Cancer* (2003) 89, 1599-1604, which review is hereby incorporated by reference in its entirety.

[0185] In a further aspect, the application provides computer programs and computer implemented products for carrying out the methods described herein. Accordingly, in one embodiment, the application provides a computer program product for use in conjunction with a computer having a processor and a memory connected to the processor, the computer program product comprising a computer readable storage medium having a computer mechanism encoded thereon, wherein the computer program mechanism may be loaded into the memory of the computer and cause the computer to carry out the methods described herein.

[0186] In another embodiment, the application provides a computer implemented product for predicting a prognosis or classifying a subject with NSCLC comprising:

(a) a means for receiving values corresponding to a subject expression profile in a subject sample; and

(b) a database comprising a reference expression profile associated with a prognosis, wherein the subject biomarker expression profile and the biomarker reference profile each has thirteen or fourteen values, each value representing the expression level of a biomarker, wherein each biomarker corresponds to one gene in Table 4A;

wherein the computer implemented product selects the biomarker reference expression profile most similar to the subject biomarker expression profile, to thereby predict a prognosis or classify the subject.

[0187] In yet another embodiment, the application provides a computer implemented product for determining therapy for a subject with NSCLC comprising:

(a) a means for receiving values corresponding to a subject expression profile in a subject sample; and

(b) a database comprising a reference expression profile associated with a therapy, wherein the subject biomarker expression profile and the biomarker reference profile each has thirteen or fourteen values, each value representing the expression level of a biomarker, wherein each biomarker corresponds to one gene in Table 4A;

wherein the computer implemented product selects the biomarker reference expression profile most similar to the subject biomarker expression profile, to thereby predict the therapy.

[0188] Another aspect relates to computer readable mediums such as CD-ROMs. In one embodiment, the application provides computer readable medium having stored thereon a data structure for storing a computer implemented product described herein.

[0189] In one embodiment, the data structure is capable of configuring a computer to respond to queries based on records belonging to the data structure, each of the records comprising:

(a) a value that identifies a biomarker reference expression profile of the 13 genes in Table 4A;

(b) a value that identifies the probability of a prognosis associated with the biomarker reference expression profile.

[0190] In another aspect, the application provides a computer system comprising

(a) a database including records comprising a biomarker reference expression profile of thirteen genes in Table 4A associated with a prognosis or therapy;

(b) a user interface capable of receiving a selection of gene expression levels of the 13 genes in Table 4A for use in comparing to the biomarker reference expression profile in the database; and

(c) an output that displays a prediction of prognosis or therapy according to the biomarker reference expression profile most similar to the expression levels of the thirteen genes.

[0191] In some embodiments, the application provides a computer implemented product comprising

(a) a means for receiving values corresponding to relative expression levels in a subject, of at least 13 biomarkers comprising the thirteen genes in Table 4A;

(b) an algorithm for calculating a combined scire based on the relative expression levels of the at least 13 biomarkers;

(c) an output that displays the combined score; and, optionally,

(d) an output that displays a prognosis or therapy recommendation based on the combined score.

[0192] The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples. These examples are described solely for the purpose of illustration and are not intended to limit the scope of the invention. Changes in form and substitution of equivalents are contemplated as circumstances might suggest or render expedient. Although specific terms have been employed herein, such terms are intended in a descriptive sense and not for purposes of limitation.

[0193] The following non-limiting example is illustrative of the present invention:

Example 1

Results:

[0194] Table 1 compares the demographic features of 133 patients with microarray profiling to 349 without the profiling. Stage IB patients had more representation in the observation cohort (55% vs. 42%, p=0.01), but all other factors were similarly distributed. There was no significant difference in the overall survivals of patients with or without gene profiling (Figure 2A). For these 133 patients, adjuvant chemotherapy reduced the death rate by 20% (HR 0.80, 95% CI 0.48-1.32, p=0.38; Figure 5).

Prognostic gene expression signature in JBR.10 patients

[0195] Using a p>0.005 as cut-off, 172 of 19,619 probe sets were significantly associated with prognosis in 62 observation patients (Figure 1A and Table 3). Using a method that was designed to identify the minimum expression gene set that can distinguish most patients with poor and good survival outcomes, a 15-gene prognostic signature was identified (Figure 1A and Table 4 and 4b). This signature was able to separate the 62 non-adjuvant treated patients into 31 low-risk and 31 high-risk patients for death (HR 15.020, 95% CI 5.12-44.04, p<0.0001; Figure 2B). Furthermore, stratified analysis showed that the signature was also highly prognostic in 34 stage IB patients (HR 13.32, 95% CI 2.86-62.11, p<0.0001, Figure 2C) and 28 stage II patients (HR 13.47, 95% CI 3.0-60.43, p<0.0001, Figure 2D). Multivariate analysis adjusting for tumor stage, age, gender and histology showed that the prognostic signature was an independent prognostic marker (HR 18.0, 95% CI 5.8-56.1; p<0.0001, Table 2). This did not differ following additional adjustment for surgical procedure and tumor size. Further analysis shows that 2 genes, MLANA and MYT1L, have no significant effect on the risk score (Figure 7). Consequently, a minimal 13-gene signature is defined.

Validation of general applicability of prognostic signature (Summary)

[0196] Applying the risk score algorithm (equation) established from the 62 BR.10 observation patients, the 15-gene signature was demonstrated to be an independent prognostic marker among all 169 DCC patients (HR 2.9, 95% CI 1.5-5.6, p=0.002; Table 2). Subgroup analyses showed results that were not statistically significant among patients from DCC-UM (HR 1.5, 95% CI 0.54-4.31, p=0.4; Table 2)

and HLM (HR 1.2, 95% CI 0.43-3.6, p=0.7; Table 2). The 15-gene signature was also prognostic among UM-SQ patients (HR 2.3, 95% CI 1.1-4.7, p=0.026; Table 2), and in the Duke's patients (HR 1.5, 95% CI 0.81-2.89, p=0.19; Table 2).

[0197] The prognostic value of the 15-gene signature was tested in stage I patients of the DCC (n=141) patients and was able to identify patients with significantly different survival outcome (Table 8). The same results apply to the 13-gene signature, given the lack of contribution of MLANA and MYT1L to the risk score (Figure 7).

Prediction of chemotherapy benefit

[0198] When tested on the microarray data of 71 JBR.10 patients who received adjuvant chemotherapy, the 15-gene signature was not prognostic (HR 1.5, 95% CI 0.7-3.3, p=0.28, Table 2). The signature was also not prognostic when applied separately to stage IB and stage II patients (Table 2). Among the DCC patients, 41 were identified as having received adjuvant chemotherapy with or without radiotherapy. The 15-gene signature was also not prognostic for these 41 patients (HR 1.1, 95% CI 0.5-2.5, p=0.8) (Table 2).

[0199] Stratified analysis showed that in JBR.10 patients with microarray data, only patients classified to the high-risk group derived benefit from the adjuvant chemotherapy (Figure 3C and 3D). High-risk patients showed 67% improved survival when treated by adjuvant chemotherapy compared to observation (HR=0.33, 95%CI 0.17-0.63, p=0.0005, Figure 3D), while those assigned to the low risk group did not benefit (Figure 3C). These results were reproduced when applied separately to both the stage IB (Figure 3E and 3F) and stage II (Figure 3G and 3H) patients.

[0200] Multivariate analysis showed that the decrease of survival associated with adjuvant chemotherapy was independent of the stage (HR=2.26, 95% CI 1.03-4.96, p=0.04). A Cox regression model with chemotherapy received and risk group indicator and their interaction term as independent covariates were performed to fit the overall survival data on the 133 patients with microarray data. This analysis revealed that the interaction term is highly significant (p=0.0003) with the high-risk group deriving significantly greater benefit from adjuvant chemotherapy.

[0201] The results and conclusions apply equally to the 13-gene signature.

The initial study population

[0202] The initial study population comprised a subset of the patients randomized in the JBR.10 trial. There were 169 frozen tumor samples collected from patients who had their surgery at one of the BR.10 Canadian Centres and have consented to the use of their samples for "future" studies in addition to RAS mutation analysis. The samples were harvested using a standardized protocol that was agreed upon during trial protocol development by designated pathologists from each participating centre. All tumors and corresponding normal lung tissue were collected as soon as or within 30 min after resection, and were snap-frozen in liquid nitrogen. For each frozen tissue fragment, a 1 mm cross-section slice was fixed in 10% buffered formalin and submitted for paraffin embedding. Histological evaluation of the HE stained sections revealed 166 samples that contained ≥ 20% tumor cellularity. Among the latter, gene expression profiling was completed successfully in samples from 133 patients. These included 58 patients randomized to the observation (OBS) arm and 75 to the adjuvant chemotherapy (ACT) arm. However, 4 ACT patients refused chemotherapy, and for the purpose of this analysis, they were assigned to the OBS arm. Therefore, the final distribution included 62 OBS patients and 71 ACT patients (Figure 1 and 4).

Microarray data analysis

[0203] The raw microarray data from Affymetrix U133A (Affymetrix, Santa Clara, CA) were pre-processed using RMAexpress v0.32, then were twice log2 transformed since the distribution of additional log2 transformed data appeared more normal. Probe sets were annotated using NetAffx v4.2 annotation tool and only grade A level probe sets 3 (NA24) were included for further analysis. Affymetrix U133A chip contains 22,215 probe sets (19,619 probe sets with grade A annotation). Since the microarray hybridizations were performed in two batches at two separate occasions (January 2004, and June 2005), and unsupervised clustering showed that a batch difference was significant (Figure 6), a distance-weighted discrimination (DWD) algorithm (https://genome.unc.edu/pubsup/dwd/index.html) was applied to homogenize the two batches. The DWD algorithm first finds a hyperplane that separates the two batches and adjusts the data by projecting the different batches on the DWD plane, finds the batch mean, and then subtracts out the DWD plane

multiplied by this mean. In addition, the data were Z score transformed which made the validation across different datasets possible.

Univariate analysis

[0204] The association of the expression of the individual probe set with overall survival (date of randomization to date of last follow up or death) was evaluated by Cox proportional hazards regression. The expression data for 62 patients in observation arm revealed 1312 probe sets that were associated with overall survival at $p<0.05$. Using a more stringent selection criteria of $p<0.005$, 172 probe sets with grade A annotation were prognostic.

Gene set signature selection

[0205] To generate the gene expression signature, an exclusion selection procedure was firstly applied and followed by an inclusion process. The MAximizing R Square Algorithm (MARSA) included 3 sequential steps: a) probe set pre-selection; b) signature optimization; and c) leave-one-out-cross-validation. First, the candidate probe sets were pre-selected by their associations with survival at $p<0.005$ level. To remove the cross platform variation, expression data was z score transformed and risk score (z score weighted by the coefficient of the univariate Cox regression) was used to synthesize the information of the probe set combination. The candidate probe sets were then subjected to an exclusion followed by an inclusion selection procedure. For the preselected 172 probe sets, the exclusion procedure excluded one probe at a time, summed up the risk score of the remaining 171 probes, the calculated the R square (R^2, Goodness-of-fit) of the Cox model[5,6]. Risk score was dichotomized by an outcome-orientated optimization of cutoff macro based on log-rank statistics (http://ndc.mayo.edu/mayo/research/biostat/sasmacros.cfm) before being introduced to the Cox proportional hazards model. A probe set was excluded if its exclusion resulted in obtaining the largest R^2. The procedure was repeated until there was only one probe set left. An inclusion procedure was followed using the probe set left by the exclusion procedure as the starting probe set. It included one probe set at a time, summed up the risk score of the included probe sets and risk score was dichotomized and R^2 was calculated. The probe set was included if its inclusion resulted in obtaining the largest R^2. The exclusion procedure produced a largest R square of 0.67 by a minimal 7 probe combination and the inclusion procedure

generated a largest R^2 of 0.78 by a 15 probe combination (Figure 1B). Then, the 15-gene signature disclosed in WO 2009/137921, published on November 19, 2009, was established after passing the internal validation by leave-one-out-cross-validation (LOOCV) and external validation on other datasets (listed below). All statistical analyses were performed using SAS v9.1 (SAS Institute, CA). The 13-gene signature of the present disclosure proved to perform as well as the 15-gene signature (Figure 7; Table 4).

Prognostic modeling by principal component analysis of signature genes

[0206] Principal components analysis (PCA) (based on correlation matrix) was carried out to synthesize the information across the chosen gene probe sets and reduce the number of covariates in building the prognostic model. The eigenvalue of greater than or equal to 1 was used as cutoff point in determining how many proponents to include in the model, and those significantly correlated to disease-specific survival (DSS) were included in the final multivariable model. The PCA analysis was done based on all 133 patients with microarray data. When correlated to the DSS based on the 62 observation patients, the first 4 principal components were found to satisfy the criteria and were included in the prognostic model. Table 10 lists the four principal components for each of the 13 genes in the 13-gene signature. The same analysis can be applied to derive principal component coefficients for additional genes selected from the 172 genes listed in Table 3, such as for example, RGS4, UGT2B4, and/or MCF2. Furthermore, one of skill will appreciate from the above description how to obtain the first four principal component coefficients for any of the genes listed in Table 3.

[0207] To determine the gene signature prognostic group, multivariate Cox regression model with the first 4 principal components were fitted to the disease specific survival of the 62 observation patients. The linear prognostic scores were calculated by the sum of the multiplication of the estimated coefficient from Cox model and the corresponding principal component value. Using the prognostic score, patients were divided into low and high risk group based on the median of the prognostic score, i.e., those with prognostic score less than the median as low risk group, while those with score no less than the median as high risk group. For the 62 observation patients with microarray data, 31 patients were classified in each group.

Applying the same rule to the 71 chemo-treated patients, 35 patients were classified in low risk group and 36 patients in high-risk group.

Validation of general applicability of prognostic signature

[0208] Validation of the 15-gene signature was carried out on stage I-II cases from Duke, UM-SQ, and DCC who did not receive adjuvant chemotherapy. When the risk score was dichotomized using the cutoff determined from the BR.10 training set, the 15-gene signature was able to separate 38 cases of low risk from 47 cases of high risk (log rank p=0.226) of NSCLC in the Duke dataset. Multivariate analysis (adjusted for stage, histology and patients' age and gender) showed that the 15-gene signature was an independent prognostic factor (HR=1.5, 95%CI 0.81-2.89, p=0.19, Table 2). Raponi contains squamous cell carcinoma only and the cases have the worst survival rate. However, the 15-gene signature was still able to separate 50 cases of low risk from 56 cases with high risk (log rank p=0.0447) and this separation was independent of stage and patients' age and gender (HR= 2.3, 95% CI 1.1-4.7 p=0.026, Table 2). The DCC dataset contained only adenocarcinoma cases. Applying the 15-gene signature on DCC stage I and II, was able to separate 87 low risk cases from the 82 high risk cases (log rank p=0.0002, Figure 2E). Multivariate analysis (adjusted for stage and patients' age and gender) showed that the prognostic value of the 15-gene signature was independent prognostic factor (HR=2.9, 95%CI 1.5-5.6, p=0.002, Table 2). There were 67 stage IB-II cases without chemotherapy in MI, the 15-gene signature was able to separate 44 low risk cases from the 23 high risk cases (log rank p=0.013). Multivariate analysis (adjusted for stage and patients' age and gender) showed that the prognostic value of the 15-gene signature was independent prognostic factor (HR=1.5, 95%CI 0.54-4.31, p=0.4, Table 2). Cases from MSKCC had a significantly better 5-year overall survival compared to other datasets. However, the 15-gene signature was able to separate 32 cases of low risk from 32 cases of high risk in MSKCC (log rank p=0.16). Multivariate analysis (adjusted for stage) revealed that the 15-gene signature was an independent prognostic factor. Validation of the 15-gene signature on HLM revealed that the 15-gene signature was able to separate 26 cases of low risk from 24 cases of high risk (log rank p=0.0084). Multivariate analysis (adjusted for stage) showed that there was a trend to separation by the 15-gene signature (HR=1.2, 95%CI 0.43-3.6, p=0.7). These validation data confirm that the 15-gene signature is a strong

prognostic signature and its power of predicting the outcome of NSCLC is independent of and superior to that of stage. Since MLANA and MYT1L proved to have no significant effect on the risk score (Figure 7), validation data thus confirm that the 13-gene signature is a strong prognostic signature and its power of predicting the outcome of NSCLC is independent of and superior to that of stage.

The benefit of chemotherapy was limited to high risk patients

[0209] A total of 30 deaths were observed in the ACT. Six of them were due to other malignancies. The 15-gene signature was unable to separate the good/bad outcome patients (p=0.83, data not shown) in the ACT. However, stratified analysis showed that only patients with high risk derived benefit from adjuvant chemotherapy (Figure 3D). Upon receiving adjuvant chemotherapy, the survival rate of the 36 high-risk patients was significantly improved (HR=0.33, 95%CI 0.17-0.63, p=0.0005, Figure 3D). On the other hand, the application of chemotherapy on low risk patients resulted in a decrease in survival rate (HR=3.67, 95%CI 1.22-11.06, p=0.0133, Figure 3C). Death was evenly distributed between the low and high risk groups in the ACT arm (15 deaths in low and high risk group, respectively). Each of these two groups contained 3 deaths that were not due to lung cancer. Stratification by risk group and stage showed that the survival rate of high risk patients from both stage IB and stage II was significantly improved by chemotherapy (Figure 3F and H). Moreover, for low risk patients of stage II, chemotherapy was associated with significantly decreased survival (Figure 3E and G). A Cox regression model with chemotherapy received and risk group indicator and their interaction term as independent covariates was performed to fit the overall survival data on the 133 patients with microarray data. This analysis revealed that the interaction term is highly significant (p = 0.0002) with the high-risk group deriving significantly greater benefit from adjuvant chemotherapy.

Discussion:

[0210] Gene expression signature is thought to represent the altered key pathways in carcinogenesis and thus is able to predict patients' outcome. However, being able to faithfully represent the altered key pathways, the signature must be generated from genome-wide gene expression data. The present study used all information generated by Affymetrix U133A chip on NSCLC samples from a randomized clinical trial to derive a 15-gene signature. The 15-gene signature was

able to identify 50% (31/62) stage IB-II NSCLC patients had relative good outcome. Multivariate analysis indicated that the 15-gene signature was an independent prognostic factor. Moreover, its independent prognostic effect had been *in silico* validated on 169 adenocarcinomas without adjuvant chemo- or radio-therapy from DCC and 85 NSCLC from Duke and 106 squamous cell carcinomas of the lung from the University of Michigan (UM-SQ). Importantly, the 15-gene signature was able to predict the response to adjuvant chemotherapy with high-risk patients across the stages being benefited from adjuvant chemotherapy. This finding was also validated on DCC dataset.

[0211] When attempting to migrate the detection platform to quantitative PCR, two genes (MLANA and MYT1L) were difficult to detect reproducably, and after further analysis were discovered to have no significant effect on the risk score (Figure 7). Thus, a 13-gene signature also predicts the response to adjuvant chemotherapy with high-risk patients across the stages being benefited from adjuvant chemotherapy.

[0212] Adjuvant chemotherapy for completely resected early stage NSCLC was a research question until the results of a series of positive trials[2, 4], including BR.10 [3], were published. However, whether chemotherapy played a beneficial role in stage IB remained to be clarified[2-6]. The present study showed that the stage IB patients were potentially able to be separated into low (49.3%, 36/73) and high (50.7%, 37/73) risk groups using the 15-gene signature. Upon administering the adjuvant chemotherapy to stage IB patients, the survival rate of patients with high risk was significantly improved (p=0.0698, Figure 3F) whereas patients with low risk did not experience a benefit in survival (p=0.0758, Figure 3E). Therefore the effect of chemotherapy on stage IB NSCLC was neutralized and thus gave an incorrect impression that no beneficial effect existed[3]. Based on the evidence provided here and from the meta-analysis[6], it may be concluded that 50.7% (37/73) stage IB NSCLC patients have the potential to benefit from adjuvant chemotherapy. Since MLANA and MYTL1 proved to have no significant effect on the risk score (Figure 7), the 13-gene signature is also a predictor that 50.7% (37/73) stage IB NSCLC patients have the potential to benefit from adjuvant chemotherapy.

[0213] Another significance of the present study was that the signature was able to identify a subgroup (50%, 30/60) of patients from stage II who did not benefit

from adjuvant chemotherapy (p=0.1498, Figure 3G). In current practice, adjuvant chemotherapy is recommended for all patients. However, the 15-gene signature suggests that about a half of the stage II patients may not benefit from adjuvant chemotherapy. Since MLANA and MYT1L have no significant effect on the risk score (Figure 7), the 13-gene signature also serves as a basis to conclude that about half of the stage II patients may not benefit from adjuvant chemotherapy.

[0214] The gene ontology analysis showed that in the 13-gene signature, 4 genes (FOSL2, HEXIM1, IKBKAP, and ZNF236) were involved in the regulation of transcription. EDN3 and STMN2 played a role in signal transduction. Transformed 3T3 cell double minute 2 (MDM2), an E3 ubiquitin ligase, which targets p53 protein for degradation, plays a key role in cell cycle and apoptosis. Dworakowska D. et al.[24] reported that overexpression of MDM2 protein was correlated with low apoptotic index, which was associated with poorer survival. Myoglobin (MB) palyed a role in response to hypoxia and Uridine monophosphate synthetase (UMPS) participated in the 'de novo' pyrimidine base biosynthetic process, however, none of them has not been explored in lung cancer. The L1 cell adhesion molecule (L1CAM) involved in cell adhesion whose overexpression was associated with tumor metastasis and poor prognosis[25-28]. ATPase, Na+/K+ transporting, beta 1 polypeptide (ATP1B1) was involved in ion transport which was reported recently to be able to discriminate the serous low malignant potential and invasive epithelial ovarian tumors[29]. These findings indicated that cellular transcription, cell cycle and apoptosis, cell adhesion and response to hypoxia were important for lung cancer progression.

[0215] The range of expression levels of members of the 13-gene signature was broad, from very low expression level such as MDM2 and ZNF236 to fairly high expression such as TRIM14 or very high expression such as ATP1B1 (Table 4). Least variable gene (<5%), such as UMPS (Table 4), was also a member of the signature. These data suggested that it may not be a good practice to exclude low expressed and least variable probe set in the data pre-selection process in an arbitrary way. The signature generated using the present strategy performed better than that of Raponi's method of using the top 50 genes. There are only 3 genes (IKBKAP, L1CAM, and FAM64A) whose significance in association with survival is in the top 50 genes (Table 4).

Materials and Methods:

Patients and Samples

[0216] Included in the JBR.10 protocol was the collection of snap-frozen or formalin-fixed paraffin embedded tumor samples for KRAS mutation analysis and tissue banking for future laboratory studies[3]. Altogether 445 of 482 randomized patients consented to banking. Snap-frozen tissues were collected from 169 Canadian patients (Figure 4). Histological evaluation of the HE section from the snap-frozen tumor samples revealed 166 that contained an estimated >20% tumor cellularity; gene expression profiling was completed in 133 of these patient samples, using the U133A oligonucleotide microarrays (Affymetrix, Santa Clara, CA). Profiling was not completed in 33 patient samples. Of 133 patients with microarray profiles, 62 did not received post-operative adjuvant chemotherapy and were group as observation patients, while 71 patients were received chemotherapy. University Health Network Research Ethics Board approved the study protocol.

RNA isolation and microarray profiling

[0217] Total RNA was isolated from frozen tumor samples after homogenization in guanidium isothiocyanate solution and acid phenol-chloroform extraction. The quality of isolated RNA was assessed initially by gel electrophoresis, followed by the Agilent Bioanalyzer. Ten micrograms of total RNA was processed, labeled, and hybridized to Affymetrix's HG-U133A GeneChips. Microarray hybridization was performed at the Center for Cancer Genome Discovery of Dana Farber Cancer Institute.

Microarray data analysis and gene annotation

[0218] The raw microarray data were pre-processed using RMAexpress v0.3[22]. Probe sets were annotated using NetAffx v4.2 annotation tool and only grade A level probe sets[23] (NA22) were included for further analysis. Because the microarray profiling was done in two separate batches at different times and unsupervised heuristic K-means clustering identified a systematic difference between the two batches (Figure 6), the distance-weighted discrimination (DWD) method (https://genome.unc.edu/pubsup/dwd/index.html) was used to adjust the difference. The DWD method first finds a separating hyperplane between the two batches and adjusts the data by projecting the different batches on the DWD plane, discover the batch mean, and then subtracts out the DWD plane multiplied by this

mean. The data were then transformed to Z score by centering to its mean and scaling to its standard deviation. This transformation was necessary for validation on different datasets in which different expression ranges are likely to exist, and for validation on different platforms, such as qPCR where the data scale is different.

Derivation of signature

[0219] The pre-selected probe sets by univariate analysis at $p<0.005$ were selected by an exclusion procedure. The exclusion selection excluded one probe set at a time based on the resultant R square (R^2, Goodness-of-fit [15, 16]) of the Cox model. It kept repeating until there was only one probe set left. The procedure was repeated until there was only one probe set left. An inclusion procedure was followed using the probe set left by the exclusion procedure as the starting probe set. It included one probe set at a time based on the resultant R^2 of the Cox model. Finally, the R^2 was plotted against the probe set and a set of minimum number of probe sets yet having the largest R^2 was chosen as candidate signature. Gene signature was established after passing the internal validation by leave-one-out-cross-validation (LOOCV) and external validation on other datasets (listed below). All statistical analyses were performed using SAS v9.1 (SAS Institute, CA).

Validation in separate microarray datasets

[0220] The prognostic value of this 15-gene signature was tested on separate microarray datasets. Three represented subsets of microarray data from the NCI Director's Challenge Consortium (DCC) for the Molecular Classification of Lung Adenocarcinoma (Nature Medicine, in review/in press). In total, the Consortium analyzed the profiles of 442 tumors, including 177 from University of Michigan (UM), 79 from H. L. Moffitt Cancer Centre (HLM), 104 from Memorial Sloan-Kettering Cancer Centre (MSK), and 82 from our group. As 39 of the latter tumors overlap with samples used in this study, only data from the first 3 groups were used for validation. In addition, patients who were noted as either unknown or having received adjuvant chemotherapy and/or radiotherapy were excluded. Therefore, the DCC dataset used in this validation study included only 169 patients: 67 from UM, 46 from HLM, 56 from MSK. Two additional published microarray datasets were also used for validation: the Duke's University dataset of 85 non-small cell lung cancer patients (Potti, et al, NEJM), and the University of Michigan dataset of 106 squamous cell carcinomas patients (UM-SQ) (Raponi et al). Raw data of these microarray studies

were downloaded and RMA pre-processed. The expression levels were Z score transformed after double log2 transformation. Risk score was the Z score weighted by the coefficient of the Cox model from the OBS. Demographic data of the DCC cohort was listed in Table 5.

[0221] The prognostic value of 2 genes, MLANA and MYT1L was tested on the same microarray datasets. The DCC dataset used in this validation study included 169 patients: 67 from UM, 46 from HLM, 56 from MSK. Two additional published microarray datasets were also used for validation: the Duke's University dataset of 85 non-small cell lung cancer patients (Potti, et al, NEJM), and the University of Michigan dataset of 106 squamous cell carcinomas patients (UM-SQ) (Raponi et al). Raw data of these microarray studies were downloaded and RMA pre-processed. The expression levels were Z score transformed after double log2 transformation. Figure 7 represents the lack of these 2 genes to contribute significantly to the signature.

Statistical analysis

[0222] Risk score was the product of coefficient of Cox proportional model and the standardized expression level. The univariate association of the expression of the individual probe set with overall survival (date of randomization to date of last followup or death) was evaluated by Cox proportional hazards regression. A stringent $p<0.005$ was set as a selection criteria in order to minimize the possibility of false-positive results.

[0223] While the present invention has been described with reference to what are presently considered to be the preferred examples, it is to be understood that the invention is not limited to the disclosed examples. To the contrary, the invention is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

[0224] While the present invention has been described with reference to what are presently considered to be the preferred examples, it is to be understood that the invention is not limited to the disclosed examples. To the contrary, the invention is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

Example 2:

[0225] This application describes a gene expression-based prognostic signature in early-stage (stage I and II) non-small-cell lung cancer (NSCLC) for identifying, more selectively than the current clinicopathological criteria (i.e. TNM staging, gender, age, etc.), patients with significantly different prognoses. The expression of signature genes is measured from a patient's tumour sample and converted into a risk score which then classifies the patient into distinct prognostic risk groups. The higher/lower risk of mortality stratification of patients is independent of clinicopathological criteria, including stage, histology, gender and age, and may assist in guiding the post-surgical treatment of early-stage NSCLC patients, i.e. suggesting chemotherapy for higher risk stage I patients but avoiding it for stage II patients with a lower risk profile.

[0226] Reverse transcription quantitative PCR (RT-qPCR) may have advantages for clinical settings, and moreover, the majority of NSCLC tumour specimens resected in hospitals are fixed in formalin and subsequently embedded in paraffin (formalin-fixed, paraffin-embedded: FFPE). This preservation method also allows for long-term storage of the tissue samples. Therefore, if the signatures could be migrated to an RT-qPCR platform and used with RNA derived from FFPE-preserved NSCLC specimens, the test may be more easily integrated into the clinical process.

[0227] This example describes the validation of a gene signature to fresh frozen and FFPE samples, using a quantitative RT-PCR platform as compared to a microarray platform.

Materials and Methods

Tissue samples

[0228] Resected tumour samples from untreated stage I and stage II NSCLC patients, fixed in formalin and embedded in paraffin (FFPE) were secured from the Ontario Tumour Bank (OTB) in Toronto and the Ohio State University as part of the Cooperative Human Tissue Network (CHTN). In addition, a set of matched fresh frozen and FFPE-preserved NSCLC tissue specimens was received from OTB.

Total RNA preparation

[0229] Total RNA was isolated from FFPE-tissue sections as curls or mounted onto slides using the RecoverAll kit following standard procedures (Ambion). After

elution in water, the total RNA concentration was assessed with the 260/280nm ratio using the NanoDrop spectrophotometer (Thermo Scientific). If the concentration was below 70 ng/μL the total RNA was concentrated using Amicon Ultra Centrifugal Filters (Millipore). The quality of the isolated total RNA was evaluated using the 2100 Bioanalyzer (Agilent Technologies). RNA integrity numbers (RIN) between 2-2.5 were detected. RNA from a cohort of matched frozen tissue specimens was isolated using the RNeasy Plus Mini kit (Qiagen) and subjected to the same methods to assess quantity and quality.

ABI TaqMan® assays

[0230] TaqMan® gene expression assays (Applied Biosystems) suitable for the detection of RNA derived from FFPE-preserved tissue samples were selected for the 15 signature genes. Five previously published reference genes for the use in either lung cancer and/or gene expression from FFPE tissue samples were selected (BAT1 and TBP (Barsyte-Lovejoy, Lau et al. 2006), ESD (Saviozzi, Cordero et al. 2006), IPO8 (Nguewa, Agorreta et al. 2008), and TFRC (Drury, Anderson et al. 2009)).

Custom designed gene expression assays

[0231] Up to six hydrolysis probe-based assays optimized for the detection of RNA from FFPE-preserved and fresh frozen tumor samples were designed for each of the 15 signature genes (IDT Technologies and Roche Applied Science).

cDNA synthesis, cDNA amplification and qPCR analysis

[0232] Total RNA was reverse transcribed using the iScript cDNA Synthesis kit according to the manufacturer's protocol (BioRad). The TaqMan® PreAmp Master Mix kit (Applied Biosystems) was used to amplify 12.5 ng of respective cDNA with gene-specific assays according the manufacturer's protocol. The real-time PCR reactions were on the 7900 HT Fast Real-Time PCR System (Applied Biosystems, 384 well plate format) in a total reaction volume of 12μl.

Normalization

[0233] The reactions were run in quadruplicates and the relative expression was calibrated on either a pancreas (only for the gene EDN3) or a universal human reference RNA (UHRR) standard curve. To normalize the expression of the genes in

the signature, the geometric mean of the selected reference genes for either frozen or FFPE-derived RNA was used.

Identification of reference genes

[0234] The GeNorm (Vandesompele, De Preter et al. 2002) and NormFinder (Andersen, Jensen et al. 2004) algorithms were used to determine the most suitable (most stable and least variable) combination of reference genes for the use with either fresh frozen or FFPE-derived RNA.

Statistical analysis

[0235] The intraclass correlation (ICC) (Shrout and Fleiss 1979) was used to assess the agreement of the two risk score calculations. Values of this statistic range from -1 to 1, with 1 indicating exact agreement. Intermediate values of agreement were defined by Landis and Koch (1977) as following: Moderate (0.41-0.60), substantial (0.61-0.80), and almost perfect (0.81-1.00) (Landis and Koch 1977).

Results

Testing pre-validated ABI assays on RNA derived from FFPE-preserved NSCLC tissue samples

[0236] Pre-validated TaqMan assays for each of the 15 signature genes were run on cDNA synthesized from the total RNA of 20 FFPE-preserved NSCLC tissue specimens. The most stable reference gene, TBP, was identified using the NormFinder algorithm from a set of previously published references genes for the use with either NSCLC or FFPE. Normalization using TBP and calculation of the relative gene expression revealed that 13 of the 15 signature genes were reliably detectable across a set of 20 different FFPE NSCLC tissue samples.

[0237] Up to 6 hydrolysis-based assays were designed for each of the 15 signature genes for the use with RNA derived from FFPE-preserved samples (amplicon size <100nt) using two different probe chemistries (i.e., regular TaqMan assays or locked nucleic acid assays (LNA). These assays were run on serially diluted cDNA derived from FFPE-preserved or fresh frozen human reference RNA (UHRR, or pancreas RNA (EDN3)). Standard curves were generated to estimate the PCR efficiency and R2 to select the best assay for each gene. An assay was

chosen if the Ct value was below 36, the PCR efficiency was between 90-110% and R2 > 0.8 (Table 15). For EDN3 only one assay of the six designed assays showed sufficient performance on FFPE samples, with a slightly higher PCR efficiency (116%).

[0238] An analysis of in silico microarray data revealed that the two genes MYT1L and MLANA did not significantly contribute to the risk score to predict outcome. When the C-index with and without these two genes was calculated and compared (Figure 7), there was no statistically significant difference between the risk score calculated from the 13 or the 15 genes. All subsequent experiments were therefore performed with 13 signature genes.

Selection of normalization genes and testing custom assays on a set of matched FFPE and frozen NSCLC samples

[0239] A set of matched frozen and FFPE-preserved early-stage NSCLC samples were secured from the Ontario Tumour Bank to select reference gene(s) for use with RNA derived from FFPE or fresh frozen preserved NSCLC tissue specimens and to test the reliability of the custom assays in detecting gene expression. To identify the best reference genes for use with RNA derived from FFPE or fresh frozen lung cancer samples, the expression of 14 previously published reference genes was evaluated with the GeNorm algorithm. A group of four reference genes ("normalization genes") was found to be stable and substantially lacking in variable for the use with RNA derived from FFPE (BAT1, TBP, PP1A and GUSB) or frozen NSCLC samples (BAT1, TBP, PP1A and IPO8). Three of the four genes overlapped (BAT1, TBP and PP1A) (Figure 8).

[0240] After successful identification of normalization genes and normalization to adjust for input differences, the relative gene expression of the signature genes was calculated for both the frozen and FFPE sample sets. Pearson correlation analysis was then performed to measure the linear dependence between the FFPE and frozen expression values of the matched samples. A Pearson correlation coefficient of 0.75 was found in 13 of the 18 matched NSCLC samples (Figure 9).

Risk score calculations based on RT-qPCR expression data and correlation with microarray classifier

A strong correlation was found on a gene by gene basis between fresh frozen and FFPE-preserved tissue samples. To examine the predictive power, the risk score was calculated, using expression data from total RNA from 30 frozen tissue samples from the UHN183 cohort. RT-qPCR analysis on this sample subset was performed using the custom assays and the reference genes selected for the use with frozen tissue specimens. After normalization, the risk scores were calculated and correlated to the risk scores calculated from the previous microarray analysis. An interclass correlation coefficient (ICC) of 0.66 was found, indicating substantial correlation between risk scores calculated based on microarray and RT-qPCR data. Association with clinical outcome data and univariate survival analysis revealed a hazard ratio (HR, high risk versus low risk) of 1.76 for those 30 samples analyzed on RT-qPCR and a HR of 2.1 for the same sample set analyzed with microarray. The p values were not significant in either case due to the small sample size. The algorithm was the same as used in Example 1.

Discussion

[0241] As shown herein, gene expression detected with custom designed assays carry prognostic information for both FFPE- and frozen-preserved tissues. Furthermore, risk scores calculated from RT-qPCR-derived expression data correlate substantially with the risk scores calculated from microarray data and are useful to provide a risk classification score for mortality stratification in NSCLC patients.

[0242] All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

Table 1. Baseline factors of BR.10 patients with and without microarray profiles

Factor	All Patients (n=482)	Microarray profiled (n=133)		No microarray profiled (n=349)		P value
		n	%	n	%	
Treatment received						
ACT	231	71	53%	160	46%	0.14
OBS	251	62	47%	189	54%	
Age						
< 65	324	87	65%	237	68%	0.6
≥ 65	158	46	35%	112	32%	
Gender						
Male	314	91	68%	223	64%	0.35
Female	168	42	32%	126	36%	
Performance Status						
0	236	67	50%	169	49%	0.72
1	245	66	50%	179	51%	
Stage of Disease						
IB	219	73	55%	146	42%	0.01
II	263	60	45%	203	58%	
Surgery						
Pneumonectomy	113	33	25%	80	23%	0.66
Other Resection	369	100	75%	269	77%	
Pathologic type						
Adenocarcinoma	256	71	53%	185	53%	0.56
Squamous	179	52	39%	127	36%	
Other	47	10	8%	37	11%	
Ras Mutation Status						
Present	117	28	21%	89	26%	0.12*
Absent	333	105	79%	228	65%	
Unknown	32	0	0%	32	9%	

*P-value: Without include those missing or unknown.

Table 2. Comparison of 5-yr Survival (multivariate) of High and Low Risk Groups in Untreated Patients and Patients who Received Adjuvant Chemotherapy.

Observation/untreated Patients				
	n	HR*	95% CI	p value
JBR.10 (randomized with microarray)	62	18.0	5.8 -56.1	<0.0001
Stage IB	34	29.9	4.5 – 197.4	0.0004
Stage II	28	16.4	3.0 – 88.1	0.001
DCC (no adjuvant therapy)	169	2.9	1.5 – 5.6	0.002
UM	67	1.5	0.54 -4.31	0.4
HLM	46	1.2	0.43-3.60	0.7
MSK	56	NA**	NA	
Duke	85	1.5	0.81 – 2.89	0.19
UM-Squamous	106	2.3	1.1 – 4.7	0.026
Patients Treated With Adjuvant Chemotherapy				
BR.10 (randomized with microarray)	71	1.5	0.7- 3.3	0.28
BR.10 Stage I	39	1.7	0.5 – 5.6	0.36
BR.10 Stage II	32	1.2	0.4 – 3.6	0.8
DCC (not randomized)	41	1.1	0.5-2.5	0.8

n: number of patients; HR: hazard ratio; CI: confidence interval

*HR compares the survival of the poor prognostic group to that of the good prognostic group as determined by the 15-gene signature with the adjustment of stage and patients' age and gender. For BR.10, and Duke, the effect of histology was also adjusted

** All events were in high risk group and female patients.

Table 3. 172 U133A probe sets that were prognostic at p<0.005 for the 62 BR.10 observation arm patients.

Probe Set ID	Representative Public ID	UniGene ID	Gene Symbol	Coeffi-cients	HR	HRL	HRH	p value
200878_at	AF052094	Hs.468410	EPAS1	-0.58	0.56	0.37	0.84	0.0048
201228_s_at	NM_006321	Hs.31387	ARIH2	0.47	1.60	1.17	2.18	0.0029
201242_s_at	BC000006	Hs.291196	ATP1B1	-0.69	0.50	0.35	0.71	0.0001
201243_s_at	NM_001677	Hs.291196	ATP1B1	-0.54	0.58	0.41	0.83	0.0028
201301_s_at	NM_001153	Hs.422986	ANXA4	-0.55	0.58	0.40	0.83	0.0028
201502_s_at	NM_020529	Hs.81328	NFKBIA	-0.62	0.54	0.36	0.79	0.0016
202023_at	NM_004428	Hs.516664	EFNA1	-0.67	0.51	0.35	0.76	0.0009
202035_s_at	AF017987	Hs.213424	SFRP1	0.69	1.99	1.39	2.86	0.0002
202036_s_at	AF017987	Hs.213424	SFRP1	0.84	2.31	1.56	3.44	0.0000
202037_s_at	AF017987	Hs.213424	SFRP1	0.74	2.09	1.43	3.07	0.0002
202490_at	AF153419	Hs.494738	IKBKAP	0.42	1.53	1.17	1.99	0.0018
202707_at	NM_000373	Hs.2057	UMPS	0.60	1.81	1.24	2.66	0.0023
202814_s_at	NM_006460	Hs.15299	HEXIM1	0.59	1.80	1.20	2.70	0.0045
203001_s_at	NM_007029	Hs.521651	STMN2	0.55	1.73	1.21	2.47	0.0027
203147_s_at	NM_014788	Hs.575631	TRIM14	-0.56	0.57	0.39	0.82	0.0028
203438_at	AI435828	Hs.233160	STC2	0.67	1.96	1.29	2.96	0.0015
203444_s_at	NM_004739	Hs.173043	MTA2	0.38	1.46	1.12	1.89	0.0046
203475_at	NM_000103	Hs.511367	CYP19A1	0.56	1.76	1.23	2.52	0.0021
203509_at	NM_003105	Hs.368592	SORL1	-0.58	0.56	0.39	0.81	0.0020
203928_x_at	AI870749	Hs.101174	MAPT	0.44	1.55	1.15	2.10	0.0044
203973_s_at	M83667	Hs.440829	CEBPD	-0.61	0.54	0.38	0.77	0.0005
204179_at	NM_005368	Hs.517586	MB	0.47	1.60	1.16	2.22	0.0044
204267_x_at	NM_004203	Hs.77783	PKMYT1	0.63	1.87	1.28	2.73	0.0011
204338_s_at	AL514445	Hs.386726	RGS4	0.57	1.77	1.23	2.53	0.0021
204531_s_at	NM_007295	Hs.194143	BRCA1	0.60	1.82	1.21	2.75	0.0043
204584_at	AI653981	Hs.522818	L1CAM	0.56	1.75	1.30	2.35	0.0002
204684_at	NM_002522	Hs.645265	NPTX1	0.48	1.61	1.18	2.19	0.0024
204810_s_at	NM_001824	Hs.334347	CKM	0.46	1.58	1.20	2.09	0.0012
204817_at	NM_012291	---	ESPL1	0.53	1.70	1.24	2.34	0.0010
204933_s_at	BF433902	Hs.81791	TNFRSF11B	0.51	1.67	1.27	2.20	0.0003
204953_at	NM_014841	Hs.368046	SNAP91	0.59	1.81	1.31	2.49	0.0003
205046_at	NM_001813	Hs.75573	CENPE	0.62	1.86	1.28	2.70	0.0012
205189_s_at	NM_000136	Hs.494529	FANCC	0.53	1.70	1.21	2.40	0.0023
205217_at	NM_004085	Hs.447877	TIMM8A	0.64	1.90	1.26	2.85	0.0020
205386_s_at	NM_002392	Hs.567303	MDM2	0.49	1.63	1.19	2.23	0.0025
205433_at	NM_000055	Hs.420483	BCHE	0.58	1.79	1.23	2.62	0.0024
205481_at	NM_000674	Hs.77867	ADORA1	0.49	1.63	1.20	2.23	0.0020
205491_s_at	NM_024009	Hs.522561	GJB3	0.46	1.58	1.18	2.11	0.0021
205501_at	AI143879	Hs.348762	---	0.40	1.49	1.13	1.97	0.0043
205825_at	NM_000439	Hs.78977	PCSK1	0.59	1.81	1.24	2.65	0.0023
205893_at	NM_014932	Hs.478289	NLGN1	0.40	1.49	1.13	1.97	0.0048
205938_at	NM_014906	Hs.245044	PPM1E	0.52	1.68	1.22	2.31	0.0013
205946_at	NM_003382	Hs.490817	VIPR2	0.50	1.65	1.17	2.33	0.0043
206043_s_at	NM_014861	Hs.6168	ATP2C2	-0.55	0.57	0.39	0.84	0.0044
206096_at	AI809774	Hs.288658	ZNF35	0.55	1.73	1.20	2.49	0.0034

Table 3 (Continued)

Probe Set ID	Representative Public ID	UniGene ID	Gene Symbol	Coeffi-cients	HR	HRL	HRH	p value
206228_at	AW769732	Hs.155644	PAX2	0.50	1.65	1.27	2.15	0.0002
206232_s_at	NM_004775	Hs.591063	B4GALT6	0.44	1.56	1.17	2.07	0.0021
206401_s_at	J03778	Hs.101174	MAPT	0.39	1.48	1.13	1.94	0.0049
206426_at	NM_005511	Hs.154069	MLANA	0.63	1.87	1.26	2.77	0.0018
206496_at	NM_006894	Hs.445350	FMO3	0.53	1.70	1.22	2.37	0.0018
206505_at	NM_021139	Hs.285887	UGT2B4	0.61	1.84	1.26	2.69	0.0017
206524_at	NM_003181	Hs.389457	T	0.78	2.18	1.35	3.53	0.0015
206552_s_at	NM_003182	Hs.2563	TAC1	0.97	2.63	1.53	4.53	0.0005
206619_at	NM_014420	Hs.159311	DKK4	0.54	1.72	1.20	2.45	0.0029
206622_at	NM_007117	Hs.182231	TRH	0.53	1.70	1.23	2.37	0.0015
206661_at	NM_025104	Hs.369998	DBF4B	0.55	1.73	1.27	2.36	0.0005
206672_at	NM_000486	Hs.130730	AQP2	0.37	1.45	1.13	1.84	0.0030
206678_at	NM_000806	Hs.175934	GABRA1	0.39	1.48	1.16	1.89	0.0014
206799_at	NM_006551	Hs.204096	SCGB1D2	0.41	1.51	1.15	1.99	0.0032
206835_at	NM_003154	Hs.250959	STATH	0.46	1.59	1.16	2.18	0.0042
206940_s_at	NM_006237	Hs.493062	POU4F1	0.54	1.72	1.23	2.40	0.0017
206984_s_at	NM_002930	Hs.464985	RIT2	0.47	1.59	1.16	2.20	0.0045
207003_at	NM_002098	Hs.778	GUCA2A	0.62	1.85	1.23	2.79	0.0032
207028_at	NM_006316	Hs.651453	MYCNOS	0.48	1.61	1.19	2.18	0.0020
207208_at	NM_014469	Hs.121605	HNRNPG-T	0.51	1.66	1.23	2.26	0.0010
207219_at	NM_023070	Hs.133034	ZNF643	0.60	1.82	1.27	2.60	0.0011
207529_at	NM_021010	---	DEFA5	0.65	1.91	1.38	2.64	0.0001
207597_at	NM_014237	Hs.127930	ADAM18	0.63	1.87	1.36	2.58	0.0001
207814_at	NM_001926	Hs.711	DEFA6	0.61	1.85	1.21	2.81	0.0041
207843_x_at	NM_001914	Hs.465413	CYB5A	-0.55	0.58	0.39	0.84	0.0047
207878_at	NM_015848	---	KRT76	0.41	1.51	1.17	1.95	0.0017
207937_x_at	NM_023110	Hs.264887	FGFR1	0.43	1.54	1.14	2.08	0.0045
208157_at	NM_009586	Hs.146186	SIM2	0.45	1.56	1.19	2.05	0.0013
208233_at	NM_013317	Hs.468675	PDPN	0.54	1.72	1.18	2.49	0.0043
208292_at	NM_014482	Hs.158317	BMP10	0.44	1.55	1.17	2.05	0.0025
208314_at	NM_006583	Hs.352262	RRH	0.56	1.75	1.19	2.58	0.0044
208368_s_at	NM_000059	Hs.34012	BRCA2	0.62	1.86	1.26	2.73	0.0018
208399_s_at	NM_000114	Hs.1408	EDN3	0.48	1.61	1.18	2.20	0.0028
208511_at	NM_021000	Hs.647156	PTTG3	0.49	1.63	1.17	2.29	0.0043
208684_at	U24105	Hs.162121	COPA	-0.52	0.59	0.41	0.85	0.0041
208992_s_at	BC000627	Hs.463059	STAT3	-0.67	0.51	0.34	0.77	0.0012
209434_s_at	U00238	---	PPAT	0.43	1.54	1.15	2.06	0.0033
209839_at	AL136712	Hs.584880	DNM3	0.54	1.72	1.18	2.50	0.0049
209859_at	AF220036	Hs.368928	TRIM9	0.45	1.57	1.16	2.12	0.0032
210016_at	BF223003	Hs.434418	MYT1L	0.60	1.82	1.31	2.52	0.0003
210247_at	AW139618	Hs.445503	SYN2	0.64	1.89	1.30	2.75	0.0008
210302_s_at	AF262032	Hs.584852	MAB21L2	0.59	1.81	1.34	2.44	0.0001
210315_at	AF077737	Hs.445503	SYN2	0.66	1.94	1.31	2.87	0.0009
210455_at	AF050198	Hs.419800	C10orf28	0.57	1.76	1.24	2.50	0.0015
210758_at	AF098482	Hs.493516	PSIP1	0.42	1.52	1.17	1.97	0.0015
210918_at	AF130075	---	---	0.46	1.59	1.24	2.04	0.0003
211204_at	L34035	Hs.21160	ME1	0.54	1.72	1.26	2.33	0.0006
211264_at	M81882	Hs.231829	GAD2	0.53	1.71	1.19	2.44	0.0034
211341_at	L20433	Hs.493062	POU4F1	0.57	1.77	1.21	2.58	0.0031
211516_at	M96651	Hs.68876	IL5RA	0.60	1.82	1.26	2.62	0.0013

Table 3 (Continued)

Probe Set ID	Representative Public ID	UniGene ID	Gene Symbol	Coeffi-cients	HR	HRL	HRH	p value
211772_x_at	BC006114	Hs.89605	CHRNA3	0.52	1.69	1.22	2.33	0.0014
212359_s_at	W89120	Hs.65135	KIAA0913	-0.53	0.59	0.42	0.82	0.0019
212528_at	AI348009	Hs.633087	---	-0.79	0.45	0.29	0.70	0.0004
212531_at	NM_005564	Hs.204238	LCN2	-0.57	0.56	0.38	0.84	0.0049
213197_at	AB006627	Hs.495897	ASTN1	0.66	1.93	1.36	2.74	0.0002
213260_at	AU145890	Hs.599993	---	0.51	1.67	1.18	2.35	0.0036
213458_at	AB023191	---	KIAA0974	0.43	1.54	1.19	1.99	0.0010
213482_at	BF593175	Hs.476284	DOCK3	0.53	1.70	1.19	2.42	0.0032
213603_s_at	BE138888	Hs.517601	RAC2	-0.62	0.54	0.37	0.79	0.0017
213917_at	BE465829	Hs.469728	PAX8	0.52	1.69	1.21	2.36	0.0022
214457_at	NM_006735	Hs.592177	HOXA2	0.72	2.06	1.40	3.03	0.0002
214608_s_at	AJ000098	Hs.491997	EYA1	0.55	1.73	1.24	2.42	0.0013
214665_s_at	AK000095	Hs.406234	CHP	-0.52	0.59	0.43	0.82	0.0014
214822_at	AF131833	Hs.495918	FAM5B	0.54	1.72	1.23	2.41	0.0017
215102_at	AK026768	Hs.633705	DPY19L1P1	0.49	1.64	1.22	2.20	0.0011
215180_at	AL109703	Hs.651358	---	0.43	1.54	1.16	2.06	0.0029
215289_at	BE892698	---	ZNF749	0.46	1.58	1.19	2.09	0.0017
215356_at	AK023134	Hs.646351	ECAT8	0.46	1.58	1.15	2.17	0.0048
215476_at	AF052103	Hs.159157	---	0.49	1.63	1.21	2.21	0.0016
215705_at	BC000750	---	PPP5C	0.52	1.68	1.22	2.32	0.0016
215715_at	BC000563	Hs.78036	SLC6A2	0.75	2.12	1.37	3.29	0.0008
215850_s_at	AK022209	Hs.651219	NDUFA5	0.48	1.62	1.18	2.23	0.0030
215944_at	U80773	---	---	0.49	1.64	1.20	2.24	0.0019
215953_at	AL050020	Hs.127384	DKFZP564C196	0.47	1.59	1.16	2.19	0.0038
215973_at	AF036973	---	HCG4P6	0.55	1.74	1.30	2.32	0.0002
216050_at	AK024584	Hs.406847	---	0.44	1.55	1.15	2.08	0.0035
216066_at	AK024328	Hs.429294	ABCA1	0.50	1.65	1.22	2.22	0.0010
216240_at	M34428	Hs.133107	PVT1	0.46	1.58	1.15	2.18	0.0046
216881_x_at	X07882	Hs.528651	PRB4	0.41	1.51	1.14	1.99	0.0042
216989_at	L13779	Hs.121494	SPAM1	0.46	1.58	1.15	2.16	0.0044
217004_s_at	X13230	Hs.387262	MCF2	0.39	1.48	1.14	1.91	0.0032
217253_at	L37198	Hs.632861	---	0.51	1.66	1.17	2.35	0.0041
217995_at	NM_021199	Hs.511251	SQRDL	-0.82	0.44	0.29	0.66	0.0001
218768_at	NM_020401	Hs.524574	NUP107	0.63	1.88	1.31	2.70	0.0006
218881_s_at	NM_024530	Hs.220971	FOSL2	-0.52	0.60	0.42	0.85	0.0044
218980_at	NM_025135	Hs.436636	FHOD3	0.63	1.88	1.29	2.74	0.0011
219000_s_at	NM_024094	Hs.315167	DCC1	1.06	2.90	1.89	4.44	0.0000
219171_s_at	NM_007345	Hs.189826	ZNF236	0.56	1.76	1.20	2.56	0.0035
219182_at	NM_024533	Hs.156784	FLJ22167	0.48	1.62	1.18	2.22	0.0027
219425_at	NM_014351	Hs.189810	SULT4A1	0.74	2.11	1.41	3.14	0.0003
219520_s_at	NM_018458	Hs.527524	WWC3	-0.49	0.61	0.44	0.84	0.0029
219537_x_at	NM_016941	Hs.127792	DLL3	0.55	1.73	1.23	2.44	0.0018
219617_at	NM_024766	Hs.468349	C2orf34	0.53	1.70	1.19	2.43	0.0035
219643_at	NM_018557	Hs.470117	LRP1B	0.55	1.73	1.30	2.30	0.0001
219704_at	NM_015982	Hs.567494	YBX2	0.75	2.12	1.42	3.16	0.0002
219882_at	NM_024686	Hs.445826	TTLL7	0.51	1.66	1.18	2.35	0.0038
219937_at	NM_013381	Hs.199814	TRHDE	0.54	1.71	1.23	2.38	0.0015
219955_at	NM_019079	Hs.562195	L1TD1	0.60	1.82	1.25	2.65	0.0018
220029_at	NM_017770	Hs.408557	ELOVL2	0.52	1.68	1.18	2.40	0.0038

Table 3 (Continued)

Probe Set ID	Representative Public ID	UniGene ID	Gene Symbol	Coeffi-cients	HR	HRL	HRH	p value
220076_at	NM_019847	Hs.156727	ANKH	0.77	2.17	1.53	3.07	0.0000
220294_at	NM_014379	Hs.13285	KCNV1	0.45	1.56	1.16	2.11	0.0036
220366_at	NM_022142	Hs.104894	ELSPBP1	0.53	1.69	1.19	2.41	0.0034
220394_at	NM_019851	Hs.199905	FGF20	0.61	1.84	1.30	2.60	0.0006
220397_at	NM_020128	Hs.591036	MDM1	0.41	1.51	1.17	1.95	0.0015
220541_at	NM_021801	Hs.204732	MMP26	0.50	1.64	1.24	2.18	0.0006
220653_at	NM_015363	---	ZIM2	0.60	1.83	1.33	2.53	0.0002
220700_at	NM_018543	Hs.188495	WDR37	0.59	1.80	1.22	2.66	0.0029
220703_at	NM_018470	Hs.644603	C10orf110	0.59	1.80	1.26	2.58	0.0012
220771_at	NM_016181	Hs.633593	LOC51152	0.60	1.81	1.23	2.67	0.0025
220817_at	NM_016179	Hs.262960	TRPC4	0.47	1.60	1.19	2.14	0.0019
220834_at	NM_017716	Hs.272789	MS4A12	0.52	1.68	1.27	2.22	0.0003
220847_x_at	NM_013359	Hs.631598	ZNF221	0.50	1.65	1.19	2.28	0.0025
220852_at	NM_014099	Hs.621386	PRO1768	0.48	1.62	1.19	2.20	0.0022
220970_s_at	NM_030977	Hs.406714	KRTAP2-4 / LOC644350	0.49	1.64	1.16	2.31	0.0050
220981_x_at	NM_022053	Hs.648337	NXF2	0.45	1.56	1.19	2.05	0.0014
220993_s_at	NM_030784	Hs.632612	GPR63	0.38	1.46	1.13	1.88	0.0041
221018_s_at	NM_031278	Hs.333132	TDRD1	0.81	2.25	1.51	3.37	0.0001
221077_at	NM_018076	Hs.127530	ARMC4	0.56	1.76	1.25	2.47	0.0013
221137_at	AF118071	---	---	0.46	1.59	1.15	2.20	0.0049
221168_at	NM_021620	Hs.287386	PRDM13	0.68	1.96	1.33	2.91	0.0007
221258_s_at	NM_031217	Hs.301052	KIF18A	0.62	1.86	1.34	2.58	0.0002
221319_at	NM_019120	Hs.287793	PCDHB8	0.40	1.49	1.14	1.96	0.0041
221393_at	NM_014627	---	TAAR3	0.50	1.64	1.17	2.31	0.0043
221591_s_at	BC005004	Hs.592116	FAM64A	0.72	2.05	1.38	3.05	0.0004
221609_s_at	AY009401	Hs.29764	WNT6	0.40	1.50	1.15	1.95	0.0028
221718_s_at	M90360	Hs.459211	AKAP13	-0.64	0.53	0.36	0.78	0.0013
221950_at	AI478455	Hs.202095	EMX2	0.67	1.96	1.41	2.72	0.0001

Table 4. Features of 13 probe sets in the gene signature

Probe Set	Gene Symbol	Gene Title	Entrez Gene ID	Coef.*	Rank of expression [n=19619 (%)]	Rank of variation [n=19619 (%)]	Rank of significant [n=172 (%)]
201243_s_at	ATP1B1	ATPase, Na+/K+ transporting, beta 1 polypeptide	481	-0.54	517 (2.6)	2224 (11.3)	111 (64.5)
203147_s_at	TRIM14	Tripartite motif-containing 14	8518	-0.56	3532 (18.0)	9499 (48.4)	112 (65.1)
221591_s_at	FAM64A	Family with sequence similarity 64, member A	7372	0.72	6171 (31.5)	6108 (31.1)	29 (16.9)
218881_s_at	FOSL2	FOS-like antigen 2	10614	-0.52	6526 (33.3)	12445 (63.4)	155 (90.1)
202814_s_at	HEXIM1	Hexamethylene bis-acetamide inducible 1	11075	0.59	7415 (37.8)	9026 (46.0)	161 (93.6)
204179_at	MB	Myoglobin	9830	0.47	7703 (39.3)	7942 (40.5)	156 (90.7)
204584_at	L1CAM	L1 cell adhesion molecule	4151	0.56	9327 (47.5)	3329 (17.0)	17 (9.9)
202707_at	UMPS	Uridine monophosphate synthetase	3897	0.60	12311 (62.8)	18737 (95.5)	101 (58.7)
208399_s_at	EDN3	Endothelin 3	4193	0.48	16344 (83.3)	8234 (42.0)	110 (64.0)
203001_s_at	STMN2	Stathmin-like 2	2315	0.55	16948 (86.4)	5690 (29.0)	109 (63.4)
202490_at	IKBKAP	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	23040	0.42	18769 (95.7)	10412 (53.1)	84 (48.8)
205386_s_at	MDM2	Mdm2, transformed 3T3 cell double minute 2	7776	0.49	19251 (98.1)	14275 (72.8)	104 (60.5)
219171_s_at	ZNF236	Zinc finger protein 236	54478	0.56	19383 (98.8)	17046 (86.9)	132 (76.7)

*Coefficient of the Cox model

Table 4b. Features of the 2 probe sets that were removed from the gene signature

Probe Set	Gene Symbol	Gene Title	Entrez Gene ID	Coef.*	Rank of expression [n=19619 (%)]	Rank of variation [n=19619 (%)]	Rank of significant [n=172 (%)]
210016_at	MYT1L	Myelin transcription factor 1-like	1908	0.60	17902 (91.2)	18637 (95.0)	27 (15.7)
206426_at	MLANA	Melan-A	2355	0.63	19159 (97.7)	17172 (87.5)	81 (47.1)

*Coefficient of the Cox model

Table 5. Demographic distributions of patients in validation sets

Clinical Factors	DCC, All n=360 (%)	DCC, UM n=177 (%)	DCC, HLM n=79 (%)	DCC, MSK n=104 (%)	Duke n=89 (%)	UM-SQ n=129 (%)
Pathology Type						
Adeno	360 (100)	177 (100)	79 (100)	104 (100)	43 (48)	0
Non-Adeno	0 (0)	0 (0)	0 (0)	0 (0)	46 (52)	129 (100)
Disease stage						
I	220 (61)	116 (66)	41 (52)	63 (61)	67 (75)	73 (57)
II	69 (19)	29 (16)	20 (25)	20 (19)	18 (20)	33 (25)
III	69 (19)	32 (18)	16 (20)	21 (20)	3 (3)	23 (18)
IV	0 (0)	0 (0)	0 (0)	0 (0)	1 (2)	0 (0)
Unknown	2 (1)	0 (0)	2 (3)	0 (0)	0 (0)	0 (0)
Adjuvant chemotherapy						
No	210 (58)	76 (43)	61 (77)	73 (70)	89 (100)	NS
Yes	64 (18)	17 (10)	16 (20)	31 (30)	0 (0)	NS
Unknown	86 (24)	84 (47)	2 (3)	0 (0)	0 (0)	NS

Table 5 (Continued)

Adjuvant radiotherapy						
No	209 (58)	76 (43)	57 (72)	76 (73)	89 (100)	NS
Yes	64 (18)	17 (10)	19 (24)	28 (27)	0 (0)	NS
Unknown	87 (24)	84 (47)	3 (4)	0 (0)	0 (0)	NS
Age (year)						
<65	163 (45)	87 (49)	17 (34)	49 (47)	33 (37)	52 (40)
≥65	197 (55)	90 (51)	25 (66)	55 (53)	56 (63)	77 (60)
Gender						
Male	177 (49)	100 (56)	40 (51)	37 (36)	54 (61)	82 (64)
Female	183 (51)	77 (44)	39 (49)	67 (64)	35 (39)	47 (36)

DCC: Directors' Challenge Consortium; UM: University of Michigan; HLM: H. Lee Moffitt Cancer Center; MSK: Memorial Sloan-Kettering Cancer Center; NS: Not specified

Table 6. Adjuvant therapies in the Director's Challenge Consortium (DCC) Patients

Adjuvant Chemotherapy	**Adjuvant radiotherapy**			
All	**No**	**Yes**	**Unknown**	**Total**
No	190	20	0	210
Yes	19	44	1	64
Unknown	0	0	86	86
University of Michigan (UM)				
No	76	0	0	76
Yes	0	17	0	17
Unknown	0	0	84	84
H. Lee Moffitt (HLM)				
No	51	10	0	61
Yes	6	9	1	16
Unknown	0	0	2	2
Memorial Sloan-Kettering (MSK)				
No	63	10	0	73
Yes	13	18	0	31
Unknown	0	0	0	0

Table 7. Primers for qPCR validation

Gene	SEQ ID NO	Forward	Amplicon Length	Tm	SEQ ID NO	Reverse	Amplicon Length	Tm
FAM64A	173	TCCTCAACAGAGCCCCTC	18	62.6	186	TTCTGGAGATGCTGAATTCCC	21	62
MB	174	AGTTGGTGCTGAACGTCTG	19	62.2	187	GGTGACCCTTAAAGAGCCTG	20	62
EDN3	175	GTGTCTACTATTGCCACCTGG	21	62	188	GCTTCCTCTGTAGTTGGACAG	21	61.9
ZNF236	176	ACGTAGACCAGTTTGAAGAGC	21	62	189	GGCTGCTGATGGTACGTATC	20	62
FOSL2	177	ACGCCGAGTCCTACTCCA	18	59	190	TGAGCCAGGCATATCTACCC	20	60
L1CAM	178	GCCCCCCAGGTTCAGTACCGCG	22	59	191	CTGACAATCTGCTCCTGCCA	20	60
TRIM14	179	CCTCAGCATCCTGAATACATCA	22	60	192	TGTGTCAGGTTCCTGTGCTG	20	60
SEMN2	180	GAACCTCGCAACATCAACATC	21	61.7	193	GGCTTCAAGATCAGCTCAAAAG	22	61.8
UMPS	181	ACACAGTGAAAAAGCAGTATGAAG	24	61.9	194	TTTCACAACTCCTGAGCCTG	20	62.1
ATP1B1	182	CCAAATGTCCTTCCCGTTCAG	21	63	195	TGCCCAGTCCAAAATACTCC	20	62
HEXIM1	183	TCGAGGACTCTACTAGCCATG	21	62.1	196	GCTCTTCCTGGACAGCAG	18	61.7
IKBKAP	184	GGAAGCAAGCCCTCTGTGT	19	60	197	TCTGCTCAACCAGCTTTCCT	20	60
MDM2	185	CATGTCTGTACCTACTGATGGTG	23	62	198	GCAATGGCTTTGGTCTAACC	20	61.5

Table 8. Risk group based on 15-gene signature in stage I patients

	n	HR	95% CI	p value
BR.10 Observation arm	34	13.3	2.9 – 62.1	<0.0001
DCC No adjuvant therapy	141	3.3	1.5-7.4	0.002
UM	57	1.9	0.6-6.1	0.28
HLM	37	2.5	0.9 – 6.9	0.07
MSK	47	NA	NA	0.05
Duke	67	1.06	0.5-2.2	0.88
UM-SQ	73	1.4	0.6-3.1	0.44

n: number of patients; HR: hazard ratio; CI: confidence interval
* HR and CI cannot be calculated as no death occurred in the good prognosis group, p value the score test.

Table 9 Probe set target sequences of the 13-gene signature

SEQ ID NO:	Probe set ID	Target sequence
35	205386_S_AT	tttcccctagttgacctgtctataagagaattatatatttctaactatataaccctaggaatttagacaacctgaaatttattcacatatatcaaagtgagaaaatgcctcaattcacatagatttcttctctttagtataattgacctactttggtagtggaatagtgaatacttactataatttgacttgaatatgtagctcatcctttacaccaactcctaattttaaataatttctactctgtcttaaatgagaagtacttggttttttttttcttaaatatgtatatgacatttaaatgtaacttattatttttttgagaccgagtcttgctctgttacccaggctggagtgcagtgggtgatcttggctcactgcaagctctgccctccccgggttcgcaccattctcctgcctcagcctcccaattagcttggcctacagtcatctgcc
78	208399_S_AT	ccgagccgagcttactgtgagtgtggagatgttatcccaccatgtaaagtcgcctgcgcaggggagggctgcccatctccccaacccagtcacagagagataggaaacggcatttgagtgggtgtccagggccccgtagagagacatttaagatggtgtatgacagagcattggccttgaccaaatgttaaatcctctgtgtgtatttcataagttattacaggtataaaagtgatgacctatcatgaggaaatgaaagtggctgatttgctggtaggattttgtacagtttagagaagcgattatttattgtgaaactgttctccactccaactcctttatgtggatctgttcaaagtagtcactgtatatacgtatagagaggtagataggtaggtagattttaaattgcattctgaatacaaactcatactccttagagcttgaattacatttttaaaatgcatatgtgctgtttggcaccgtggcaagatggtatcagagagaaacccatcaattgctcaaatactc
4	201243_S_AT	ggtgatgggttgtgttatgcttgtattgaatgctgtcttgacatctcttgccttgtcctccggtatgttctaaagctgtgtctgagatctggatctgcccatcactttggcctagggacagggctaattaatttgctttatacattttcttttactttccttttttcctttctggaggcatcacatgctggtgctgtgtctttatgaatgttttaaccattttcatggtggaagaattttatatttatgcagttgtacaatttattttttttctgcaagaaaaagtgtaatgtatgaaataaaccaaagtcacttgtttgaaaataaatctttattttgaactttataaaagcaatgcagtaccccatagactggtgttaaatgttgtctacagtgcaaaatccatgttctaacatatgtaataattgccaggagtacagtgctcttgttgatcttgtattcagtcaggttaaaa
22	204179_AT	tgttccggaaggacatggcctccaactacaaggagctgggcttccagggctaggcccctgccgctcccacccccacccatctgggccccgggttcaagagagagcggggtctgatctcgtgtagccatatagagtttgcttctgagtgtctgctttgtttagtagaggtgggcaggaggagctgaggggctggggctggggtgttgaagttggctttgcatgcccagcgatgcgcctccctgtgggatgtcatcaccctgggaaccgggagtgcccttggctcactgtgttctgcatggtttggatctgaattaattgtcctttcttctaaatcccaaccgaacttcttccaacctccaaactggctgtaaccccaaatccaagccattaactacacctgacagtagcaattgtctgattaatcactggccccttgaagacagcagaatgtccctttgcaatgaggaggagatctgggctgggcgggccagctggggaagcatttgactatctggaacttgtgtgtgcctcctcaggtatggca
169	221591_S_AT	cacatctggacccatcagtgactgcctgccatagcctgagagtgtcttggggagaccttgcagaggggagaattgttccttctgctttcctaggggactcttgagcttagaaactcatcgtacacttgaccttgagccttctatttgcctcatctataacatgaagtgctagcatcagatattttgagagctcttagctctgtacccgggtgcctggtttttggggagtcatccgcagagtcactcacccactgtgtttctggtgccaaggctcttgagggccccactctcatccctcctttcccctaccagggactcggaggaaggcataggagatatttccaggcttacgaccctgggctcacgggtacctatttatatgctcagtgcagagcactgtggatgtgccaggaggggtagccctgttcaagagcaatttctgccctttgtaaattatttaagaaacctgctttgtcattttattagaaagaaaccagcgtgtgacttttcctagataacactgctttc
15	203147_S_AT	accaatcacgcctacagtgctttgaaggtttcctctcctaggctagtttcaaacaggccctaaacaagtctgctgctgccctctcatcagacctccgcaccctcaccccaccatcactta

Table 9 (Continued)

SEQ ID NO:	Probe set ID	Target sequence
		nactactttaatccagttccttcaaagtgataccccacaggtaagccctcagcatcctg aatacatcatccgcagcctgggaaccttctccctcgtacagcacaggaacctgacaca tagtaggcacacagtaaacgtttgtgaatgaatgggagtcatccagtcctgactcttctgt ctcttgaggtcccttgaatcttccgcttcctccccaccgatttcagcgtgtccacatcacag ctccctccagaagctgcaagagcttcttagcagttcctggtctgaaccctctcccagtcct catcttccaccctaaaactagagtgatcttcctaaaacttcacttaacccctcagctatga aaaggcttccaggagtttccatgaa
130	218881_ S_AT	aggtcacagtatcctcgtttgaaagataattaagatcccccgtggagaaagcagtgac acattcacacagctgttccctcgcatgttatttcatgaacatgacctgttttcgtgcactaga cacacagagtggaacagccgtatgcttaaagtacatgggccagtgggactggaagtg acctgtacaagtgatgcagaaaggagggtttcaaagaaaaaggattttgtttaaaatac tttaaaaatgttatttcctgcatcccttggctgtgatgcccctctcccgatttcccaggggctc tgggagggaccttctaagaagattgggcagttgggtttctggcttgagatgaatccaag cagcagaatgagccaggagtagcaggagatgggcaaagaaaactggggtgcactc agctctcacaggggtaatca
11	202490_ AT	gaggatggcacaagcgattcacgtaggatctgcccctgtgaccaaaacacctcccatt gggccccacttccaacactggtgatcacatttcaacatgaggtttagggaaacaaatgc ctaaactacagcactgtacataaactaacaggaaatgctgcttttgatcctcaaagaagt gatatagccaaaattgtaatttaagaagcctttgtcagtatagcaagatgttaactataga atcaatctaggagtattcactgtaaaattcaacttttctgtatgtttgaacattttcacaatctc ataggagtttttaaaaagaagagaaagaagatatactttgctttggagaaatctacttttttg acttacatgggtttgctgtaattaagtgcccaatattgaaaggctgcaagtactttgtaatc actctttggcatgggtaaataagcatggtaacttatattgaaatatagtgctcttgctttggat aactgtaaagggacccatgctgatagactggaaa
12	202707_ AT	aagttcattcttaagcttgctttttttgagactggtgtttgttagacagccacagtcctgtctgg gttagggtcttccacatttgaggatccttcctatctctccatgggactagactgctttgttattc tatttatttttttaatttttttcgagacaggatctcactctgttgcccaggatggagtgcagtggt gagatcacggctcattgcagcctcgacctcccaggtgatcctcccacctcagcttccag attagctggtgctataggcatgcaccaccacgtccatctaaatttctttattatttgtagagat gaggtcttgccatgttacccaggctggtctcaactcctgggctcaagcgatcctcctgcct cagtctctcaaagtgctgggattacaggtgtgagccactgtgcccagcctaattgcagta agacaa
14	203001_ S_AT	acctcgcaacatcaacatctatacttacgatgatatggaagtgaagcaaatcaacaaa cgtgcctctggccaggcttttgagctgatcttgaagccaccatctcctatctcagaagccc cacgaactttagcttctccaaagaagaaagacctgtccctggaggagatccagaaga aactggaggctgcaggggaaagaagaaagtctcaggaggcccaggtgctgaaaca attggcagagaagagggaacacgagcgagaagtccttcagaaggctttggaggaga acaacaacttcagcaagatggcggaggaaaagctgatcctgaaaatggaacaaatt aaggaaaaccgtgaggctaatctagctgctattattgaacgtctgcaggaaaaggaga ggcatgctgcggaggtgcgcaggaacaaggaactccaggttgaactgtctggctgaa gcaagggagggtctggcacgcc
13	202814_ S_AT	tgcctctcgcgcatggaggacgagaacaaccggctgcggctggagagcaagcggct gggtggcgacgacgcgcgtgtgcgggagctggagctggagctggaccggctgcgcg ccgagaacctccagctgctgaccgagaacgaactgcaccggcagcaggagcgagc gccgctttccaagtttggagactagactgaaacttttttgggggaggggcaaagggga ctttttacagtgatggaatgtaacattatatacatgtgtatataagacagtggacctttttatg acacataatcagaagagaaatccccctggctttggttggtttcgtaaatttagctatatgta gcttgcgtgctttctcctgttcttttaattatgtgaaactgaagagttgcttttcttgttttccttttta

Table 9 (Continued)

SEQ ID NO:	Probe set ID	Target sequence
		gaagtttttttccttaatgtgaaagtaatttgaccaagttataatgcatttttgtttttaacaaat ccccctccttaaacggagctataaggtggccaaatctga
133	219171_S_AT	cttttgttcttgctgggttatttattttgattttagcattaaatgtcatctcaggatatctctaaaag gggttgtttaattcctaattgtatagaaagctagtttggtgaattgtattggttaattgactgttt aaggccttaacaggtgaatctagagcctacttttattttggttaaagaaaaagaaaatatc aataattcaattttgtgtcttttctcaatttattagcaaacacaagacattttatgtattatttcga tttacttcctaattataaaagctgctttttgcagaacattccttgaaaatataaggttttgaaa agacataattttacttgaatctttgtggggtacaggttgatctttatattttactggttgttttaaa aattctagaaaagagatttctaggcctcatgtataaccagggttttgaggataaagaact gtatttttagaactatctcatcatagcatatctgctttggaataactat
26	204584_AT	cctccctatcgtctgaacagttgtcttcctcagcctcctcccgcccccaccttgggaatgta aatacaccgtgactttgaaagtttgtacccctgtccttccctttacgccactagtgtgtaggc agatgtctgagtccctaggtggtttctaggattgatagcaattagctttgatgaacccatcc caggaaaaataaaaacagacaaaaaaaaaggaaagattggttctcccagcactgct cagcagccacagcctccctgtatgcctgtgcttggtctactgataagccctctacaaaa

Table 10. Coefficient of individual genes in 13-gene signature: Principal Component values

Gene	Gene Symbol	Probe set	pc1	pc2	pc3	pc4
1	ATP1B1	201243_s_at	-0.189	-0.423	0.229	0.059
2	IKBKAP	202490_at	0.364	0.070	-0.357	-0.120
3	UMPS	202707_at	0.353	-0.009	0.136	0.011
4	HEXIM1	202814_s_at	-0.108	0.504	0.265	0.279
5	STMN2	203001_s_at	0.326	0.044	-0.100	-0.122
6	TRIM14	203147_s_at	-0.148	0.212	0.132	-0.368
7	MB	204179_at	0.197	0.028	0.548	-0.161
8	L1CAM	204584_at	0.042	0.510	0.077	0.276
9	MDM2	205386_s_at	0.180	0.081	0.325	-0.500
10	EDN3	208399_s_at	0.413	0.042	-0.188	-0.260
11	FOSL2	218881_s_at	0.036	-0.209	-0.225	0.190
12	ZNF236	219171_s_at	0.188	-0.313	0.297	0.332
13	FAM64A	221591_s_at	0.283	0.216	-0.174	0.324
Eigenvalues of principal components			**3.33**	**1.82**	**1.37**	**1.32**
Weight of each PC for risk score			**0.557**	**0.328**	**0.430**	**0.335**

Risk score = 0.557*PC1 + 0.328*PC2 + 0.43*PC3 +0.335*PC4

where

PC1 = Sum $[pc1*(\text{expression data})]_{\text{Gene 1-13}}$
PC2= Sum $[pc2*(\text{expression data})]_{\text{Gene 1-13}}$
PC3 = Sum $[pc3*(\text{expression data})]_{\text{Gene 1-13}}$
PC4 = Sum $[pc4*(\text{expression data})]_{\text{Gene 1-13}}$

Patients classified as high risk or lower risk according to risk score ≥ -0.1 or < -0.1.

Table 11 Probe set target sequences for 172 genes

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
1	200878_at	EPAS1	cactttgcaactccctgggtaagagggacgacacctctggtttttcaataccaattac atggaacttttctgtaatgggtacnaatgaagaagtttctaaaaacacacacaaagc acattgggccaactatttagtaagcccggatagacttattgccaaaaacaaaaaata gctttcaaaagaaatttaagttctatgagaaattccttagtcatggtgttgcgtaaatc atattttagctgcacggcattaccccacacagggtggcagaacttgaagggttactg acgtgtaaatgctggtatttgatttcctgtgtgtgttgccctggcattaagggcatttta cccttgcagttttactaaaacactgaaaaatattccaagcttcatattaaccctacctg tcaacgtaacgat
2	201228_s_at	ARIH2	cctacccacctcaaaatgtctgtactgcaagagggccctgggcctctgctttccatatt cacgtttggccagagttgtagtcccaaagaagagcatgggtggcagatggtaggga attgaactggcctgtgcaatgggcatggagcacaaggggtcacagcatgcctcctgc cttaccgtggcagtacggagacagtccagaacatggtcttcttgccacggggtgttgt tgtctctggtggtgctgcatgtctgtggctcacctttattcttgaaactgaggtttacct ggatctggctactgaggctagagcccacagcagaatggggttgggcctgtggccccc caaactagggggtgtgggttcatcacagtgttgccttttgtctcctaaagatagggat ctacttttgaagggaattgttcctcccaaata
3	201242_s_at	ATP1B1	agagctgatcacaagcacaaatctttcccactagccatttaataagttaaaaaaaga tacaaaaacaaaaacctactagtcttgaacaaactgtcatacgtatgggacctacac ttaatctatatgctttacactagctttctgcatttaataggttagaa
4	201243_s_at	ATP1B1	ggtgatgggttgtgttatgcttgtattgaatgctgtcttgacatctcttgccttgtcctcc ggtatgttctaaagctgtgtctgagatctggatctgcccatcactttggcctagggaca gggctaattaatttgctttatacattttcttttacttttcctttttttcctttctggaggcatca catgctggtgctgtgtctttatgaatgttttaaccattttcatggtggaagaattttatat ttatgcagttgtacaattttatttttttctgcaagaaaaagtgtaatgtatgaaataaa ccaaagtcacttgtttgaaaataaatctttatttttgaacttttataaaagcaatgcagta ccccatagactggtgttaaatgttgtctacagtgcaaaatccatgttctaacatatgta ataattgccaggagtacagtgctcttgttgatcttgtattcagtcaggttaaaa
5	201301_s_at	ANXA4	ggtgaaatttctaactgttctctgttcccggaaccgaaatcacctgttgcatgtgtttg atgaatacaaaaggatatcacagaaggatattgaacagagtattaaatctgaaaca tctggtagctttgaagatgctctgctggctatagtaaagtgcatgaggaacaaatctg catattttgctgaaaagctctataaatcgatgaagggcttgggcaccgatgataaca ccctcatcagagtgatggtttctcgagcagaaattgacatgttggatatccgggcaca cttcaagagactctatggaaagtctctgtactcgttcatcaagggtgacacatctgga gactacaggaaagtactgcttgttctctgtggaggagatgattaaaataaaaatccc agaaggacaggaggattctcaacactttgaatttttttaacttcatttttctacactgct attatcattatctc

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
6	201502_s_at	NFKBIA	ccaactacaatggccacacgtgtctacacttagcctctatccatggctacctgggcat cgtggagcttttggtgtccttgggtgctgatgtcaatgctcaggagccctgtaatggcc ggactgcccttcacctcgcagtggacctgcaaaatcctgacctggtgtcactcctgtt gaagtgtggggctgatgtcaacagagttacctaccagggctattctccctaccagctc acctggggccgcccaagcacccggatacagcagcagctgggccagctgacactaga aaaccttcagatgctgccagagagtgaggatgaggagagctatgacacagagtcag agttcacggagttcacagaggacgagctgccctatgatgactgtgtgtttggaggcc agcgtctgacgttatgag
7	202023_at	EFNA1	ccaccttcacctcggagggacggagaaagaagtggagacagtcctttcccaccattc ctgcctttaagccaaagaaacaagctgtgcaggcatggtcccttaaggcacagtggg agctgagctggaaggggccacgtggatgggcaaagcttgtcaaagatgccccctcc aggagagagccaggatgcccagatgaactgactgaaggaaaagcaagaaacagtt tcttgcttggaagccaggtacaggagaggcagcatgcttgggctgacccagcatctc ccagcaagacctcatctgtggagctgccacagagaagtttgtagccaggtactgcat tctctcccatcctggggcagcactccccagagctgtgccagcaggggggctgtgcca acctgttcttagagtgtagctgtaagggcagtgcccatgtgtacattctgcctagagtg tagcctaaagggcagggcccacgtgtatagtatctgta
8	202035_s_at	SFRP1	tcggccagcgagtacgactacgtgagcttccagtcggacatcggcccgtaccagagc gggcgcttctacaccaagccacctcagtgcgtggacatccccgcggacctgcggctg tgccacaacgtgggctacaagaagatggtgctgcccaacctgctggagcacgagac catggcggaggtgaagcagcaggccagcagctgggtgcccctgctcaacaagaact gccacgccggcacccaggtcttcctctgctcgctcttcgcgcccgtctgcctggaccg gcccatctacccgtgtcgctggctctgcgaggccgtgcgcgactcgtgcgagccggtc atgcagttcttcggcttctactggcccgagatgcttaagtgtgacaagttccccgagg gggacgtctgcatcgccatgacgccgcccaatgccaccgaagcctccaagccccaa ggcacaacggtgtgtcctccctgtgacaacgagttgaaatctgaggccatcattgaa catctctgt
9	202036_s_at	SFRP1	gacaaaccatttccaacagcaacacagccactaaaacacaaaaaggggattggg cggaaagtgagagccagcagcaaaaactacattttgcaacttgttggtgtggatcta ttggctgatctatgcctttcaactagaaaattctaatgattggcaagtcacgttgttttc aggtccagagtagtttctttctgtctgctttaaatggaaacagactcataccacactta caattaaggtcaagcccagaaagtgataagtgcagggaggaaaagtgcaagtcca ttatgtaatagtgacagcaaaggcccaggggagaggcattgccttctctgcccacag tctttccgtgtgattgtctttgaatctgaatcagccagtctcagatgccccaaagtttcg gttcctatgagcccggggcatgatctgatccccaagacatg
10	202037_s_at	SFRP1	taacacttggctcttggtacctgtgggttagcatcaagttctccccagggtagaattca atcagagctccagtttgcatttggatgtgtaaattacagtaatcccatttcccaaacct aaaatctgtttttctcatcagactctgagtaactggttgctgtgtcataacttcatagat gcaggaggctcaggtgatctgtttgaggagagcaccctaggcagcctgcagggaat aacatactggccgttctgacctgttgccagcagatacacaggacatggatgaaattc ccgtttcctctagtttcttcctgtagtactcctcttttagatcc

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
11	202490_at	IKBKAP	gaggatggcacaagcgattcacgtaggatctgcccctgtgaccaaaacacctcccat tgggccccacttccaacactggtgatcacatttcaacatgaggtttagggaaacaaa tgcctaaactacagcactgtacataaactaacaggaaatgctgcttttgatcctcaaa gaagtgatatagccaaaattgtaatttaagaagcctttgtcagtatagcaagatgtta actatagaatcaatctaggagtattcactgtaaaattcaacttttctgtatgtttgaac attttcacaatctcataggagtttttaaaaagaagagaaagaagatatactttgcttt ggagaaatctacttttttgacttacatgggtttgctgtaattaagtgcccaatattgaaa ggctgcaagtactttgtaatcactctttggcatgggtaaataagcatggtaacttatat tgaaatatagtgctcttgctttggataactgtaaagggacccatgctgatagactgga aa
12	202707_at	UMPS	aagttcattcttaagcttgctttttttgagactggtgtttgttagacagccacagtcctg tctgggttagggtcttccacatttgaggatccttcctatctctccatgggactagactgc tttgttattctatttatttttttaatttttttcgagacaggatctcactctgttgcccaggat ggagtgcagtggtgagatcacggctcattgcagcctcgacctcccaggtgatcctccc acctcagcttccagattagctggtgctataggcatgcaccaccacgtccatctaaatt tctttattatttgtagagatgaggtcttgccatgttacccaggctggtctcaactcctgg gctcaagcgatcctcctgcctcagtctctcaaagtgctgggattacaggtgtgagcca ctgtgcccagcctaattgcagtaagacaa
13	202814_s_at	HEXIM1	tgcctctcgcgcatggaggacgagaacaaccggctgcggctggagagcaagcggct gggtggcgacgacgcgcgtgtgcgggagctggagctggagctggaccggctgcgcg ccgagaacctccagctgctgaccgagaacgaactgcaccggcagcaggagcgagc gccgctttccaagtttggagactagactgaaacttttttgggggaggggcaaaggg gacttttttacagtgatggaatgtaacattatatacatgtgtatataagacagtggacc tttttatgacacataatcagaagagaaatccccctggctttggttggtttcgtaaattt agctatatgtagcttgcgtgctttctcctgttctttttaattatgtgaaactgaagagttg cttttcttgttttcctttttagaagttttttttccttaatgtgaaagtaatttgaccaagtta taatgcatttttgtttttaacaaatcccctccttaaacggagctataaggtggccaaat ctga
14	203001_s_at	STMN2	acctcgcaacatcaacatctatacttacgatgatatggaagtgaagcaaatcaacaa acgtgcctctggccaggcttttgagctgatcttgaagccaccatctcctatctcagaag ccccacgaactttagcttctccaaagaagaaagacctgtccctggaggagatccaga agaaactggaggctgcaggggaaagaagaaagtctcaggaggcccaggtgctgaa acaattggcagagaagagggaacacgagcgagaagtccttcagaaggctttggag gagaacaacaacttcagcaagatggcggaggaaaagctgatcctgaaaatggaac aaattaaggaaaaccgtgaggctaatctagctgctattattgaacgtctgcaggaaa aggagaggcatgctgcggaggtgcgcaggaacaaggaactccaggttgaactgtct ggctgaagcaagggagggtctggcacgcc
15	203147_s_at	TRIM14	accaatcacgcctacagtgctttgaaggtttcctctcctaggctagtttcaaacaggcc ctaaacaagtctgctgctgccctctcatcagacctccgcaccctcaccccaccatcac ttanactactttaatccagttccttcaaagtgatacccccacaggtaagccctcagca tcctgaatacatcatccgcagcctgggaaccttctccctcgtacagcacaggaacctg acacatagtaggcacacagtaaacgtttgtgaatgaatgggagtcatccagtcctga ctcttctgtctcttgaggtcccttgaatcttccgcttcctccccaccgatttcagcgtgtc cacatcacagctccctccagaagctgcaagagcttcttagcagttcctggtctgaacc ctctcccagtcctcatcttccaccctaaaactagagtgatcttcctaaaacttcactta acccctcagctatgaaaaggcttccaggagtttccatgaa

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
16	203438_at	STC2	gtccacattcctgcaagcattgattgagacatttgcacaatctaaaatgtaagcaaa gtagtcattaaaaatacaccctctacttgggctttatactgcatacaaatttactcatg agccttcctttgaggaaggatgtggatctccaaataaagatttagtgtttatttttgagc tctgcatcttaacaagatgatctgaacacctctcctttgtatcaataaatagccctgtt attctgaagtgagaggaccaagtatagtaaaatgctgacatctaaaactaaataaat agaaaacaccaggccagaactatagtcatactcacacaaagggagaaatttaaact cgaaccaagcaaaaggcttcacggaaatagcatggaaaaacaatgcttccagtggc cacttcctaaggaggaacaaccccgtctgatctcagaattggcaccacgtgagcttg ctaagtgataatatctgtttctactacggatttaggcaacaggacctgtacattgtcac attgcat
17	203444_s_at	MTA2	cacaaaggataccagggccctacggaaggctctgacccatctggaaatgcggcgag ctgctcgccgacccaacttgcccctgaaggtgaagccaacgctgattgcagtgcggc cccctgtccctctacctgcaccctcacatcctgccagcaccaatgagcctattgtcctg gaggactgagcacctgtggggaagggaggtgggctgagaggtagagggtggatgc ccagggcacccaaacctcccttcccttttcgtgtcgaagggagtgaggagtgaattaa ggaagagagcaagtgagtgtgtgtccctggaggggttgggcgccctctggtgttacc acctcgagacttgtctcatgcctccatgcttgccgatggaggacagactgcaggaact tggcccatgtgggaacctagcctgttttgggggtaggacccacagatgtcttggac
18	203475_at	CYP19A1	gaaattctttcccagtctgtcgatttatgcctcagccacttgcctgtgctacaattcatt gtgttacctgtagattcaggtaatacaaaccatatataatcatcaagtaatacaaact aatttagtaatagcctgggttaagtattattagggccctgtgtctgcatgtagaaaaa aaaattcacatgatgcacttcaaattcaaataaaaatccttttggcatgttcccatttttt gcttagctcaattagtgtggctaaccaagagataactgtaaatgtgacattgatttgc tcttactacagctacagtgattgggggaggaaaagtcccaacccaatgggctcaaac ttctaaggggtactcctctcatcccttatccttctccctcgacattttctccctctttctt cccatgaccccaaagccaagggcaacagatcagtaaagaacgtggtcagagtaga acccctg
19	203509_at	SORL1	gaatatcacagcttaccttgggaatactactgacaatttctttaaaatttccaacctga agatgggtcataattacacgttcaccgtccaagcaagatgccttttggcaaccagat ctgtggggagcctgccatcctgctgtacgatgagctggggtctggtgcagatgcatct gcaacgcaggctgccagatctacggatgttgctgctgtggtggtgcccatcttattcct gatactgctgagcctgggggtggggtttgccatcctgtacacgaagcaccggaggct gcagagcagcttcaccgccttcgccaacagccactacagctccaggctggggtccgc aatcttctcctctggggatgacctgggggaagatgatgaagatgcccctatgataact ggattttcagatgacgtccccatggtgatagcctgaaagagctttcctcactagaaac ca
20	203928_x_at	MAPT	gagtccagtcgaagattgggtccctggacaatatcacccacgtccctggcggaggaa ataaaaagattgaaacccacaagctgaccttccgcgagaacgccaaagccaagac agaccacggggcggagatcgtgtacaagtcgccagtggtgtctggggacacgtctcc acggcatctcagcaatgtctcctccaccggcagcatcgacatggtagactcgcccca gctcgccacgctagctgacgaggtgtctgcctccctggccaagcagggtttgtgatca ggcccctggggcggtcaataatngtggagaggagagaatgagagagtgtggaaaa aaaaagaataatgacccggcccccgccctctgcccccagctgctcctcgcagttcgg ttaattggttaatcacttaacctgcttttgtcactc

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
21	203973_s_at	CEBPD	aagcggcgcaaccaggagatgcagcagaagttggtggagctgtcggctgagaacg agaagctgcaccagcgcgtggagcagctcacgcgggacctggccggcctccggcag ttcttcaagcagctgcccagcccgcccttcctgccggccgccgggacagcagactgc cggtaacgcgcggccggggcgggagagactcagcaacgacccatacctcagaccc gacggcccggagcggagcgcgccctgccctggcgcagccagagccgccgggtgccc gctgcagtttcttgggacataggagcgcaaagaagctacagcctggacttaccacca ctaaactgcgagagaagctaaacgtgtttatttttcccttaaattatttttgtaatggta gctttttctacatcttactcctgttgatgcagctaaggtacatttgtaaaaagaaaaaa aaccagacttttcagacaaacccttttgtattgtagataagaggaaaagactgagcat gctcacttttttatattaa
22	204179_at	MB	tgttccggaaggacatggcctccaactacaaggagctgggcttccagggctaggccc ctgccgctcccacccccacccatctgggccccgggttcaagagagagcggggtctga tctcgtgtagccatatagagtttgcttctgagtgtctgctttgtttagtagaggtgggca ggaggagctgaggggctggggctggggtgttgaagttggctttgcatgcccagcgat gcgcctccctgtgggatgtcatcaccctgggaaccgggagtgcccttggctcactgtg ttctgcatggtttggatctgaattaattgtcctttcttctaaatcccaaccgaacttcttc caacctccaaactggctgtaaccccaaatccaagccattaactacacctgacagtag caattgtctgattaatcactggccccttgaagacagcagaatgtccctttgcaatgag gaggagatctgggctgggcgggccagctggggaagcatttgactatctggaacttgt gtgtgcctcctcaggtatggca
23	204267_x_at	PKMYT1	ctgtggtgcatggcagcggaggccctgagccgagggtgggccctgtggcaggccct gcttgccctgctctgctggctctggcatgggctggctcaccctgccagctggctacag cccctgggcccgccagccaccccgcctggctcaccaccctgcagtttgctcctggaca gcagcctctccagcaactgggatgacgacagcctagggccttcactctcccctgagg ctgtcctggcccggactgtggggagcacctccacccccggagcaggtgcacaccca gggatgccctggacctaagtgacatcaactcagagcctcctcggggctccttcccctc ctttgagcctcggaacctcctcagcctgtttgaggacaccctagacccaacctgagcc ccagactctgcctctgcacttttaacctttatcctgtgtctctcccgtcgcccttgaaa gctggggcccctcgggaactcccatggtcttctctgcctggccgtgtctaataa
24	204338_s_at	RGS4	gaaacatcggctaggtttcctgctgcaaaaatctgattcctgtgaacacaattcttccc acaacaagaaggacaaagtggttatttgccagagagtgagccaagaggaagtcaa gaaatgggctgaatcactggaaaacctgattagtcatgaatgtgggctggcagcttt caaagctttcttgaagtctgaatatagtgaggagaatattgacttctggatcagctgt gaagagtacaagaaaatcaaatcaccatctaaactaagtcccaaggccaaaaaga tctataatgaattcatctcagtccaggcaaccaaagaggtgaacctggattcttgcac cagggaagagacaagccggaacatgctagagcctacaataacctgctttgatgagg cccagaagaagattttcaacctgatggagaaggattcctaccgccgcttcctcaagtc tcgattctatcttgatttggtcaacccgtcca
25	204531_s_at	BRCA1	ttcaagaaccggtttccaaagacagtcttctaattcctcattagtaataagtaaaatgt ttattgttgtagctctggtatataatccattcctcttaaaatataagacctctggcatga atatttcatatctataaaatgacagatcccaccaggaaggaagctgttgctttctttga ggtgatttttttccttgctccctgttgctgaaaccatacagcttcataaataatttttgct tgctgaaggaagaaaaagtgtttttcataaacccattatccaggactgtttatagctg ttggaaggactaggtcttccctagccccccagtgtgcaagggcagtgaagacttga ttgtaca

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
26	204584_at	L1CAM	cctccctatcgtctgaacagttgtcttcctcagcctcctcccgccccccaccttgggaat gtaaatacaccgtgactttgaaagtttgtacccctgtccttccctttacgccactagtgt gtaggcagatgtctgagtccctaggtggtttctaggattgatagcaattagctttgatg aacccatcccaggaaaaataaaaacagacaaaaaaaaaggaaagattggttctcc cagcactgctcagcagccacagcctccctgtatgcctgtgcttggtctactgataagc cctctacaaaa
27	204684_at	NPTX1	ttccttttgtagattcccagtttattttctaagactgcaaagatcactttgtcaccagcc ctgggacctgagaccaaggggtgtcttgtgggcagtgaggggtgaggagaggct ggcatgaggttcagtcattccagtgagctccaaagaggggccacctgttctcaaaag catgttggggaccaggaggtaaaactggccatttatggtgaacctgtgtcttggagct gacttactaagtggaatgagccgaggatttgaatatcagttctaaccttgatagaag aaccttgggttacatgtggttcacattaagaggatagaatcctttggaatcttatggc aaccaaatgtggcttgacgaagtcgtggtttcatctctt
28	204810_s_at	CKM	gcaagcaccccaagttcgaggagatcctcacccgcctgcgtctgcagaagaggggt acaggtgcggtggacacagctgccgtgggctcagtatttgacgtgtccaacgctgat cggctgggctcgtccgaagtagaacaggtgcagctggtggtggatggtgtgaagctc atggtggaaatggagaagaagttggagaaaggccagtccatcgacgacatgatccc cgcccagaagtaggcgcctgcccacctgccaccgactgctggaaccccagccagtg ggagggcctggcccaccagagtcctgctccctcactcctcgccccgcccctgtccca gagtccacctgggggctctctccacccttctcagagttccagtttcaaccagagttcca accaatgggctccatcctctggattctggccaatgaaatatctccctggcagggtcct cttcttttcccagagctcctccccaaccaggagctctagttaatg
29	204817_at	ESPL1	tgtttggctgtagcagtgcggccctggctgtgcatggaaacctggaggggctggca tcgtgctcaagtacatcatggctggttgccccttgtttctgggtaatctctgggatgtga ctgaccgcgacattgaccgctacacggaagctctgctgcaaggctggcttggagcag gcccaggggccccccttctctactatgtaaaccaggcccgccaagctccccgactca agtatcttattggggctgcacctatagcctatggcttgcctgtctctctgcggtaaccc catggagctgtcttattgatgctagaagcctcataactgttctacctc
30	204933_s_at	TNFRSF 11B	gataaaacggcaacacagctcacaagaacagactttccagctgctgaagttatgga aacatcaaaacaaagcccaagatatagtcaagaagatcatccaagatattgacctc tgtgaaaacagcgtgcagcggcacattggacatgctaacctcaccttcgagcagctt cgtagcttgatggaaagcttaccgggaaagaaagtgggagcagaagacattgaaa aaacaataaaggcatgcaaacccagtgaccagatcctgaagctgctcagtttgtggc gaataaaaaatggcgaccaagacaccttgaagggcctaatgcacgcactaaagca ctcaaagacgtaccactttcccaaaactgtcactcagagtctaaagaagaccatcag gttccttcacagc
31	204953_at	SNAP91	agagaggtgctattcaagtgattctgaaggcaccccaaggtatatctgtaatttaaag attactgcaaatatctttactttactgtgggtttttagtacatctgttaatttagtgtttct ttgtgtgttttgtagactagtgttcttccatccttcaactgagctcaaagtaggttttgtt gtaacattgtgattaggatttaaactaattcagagaattgtatcttttactgtacatact gtattctttaagttttaatttgttgtcatactgtctgtgctgatggcttggcttaagatttt gatgcataaatgaggtcactgttgatcagtgttgctagtagcttggcagctcttcataa aagcatattgggttggaaaggtgtttgcctatttttca

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
32	205046_at	CENPE	aatcagcatctttccaatgaggtcaaaacttggaaggaaagaaccctaaaagaga ggctcacaaacaagtaacttgtgagaattctccaaagtctcctaaagtgactggaac agcttctaaaaagaaacaaattacaccctctcaatgcaaggaacggaatttacaag atcctgtgccaaaggaatcaccaaaatcttgttttttgatagccgatcaaagtcttta ccatcacctcatccagttcgctattttgataactcaagtttaggcctttgtccagaggt gcaaaatgcaggagcagagagtgtggattctcagccaggtccttggcacgcctcctc aggcaaggatgtgcctgagtgcaaaactcagtagactcctctttgtcacttctctgga gatccagcattccttatttggaaatgactttgtttatgtgtctatccctggtaatgatgtt gtagtgcagcttaatttcaattcagtctttactttgccactag
33	205189_s_at	FANCC	ttccctccacctccaagacaggtggcggccgggcaggcactcttaagcccacctccc cctcttgttgccttcgatttcggcaaagcctgggcaggtgccaccgggaaggaatggc atcgagatgctgggcggggacgcggcgtggcgagggggcttgacggcgttggcggg gctgggcacaggggcagccgcagggaggcagggatggcaaggcgtgaagccacc ctggaaggaactggaccaaggtcttcagaggtgcgacagggtctggaatctgacctt actctagcaggagtttttgtagactctccctgatagtttagtttttgataaagcatgctg gtaaaaccactaccctcagagagagccaaaaatacagaagaggcggagagcgccc ctccaaccaggctgttattcccctggactc
34	205217_at	TIMM8A	gtacatgggactatgcttttctcaaagccccattaactgcttcctataatttgatagtg ggaccacatacgtaaaaatctctcatttgtgtggagtcatttctgatttcaggggagat ccttgtgtttatcagaaagggcagaagtaggggaagaataatttggtatccttatcta gtgtttgattgtcaatgctggagaaaaatatctgtaagagtgtttatacagtacacttc agttatcttgatctccctttcctatatgatgatttgcttaaatatccatattaagtaagtc tcaaggtagggtaggcagcctgagagtctagaggcctttagttataaaggaatctag ccagtgaacataattcttattactagactgccacaaggaagaaattaacttaccctgt atatcagggtacaaaaaattcagtgatgtgcctaaataagttataaagatttaggcc aatcagaagctaacagcagtttcaggtagaggtgcatgcctaatgttagttagtgta gattccatttactgcattctt
35	205386_s_at	MDM2	tttcccctagttgacctgtctataagagaattatatatttctaactatataaccctagga atttagacaacctgaaatttattcacatatatcaaagtgagaaaatgcctcaattcac atagatttcttctctttagtataattgacctactttggtagtggaatagtgaatacttac tataatttgacttgaatatgtagctcatcctttacaccaactcctaattttaaataattt ctactctgtcttaaatgagaagtacttggttttttttttcttaaatatgtatatgacattt aaatgtaacttattatttttttttgagaccgagtcttgctctgttacccaggctggagtgc agtgggtgatcttggctcactgcaagctctgccctccccgggttcgcaccattctcctg cctcagcctcccaattagcttggcctacagtcatctgcc
36	205433_at	BCHE	ggaaagcaggattccatcgctggaacaattacatgatggactggaaaaatcaattta acgattacactagcaagaaagaaagttgtgtgggtctctaattaatagatttacccctt tatagaacatattttcctttagatcaaggcaaaaatatcaggagctttttacacacct actaaaaaagttattatgtagctgaaacaaaaatgccagaaggataatattgattcc tcacatctttaacttagtattttacctagcatttcaaaacccaaatggctagaacatgt ttaattaaatttcacaatataaagttctacagttaattatgtgcatattaaaacaatgg cctggttcaatttctttctttccttaataaatttaagtttttttccccccaaaattatcagtg ctctgcttttagtcacgtgtattttcattaccactcgtaaaaaggtatctttttttaaatga attaaatattgaaacactgtacaccatagtttaca

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
37	205481_at	ADORA1	gaggagaacactagacatgccaactcgggagcattctgcctgcctgggaacggggt ggacgagggagtgtctgtaaggactcagtgttgactgtaggcgcccctggggtgggt ttagcaggctgcagcaggcagaggaggagtacccccctgagagcatgtgggggaa ggccttgctgtcatgtgaatccctcaataccctagtatctggctgggttttcaggggc tttggaagctctgttgcaggtgtccgggggtctaggactttagggatctgggatctgg ggaaggaccaacccatgccctgccaagcctggagcccctgtgttgggggcaaggt gggggagcctggagcccctgtgtgggagggcgaggcgggggagcctggagcccct gtgtgggagggcgaggcgggggatcctggagcccctgtgtcgggggcgagggag gggaggtggccgtcggttgaccttctgaacatgagtgtcaactccaggacttgcttcc aagcccttccctctgttggaaattgggtgtgccctggctcc
38	205491_s_at	GJB3	tgcttccagccttcgtaattagacttcaccctgagtacacacacaatcactgccactct cactatagacaaaccacactccctcctctgtcacccagtcactgccatctcaacacac atccccaccctgtgtacacacaatctctgttattcatactctcactccttatgcgcactc tcaacagggcatgtagtctgcactcaagcatgccatcccagcctcaccctgcatttta ttcggctcatcccattttccctgaacattttcgctgaactagggccctggcaggatgct gggactgtgcaaggaggtaggacctatgcccacggagctaagagacaggaacaca ggctcatctcccgcactaaccaaccccctgggatggctcacagcctgctcccagtgctg tgtcatgacctgaa
39	205501_at	PDE10A	atgcttgcccaacacactgtgaaatagttaccaaaatttgtacaaatgcagcatcttc attctttctgagaagacaagatggttttctttacatgaacaaatgaacaaaagagatc ctagatccataacgtagctaaggcatctaagagtttgctgttgataatcttgctgacc aaaaactactggagagtaacacaggttatatgccatcacaaatacaatgctcatga agaactgatttgtagagtcaatgaacctgtgtccagaattttaataggctctctattgg aaggagaaagaatttcaagttaacagtatctaactttatcatagttgatgttagtaaa ttttaaaaaatgattttatatgtatgacaaaaatctttgtaaaatgcgcaagtgcaat aatttaaagaggtcttaactttgcatttataaattataaatattgtacatgtgtgtaatt ttttcatgtattcatttgcagtctttgtatttaaaa
40	205825_at	PCSK1	tttccattcccaatctagtgctagatgtataaatctttcttttgattcttcctaacaaaat attttctgggttaaaaccccagccaactcattgggttgtagccaaaggttcactctca agaagctttaatatttaaataaaatcatattgaatgtttccaacctggagtataatatt cagatataaaacagttttgtcagtctttcttagtgcctgtgtggatttttgtgaaaatgt caaagagaaaacttatatactatttcccttgaaattttaaactatattttctttacaggt atttataatataccaatgcttttatcaaacagaattttaaagagcataataaattatat taaagaaccaaaagttttcctgagaataagaaagtttcacccaataaaatatttttga aaggcatgttcctctgtcaatgaaaaaaagtacatgtatgtgttgtgatattaaaagt gacatttgtctaatagcctaatacaacatgtagctgagtttaacatgtgtggtcttg
41	205893_at	NLGN1	gaacctaggagagtcaacatctggaggattttagtctttcttacacatatgtgtgattt taaacgaatattctcagaccacaggaaactcttcatccccctgttgtttaccagtaac agtatatcacagacctttccaaatgtttgtatatgtaatcagatgtacatttatattga aaaacaaatgagatggacttaaagagcacatcctgataaatactttctctctcacctg tactatatttctattagactaaagttatgtgatttttttttacatttttcagatgactag caattttgatagtttataagataatgcaaagaactttctctgacaaactaactgcagt aacagaaacctttctttttcagttactctttttcaagaatgaaagattattatacaaaaa attgtatactacttgatggaaccaactttgtacatcttggccatgtcactggtcattg

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
42	205938_at	PPM1E	catgctaggctttctcagtggggaaaaaaatggctggatagaactgggacaaacac agacccatctttaggggtctggattttgtaggtccgactacacagcagtgttaactcat ttctcatgccattagctctctacaaaataaagcaaagtagttctagtgtggtcgttata aaccaatattgtgaaaaatagcaactattcatttgttcacaacatgcgtatttataga gtagttaggtaccatttgtaaggtaaatcctttaaaattctataatacatactaaaata gtggttattggtctgatatatgctgctcttggttctataaactagataaaagcagtgct ttgtgaaatgcagtgttctctcttaacgccactggtgataggaagtagttcccttcagt tcaaatc
43	205946_at	VIPR2	ttcctcccctgtagggtttggacagacccacccccagccttgcccagctttcaaagga caaaagggagcatcccccacctactctcaggtttttgaggaaacaaagatttgtggt aactgaaggtgttgggtcagtggccaggtgccgacactgagctgtgacccagaggg gacgctgaggaagtgggcgtgagtggacntgtcaggtggttaccaggcactggttgt tgatggtcggtggttgggtgtgggcagtcatcagtcatcaggtgtgctcaggggaca atctcccctcaaccgcacatgtgccactgttcagcggagctgactggtttcnncctggt agagggnccggctgtttcctgacagatgcctggtgagcaggggaagcaggacccag tggtcancaggtgtctttaactgtcattgtgtgtggaatgtcgcagactcctccacgtg gcgggaatgagct
44	206043_s_at	ATP2C2	gcaccacgacgatgacgttcacttgttttgtgtttttcgatctcttcaacgccttgacct gccgctctcagaccaagctgatatttgagatcggctttctcaggaaccacatgttcctc tactccgtcctggggtccatcctggggcagctggcggtcatttacatccccccgctgc agagggtcttccagacggagaacctgggagcgcttgatttgctgtttttaactggatt ggcctcatccgtcttcattttgtcagagctcctcaaactatgtgaaaaatactgttgca gccccaagagagtccagatgcaccctgaagatgtgtagtggaccgcactccgcggc accttccctaatcatctcgatctggttgtgactgtggcccctgccgtgtctcctcgtcag gggagacttttaggaggccgcagccttccatcaccggatcagtttttcctcttaggaa agctgcaggaacctcgtgggc
45	206096_at	ZNF35	gtggctttcctaggaatgggtcgtacaaagctaagtggtaatgatgctatttggggaa aggtctttttgcttaantttgtttttaaaactctgatgattncttgagcaacaggcag gttatctgcctggttgaattctggttgaaccgtgtattctaatatttctggttaagtggt gactgggtaaggaaaccacttggggtagcagttcaacaattcacttacgaatgtttat aagctttccatttcctaggtaatttttaaaagccagtcaaaacaaaaactttactgaa aatggacagaaataggaaatggacttttccttactgtctatacctcctgaaccttggt attgtaaagatctggggacctctgggtctgttctgaccattccctagtctccatggcca agcactcaaggattgatggacaccacacaccagctatattcatttgccaagatcaac agctccttctccaaacaactcaagccccaattccnatcgcattcnnttngggtgag atgcaactaacagcccctt
46	206228_at	PAX2	gcaggctagatccgaggtggcagctccagcccccgggctcgcccccctngcgggcgt gccccgcgcgcccgggcggccgaaggccgggccgccccgtcccgccccgtagttg ctctttcggtagtggcgatgcgccctgcatgtctcctcacccgtggatcgtgacgactc gaaataacagaaacaaagtcaataaagtgaaaataaataaaaatccttgaacaaa tccgaaaaggcttggagtcctcgcccagatctctctcccctgcgagccctttttatttg agaaggaaaaagagaaaagagaatcgtttaagggaacccggcgcccagccaggc tccagtggcccgaacggggcggcgagggcggcgagggcgccgaggtccggcccat cccagtcctgtggggctggccgggcagagaccccggacccaggcccaggcctaacc tgctaaatgtccccggacggttctggtctcctcggccactttcagtgcgtcggttcgttt tgattctttt

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
47	206232_s_at	B4GALT6	tgcagttttgcatgtaatcggttatacctttattggacttttatagacatttttttatttgca tgaaaaaaactcactaaatttacatcactaaacaaaggttaacccttgtgtgaaatg aaggaactgtcaataattgacagccaactaatacagtaaactgttatactagttttga gctttagacctcagccttttgtgtggaagaagtcacagctttcttaggctttaaaggaa aagaaggaaggacttaaatagcttttcttcctaccgggattacctatgtttttccttgct tgcaatctcatctgattttgctagaaatcacaaccatattgtttatgcatattgcatga gtattaccaagaaaaaaatctttaaaagttgtgatgtgacatgatataaaggatctct ttatgttaaatgtctttccatgtacctctggtgtgtcagggattttgtgcctcaaaaaat gtttccaaggttgtgtgtttatactgtgtatttttttttaaattcacggtgaacagcactttt tattatttcca
48	206401_s_at	MAPT	aggtggcagtggtccgtactccacccaagtcgccgtcttccgccaagagccgcctgc agacagcccccgtgcccatgccagacctgaagaatgtcaagtccaagatcggctcc actgagaacctgaagcaccagccgggaggcgggaaggtgcaaatagtctacaaac cagttgacctgagcaaggtgacctccaagtgtggctcattaggcaacatccatcata aaccaggaggtggccaggtggaagtaaaatctgagaagcttgacttcaaggacag agtccagtcgaagattgggtccctggacaatatcacccacgtccctggcggaggaaa taaaaagattgaaacccacaagctgaccttccgcgagaacgccaaagccaagaca gaccacggggcggagatcgtgtacaagtcgccagtggtgtctggggacacgtctcca cggcatctcagcaatgtctcctccaccggcagcatcgacatggtagactcgccccag ctcgccacgctagctgacgaggtgtctgcctcc
49	206426_at	MLANA	gtaaagatcctatagctctttttttttgagatggagtttcgcttttgttgcccaggctgg agtgcaatggcgcgatcttggctcaccataacctccgcctcccaggttcaagcaattc tcctgccttagcctcctgagtagctgggattacaggcgtgcgccactatgcctgacta attttgtagttttagtagagacggggtttctccatgttggtcaggctggtctcaaactcc tgacctcaggtgatctgcccgcctcagcctcccaaagtgctggaattacaggcgtga gccaccacgcctggctggatcctatatcttaggtaagacatataacgcagtctaatta catttcacttcaaggctcaatgctattctaactaatgacaagtattttctactaaacca gaaattggtagaaggatttaaataagtaaaagctactatgtactgccttagtgctgat gcctgtgtactgccttaaatgtacctatggcaatttagctctcttgggttcccaaatccc tctcacaagaatgt
50	206496_at	FMO3	aaagcccaacatcccatggctgtttctcacagatcccaaattggccatggaagtttat tttggcccttgtagtccctaccagtttaggctggtgggcccagggcagtggccaggag ccagaaatgccatgctgacccagtgggaccggtcgttgaaacccatgcagacacga gtggtcgggagacttcagaagccttgcttctttttccattggctgaagctctttgcaatt cctattctgttaatcgctgttttccttgtgttgacctaatcatcattttctctaggatttct gaaagttactgacaatacccagacaggggctttgc
51	206505_at	UGT2B4	taattacgtctgaggctggaagctgggaaacccaataaatgaactcctttagtttatt acaacaagaagacgttgtgatacaagagattcctttcttcttgtgacaaaacatcttt caaaacttaccttgtcaagtcaaaatttgttttagtacctgtttaaccattagaaatatt tcatgtcaaggaggaaaacattagggaaaacaaaaatgatataaagccatatgag gttatattgaaatgtattgagcttatattgaaatttattgttccaattcacaggttacat gaaaaaaaatttactaagcttaactacatgtcacacattgtacatggaaacaagaac attaagaagtccgactgacagtatcagtactgttttgcaaatactcagcatactttgg atccatttcatgcaggattgtgttgttttaac

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
52	206524_at	T	agcagtggaggagcacacggacctttccccagagcccccagcatcccttgctcacac ctgcagtagcggtgctgtccaggtggcttacagatgaacccaactgtggagatgatg cagttggcccaacctcactgacggtgaaaaaatgtttgccagggtccagaaacttttt ttggtttatttctcatacagtgtattggcaactttggcacaccagaatttgtaaactcca ccagtcctactttagtgagataaaaagcacactcttaatcttcttccttgttgctttcaa gtagttagagttgagctgttaaggacagaataaaatcatagttgaggacagcaggtt ttagttgaattgaaaatttgactgctctgccccctagaatgtgtgtattttaagcatatg tagctaatctcttgtgtt
53	206552_s_at	TAC1	ttcagcttcatttgtgtcaatgggcaatgacaggtaaattaagacatgcactatgagg aataattatttatttaataacaattgtttggggttgaaaattcaaaaagtgtttatttttt catattgtgccaatatgtattgtaaacatgtgttttaattccaatatgatgactccctta aaatagaaataagtggttatttctcaacaaagcacagtgttaaatgaaattgtaaaa cctgtcaatgatacagtccctaaagaaaaaaaatcattgctttgaagcagttgtgtca gctactgcggaaaaggaaggaaactcctgacagtcttgtgcttttcctatttgttttca tggtgaaaatgtactgagattttggtattacactgtatttgtatctctgaagcatgtttc atgttttgtgactatatagagatgtttttaaaagtttcaatgtgattctaatgtcttcatt tcattgtatgatg
54	206619_at	DKK4	ctgtctgacacggactgcaataccagaaagttctgcctccagccccgcgatgagaag ccgttctgtgctacatgtcgtgggttgcggaggaggtgccagcgagatgccatgtgct gccctgggacactctgtgtgaacgatgtttgtactacgatggaagatgcaaccccaat attagaaaggcagcttgatgagcaagatggcacacatgcagaaggaacaactggg cacccagtccaggaaaaccaacccaaaaggaagccaagtattaagaaatcacaag gcaggaagggacaagagggagaaagttgtctgagaacttttgactgtggccctgga ctttgctgtgctcgtcatttttggacgaaaatttgtaagccagtccttttggagggaca ggtctgctccagaagagggcataaagacactgctcaagctccagaaatcttccagcg ttgcgactgtggccctggactactgtgtcgaagccaattgaccagcaatcggcagca tgctcgat
55	206622_at	TRH	gccctcttcctttaggcatgtgagaaaatcagcctagcagtttaaaccccactttcctc cacttagcaccataggcaaggggcagatcccagagcccctctcacccccccccacc acaggcctgctccttccttagccttggctaagatggtccttctgtgtcttgcaaagact ccccaagtggacagggagcccctgggagggcagccagtgagggtggggtgggact gaagcgttgtgtgcaaatccagcttccatcccctccccaacctggcaggattctccat gtgtaaacttcacccccaggacccaggatcttctcctttctgggcatccctttgtgggt gggcagagccctgacccacagctgtgttactgcttggagaagcatatgtaggggcat accctgtggtgttgtgctgtgtctggctgtgggataaatgtgtgtgggaatattgaaac atcgcctaggaattgtggtttgtatataaccctctaagcccctatcccttgtcgatgac agtca
56	206661_at	DBF4B	accaggagtgtcagcttttagaaggatcatggtcatgtgagcttctggtcaccggaag ccagaaatactcagctgccatgttgatccacaaaggtgggaggatgtggggaaggg ggaaagcggtgaggacgcagagtgcaggctgtggcctcggcatcccgcaggaggtc cctagaacatgccgtttcatgtcacctgctacagctctcccccagctagtatgatgatc cgttttacaaatgcagaaatgatcttaatattcatgaccactggccaggcgaggtggc tcacacctgtaatcccagcactttgggaggccaaggcgggtggatcacaaggtcaa gagttcgagaccagcctgaccaacgtggtgaaaccccgtctctactaaaaatagaa gcattagccgagcctggtgg

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
57	206672_at	AQP2	gcgcagagtagctgcttcctggacgtgcgcgcccaggccagtgctgtgagcaggcg gggaggaggctgccggaggagcctgagcctggcaggttcccctgccctgaggctgt gagcagctagtggtggcttctcctgccttttcagggaactgggaaacttaggggact gagctggggagggaggcaggtgggtggtaagagggaaactctggagagcctgcac ccaggtactgagtggggagtgtacagaccctgccttgggggttctgggaatgatgca actggttttactagtgtgcaagtgtgttcatccccaagttctcttttgtcctcacatgca gagttgtgcatgcccctgagtgtgaacaggtttgcctacgttggtgca
58	206678_at	GABRA1	tggtttattgccgtgtgctatgcctttgtgttctcagctctgattgagtttgccacagtaa actatttcactaagagaggttatgcatgggatggcaaaagtgtggttccagaaaagc caaagaaagtaaaggatcctcttattaagaaaaacaacacttacgctccaacagca accagctacacccctaatttggccaggggcgacccgggcttagccaccattgctaaa agtgcaaccatagaacctaaagaggtcaagcccgaaacaaaaccaccagaaccca agaaaacctttaacagtgtcagcaaaattgaccgactgtcaagaatagccttcccgc tgctatttggaatctttaacttagtctactgggctacgtatttaaacagagagcctcag ctaaaagcccccacaccacatcaatagatcttttactcacattctgttgttcagttcctc tgcactgggaatttatttatgttctcaacgcagtaattccca
59	206799_at	SCGB1D2	tagaagtccaaatcactcattgtttgtgaaagctgagctcacagcaaaacaagccac catgaagctgtcggtgtgtctcctgctggtcacgctggccctctgctgctaccaggcca atgccgagttctgcccagctcttgtttctgagctgttagacttcttcttcattagtgaac ctctgttcaagttaagtcttgccaaatttgatgcccctccggaagctgttgcagccaag ttaggagtgaagagatgcacggatcagatgtcccttcagaaacgaagcctcattgcg gaagtcctggtgaaaatattgaagaaatgtagtgtgtgacatgtaaaaactttcatcc tggtttccactgtctttcaatgacaccctgatctt
60	206835_at	STATH	aagcttcacttcaacttcactacttctgtagtctcatcttgagtaaaagagaacccagc caactatgaagttccttgtctttgccttcatcttggctctcatggtttccatgattggagc tgattcatctgaagagaaatttttgcgtagaattggaagattcggttatgggtatggc ccttatcagccagttccagaacaaccactatacccacaaccataccaaccacaatac caacaatatacctttaatatcatcagtaactgcaggacatgattattgaggcttgatt ggcaaatacgacttctacatccatattctcatctttcataccatatcacactactacca ctttttgaagaatcatcaaagagcaatgcaaatgaaaaacactataatttactgtata ctctttgtttcaggatacttgccttttcaattgtcacttgatgatataattgcaatttaaa ctgttaagctgtgttcagtactgtttc
61	206940_s_at	LOC100131317 /// POU4F1	ggtttgttaccatcctttaatcataactaaaacattgaaaacagaacaaatgagaaa agaaaaaaaacctgccgattaacaatgacgaaaatcatgcatgatctgaaaggtgt ggaaagaaacacaattaggtctcactctggttaggcattatttatttaattatgttgta tatcattgtttgcagggcaacattctatgcattgaactgagcactaactgggctagctt ctggtagacgtttgtggctagtgcgattcacagtctactgcctgttccactgaaacatt ttgtcatattcttgtattcaaagaaaaaaggaaaaaaagattattgtaaatattttatt taatgcacacattcacacagtggtaacagactgccagtgttcatcctgaaatgtctca cggattgatctacctgtccatgtatgtctgctgagctttctccttggttatgttttt

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
62	206984_s_at	RIT2	taaagagctcatttttcaggtccgccacacctatgaaattcccctggtgctggtgggta acaaaattgatctggaacagttccgccaggtttctacagaagaaggcttgagtcttgc ccaagaatataattgtggtttttttgagacctctgcagccctcagattctgtattgatga tgcttttcatggcttagtgagggaaattcgcaagaaggagtccatgccatccttgatg gaaaagaaactgaagagaaaagacagcctgtggaagaagctcaaaggttctttga agaagaagagagaaaatatgacatgatatctttgcttttgagttcctcacgctctctg aattttattagttggacaattccatatgtagcattctgcttcaatattatctctctatgtg tctctctctctttaaatatctgcctgtaggtaaaagcaagctctgcatatctgtacctct tgagatagttttgttttgcctttaacagttggatgga
63	207003_at	GUCA2A	gaggggtcaccgtgcaggatggaaatttctccttttctctggagtcagtgaagaagct caaagacctccaggagccccaggagcccagggttgggaaactcaggaactttgcac ccatccctggtgaacctgtggttcccatcctctgtagcaacccgaactttccagaaga actcaagcctctctgcaaggagcccaatgcccaggagatacttcagaggctggagg aaatcgctgaggacccgggcacatgtgaaatctgtgcctacgctgcctgtaccggat gctagggggcttgcccactgcctgcctcccctccgcagcagggaagctcttttctcct gcagaaagggccacccatgatactccactcccagcagctcaacctaccctggtccag tcgggaggagcagcccgggggaggaactgggtgact
64	207028_at	LOC100129296 /// MYCNOS	ctccccccgagagaaggctgcaaagctgggaagcccagggtgtgctcctcccgccct tttggacccccgggcttgcaccggctgcactctgagaaccagctgcgcgcggagcgg tgcaatgcagcacccaccctgcgagcctggcaattgcttgtcattaaaagaaaaaaa aattacggagggctccgggggtgtgtgttgggggaggggagaccgatgcttctaaccc agcccccgctttgactgcgtgttgtgcagctgagcgcgaggccaacgttgagcaagg ccttgcagggaggttgctcctgtgtaattacgaaagaaggctagtccgaaggtgcaa aatagcagggagaggacgcgcccccttaggaacaagacctctggatgtttccagttt caaattgaaagaagaggggcgcccccctg
65	207208_at	RBMXL2	acagcagcagttatggccggagcgaccgctactcgaggggccgacaccgggtgggc agaccagatcgtgggctctctctgtccatggaaaggggctgccctccccagcgtgatt cttacagccggtcaggctgcagggtgcccaggggcggaggccgtctaggaggccgc ttggagagaggaggaggccggagcagatactaagcaggaacagacttgggaccaa aaatccctttcaacgaaactaacaaaaagaagaacctgttgtatggtaactaccca aggactagtacaaggaagagttgtttttaccttttaagaatttcctgttaagatcgtct ccatttttatgcttttgggagaaaaaacttaaaattcgtttagtttagttttggaattgtt aacgtttctttcaacaagctcctgttaaaagtatatgaacctgagtactagtcttctta catttacaagtagaaattcgattaatggcttcttcccttgtaaattttcttg
66	207219_at	ZNF643	cagccagagcattggactgatccagcatttgagaactcatgttagagagaaacctttt acatgcaaagactgtggaaaagcgtttttccagattagacaccttaggcaacatgag attattcatactggtgtgaaaccctatatttgtaatgtatgtagtaaaaccttcagcca tagtacatacctaactcaacaccagagaactcatactggagaaagaccatataaat gtaaggaatgtgggaaagcctttagccagagaatacatctttctatccatcagagag tccatactggagtaaaaccttatgaatgcagtcattgtgggaaagcctttaggcatga ttcatcctttgctaaacatcagagaattcatactggagaaaaaccttatgattgtaat gagtgtggaaaagccttcagctgtagttcatcccttattagacactgcaaaacacatt taagaaataccttcagcaatgttgtgtgaaatatactaaacatcaaagaatctatgtt ggagcacaagattctaaatcagtggttccctg

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
67	207529_at	DEFA5	gagtcactccaggaaagagctgatgaggctacaacccagaagcagtctggggaag acaaccaggaccttgctatctcctttgcaggaaatggactctctgctcttagaacctc aggttctcaggcaagagccacctgctattgccgaaccggccgttgtgctacccgtga gtccctctccggggtgtgtgaaatcagtggccgcctctacagactctgctgtcgctga gcttcctagatagaaaccaaagcagtgcaagattcagttcaaggtcctgaaaaaag aaaaacattttactctgtgtaccttgtgtctt
68	207597_at	ADAM18	gtgacgctcaatctacagtttattcatatattcaagaccatgtatgtgtatctatagcc actggttcctccatgagatcagatggaacagacaatgcctatgtggctgatggcacc atgtgtggtccagaaatgtactgtgtaaataaaacctgcagaaaagttcatttaatgg gatataactgtaatgccaccacaaaatgcaaagggaaagggatatgtaataattttg gtaattgtcaatgcttccctggacatagacctccagattgtaaattccagtttggttcc ccaggggtagtattgatgatggaaattttcagaaatctggtgacttttatactgaaa aaggctacaatacacactggaacaactggtttattctgagtttctgcatttttctgccg tttttcatagttttcaccactgtgatctttaaaagaaatgaaataagtaaatcatgtaa cagagagaatgcagagtataatcgtaattcatccgttgtatcag
69	207814_at	DEFA6	gagccactccaagctgaggatgatccactgcaggcaaaagcttatgaggctgatgc ccaggagcagcgtggggcaaatgaccaggactttgccgtctcctttgcagaggatgc aagctcaagtcttagagctttgggctcaacaagggctttcacttgccattgcagaagg tcctgttattcaacagaatattcctatgggacctgcactgtcatgggtattaaccacag attctgctgcctctgagggatgagaacagagagaaatatattcataatttactttatg acctagaaggaaactgtcgtgtgtcccatacattgccatcaactttgtttcctcat
70	207843_x_at	CYB5A	gctggaggtgacgctactgagaactttgaggatgtcgggcactctacagatgccagg gaaatgtccaaaacattcatcattggggagctccatccagatgacagaccaaagtta aacaagcctccagaaccttaaaggcggtgtttcaaggaaactcttatcactactattg attctagttccagttggtggaccaactgggtgatccctgccatctctgcagtggccgtc gccttgatgtatcgcctatacatggcagaggactgaacacctcctcagaagtcagcg caggaagagcctgctttggacacgggagaaaagaagccattgctaactacttcaac tgacagaaaccttcacttgaaaacaatgattttaatatatctctttctttttcttccgac attagaaacaaaacaaaaagaactgtcctttctgcgctcaaatttttcgagtgtgcct ttttattcatctacttt
71	207878_at	KRT76	gagctcaagccagcatagctccaccaagtgatctactgttccaaatctctataaccac ctgcttcccactcagcctgcaatagtgtttcccactctctgcttggcatcaatagatgc ataagggtcaaccacatttttcctcaagttccctggagaagaagctgaactcctggtt tctccatccccatgaccttcccagggccatggaggtcctgctgctggtctgggatgat gatgcccctggaaaccttcctgcaatggccccttactttggacagcaacccctgagcc caagccagttttggccttcacagcctggccggttcccactctggcccatctcccattctt actgggagttggagatttgaagccagtcatctcagcactgtctgaggagggcagagc catgggttctgtgctggagggtgcacggccaagatctccagactgctggttcccagg gaaccctccctacatctgggcttcagatcctgactcccttctgtcccctaattccctga gctgtagatcctctggt

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
72	207937_x_at	FGFR1	cgcacccgcatcacaggggaggaggtggaggtgcaggactccgtgcccgcagactc cggcctctatgcttgcgtaaccagcagcccctcgggcagtgacaccacctacttctcc gtcaatgtttcagcttgcccagatctccaggaggctaagtggtgctcggccagcttcc actccatcactcccttgccatttggacttggtactcggcttagtgattagaggccctga acaggtggtggtatccctgctctgctggagaggaacccagatgctctcccctcctcgg aggatgatgatgatgatgatgactcctcttcagaggagaaagaaacagataacacc aaaccaaaccccgtagctccatattggacatccccagaaaagatggaaaagaaatt gcatgcagtgccggctgccaagacagtgaagttcaaatgcccttccagtgggacccc aaaccccacactgcgctggttgaaaaatggcaaagaattcaaacctgaccacagaa ttggaggctacaaggtccgttatgccacctgga
73	208157_at	SIM2	ctgccctgtacatgctagttcaacagaaaggaatggcctttcaccttctcctggtggc aggcaagcagatgtcctctgcggagataccgccagctccccaggacgcagactgac tcctgtttgctcgctggaccaaccccaggcagaaggtggaaggtgggaacagaggtt tagctgcaggacatgtattcccattgcaccgagacctaactgccgctcagagtgtag accgagatggtgcagatgcctgcagtgccattaaaatgtgggtgaaggtgacatcag gattatgtgccccaggccgggctcagtggctcacacctgtaatcccagcactttggga ggccaaggtgggcggatcacctgaggtcaggagtttgcgacaagcctgccaacaag ctgaaacc
74	208233_at	PDPN	gaaatctctgatataagctgggtgtggtggctcgtgcctgtagtctcagctgctgggc aactgcagaccagcctgggcaacatagtaagaccctgtctcaaaaaaataatctctg gtacaatggtcatgttccaaagttccttacttgggcctcttgagtgcagtggctcacac ctggaatcccagtgctttgagaggctgaggaggcaggaggttcacttgtgcccagga atttgaggctgcagtgagctatgattgtgccactgcactccagcctgggtgacagagc aagactgtgctctcttaaaaataagaaagagcctcttcatcttcaaaaggactacatc tgaagtttccccagaaggacaaatgtctacttagaccttataaatttccaaaataaga gagtcagagccagaggtggcttgtaagttgacttctgttgagatctgaccacatttga tctcttgttttaattttccaactaactgaacttggaagaaaacccaaaccaagttttaa tctgatgccta
75	208292_at	BMP10	ccatgagcaacttccagagctggacaacttgggcctggatagcttttccagtggacct ggggaagaggctttgttgcagatgagatcaaacatcatctatgactccactgcccga atcagaaggaacgccaaaggaaactactgtaagaggaccccgctctacatcgactt caaggagattgggtgggactcctggatcatcgctccgcctggatacgaagcctatga atgccgtggtgtttgtaactacccccctggcagagcatctcacacccacaaagcatgc aattatccaggccttggtccacctcaagaattcccagaaagcttccaaagcctgctgt gtgcccacaaagctagagcccatctccatcctctatttagacaaaggcgtcgtcacct acaagtttaaatacgaaggcatggccgtctccgaatgtggctgtagatagaagaag agtcctatggcttatttaataactgtaaatgtgtatatttggtgttcctatttaatgaga ttatttaataagggtgtacagtaatagaggcttgctgccttcaggaa

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
76	208314_at	RRH	atgatctgcatgtttctggtggcatggtccccttattccatcgtgtgcttatgggcttctt ttggtgacccaaagaagattcctcccccatggccatcatagctccactgtttgcaaa atcttctacattctataacccctgcatttatgtggttgctaataaaaagtttcggaggg caatgcttgccatgttcaaatgtcagactcaccaaacaatgcctgtgacaagtatttt acccatggatgtatctcaaaacccattggcttctggaagaatctgaaataagagaaa aggacacgctatcaaaacactttagttttttgacaatgcttttcttttaaatatgagccc atttagatcaagtgcagacatggatcattgtcctatgagagtgtaagctcctcaagca cagctcgtgcttccgtttgtgcactctggctgctgtagtgtatgcttctctgtgtcctgat atatcaacttattgctcatctcctttgatgaattaggcatcagaggttaaggtccccttt c
77	208368_s_at	BRCA2	gaacaggagagttcccaggccagtacggaagaatgtgagaaaaataagcaggaca caattacaactaaaaaatatatctaagcatttgcaaaggcgacaataaattattgac gcttaacctttccagtttataagactggaatataatttcaaaccacacattagtactta tgttgcacaatgagaaaagaaattagtttcaaatttacctcagcgtttgtgtatcggg caaaaatcgttttgcccgattccgtattggtatacttttgcttcagttgcatatcttaaa actaaatgtaatttattaactaatcaagaaaaacatctttggctgagctcggtggctc atgcctgtaatcccaacactttgagaagctgaggtgggaggagtgcttgaggccagg agttcaagaccagcctgggcaacatagggagacccccatctttacgaagaaaaaaa aaaaggggaaaagaaaatcttttaaatctttggatttgatcactacaagt
78	208399_s_at	EDN3	ccgagccgagcttactgtgagtgtggagatgttatcccaccatgtaaagtcgcctgcg caggggagggctgcccatctccccaacccagtcacagagagataggaaacggcatt tgagtgggtgtccagggccccgtagagagacatttaagatggtgtatgacagagcat tggccttgaccaaatgttaaatcctctgtgtgtatttcataagttattacaggtataaa agtgatgacctatcatgaggaaatgaaagtggctgatttgctggtaggattttgtaca gtttagagaagcgattatttattgtgaaactgttctccactccaactcctttatgtggat ctgttcaaagtagtcactgtatatacgtatagagaggtagataggtaggtagatttta aattgcattctgaatacaaactcatactccttagagcttgaattacatttttaaaatgc atatgtgctgtttggcaccgtggcaagatggtatcagagagaaacccatcaattgctc aaatactc
79	208511_at	PTTG3	ttgtggctacaaaggatgggctgaagctggggtctggaccttcaatcaaagccttag atgggagatctcaagtttcaatatcatgttttggcaaaacattcgatgctcccacatcc ttacctaaagctaccagaaaggctttgggaactgtcaacagagctacagaaaagtc agtaaagaccaatggacccctcaaacaaaaacagccaagcttttctgccaaaaaga tgactgagaagactgttaaagcaaaaaactctgttcctgcctcagatgatggctatcc agaaatagaaaaattatttcccttcaatcctctaggcttcgagagttttgacctgcctg aagagcaccagattgcacatctccccttgagtgaagtgcctctcatgatacttgatga ggagagagagcttgaaaagctgtttcagctgggcccccccttcacctttgaagatgcc ctctccaccatggaaatccaatctgttgcagtctccctttaagcattctgttgaccctgg atg

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
80	208684_at	COPA	ggtttaaggatcagtcctctgcagtttcgctaaggccccctttgtgtgcatgggtcagt caccatatgttccccccagagaatgtgtctatatcctccttctaacagcaccttccccc tgcagctactcttcagatctggctctctgtaccctaaaacctagtatctttttctcttcta tggaaaatccgaaggtctaaacttgactttttgaggtcttctcaacttgactacagtt gtgctcataattgtccttgcctttccagcttaattattttaaggaacaaatgaaaactct gggctgggtggagtggctcatacctgtaatcccagcactttgggaggctacggtggg cagatcatctgaggccaggagttcgagacctgcctggccaacatggcaacaccccgt ctctaataaaaatataaaaattagcctggcatggtagcatgcgcctatagtcccagct gctcaggaggctgaggcatgagaatcgcttgaacctaggaggtggaggttgcattca actgagatcatacc
81	208992_s_at	STAT3	actggtctatctctatcctgacattcccaaggaggaggcattcggaaagtattgtcgg ccagagagccaggagcatcctgaagctgacccaggcgctgccccatacctgaagac caagtttatctgtgtgacaccaacgacctgcagcaataccattgacctgccgatgtcc ccccgcactttagattcattgatgcagtttggaaataatggtgaaggtgctgaaccct cagcaggagggcagtttgagtccctcacctttgacatggagttgacctcggagtgcg ctacctcccccatgtgaggagctgagaacggaagctgcagaaagatacgactgagg cgcctacctgcattctgccacccctcacacagccaaaccccagatcatctgaaactac taactttgtggttccagatttttttttaatctcctacttctgctatctttgagc
82	209434_s_at	PPAT	ttgacagctctttaagcccacatgcagcagtgggtcagataaccctgtggcagtgac acgggcaaattggcatttgaataaagccctgggaccacctcaacatgcgtagcctct tgtcttaaatgtactccccatggcagcatggaggaggcaagacctgtgggtcaatttt gaactggccttactttgatttttaaaacaagagactcagggaaagtactaaaccaaa atctctgattttactttgcgttttctgtagtttttgttttactgagatgcttttgtaaagga aaataatactgtgacagtttagtaattctacagattcttaatatttctccatcatggcct ttttacttcacaattttctgaagtctgaattcaattacaatttttttttttttaccaatttaat ctcaaatgttgtttaactgctttaaattcatatacgtagagtattataaactgcagaga tgaaaaatgtgttttcacgggatttatattgtgaactaaactaagcctacttttttgtga ct
83	209839_at	DNM3	gagacttctcacttctggttggaggtttcacatatggctcaactcaagtcattaatctct ttttaatttttactcttgaattccttaaacttcgctcattatgaaatgttttaaaattatg acaaaaattactctgtctaaccacttgccttgtctgctaccagtttgttaaaaattattc cccccaaccagtaattccaccagtactacttgatttgtgttatatttcctatgtacatgt acagcctttgttttgcttgcttgtctatttttactttccctttttgggtcaaatttttctttt gctttgtttgaagaaggaatatacagaagtaaaatcttgtcttctctgctgattcttta attaatatgagccggatactttccactgtcttcttggcactttcaggatttcttaatgct gatatatggactcttagaatggaatttttgaagaaaaatctcaaagcctgtatcgttct
84	209859_at	TRIM9	ataggttacccttgaaattcattagtttgtcataaagttttaggaaaggtaggacccg gaaagaagttctaattagttgtctaaatatttttcagtgagccaagaaattcaccatg aaaaaacaagaataacaaatagaagggaagagataggatgggaaagctaacaaa ttaaagttttggcaaaaaggaatatatgtaaatagctaattatttacttttgtgcttact ttatttagattatttctatcagttacaatctttttctagttaagtgtacctaatttatgga atgggtgctatcctgtttatgtgtgtcttggtttttcttggctacagaaaaactgttgca gggcaacactagtttgatatttgatttactctccaatgagactcaatggctgggccgt ggtagactcatagttcctcttgttctttattaaattcatcctgctaattagatttctagtg acttgtaacatgtagtttacactgaattgcaattacagatgcatacaactactatacta

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
85	210016_at	LOC100134306 /// MYT1L	ataacagcatatgcatttccccaccgcgttgtgtctgcagcttctttgccaatatagta atgcttttagtagagtactagatagtatcagttttggattcttattgttatcacctatgta caatggaaagggattttaagcacaaacctgctgctcatctaacgttggtacataatct caaatcaaaagttatctgtgactattatatagggatcacaaaagtgtcacatattaga atgctgacctttcatatggattattgtgagtcatcagagtttattataacttattgttcat attcatttctaagttaatttaagtaatcatttattaagacagaattttgtataaactatt tattgtgctctctgtggaactgaagtttgatttatttttgtactacacggcatgggtttgt tgacactttaattttgctataaatgtgtggaatcacaagttgctgtgatacttcatttttt aaattgtgaactttgtacaaattttgtcatgctggatgttaacacat
86	210247_at	SYN2	tcatgtcttattcttccctgtgaaaccaggattaatcgtggactcctggcagcttaacc tagctcagttgcagtgctaagcatgccccgccccccattcagtgatacctgtttgggaa gtatatacttccccaaaagtactcttggccctaagttttaggaactttccccgacctgg atcccttgtcatacctgtgttactgtttaaagcacacccacccaacttacaagatctta ggctgctgtggtggtgaagcaccttgagtctgctgatattcgggagaacaaggatct gcagtttccccttttctcccctctgaagagtggttcttatgtgcaatctgcagtaacctt gaactccagagctgcactatagaggagaatgcatgccactatgacagcagtatgcc aagctttgtgttcatctcctaata
87	210302_s_at	MAB21L2	atttcgttttgcttttggttgcctgaatgttgtcaccaagtgaaaaaattatttaactat atgtaaaatttctctttttaaaaaaaagttttactgatgttaaacgttctcagtgccaat gtcagactgtgctcctccctctcctgaacctctaccctcaccctgagctgtcttgttgaa aacagt
88	210315_at	SYN2	tattctcgactgtaatggcattgcagtagggccaaaacaagtccaagcttcttaaaat gattggtggttaatttttcaaagcagaaattttaagccaaaaacaaacgaaaggaaa gcggggaggggaaaacagaccctcccactggtgccgttgctgcgttctttcaatgctg actggactgtgtttttcctatgcagtgtcagctcctctgtctggttgtttacctgttcctgt tcgtgcttgtaatgctcacttatgttttctctgtataacttgtgattccagggctgtttgt caacagtatacaaaagaattgtgcctctcccaagtccagtgtgactttatcttctgggt ggtttg
89	210455_at	C10orf28	gaaatcagcgaggctcaagttccaagcaaaccattccaaaatgtggaattctgtgac ttcagtaggcatgaacctgatggggaagcatttgaagacaaagatttggaaggcag aattgaaactgataccaaggttttggagatactatatgagtttcctagagttttagtt ctgtcatgaaacctgagaatatgattgtaccaataaaactaagctctgattctgaaat tgtacaacaaagcatgcaaacatcagatggaatattgaatcccagcagcggaggca tcaccactacttctgttcctggaagtccagatggtgtctttgatcaaacttgcgtagatt ttgaagttgagagtgtaggtggtatagccaatagtacaggtttcatcttagatcaaaa gatacagattccattcctgcaactatgggtcacatctctctgtcagagagcacaaatg acactgttagtccagtaatgattagagaatgtgagaagaatgacagcactgctgatg agttacatgtaaagcacgaacctcctgatacag

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
90	210758_at	PSIP1	gggctcaaagcattaatccagttactgaaaagagaatacaagtggagcaaacaag agatgaagatcttgatacagactcattggactgaatttccccttcccccatgatgg aagaatgttcagattctaaattgaggacttcattattaatggcattactgtgttatgatt aacaaatttcttgtaaggtacacactacatactaaggtcggccatcattccgttttttttt ttttttttttttttaaccaagcttaaaatgaagcttaaaatgaagctttgtgtttgaaag taataacaagctcagacgaagatggtggttgtacattattcatctagaaaatataaa aattcattttgttttgaagctagttattaaactggaatagcagttatatccctgagaat ggggccctt
91	210918_at	---	gctgctgttttcttctaactgcagggaaaatgctgtctaaaagaaaataataaatttgt atctgctgagttctcttagcataaggcaccaacaaaacaaccttcaggaagggaga agaaaccatcctcccactcatccttcagaggatttagataaagtgaagggaagaatc gttctccagctccttcggaatttacgccggcatcagggcaggcttgttactgctggatc cattgtctgctcaaggttacttattccactaagacgtacatcctaccacggaccacgg ctttgtagctagccaggctctgagtgtgtgtgtagatgaaccatttctctctccagtaa atgaatgacagtctttctagggctcttgtcttctgctgggaggcag
92	211204_at	ME1	agtcactctcccagatggacggactctgtttcctggccaaggcaacaattcctacgtg ttccctggagttgctcttggggtggtggcctgcggactgagacacatcgatgataagg tcttcctcaccactgctgaggtcatatctcagcaagtgtcagataaacacctgcaaga aggccggctctatcctcctttgaataccattcgagacgtttcgttgaaaattgcagtaa agattgtgcaagatgcatacaaagaaaagatggccactgtttatcctgaaccccaaa acaaagaagaatttgtctcctcccagatgtacagcactaattatgaccagatcctacc tgattgttatccgtggcctgcagaagtccagaaaatacagaccaaagtcaaccagta acgcaacagcta
93	211264_at	GAD2	gttccacttctctaggtagacaattaagttgtcacaaactgtgtgaatgtatttgtagtt tgttccaaagtaaatctatttctatattgtggtgtcaaagtagagtttaaaaattaaac aaaaaagacattgctccttttaaaagtcctttcttaagtttagaatacctctctaagaa ttcgtgacaaaaggctatgttctaatcaataaggaaaagcttaaaattgttataaata cttcccttacttttaatatagtgtgcaaagcaaactttattttcacttcagactagtagg actgaatagtgccaaattgcccctgaatcataaaaggttctttggggtgcagtaaaa aggacaaagtaaatataaaatatatgttgacaataaaaactcttgccttttcatagt attagaaaaaaatttctaatttacctatagcaacatttcaaat
94	211341_at	LOC100131317 /// POU4F1	gcatttgaaactgagcactaaactgggctagctttctggtagaccgttttgtggctagt gcgatttcacagtctactgcctgtttccactgaaaacatttttgtcatattcttgtattca aagaaaacaggaaaaaagttattgtaaatattttatttaatgcacacattcacacag tggtaacagactgccagtgttcatcctgaaatgtctcacggattgatctacctgtctat gtatgtctgctgagctttctccttggttatgtttttctctttaccttctcctcccttactt ctatcagaaccaattctatgcgccaaatacaacaggggatgtgtcccagtacactt acaaaataaaacataactgaaagaagagcagttttatgatttgggtgcgtttttgtgt ttatactgggccaggtcctg
95	211516_at	IL5RA	ggcagccttccttgtgatcaaaaaaggtaatcccagaaacgtacccgttcactcgtg ggtcttaaaatggtttcatatctctattgtgactaattttctctcggtctactgccttttc aatcaggaatagatttgccatgaagccagtgaagtttttaagtgtctaggcttctcatt agtgccaactctcctagacctggtgcctgtttttttttccaagttttgtttctacttctatcc atttttttaaattaaacttttttattttgaaataattatcacactcacaagctgtgggaaga aataatagagatcctgtgtctctttcatccagttttcctcaagggtaacatct

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
96	211772_x_at	CHRNA3	tgctcaacgtgcactacagaaccccgacgacacacacaatgccctcatgggtgaag actgtattcttgaacctgctccccagggtcatgttcatgaccaggccaacaagcaacg agggcaacgctcagaagccgaggcccctctacggtgccgagctctcaaatctgaatt gcttcagccgcgcagagtccaaaggctgcaaggagggctacccctgccaggacggg atgtgtggttactgccaccaccgcaggataaaaatctccaatttcagtgctaacctca cgagaagctctagttctgaatctgttgatgctgtgctgtccctctctgctttgtcaccag aaatcaaagaagccatccaaagtgtcaagtatattgctgaaaatatgaaagcacaa aatgaagccaaagaggaacaaaaagcccaagagatccaacaattgaaacgaaaa gaaaagtccacagaaacatccgatcaagaacctgggctatgaatttccaatcttcaa caacctgtt
97	212359_s_at	KIAA0913	cagcgctgccagcaggcatacatgcagtacatccaccaccgcttgattcacctgact cctgcggactacgacgactttgtgaatgcgatccggagtgcccgcagcgccttctgcc tgacgcccatgggcatgatgcagttcaacgacatcctacagaacctcaagcgcagc aaacagaccaaggagctgtggcagcgggtctcactcgagatggccaccttctccccc tgagtctttcacccttagggtcctatacagggacccaggcctgtggctatgggggccc ctcacacaggggagtgaaacttggctggacagatcatcctcactcagttccctggt agcacagactgacagctgctcttgggctatagcttggggccaagatgtctcacaccct agaagcctagggctgggggagacagccctgtctgggaggggcgttgggtggcctc tggtatttattt
98	212528_at	---	gtcactcatttccttgaacagcacccccctttatactagcagccatttgtgccattgcct gtgccctagggtttgtggggagagagcgagggatcactgagcagttttcccagagct ccatgggaaggcaagctctccctcccaatgggagccccactgtcactaactgtaaac tcaggctcaggcttcaactgcctacccccatcctcatatttctgtctgtcccagcacctc aggagcattctcattgtggccggctaactccgcctggatgtgaacaggcaagcacag tgggaaatgagtcacgtacttgtattgcacagtggacacctctagaggtccattggtt taaagggatagggaaggaggagggatgagaccatcacccctcccagaagtaaat ctagtatctgagttttctttat
99	212531_at	LCN2	caagagctacaatgtcacctccgtcctgtttaggaaaaagaagtgtgactactggat caggacttttgttccaggttgccagcccggcgagttcacgctgggcaacattaagagt taccctggattaacgagttacctcgtccgagtggtgagcaccaactacaaccagcat gctatggtgttcttcaagaaagtttctcaaaacagggagtacttcaagatcaccctct acgggagaaccaaggagctgacttcggaactaaaggagaacttcatccgcttctcc aaatctctgggcctccctgaaaaccacatcgtcttccctgtcccaatcgaccagtgtat cgacggctgagtgcacaggtgccgccagntgccgcaccagcccgaacaccattgag gga
100	213197_at	ASTN1	tttccccttggaagacactattgatctcaacctgctgacttttcctaatgcttacctgaa ggaacccatcctggctagaaagggtgatggtactggaccggtattcaaccttgagttt tcaagctgccaaacaggtcttaagggaggtgcttatatcccaccaacactctcccag ctcccatgtccccaagacctctggagtttcctcttgaatgtacatgaaccactgtaata gcattagacttttaattgagtgtgcaatcgttttccatggagtttggtccgttcattattt tttagttaactacacttcttgatattcaaatgttctattaaaaaaactgagtatgaaga aaaacactttactactgcagaa

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
101	213260_at	FOXC1	tcccccatttacaatccttcatgtattacatagaaggattgctttttaaaaatatactg cgggttggaaagggatatttaatctttgngaaactattttagaaaatatgtttgtaga acaattatttttgaaaaagatttaaagcaataacaagaaggaaggcgagaggagca gaacattttggtctagggtggtttctttttaaaccatttttttcttgttaatttacagttaa acctaggggacaatccggattggccctcccccttttgtaaataacccaggaaatgta ataaattcattatcttagggtgatctgccctgccaatcagactttggggagatggcga tttgattacagacgttcgggggggtgggggggcttgcagtttgttttggagataataca gtttcctgctatctgccgctcctatctagaggcaacacttaagcagtaattgctgttgc ttgttgtca
102	213458_at	FAM149B1	agcctgaaacaggaactcacatgagactcagggccaccaggaaatgcttaaaatac atactctttcccaaaagcaaatctataattctgtttcaattttatgaatatatgaatag acaaaatgaatcgaattacataactatgtcattcattaaatggcaacaatgctgaca gcaagcagtagatcctctgattccaattaccatttgtttttacccaattctatttgcta gaggtagtaagtactctggcactcataaatcacatgatgataaaaaggaacatgag gccgggtatggtggctcacaactgtaatccccataccttggg
103	213482_at	DOCK3	tatgggtcagttacagcagccctcacctcaaagggctggcctgcttctcagcctacat tcatttgcaagcttcaatctctggaccatctggtgttcacaggtgttagagggttaggg gttaggggctagttttggatttgattcataggtaggagggcttagattttaaggcactt ctgaaagtcaatccctggacaaggcagtcatcacataagaacagctaccttctccac ttggtggcacaagaggtagggaggggagtatgggttcatttgncttcgcattatgca aggtgaaaccgtttgttttccctctccatttttccctaactaaatgaaaaggacacattc tgaaatccctttttgttggagaataagtcagtctgaggggaaatgggaggccagagat gagaaccctttgaaaagattgtaaaatactgattttcattctttcaagcttatttgtaa atacctatttgaatgctgtgtatttgtacaggaatttgagcaaaaaatgtatagagtgt gatgtccaattggtattcagcactat
104	213603_s_at	RAC2	gagcttcgttgatggtcttttctgtactggaggcctcctgaggcnnnnnnagcccca ggacccattaagccacccccgtgttcctgccgtcagtgccaactnnnnnatgtggaa gcatctacccgttcactccagtcccaccccacgcctgactcccctctggaaactgcag gccagatggttgctgccacaacttgtgtaccttcagggatggggctcttactccctcct gaggccagctgctctaatatcgatggtcctgcttgccagagagttcctctacccagca aaaatgagtgtctcagaagtgtgctcctctggcctcagttctcctcttttggaacaaca taaaacaaatttaattttctacgcctctggggatatctgctcagccaatggaaaatct gggttcaaccagcccctgccatttcttaagactttctgctccactcacaggatcctgag ctgcacttacctgtgagagtcttcaaacttttaaaccttgccagtcaggacttttgctat tgcaaatagaaaacccaactcaacctgctt
105	213917_at	PAX8	ctgcctggttaccgtggcgatgtgcttaatgcagcgttgaaaatacagaatactgact cctctgtccctcctggccccggactccctccctccctccttcctcttctggagcgtgaa atgagattggtcaagataaaaaaggaaaagattcggttatttttttaagagtgtggat aatggggcctctcaatcaaaatcccagtctccagtcggttccccccattcccttccaa cccctccaccttcccctgccgcctgcttagaggaggaggaagaaacataaagcaca aggcttttctcttaattatgaatcattccctgagggcaggcccagggcaaggggttcc tggggcccagagtctgacctgtgaggtagctagaaggcttgagcctctcatcaaagt cc

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
106	214457_at	HOXA2	ctttgcaggactttagcgttttctccacagattcctgcctgcagctttcagatgcagttt cacccagtttgccaggttccctcgacagtcccgtagatatttcagctgacagcttaga ctttttacagacacactcaccacaatcgacttgcagcatctgaattactaaaaacat taaagcaaaacaaagcatcaccaaacaaaaactcctttgaccaggtggttttgcctt cttttatttgggagtttatttttattttcttcttgacctacccctttccctcctttaagtgtt gaggattttctgtttagtgattccctgacccagtttcaaacagagccatcttttacaga ttattttggagttttagttgttttaaacctaactcaacaacccttatgtgattcctgaga gc
107	214608_s_at	EYA1	gtcaccctgaggaaggttcattgccattgtcatcaccatggaaacaacgttcctctcc acctgcattatgtactacatgacaggcatcaatctggggaaataataaaattatcac ctttgtcagaccataagagtttctccaaaagtggtcagtttggctgggcaatatttnct ctcatctaacaaacacaatccattgtcatgaaattacccttaggatgagtcttctttaa tcaatcatatattgggcggaaaaaacaccagctttgacccgaagtagttgaagagct acttcattctttctgaagttgtgtgttgctgctagaaatagtcatttgtgaattatcca aattgtttaaattcacaattgaattagttttttcttccttttgcttgaagcaaacagttg acaatttttaaccttttcattttatgtttttgtactctgcagactgaaaagacaaagttt atcttggccttactgtataaaggtgtgctgtgtccaccgttgtgtacaga
108	214665_s_at	CHP	gaggtctggcactagtagcacaacctaaggtggcattacagatctttgagcgagcca cagcaacttttctgccaagtcagcttnagttnagacttcagtgaatcaggntattgct atcctaatgtatgtctctatgagtgtatntagccacanantctgcccttggttganttt ctgactcattgcttgcttgcttgtttccttgctttggaaaactatnnaagattgctaaaa aataccactgcaaagtgatggaaaagggtggagaacaggggagtagccaggctgg atggctcaaatataaatgaatgaggaattctttatgaagtatcagtcagattttatga ttaagtgatgtaatataggaattatgtaaaagggaagaatgtctgatactgatctatt agagaggtactttagaggcttcttgattggcataaagttcctaaggttatagattttcc ccccttttggctgtatagcaaagtgttttaatccacggttgtgccttattgttccattaa aa
109	214822_at	FAM5B	caatgggaggggtcggagctcttccttcccctctgtggagtcacttttgtattctttta accagatttcttaaaatgttgttgttttgtgaatcctgacattggttcttacttttgtatg ctgcctcctctgtgccctcccagacgctgactgggaaacacaagaagtacaaccaac aggaaccagcgccaagggcaggcagcggcctccttgctcccctcccttactcctccct ctgctgcctcctcccccccaccaagtttcagggccctggattgttcccagttcccattgtg gtcccttcagagctcctttccaacagcatctctctgtcgaagaaagaagctctgtcaa gttagagagagacaatgtgtaggaaatgttctttttaaaaaaaaataacaaaaaca aaacaaaactatnnannntgtgattgttttccttgttaatctgctccaaccacctgaa catctaagta
110	215102_at	DPY19L1P1	gagacgggagtttaccccgatcacagaaaccataccaactgaaagacaaatcagc atcttgctggacgacccctcacagagctcctagatccttgaagtgtgaacttcagcag ctgagagagatggggtctcactatgttgcccaggctggtcttgaactcctggactcaa gcaatcctctcacctcagcctcccaaagtgctgggattacagattttataaatattgtt gatctttttgaaaaaccaactgttggcttcattttnttattgtgtaatactaccttaga ggacagcagttcctaatacctacttttattatgagtctctgccatttataaagaactgt ggacagcacagggaatgggggaagaaaactctggtgcagcttgaatcttggtagca aaacagtgacttcatcagaaaattttgtcactctctattagatataatggagtttgacc atttggaatttggaatttttcaaatgaatatgacaaaaatttaaaaaactcttgtatta ctatgtgataacacagatctttacaacttta

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
111	215180_at	---	aagccttcaccagatggtcaagcagatgctggtgccatgcccttgancntcncncca ccatcccccacctagccactatatgggttgttagatattttgaccacctcctcttcnctc actccactattcaactcactgcatcatcaatgtacttattacaaacctgtcacaagcca ggtcttatgctaggtgctcctctcaacaggttcttgagctggcaggggagagagaga cattcaaacaccaaggattaatataccattacaggtttaaagacagaggcctataag ggtcccctggcagtgccatggaggtagggcatggtcggctgtacctgtagaggtgtct aaagggaggcttgcaagctgccccttgaaggacgagcagaaaattgtacatgagga caagtaggaaaggaattccaggaggagggatcagcatgtgca
112	215289_at	HLA-DRB1 /// HLA-DRB2 /// HLA-DRB3 /// HLA-DRB4 /// HLA-DRB5 /// LOC100133484 /// LOC100133661 /// LOC100133811 /// LOC730415 /// RNASE2 /// ZNF749	ggactaaatcgagccttattatacatcagcagtctcacactggagaaagtccttttaa gttaagggannggnnnnnnannntnnancaaatgtaatactggtcagcgccaaa aaactcacactggagaaaggtcttatgagtgtggtgaatccagcaaagtgtttaaat acaactccagcctcattaaacatcagataattcatactggaaaaaggccttagtgga gtgaatgcaggaaagtcaccaaaactgtcacctcattcagcaccaaaaggttcacat cggaccaagaacctattaatatatgtaaatctaatgttgaaagagttcagatggaaa tctgcgaggatttcctgctgggaactacatta
113	215356_at	TDRD12	aattgggcaggctcttgggaagtagaaagttctggtgtttttgctggtgaaggttttga ctgtggagctcttctaacacccatatcagtgtctgtttctctgcatgtggctgctgccct gttggtggagctctgggggcagagaccaggccgccgtccagtggcgcnccgtgcgc accagctgcctgctgtttacacccaggtgcgccgagtctctttcatacagcacagcaa atgataatagctagtgacaatgtgtttcctgtgcactcgtgaaaatgcagggaggac aactgcatgcttagatctgtttctttttttcagacattcaaatgttctaatatctgaagct aacattttgtaggatataggatgctgattatgtgaacaattagtcattggttttctgtac tgctatgaatatgtctgatttcaagttttggtcaaatatctaaaatgcaaggtgaaagt gcctttgtctctatgcttctaaaatcgctcatgcttagttgtggtatggatgtcttccgc agtg

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
114	215476_at	---	cttggtaagccttgcctgtagcggctccgctgccgagtgctttgacaccaggcgctcc cagagctctgccccccactgccaagcggcagctgctccggagggcacgggggctgg atttggctgtggcttctccagctctgcacaagagcccccctccctggccctgctgcag catgactgcctcctggctcgtgtcacccactctgtctctgtctctcttcatacgtttccag ctgagctgggatccatagtctgtttccctctccacgaccaatctatttatcttctctgga acttcttgtaatgccgggagtgcagagcttacaagttggggcaggaagctttagaag cccaggnagccctgagaggctctttccttgtaagtgggtctctccccaggagcctctt ggaatatttagcagggacttttacccatgctgggtctagagaccctcccgcccctctgt ttcctgccctcctacttagactgggatctggtttccctcagctggttcccttgctagcgt gtgactctgtgtgtct
115	215705_at	PPP5C	gttcacagcagtgggtaggcccagcagtggttcttgacatcacacgatgaggcgngc atctcccgtcatccagggagaccagaggacccttgtctcactcccagttggctnttag tcacagccccgctttgtctttgacatggacgtttgtgatgatcacgttcctcccgctccc cgtgtntgaagagtgctccctgactggctgccgtctcctccctgtcgggtctggctggg ttctccanagggagtgctgcggaggggacacagcanaggccccatgctcgtgatgt atgttgcagatcattttcccccattctgtccttttttgttaaattgtggtaaaaagcaca taacataaactgtaccncccttaaccatttgaaagtatatatcccagactgtcttttatc tttagacttcacttgtggtttgttgcc
116	215715_at	SLC6A2	tcccctggaagttgtcctttctgatcctctcttcttttcccatttacaaatgatttcgtga ctgtagtttttgttcaccttctgtgcatctggcctgggggctgttagctcagaggagag gagcaaacaggaaaatgacttctgttctgtccccgctgttttgggggaagtctctccc actttgggatcctgctgaagctaggttcatgaggtcggaaatccccaccacatttgcc tagactttgggcacaggagttcttagtccaccaaatcaga
117	215850_s_at	NDUFA5	cattttctctaactttatctcctatgcatttccttatgtgtcctgtacagcagtatattcc aaaatccccagtggatgtctgaaaaccacatatagtaccaaactgtatatatgctat gttttgtttcatacatacctataataaagtttaatttatgaattaggcacaataagaga taagcaggctggacgtgctggctcacgcctgtaatcccagcactttgggaggctgag gcgggtggattgctttagcccaggagtttaagaccagcctggccaacatggcaaaac cccgtctctataaaaaatgtggaaattaatcaggtgtggt
118	215944_at	---	gagatgaccgaaaacttcaacccctgcagtcagcaatggtcaacagaaagggccca attctccacgacaatgcatgatcgcacattacacaactaaagcttcaaaagttgaac taactgggctacgaagttttgcctcatccaccatattcacctgacctcccgccaaccg actaccacttcttcaatcatctcgacaacttttgcaaggaaaacacttccacaacca gtagaatgcaaaaagtgctttccaagagttcactgaatcctgaagcacggatttttat gctacaggaataaacaaacttatttttcattggtaaaaatgtgttgattgtaatggatc ctattttgattaatgaagatgtgtttgagcctagttataatgatttaaaattcacgatcc aaaaccgcaattacttttgcatcagcctaatatgaggaagtaatagttgaacagaat aattctttcctggaagtct

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
119	215953_at	DKFZP564C196	ttggtttggtctggtttggctacctgattcctgctgtctttttctacgccaggtgaagag gcactttcaagatccttctctgagacctgcaccaataagactataccaatgttcagttg aaacatcaggtataagtttagcggaaacgaaagtacaacctgctttgaaataaattc caaggacagattgtcattaacgaaatagaaagtggactatgcccctcatgctgccag cgcctggtatgatgcggcgtgacacgcagcgcttgcggcagtacaatgcccccaatc acccgccccgccccgacgcgccgcccactcacggcaaagagagccacctagtgagg gattattctcatttccgcggtggggttctgcttttctttctaccatgagcgcccaaggat agacactcctactacctattacctcaaatagcctacatttctttccgaa
120	215973_at	HCG4P6	agaacactgagcgaggctctgtagatggatgtaataaaaatctataaaacaatgtgt ttaaacctaagaattctactgctttccaattccttccctctgctccttttcctaacctcct gcttctccagcccttccctctgtccctttcancctcaggccctcctctcccctagtccc caccaccctgtcacttctaaattgtggctctagcattgtcccattacctgctangtgac tgttctctccacagtggtcctgctcctgtgagtcagagtgtgtcatttcctcacctaaaa cactccagtggctccacctcggtcttgtgaagcttctagaatgtcaggcacgtgagca tatgagggcatacctggttcatcttaggcactaaattnnnntttgttgactgaatgaa tgaaatatgaatgtattaaattgcatcacagaaagttataaaatgtaaaacactgaa aaattaagaaatattttatnttatgtaactagtgtgcatatcaattcattccgagtctg ttgagcctgtgtat
121	216050_at	---	aatgattcaactcatgtgatccagtgttacattcagtgtggtaatgaagaacagtcaa aacaggcttttgaagaattgggagataatttggttgaattaagtaaagccaaatact ccagaaatattttaaagaaatgtctcacgttgtgaacatgtaccctagaacttaaagt ataataaaaaaaaaaaaaaannggaaagtatcttgcacaagctcacgtagctggta agttacatagttgggatctgaattcagttgtggcttcatgcctgagcttttaactactac tactaaactgagaaggcacttgcttgagtaaattatgtcatcctcttaat
122	216066_at	ABCA1	gatgtggcatgtgatgacattgcacatggncagttaantgngccaagaagngcagc agtagcagcaacnggagatgcaaagcccaacatgatggggagagaaantnttctt tcaatatgtgcttctgtaccaaaagtggaatttcacgagagacatattttggaacattt ttccttttgtgtgtgcgtgagtgtttccctgtttccagccaagggtattgtgagtttctcc tgggcctccttcagaatctgggtgctctggaaagcagtgttttggcaacatggggaaa gtatggcagtgtgggagggtcagctgggtctgggtttgaatattgcatttgaatatttt accagcattgatgtcggataaattatttagtccctgtaagcctcagttttntcttnttct acatacacataatatatttgactctttgttgtgat
123	216240_at	PVT1	tttcctaactttctgatcccttggaggtgataatcaaatattctagtctgaggcattggg atacatggtgctaggttctgagactctgcgtcaggcctgaaccctgcattttgtggag gtgggtgggagaatgtnccccctggggaacatgcctagacacgggggacaacagttg ccctcatggggaggtacctgtttactcgctgttatgggaccgctttcacaaaaccact gcaggtgagtgagttcctgctgaatatcaggcctggtgtctctagactcattattnccc ccacccaaccccctatgttagttcatctcgagccacatttttattgccataatccaggcc tggacaggccaagatcttttaacaattttaattactgaaaataataactgcatttttttt naaagcccaacttttnggtanagtcagcccaaaatacagtctttgtgttgccatctgg gaactggatttggaattgttcttccatgagactgcagagcag

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
124	216881_x_at	PRB1 /// PRB4 /// PRH1 /// PRH2 /// PRR4	ccacctcctccaggaaagccagaaagaccacccccacaaggaggtaaccagtccc aaggtcccccacctcatccaggaaagccagaaggaccacccccacaggaaggaaa caagtcccgaagtgcccgatctcctccaggaaagccacaaggaccaccccaacaag aaggcaacaagcctcaaggtcccccacctcctggaaagccacaaggcccacccccca gcaggaggcaatccccagcagcctcaggcacctcctgctggaaagccccaggggcc acctccacctcctcaagggggcaggccacccagacctgcccagggacaacagcctc cccagtaatctaggattcaatgacaggaagtgaataagaagatatcagtgaattca aataattcaattgctacaaatgccgtgacattggaacaaggtcatcatagctctaac
125	216989_at	SPAM1	gtttgatgtctattatctcacttcatcctcaccaggaccccatccgagccttaatttcag ttgacagtaactattggatccccaggaatatgtttgcatatttggggagaaaatacta ttggaggggaacagaaatgctactaagggtctcactgtgtcacccaggctggagtcc atcaaagctcactgcagccttaaccttctgtgctcaagggatcctcccacttaagcctc ctgagtagctggaactacaggcatatgccaccgagcctggctaatctttgatttttttg tacagattgtgtctccttatgttgctcaggctggactcaaacttctggtctcaagcgat ctttccatcttagcttcccaaattgttggaattatggacatgagccagtgtgcttggcct gatttttttttttttttaatgagaaaaacgttccttaagaaaagtttcattgtaagacg aggacttgctatgttgccagtttggtcttgaactcggtctcaagtgattctcctgccttg ggttcccaaagcgtttgggccggcagatgt
126	217004_s_at	MCF2	ctgaattggaacacaccagcactgtggtggaggtctgtgaggcaattgcgtcagttc aggcagaagcaaatacagtttggactgaggcatcacaatctgcagaaatctctgaa gaacctgcggaatggtcaagcaactatttctaccctacttatgatgaaaatgaagaa gaaaataggcccctcatgagacctgtgtcggagatggctctcctatattgatgaagct actatgtcaaatggcaagtagctctttcctgcctgcttctcagctcatttggaaaaata ctgcgcaaaagacattgagctcaaatgatgcagatgttgttttcaggttaatggacac gcaaagaaaccacagcacatacttctttttctttcatttaataaagcttttaattatggt acgctgtctttttaaaatcatgtatttaatgtgtcagatattgtgcttgaaagattctca tctcagaatacttttggact
127	217253_at	SH3BP2	gagtgtcttgactattctggctctttgtattttcatgtaaggtttttctcccatataagttt taaaatcagcttgtcaattccaacaacaatgatgcacttgatagtttgggaatttatta tagctatcaatcagttttgggaaaattgacgtctttacaatattgagttttctgattcat gaacatggtttacctctcttcccatgggggtctcctttaaggtttaccaataggatttta tatttggggccattgnggtcttgcttatcttaagtnnnnnnnnnnnnnnnaaatct cttgaccncatgatctgcccgccttgtcctcccaaagtgctgggattacaggcgtgag ccaccgcacctggcctgcaatacagtattgttaaccgtcttcaccatgttgtacgttag agctccagaaattatttancatgcataactgaaactttatactctttgaacaccacctc cccatttccctctcccggcagccatttgtgcctctcggttctctttattagcttccattttg tgggtcagt

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
128	217995_at	SQRDL	tacgtcaaagaccgctgctgcagtagctgcccagtcaggaatacttgataggacaat ttctgtaattatgaagaatcaaacaccaacaaagaagtatgatggctacacatcatg tccactggtgaccggctacaaccgtgtgattcttgctgagtttgactacaaagcagag ccgctagaaaccttccccctttgatcaaagcaaagagcgcctttccatgtatctcatga aagctgacctgatgcctttcctgtattggaatatgatgctaaggggttactggggagg accagcgtttctgcgcaagttgtttcatctaggtatgagttaaggatggctcagcactt gctcatcttggatggcttctgggccaaaactgcagtcactgaatgaccaagagcagc acgaaggacttggaacctatccttgtaaagagttccttgatgggtaatggtgaccaa atgcctccctttcagtacctttgaacagcaaccatgtgggctactcatgatgggcttg at
129	218768_at	NUP107	ttggatgccctaactgctgatgtgaaggagaaaatgtataacgtcttgttgtttgttga tggagggtggatggtggatgttagagaggatgccaaagaagaccatgaaagaaca catcaaatggtcttactgagaaagctttgtctgccaatgttgtgttttctgcttcatacg atattgcacagtactggtcagtatcaggaatgcctacagttagcagatatggtatcct ctgagcgccacaaactgtacctggtattttctaaggaagagctaaggaagttgctgc agaagctcagagagtcctctctaatgctcctagaccagggacttgacccattagggt atgaaattcagttatagtttaatctttgtaatctcactaattttcatgataaatgaagtt tttaataaaatatacttgttattagtaatttttttcttttgcattaccatgtaaaatttaga catttgaattttgtacttttcagaatattatcgtgacactttcaacatgtagggatatca gcgtttctctgtgtgct
130	218881_s_at	FOSL2	aggtcacagtatcctcgtttgaaagataattaagatcccccgtggagaaagcagtga cacattcacacagctgttccctcgcatgttatttcatgaacatgacctgttttcgtgcac tagacacacagagtggaacagccgtatgcttaaagtacatgggccagtgggactgg aagtgacctgtacaagtgatgcagaaaggagggtttcaaagaaaaaggattttgttt aaaatactttaaaaatgttatttcctgcatcccttggctgtgatgcccctctcccgattt cccaggggctctgggagggacccttctaagaagattgggcagttgggtttctggcttg agatgaatccaagcagcagaatgagccaggagtagcaggagatgggcaaagaaa actggggtgcactcagctctcacaggggtaatca
131	218980_at	FHOD3	gcacctcggagttgcagctgtgacactcataggttactcccaggagtgtgctgagca gaaggcaagctcttgctggatgaaacccctccaggtggggttggggagacttgatat tcacatccaacagtttgaaaagggagagctcaattcccagcgtcaccccatggcttgt gttgcctgctacgcattgacttggatctccaggagtcccctgcacataccttctccatc gtgtcagctgtgtttctcttgattccgtgacacccggtttattagttcaaaagtgtgaca ccttttctgggcaaggaacagccccttaaggagcaaatcacttctgtcacagttatt atggtaatatgaggcaatctgattagcttcacagactgagtctccacaacacc
132	219000_s_at	DSCC1	tcaagtgagtgagttcccctctacttttagccttccacccaaactggaagcctctaggt gctatcaattatttatatccatcgtttacatccatgaaattggctgaataattactcctc tgcctggcgtagacatgtgctttgggaaaaaaacgagtttataatcctataatgaag aatactggcacaggcaatgctcactcgaaaacttcaagtaatttctagttggttttgg aatgcttgataaagttcctttacagctttattttcctgatttgttttggtttagatcaaag ttcaaattaattttaacttagctaatgaactcatcaccaggacagttggaggggtag gccgaggttaaatggtccacgtttcaaaaatgttaat

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
133	219171_s_at	ZNF236	cttttgttcttgctgggttatttattttgattttagcattaaatgtcatctcaggatatctc taaaaggggttgtttaattcctaattgtatagaaagctagtttggtgaattgtattggtt aattgactgtttaaggccttaacaggtgaatctagagcctacttttatttttggttaaag aaaaagaaaatatcaataattcaattttgtgtcttttctcaatttattagcaaacacaa gacattttatgtattatttcgatttacttcctaattataaaagctgctttttgcagaaca ttccttgaaaatataaggttttgaaaagacataattttacttgaatctttgtggggtac aggttgatctttatattttactggttgttttaaaaattctagaaaagagatttctaggcc tcatgtataaccagggttttgaggataaagaactgtatttttagaactatctcatcata gcatatctgctttggaataactat
134	219182_at	FLJ22167	ttaccctcgtggctaagcaagtgtctgcaggagcagagatggctggaaggggcctct gcacacggaagatggcttgttcagcccattcacctcctgaggatgtgggcagtctcct ccaagaacacatggagctgcttcctgatcccaagcaggtcattgccactggaaggac atggccccggtgatccatgcttcatgcccacccagaaacacacccctcagtgtgtgcc tcagtttactttggagatcagttgtcgtttttagtgctcctttaggcttactaaaacagtt ttggaaacaaagctattttgaagtattcaagcagaggaattccctaacactgacc
135	219425_at	SULT4A1	gaccattttgcgagtgtagccctgtttcactcggatcaggttggcacggccgcctgcgt gtctgtccacctcatccctccgtgtatctgagggagtaaaggtgaggtctttattgctt cactgcctaattttctcacccacattcgctgaagcgatggagagtcgggggccagta gccagccaaccccgtggggaccggggttgtctgtcatttatgtggctggaaagcacc caaagtggtggtcaggagggtcgctgctgtggaaggggtctccgttcttggtgctgta tttgaaacgggtgtagagagaagcttgtgtttttgtttgtaatggggagaagcgtggc caggcagtggcacgtggcatcgcatggtgggctcggcagcaccttgcctgtgtttctg tgagggaggctgctttctgtgaaatttctttatatttttctatttttagtactgtatggat gttactgagcactacacatgatccttctgtgcttgcttg
136	219520_s_at	WWC3	aaggaaggccagagagccgcgcagttctctgcaggtgcagatgcaggcagtggag gtggcctgagcaggcagaaggacaccaagcgccctatgttgcttgtcattcatgacg tggtcttggagcttctgactagttcagactgccacgccaaccccagaaaataccccac atgccagaaaagtgaagtcctaggtgtttccatctatgtttcaatctgtccatctacca ggcctcgcgataaaaacaaaacaaaaaaacgctgccaggttttagaagcagttctg gtctcaaaaccatcaggatcctgccaccagggttcttttgaaatagtaccacatgtaa aagggaatttggctttcacttcatctaatcactga
137	219537_x_at	DLL3	tcccggctacatgggagcgcggtgtgagttcccagtgcaccccgacggcgcaagcg ccttgcccgcggccccgccgggcctcaggcccggggaccctcagcgctacctttgcc tccggctctgggactgctcgtggccgcgggcgtggccggcgctgcgctcttgctggtc cacgtgcgccgccgtggccactcccaggatgctgggtctcgcttgctggctgggaccc cggagccgtcagtccacgcactcccggatgcactcaacaacctaaggacgcaggag ggttccggggatggtccgagctcgtccgtagattggaatcgccctgaagatgtagac cctcaagggatttatgtcatatctgctccttccatctacgctcgggaggtagcgacgcc ccttttccccccgctacacactgggcgcgctgggcagaggcagcacctgctttttccct acccttcctcgattctgtccgtgaaatgaattgggtagagtctctggaaggttttaagc ccattttcagttctaacttactttcatcctattttgcatccc

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
138	219617_at	C2orf34	tgaagaaaaccttcattacccgcttctgcttattttgaccaaacatggatagaagatt aagcttctcaaagacgaagaaacgtatcaagtgcatagggaatatttttacaaaaac ggaaatctgtaaggggtataatcgcctgcctgcgccctttgcagcatttcacgtgtgg gctatggactccacctgtcctcacccacgttattccccagctgccctctccagctccct ccccgcctcttttttacactctgcttgttgctcgtcctgccctaaacctttgtttgtctttaa atgtgtataagctgcctgtctgtgacttgaatttgactggtgaacaaactaaatatttt tccctgtaattgagacagaatttcttttgatgatacccatccctccttcatttttttttttt ttttggtctttgttctgttttggtggtggtagtttttaatcagtaaacccagcaaatatca tgattctttcctggttagaaaaataaataaagtgtatctttttatctccctc
139	219643_at	LRP1B	tattcacaagttttggagggctttttgttcctctgatagacatgactgacttttagctgt cataatgtattaacctaacagatgaaatatgttaaatatgtggttgctctttatcccttt gtacaagcattaaaaaaactgctgttttataagaagacttttgttgtactatgtgcat gcatactacctatttctaaactttgccatattgaggcctttataaactattgatttatgt aatactagtgcaattttgcttgaacaatgttatgcatatcataaacttttcaggttctt gtttaagtacatttttttaaattgaacagtatttttcattttggttataatatagtcattttg cctatgtttc
140	219704_at	YBX2	ctcagcccctgtcaacagtggggaccccaccaccaccatcctggagtgattccaact caactcaaaggacacccagagctgccatctggtatctgccagtttttccaaatgacct gtaccctacccagtaccctgctccccctttcccataattcatgacatcaaaacaccag cttttcacctttttccttgagactcaggaggaccaaagcagcagccttttgctttttctttt ttcttccctcccccttatcaagggttgaaggaagggagccatccttactgttcagagac agcaactccctcccgtaactcaggctgagaag
141	219882_at	TTLL7	gtttctgtgattcaggatcctcttgggagagtatattcaataaaagcccggaggtggt gactcctttgcagctccagtgttgccagcgcctagtggagctttgtaaacagtgcctg ctagtggtttacaaatatgcaactgacaaaagaggatcactttcaggcattggtcctg actggggtaattccaggtatttactaccagggagcacccaattcttcttgagaacacc aacctacaacttgaagtacaattcacctggaatgactcgctccaatgttttgtttacat ccagatatggccatctgtgaaacagaagggaagatcgccattggttat
142	219937_at	TRHDE	ggaggtcccaaatatgtggtctatcaccactgaattcatgtaatagataagaaaaaa attagaggtggatgtcttgttttgtgtcatgaattactaaaatctcttagtagttgtggt atatttttgagtaaaattaccatttccagatttgagtttgaagggcttttatagttgtatt ttcctcctcactgttaataatcataatccttttttcagtattttagtggccttgaacaactg gtttatctacaatctcaaatcctaagtgtataattatgtgcaatgttcaatacctcatat aatacttgctcaacagtatagtggtaccaatggcattaagatggtgtttttgttctaca tatttttcaataatttattctttctaatgttgaaattatatcaggctttaccggtt
143	219955_at	L1TD1	gaagttgcaacattcgtttgataggaattccagaaaaggagagttatgagaatagg gcagaggacataattaaagaaataattgatgaaaactttgcagaactaaagaaag gttcaagtcttgagattgtcagtgcttgtcgagtacctagtaaaattgatgaaaagag actgactcctagacacatcttggtgaaattttggaattctagtgataaagagaaaata ataagggcttctagagagagaagagaaattacctaccaaggaacaagaatcaggtt gacagcagacttatcactggacacactggatgctagaagtaaatggagcaatgtctt caaagttctgctggaaaaaggctttaatcctagaatcctatatccagccaaaatggc atttgattttaggggcaaaacaaaggtatttcttagtattgaagaatttagagattatg ttttgcatatgcccaccttgagagaattactggggaataatataccttagcacgccag ggtgactaca

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
144	220029_at	ELOVL2	gttatacagatgccatgctccacaccacgagcagtgtacaaatctggctgcccgttta ctttctgagcaagcactggagtccactccgacctttttctttgaacatgcatgctgctg gaatatgtataaatcagaactagcagaagtagcagagtgatgggagcaaaatagg cactgaattcgtcaactctttttttgtgagcctacttgtgaatattacctcagatacctgt tgtcactcttcacaggttatttaagttcttgaagctgggaggaaaaagatggagtagc ttggaaagattccagcactgagccgtgagccggtcatgagccacgataaaaaatgc cagtttggcaaactcagcactcctgttccctgctcaggtatatgcgatctctactgaga agcaagcacaaaagtagaccaaagtattaatgagtatttccttctccataagtgca ggactgttactcactactaaactct
145	220076_at	ANKH	gaacgtcgtatgagatcctacaatggaagaataaaatcacctcattcttcatttcaga tctgaacattagcagtgatctagatttttttttttttaaacaaaattaagtgtgcttaga gtcatccctctacatgggctgtggctgtcagcccataggtttgtcagtttcacatcaaa actgtgggtataaactgttgaaaccaatcacattaaaatatttagctgggcacagtg gtgtgcatctgtagtcccagctacttgggaggctgaggcaggaggatcgcttaagca caggagttggaatccagcctgagcaacagagcaaaaccccgtctctaaaatacaaa taaaatatttgtgtagtttttgattaaaattgactacagcggtcagtataaaatacatg tcgcttttaaggaagtgctctttatgtatctaacagatggaagtttttgcattggtaag agcatttatatatgctttgtttcagggtttatggatttgtattcatatattgtcaaatagg tttcatactctaattttactt
146	220294_at	KCNV1	agattatatccctatcttctttttcatgtaaaccactggtcacaaatgaactgatctctg tatcccattattactataagaggtgggaatcccaaaactgcttagattgcagtacatg agtttacacaaagacttcaacaattgcacatcttcattctcccaactgagtgtagtatg tggagcataaaacagcatattcttagtatttcatgaatatcagatggtctttaaatgtc tctttatggatgtattgttcacattatggctttaaaataatgaatatgtaaaagtgagg tagtgaacatcctaaatttctacactggaattactaaataatcttatttcataaaatgg gaaatatatgttaaatgacatcactggatgaacttgaagatcttttacttgttaacaa aaaaatactatggacagctttctgattgttggggtaaatagcaaatgttcaaactttg caggcattttgacattcatcataacaacacaattcctagacatt
147	220366_at	ELSPBP1	ttaggcagtctgtggtgctcagtcacctctgtcttcgatgagaaacagcagtggaaat tctgtgaaacgaatgagtatgggggaaattctctcaggaagccctgcatcttcccctc catctacagaaataatgtggtctctgattgcatggaggatgaaagcaacaagctctg gtgcccaaccacagagaacatggataaggatggaaagtggagtttctgtgccgaca ccagaatttccgcgttggtccctggctttccttgtcactttccgttcaactataaaaaca agaattattttaactgcactaacaaaggatcaaaggagaaccttgtgtggtgtgcaa cttcttacaactacgaccaagaccacacctgggtgtattgctgatgctgaggaaagg agaaatatcttcagaggaagactgccgccatactgaggctgagcacagatttgtcttt ttcattgcatctgtcaa
148	220394_at	FGF20	gtgtggcagtgggactggtcagtattagaggtgtggacagtggtctctatcttggaat gaatgacaaaggagaactctatggatcagagaaacttacttccgaatgcatctttag ggagcagtttgaagagaactggtataacacctattcatctaacatatataaacatgg agacactggccgcaggtattttgtggcacttaacaaagacggaactccaagagatg gcgccaggtccaagaggcatcagaaatttacacatttcttacctagaccagtggatc cagaaagagttccagaattgtacaaggacctactgatgtacacttgaagtgcgatag tgacattatggaagagtcaaaccacaaccattctttcttgtcatagttcccatcataaa ataatgacccaagcagacgttcaaa

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
149	220397_at	MDM1	tatgcatttttttaccacaatttttaaaaagtttgaatagaaatttttaatgtctttgagtg gattttgtttttttgaacagttggatagacttctgcgtaagaaagctggattgactgttgt tccttcatataatgccttgagaaattctgaatatcaaaggcagtttgtttggaagactt ctaaagaaactgctccagcttttgcagccaatcaggtagcttaatggatgtaatacat ttctgagtaccattatcttatctagtaatgtagatttacatagaattaagagttgaaag aaattaagtacttaagtagcctggaggtaggttctagaaaaccaaaatgagagtttt gctaaaatcatcctattacttatgatttatggtagtaatattatactgtcctaggcttct gatgatcattgttgccagatgcagcacatatactaaatatgagacagggtaatgaaa acttggggaactggtaagtttttgcatgctac
150	220541_at	MMP26	tgacccctttgatattccagcaagtgcagaatggagatgcagacatcaaggtttctttt ctggcagtgggcccatgaagatggttggccctttgatgggccaggtggtatcttaggc catgccttttttaccaaattctggaaatcctggagttgtccatttttgacaagaatgaaca ctggtcagcttcagacactggatataatctgttcctggttgcaactcatgagattgggc attctttgggcctgcagcactctgggaatcagagctccataatgtaccccacttactg gtatcacgaccctagaaccttccagctcagtgccgatgatatccaaaggatccagca tttgtatggagaaaaatgttcatctgacataccttaatgttagcacagaggacttattc aacctgtccttttcagggagtttattggaggatcaaagaactgaaagcactagagcag ccttggggactgctaggatgaagccctaaagaatgcaacctagtcaggttagctgaa ccgacactcaaaacgctac
151	220653_at	PEG3 /// ZIM2	aaggtagaaagccttccgtccagtgtgcgaatctctgtgaacgtgtaagaattcaca gtcaggaggactactttgaatgttttcagtgcggcaaagcttttctccagaatgtgcat cttcttcaacatctcaaagcccatgaggcagcaagagtccttcctcctgggttgtccc acagcaagacatacttaattcgttatcagcggaaacatgactacgttggagagaga gcctgccagtgttgtgactgtggcagagtcttcagtcggaattcatatctcattcagca ttatagaactcacactcaagagaggccttaccagtgtcagctatgtgggaaatgtttc ggccgaccctcatacctcactcaacattatcaactccattctcaagagaaaactgttg agtgcgatcactgttgagaaaccttttagtcacagcacacacttttctcaacattattgg cttcctcctagagtgttgtgagtgtgagaaggcctttcactagcccc
152	220700_at	---	atgttactacaaacttgattaaacttctggtggaaattccatcacattttatgcaatttt caatttatttctccaatttatttttaatgccacatggacattatattccttaaccattcttt tgcatgtgattaacatttgtgaaattaaccacttaagcaagtgtttttgctttgatgaa agaaaaatgtttaaaatcctactggatatgaaactgaaagtaatgttttgtgttttttg tttcaaatgaaagtgtaaattaagaatttgttggcagggcgtggtggctcatgcctgt aatcccagcactttgggaggccgaggtgggcagatcacctgaggtcagcagtccaa gaccaccctggccaacatggtgaagtcccgtctctactaaaaatacaaaaatcagct gggcatggtggcgggcacttgtagtcccagctactcaggaggctgaagcaggagaa tcacttgaactcaggaggcagaagttgcggttagccga
153	220703_at	C10orf110	cctctctccactctctagaaatattaaggctaggctgctgctgtatgtcagggctagtc ccctcttctatgaatccagaataactctgaagaagccgagtaacaggcatgaagtga agagaaatcgctgtaacaggaagacagcaaagcagatgctaatgaccacactattt aacgaactggaaccaacgagaaaatacggtattactgaagactgcacttccttgaa cagagtgctcttctcagcaaatcggaaatgcctacacaaatcgctttacaagaaaga ctgtttcaaagcagcacctttctcaatgttctcgttcaggtgacaattcttcttggtctc agctccaattttattgtcattttcatcaataaggatacacatctctgccaggagttgaa cctgttgcttgtcgaggtggttagtgtttatttcaggcatcattacaaaatgtctgatct gttctagaaccct

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
154	220771_at	LOC51152	aagtatctccatacaaaatacggttgaattacaaaaagaaaattgtaacattagcat ggacaaacctggcaggtactccttaactctcctaagtaataaaaactgtaaaatgca aataagccttcgatgacatttactaacctttactaaagtatcaatgatgacttggttgt ttaaacagctgacatttgggcaatttgagtatgtcaaactcaataatactggttttcat ttgcaagatccacttaaaacttaaggaggccaaaaaacatcatttaaaatacectat aaattataatcatacatatgatacgaaaaatatcctacttcag
155	220817_at	TRPC4	catacacatacgtattttccgtagtgctctgggtgggggaaaatgtttaaattgtatta gcaaatgctaacttacactttatagcatttatcagctgtggcatattacctgtaacatg tttaaattaaggcaaaggcaatcaaaaacctttttgttttgtagcctgcttttgctttca caatttgtcttacaatt
156	220834_at	MS4A12	gctggccaagactactgggccgtgctttctggaaaaggcatttcagccacgctgatg atcttctccctcttggagttcttcgtagcttgtgccacagcccatttttgccaaccaagca aacaccacaaccaatatgtctgtcctggttattccaaatatgtatgaaagcaaccctg tgacaccagcgtcttcttcagctcctcccagatgcaacaactactcagctaatgcccc taaatagtaaaagaaaaaggggtatcagtctaatctcatggagaaaaactacttgc aaaaacttcttaagaagatgtcttttattgtctacaatgatttctagtctttaaaaactg tgtttgagatttgtttttaggttggtcgctaatgatggctgtatctcccttcactgtctctt cctacattaccactactacatgctggcaaaggtgaaggatcagaggactgaaaaat gattctgcaactctcttaaa
157	220847_x_at	ZNF221	tgacatgcaccagagggtccacaggggagagcgaccctataattgtaaggaatgtg gaaagagctttggctgggcttcatgtcttttgaaacatcagagactccacagtggag aaaagccattgaaatctggagtgtgggaagagatctactcagaattcacagcttcat ttacatcagtaagtctatgtgggagaaaagccatataaatgtgagaagtgtgggaa gggctttggctgggcctcaactcatctgacccatcaattctccacagcagagaaaaa ccattcaaatatgagaactgtgggaagagctttgtacatagatcatatctttttttttt ttttgagacagagtctcactctttcacccaagcctgactgcagtggcg
158	220852_at	PRO1768	gaaaagcgccctgtgctgagtaaagcagccagtcttctcttgtcacagtaaaaggct gggagtaaaatttcccataaacacaggggaaacctacatttactcacatgccaagg aaaatggcacggaagacccacgtgtagccacagcagagtctatgcagagggcctgc aaatgcctggggtgcgagtgaatgcctggaggggcggagtttccaagataacagct attgtgttttctttttcacacttcagaagagaatcctaaggactagactccgctcagtg cattccttttttcatacactgatctcaagtacaatcacataattttgaaaatccatgtagt cctccctaaataaaattataaggataggtttctatttccttccgattacctagatacctc cgtcttctggaaaaccccaaaaagaccagtagacgaatcaggaaggtcctaggagt gattcctccaat

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
159	220970_s_at	KAP2.1B /// KRTAP2-4 /// LOC644350 /// LOC728285 /// LOC728934 /// LOC730755	tgcccccacagagcaatacactgaagcctaaacatctatctggtgtttttaaaaagtt aaaagaaaaatagatttttttttcacaaggtgacaatagtgatttttaccatctggata cagcctggtgtaagcagacgtccattaccaccctcacccacattttcaggtgtctaca tcagccttagtcattatggatagtaaatcgacctttaagaattcctggggtggactttg caaacacattctacaacctgatggtttttactgctcaaactgtcaccatcatcttttgca atgtgttgctcactgttgtcaata
160	220981_x_at	LOC650686 /// NXF2 /// NXF2B	ggacagtctcagggttctgttctcgccttcacccggaccttcattgctacccctggcag cagttccagtctgtgcatcgtgaatgacgagctgtttgtgagggatgccagcccccaa gagactcagagtgccttctccatcccagtgtccacactctcctccagctctgagccctc cctctcccaggagcagcaggaaatggtgcaggctttctctgcccagtctgggatgaa actggagtggtctcagaagtgccttcaggacaatgagtggaactacactagagctgg ccaggccttcactatgctccagaccgagggcaagatccccgcagaggccttcaagca aatctcctaaaaggagccctccgatgtcttctttgtcttcgttcacatcctctttgtttcc tcttttcaccagcctaaggcctggctgaccaggaagccaacgttaacttgcaggcca cgtgacataac
161	220993_s_at	GPR63	aagtctgcattgaatccgctgatctactactggaggattaagaaattccatgatgctt gcctggacatgatgcctaagtccttcaagtttttgccgcagctccctggtcacacaaa gcgacggatacgtcctagtgctgtctatgtgtgtggggaacatcggacggtggtgtg aatattggaactggctgacattttgggtgatgcttgttctttattgacattgaattctctt tctcatagcctctccactttatttttttttataggtttgtgtatgtatgtgtgtgagcagt gtaaagaaagaatggtaattatagttctgttaccaagaataaataataggaaagtg attacaaatattacctccagggttcaatagaaatcctcaatttagggtgaggagactt ttttttggttttggggtttttccttgattgattttgttttcatagtgggaatcaggattgtg ctttattgagcctgcagttacattgaattgtaggtgtttcgtgtgctgctaaggta
162	221018_s_at	TDRD1	gggactgtcgatgtagctgataagctagtgacatttggtctggcaaaaaacatcaca cctcaaaggcagagtgctttaaatacagaaaagatgtataggacgaattgctgctgc acagagttacagaaacaagttgaaaaacatgaacatattcttctcttcctcttaaaca attcaaccaatcaaaataaatttattgaaatgaaaaaactggtaaaaagttaagtaa gttaaatcgtatgttttcgcctcttctgtgatcaccaataggacatcttcaggcatattg gcaggatagagctaatggagtgaaacctattgtaaggctgtactttcgtgatttaatg acctgaggtttggtcataatgcttctgctgtttttgtaggtttatctgatcgttttcctttg ctactgctaatggaactgaacccccaggggtattccagttgtaatagcctttccttact gttgtttgg

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
163	221077_at	ARMC4	gttgagttgaaattctgccgcttactcaatggccttgggtgatgatgctgtaccctaat tctaaaggaagcaatgaaccccctttcagctaccttactgataagcacttatgttctg ccttctgctatcctgatggttcgggttgtctgtcttactatctacttcttgagtagagag accacattaaatttattgctgtatctcacagggcatcttgctagtgtgcacaggctcgc ctccctacctctgccccgatggtgtgaaggggagagggcgaggttccttagtggcag ggctttgctgttcttcactctcagccccctgaaagcagttcttcctgcctctgagcctgt ctttccttctgctgttaacttctttcctacttttcttgcatccctctcccttccttttcctgcc gtctttcttgtagacat
164	221137_at	---	aaaaggactaactcacatggctgcagtaagtgctggctgttagctggaagcacaac caaggctgttaacaggtgtgccttggttctcttccatatggcttctcttttgttttcagta ctctgcagtttaattatgatgcatgcaggtgtgaatttctgtttattctgcttgggatgt gttttccttctgggatctgtgaatcggtttctcattatttttgtaaaacctgaagccagtt atctcttaaaataccagctctccttg
165	221168_at	PRDM13	ctggacttcttggatgagctcaccctgaaccgcccaggcggtctgctcttggtgttcag aatcacatcaatgcgaacgtcacagcgccttcgagggcgcagattttaactgccacg tatttttaagttgtacttttctgtggaggaaattgtgccttttgaaacgacgttttgtgtg tgtatttcacgttagcatttcattgcataggcaaaacactagtcacaattgggtagat gtgacatccatatacttgtttacattttatctgttctcatgtcaaagactactccttgccc cattgaatatatagtggtagcaggtgtacaaattggtcaagttgcaattatttatgag agaataatgataaatgtaaaatatctaaagcatgaatctaagagcacgcaatatat aattttaaagaaaatattctatttggtagaatacaaatgtggtgtgtgttgttttataat gactgctgtacagtgggtatagtattttggttttggttccagattgtgcaatc
166	221258_s_at	KIF18A	gtgaagacatcaagagctcgaagtgtaaattacccgaacaagaatcactaccaaat gataacaaagacattttacaacggcttgatccttcttcattctcaactaagcattctat gcctgtaccaagcatggtgccatcctacatggcaatgactactgctgccaaaaggaa acggaaattaacaagttctacatcaaacagttcgttaactgcagacgtaaattctgg atttgccaaacgtgttcgacaagataattcaagtgagaagcacttacaagaaaaca aaccaacaatggaacataaaagaaacatctgtaaaataaatccaagcatggttaga aaatttggaagaaatatttcaaaaggaaatctaagataaatcacttcaaaaccaag caaaatgaagttgatcaaatctgcttttcaaagtttatcaataccctttcaaaaatata tttaaaatctttgaaagaagacccatcttaaagctaagtttacccaagtactttcagc aagc
167	221319_at	PCDHB8	cgggagcctgtctcagaactatcagtacgaggtgtgcctggcaggaggctcaggga cgaatgagttccagttcctgaaaccagtattacctaatattcagggccattcttttggg ccagaaatggaacaaaactctaactttaggaatggctttggtttcagccttcagttaa agta

Table 11 (Continued)

SEQ ID NO:	Probe Set ID	Gene Symbol	Target Sequence
168	221393_at	TAAR3	gaactccaccataaagcaactgctggcattttgctggtcagttcctgctctttttttctttt ggtttagttctatctgaggccgatgtttccggtatgcagagctataagatacttgttgc ttgcttcaatttctgtgcccttactttcaacaaattctgggggacaatattgttcactac atgtttctttaccccctggctccatcatggttggtatttatggcaaaatctttatcgtttcc aaacagcatgctcgagtcatcagccatgtgcctgaaaacacaaaggggcagtgaa aaaacacctatccaagaaaaaggacaggaaagcagcgaagacactgggtatagta atgggggtgtttctggcttgctggttgccttgttttcttgctgttctgattgacccatacc tagactactccactcccatactaatattggatcttttagtgtggctccggtacttcaact ctacttgcaaccctcttattcatggctttttaatccatggtttcagaaagcattcaagt acatagtgtcaggaaaaatatttagctcccattcagaaactgc
169	221591_s_at	FAM64A	cacatctggacccatcagtgactgcctgccatagcctgagagtgtcttggggagacct tgcagagggggagaattgttccttctgctttcctaggggactcttgagcttagaaactc atcgtacacttgaccttgagccttctatttgcctcatctataacatgaagtgctagcat cagatatttgagagctcttagctctgtacccgggtgcctggtttttggggagtcatccg cagagtcactcacccactgtgtttctggtgccaaggctcttgagggccccactctcatc cctcctttccctaccagggactcggaggaaggcataggagatatttccaggcttacg accctgggctcacgggtacctatttatatgctcagtgcagagcactgtggatgtgcca ggaggggtagccctgttcaagagcaatttctgccctttgtaaattatttaagaaacct gctttgtcattttattagaaagaaaccagcgtgtgactttcctagataacactgctttc
170	221609_s_at	WNT6	ccgccaggagagcgtgcagctcgaagagaactgcctgtgccgcttccactggtgctg cgtagtacagtgccaccgttgccgtgtgcgcaaggagctcagcctctgcctgtgaccc gccgcccggccgctagactgacttcgcgcagcggtggctcgcacctgtgggacctca gggcaccggcaccgggcgcctctcgccgctcgagcccagcctctccctgccaaagcc caactcccagggctctggaaatggtgaggcgaggggcttgagaggaacgcccaccc acgaaggcccagggcgccagacggccccgaaaaggcgctcggggagcgtttaaag gacactgtacaggccctccctcccctggcctctaggaggaaacagtttttagactg gaaaaaagccagtctaaaggcctctggatactgggctccccagaactgc
171	221718_s_at	AKAP13	gcgatgcagaaatgaaccaccggagttcaatgcgagttcttggggatgttgtcagga gacctcccattcataggagaagtttcagtctagaaggcttgacaggaggagctggtg tcggaaacaagccatcctcatctctagaagtaagctctgcaaatgccgaagagctca gacacccattcagtggtgaggaacgggttgactctttggtgtcactttcagaagagga tctggagtcagaccagagagaacataggatgtttgatcagcagatatgtcacagatc taagcagcagggatttaattactgtacatcagccatttcctctccattgacaaaatcc atctcattaatgacaatcagccatcctggattggacaattcacggccctt
172	221950_at	EMX2	gtaggctcagcgatagtggtcctcttacagagaaacggggagcaggacgacgggg gngctggggntggcgggggagggtgcccacaaaaagaatcaggacttgtactggg aaaaaaaccccctaaattaattatatttcttggacattccctttcctaacatcctgaggc ttaaaaccctgatgcaaacttctccctttcagtggttggagaaattggccgagttcaac cattcactgcaatgcctattccaaactttaaatctatctattgcaaaacctgaaggact gtagttagcggggatgatgttaagtgtggccaagcgcacggcggcaagttttcaagc actgagtttctattccaagatcatagacttactaaagagagtgacaaatgcttcctta atgtcttctataccagaatgtaaatatttttgtgttttgtgttaatttgttagaattctaa cacactatatacttccaa

Table 12. Validation of the independent prognostic value of the 15-gene signature in four other separate stage IB-II patient cohorts who received no adjuvant treatment

Trial/Source	Tumour Type	n	Hazard Ratio	95% CI	p value
JBR.10	All NSCLC	62	18.00	5.78-56.05	<0.0001
DCC	ADC	96	2.26	1.02 – 4.97	0.044
NLCI	All NSCLC	133	2.27	1.18 – 4.35	0.014
Duke	All NSCLC	48	1.96	0.87 – 4.42	0.11
UM-SQ	SQC	79	3.57	1.48 – 8.58	0.005

HR: hazard ratio; OBS: observation; NSCLC: non-small cell lung cancer; ADC: adenocarcinoma; SQC: squamous cell carcinoma; DCC Director's Challenge Consortium adenocarcinoma dataset; NLCI: Netherlands Cancer Institute; Duke: Duke University; UM-SQ: University of Michigan, squamous cell carcinoma dataset.

Table 13. Demographic features of patients in the four validation sets of stage IB and II patients.

	Director's Challenge (DCC)				NLCI	Duke	UM-SQ
Clinical Factors	All n=96 (%)	UM n=27 (%)	HLM n=38 (%)	MSK n=31 (%)	n=133 (%)	n=48 (%)	n=79 (%)
Pathologic subtype							
Adeno	96 (100)	27 (100)	38 (100)	31 (100)	39 (29)	18 (38)	0
Non-Adeno	0 (0)	0 (0)	0 (0)	0 (0)	94 (71)	30 (62)	79 (100)
Stage							
IB	68 (71)	17 (63)	29 (76)	22 (71)	78 (59)	30 (63)	46 (59)
II	28 (29)	10 (37)	9 (24)	9 (29)	55 (41)	18 (37)	33 (41)
Age (years)							
<65	40 (42)	14 (52)	14 (37)	12 (39)	68 (51)	20 (42)	26 (33)
≥65	56 (58)	13 (48)	24 (63)	19 (61)	65 (49)	28 (58)	53 (67)
Sex							
Male	49 (51)	16 (59)	21 (55)	12 (39)	NA	32 (67)	49 (62)
Female	47 (49)	11 (41)	17 (45)	19 (61)	NA	16 (33)	30 (38)

DCC: Directors' Challenge Consortium; UM: University of Michigan; HLM: H. Lee Moffitt Cancer Center; MSK: Memorial Sloan-Kettering Cancer Center; NLCI: Netherlands Cancer Institute.

*Only stage IB-II patients who did not receive adjuvant therapy of any type (chemotherapy or radiotherapy); NA: not available.

5

Table 14. Demographic features of patients in UHN183 validation set (stage I and II) and the training set (BR10 – OBS).

Clinical factors – A comparative table of the 2 datasets (training and current validation)

	MedBioGene N=183 N (%)	BR10 – OBS N= 62 N (%)
Age		
Median (range)	70 (40-88)	61.2 (35.4-76.7)
<65	60 (33)	44 (69)
>=65	123 (67)	19 (31)
Sex		
Women	84 (46)	18 (29)
Men	99 (54)	44 (71)
Stage		
1A	49 (27)	
1B	80 (44)	34 (55)
2A	9 (5)	28* (45)
2B	45 (25)	
3A		
Histology		
Adenocarcinoma (ADE)	130 (71)	32 (52)
Squamous (SQC)	43 (24)	26 (42)
Adenosuamous (ASQ)	2 (1)	
Large cell (LC)	8 (4)	
Other		4 (6)
15 gene signature		
Low risk	90 (49)	29 (47)
High risk	93 (51)	33 (53)

* Stage 2 or higher

Table 15. Best performing custom assays

	FFPE		UHRR/pancreas	
Assay	PCR efficiency (%)	R2	PCR efficiency (%)	R2
ATP1B1	101.5	0.99	98	0.99
TRIM14	94.5	0.97	107	0.99
FAM64A	NA	NA	102	0.98
FOSL2	98.5	0.98	95	0.99
HEXIM1	93.5	0.91	99	0.99
MB	103.34	0.91	99.2	0.98
L1CAM	93.92	0.87	87.1	0.8
UMPS	96.5	0.89	97	0.99
EDN3	116	0.82	94.6	0.98
STMN2	102	0.94	104	0.97
IKBKAP	95.5	0.96	97	0.99
MDM2	91.5	0.92	107	0.99
ZNF236	100	-.86	96	0.99

Table 16. Primer and Probe Designs

AssayID	Forward Primer*	SEQ ID NO:	Reverse Primer*	SEQ ID NO:	Probe	SEQ ID NO:	Vendor
L1CAM_LNA3	**CGGCTACTCTGGAGAGGACTAC**	203	**CGGCACTTGAGTTGAGGATT**	205	(L1CAM -LNA3 #11)(Cat#04685105001)		Roche
IDT MB_2	AGTTGGTGCTGAACGTCTG	174	GGTGACCCTTAAAGAGCCTG	187	CTGGGATGTCAGCCTCCACCTTC	207	IDT
IDT EDN3_2	GTGTCTACTATTGCCACCTGG	175	GCTTCCTCTGTAGTTGGACAG	188	CGTCTGTTCGGGAGTGTTGATCCAAA	208	IDT
IDT ATP1B1_1	CCAAATGTCCTTCCCGTTCAG	182	TGCCCAGTCCAAAATACTCC	195	ATCCTTATCTTCATCTCGCTTGCCAGTG	209	IDT
IDT FAM64_3	TCCTCAACAGAGCCCCTC	173	TTCTGGAGATGCTGAATTCCC	186	CGAGTCTGACAGTGACCTAGAGCCT	210	IDT
FOSL2_LNA3	ACGCCGAGTCCTACTCCA	177	TGAGCCAGGCATATCTACCC	190	(FOSL2 - LNA3 #70)(Cat#04688937001)		Roche
IDT HEXIM1_2	TCGAGGACTCTACTAGCCATG	183	GCTCTTCCTGGACAGCAG	196	TCAACACCAGCCTCAAACTAGCAACT	211	IDT
IKBKAP_LNA3	GGAAGCAAGCCCTCTGTGT	184	TCTGCTCAACCAGCTTTCCT	197	(IKBKAP - LNA3 #26)(cat#04687574001)		Roche
IDT MDM2_2	**GAAAACCCCGGATGGTGAG**	204	**CGAAGCTGGAATCTGTGAGG**	206	CAACATGTCTGTACCTACTGATGGTGCT	212	IDT
IDT STMN2_1	GAACCTCGCAACATCAACATC	180	GGCTTCAAGATCAGCTCAAAAG	193	AGCAAATCAACAAACGTGCCTCTGG	213	IDT
TRIM14_LNA3	CCTCAGCATCCTGAATACATCA	179	TGTGTCAGGTTCCTGTGCTG	192	(TRIM14 - LNA3 #7)(cat#04685059001)		Roche
IDT UMPS_3	ACACAGTGAAAAAGCAGTATGAAG	181	TTTCACAACTCCTGAGCCTG	194	CCACGTGAGCATTTACTAGATCTGCCC	214	IDT
IDT ZNF236_3	ACGTAGACCAGTTTGAAGAGC	176	GGCTGCTGATGGTACGTATC	189	TGCTGGTTCGAAGGACTGTTGCG	215	IDT
*Sequences in bold are different from what are shown in Table 7.							

References:

1. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer Statistics, 2007. CA Cancer J Clin 2007;57:43-66.

2. Arriagada R, Bergman B, Dunant A, Le Chevalier T, Pignon JP, Vansteenkiste J. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. N Engl J Med 2004;350:351-60.

3. Winton T, Livingston R, Johnson D, et al. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. N Engl J Med 2005;352:2589-97.

4. Douillard JY, Rosell R, De Lena M, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. Lancet Oncol 2006;7:719-27.

5. Strauss GM, Herndon JE, II, Maddaus MA, et al. Adjuvant chemotherapy in stage IB non-small cell lung cancer (NSCLC): Update of Cancer and Leukemia Group B (CALGB) protocol 9633. ASCO Meeting Abstracts 2006;24:7007-.

6. Pignon JP, Tribodet H, Scagliotti GV, et al. Lung Adjuvant Cisplatin Evaluation (LACE): A pooled analysis of five randomized clinical trials including 4,584 patients. ASCO Meeting Abstracts 2006;24:7008-.

7. Scagliotti GV, Fossati R, Torri V, et al. Randomized study of adjuvant chemotherapy for completely resected stage I, II, or IIIA non-small-cell Lung cancer. J Natl Cancer Inst 2003;95:1453-61.

8. Waller D, Peake MD, Stephens RJ, et al. Chemotherapy for patients with non-small cell lung cancer: the surgical setting of the Big Lung Trial. Eur J Cardiothorac Surg 2004;26:173-82.

9. Douillard JY, Rosell R, Delena M, Legroumellec A, Torres A, Carpagnano F. ANITA: Phase III adjuvant vinorelbine (N) and cisplatin (P) versus observation (OBS) in completely resected (stage I-III) non-small-cell

lung cancer (NSCLC) patients (pts): Final results after 70-month median follow-up. On behalf of the Adjuvant Navelbine International Trialist Association. ASCO Meeting Abstracts 2005;23:7013-.

10. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. Lancet 2000;355:479-85.

11. Nesbitt JC, Putnam JB, Jr., Walsh GL, Roth JA, Mountain CF. Survival in early-stage non-small cell lung cancer. Ann Thorac Surg 1995;60:466-72.

12. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816-24.

13. Chen HY, Yu SL, Chen CH, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 2007;356:11-20.

14. Lu Y, Lemon W, Liu PY, et al. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. PLoS Med 2006;3:e467.

15. Potti A, Mukherjee S, Petersen R, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. N Engl J Med 2006;355:570-80.

16. Raponi M, Zhang Y, Yu J, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res 2006;66:7466-72.

17. Wigle DA, Jurisica I, Radulovich N, et al. Molecular profiling of non-small cell lung cancer and correlation with disease-free survival. Cancer Res 2002;62:3005-8.

18. Bianchi F, Nuciforo P, Vecchi M, et al. Survival prediction of stage I lung adenocarcinomas by expression of 10 genes. J Clin Invest 2007;117:3436-44.

19. Sun Z, Wigle DA, Yang P. Non-overlapping and non-cell-type-specific gene expression signatures predict lung cancer survival. J Clin Oncol 2008;26:877-83.

20. Lau SK, Boutros PC, Pintilie M, et al. Three-gene prognostic classifier for early-stage non small-cell lung cancer. J Clin Oncol 2007;25:5562-9.

21. Oshita F, Ikehara M, Sekiyama A, et al. Genomic-wide cDNA microarray screening to correlate gene expression profile with chemoresistance in patients with advanced lung cancer. J Exp Ther Oncol 2004;4:155-60.

22. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185-93.

23. Affymetrix, ed. Transcript assignment for NetAffx™ annotation; 2006.

24. Dworakowska D, Jassem E, Jassem J, et al. Clinical significance of apoptotic index in non-small cell lung cancer: correlation with p53, mdm2, pRb and p21WAF1/CIP1 protein expression. J Cancer Res Clin Oncol 2005;131:617-23.

25. Allory Y, Matsuoka Y, Bazille C, Christensen EI, Ronco P, Debiec H. The L1 cell adhesion molecule is induced in renal cancer cells and correlates with metastasis in clear cell carcinomas. Clin Cancer Res 2005;11:1190-7.

26. Boo YJ, Park JM, Kim J, et al. L1 expression as a marker for poor prognosis, tumor progression, and short survival in patients with colorectal cancer. Ann Surg Oncol 2007;14:1703-11.

27. Gast D, Riedle S, Schabath H, et al. L1 augments cell migration and tumor growth but not beta3 integrin expression in ovarian carcinomas. Int J Cancer 2005;115:658-65.

28. Thies A, Schachner M, Moll I, et al. Overexpression of the cell adhesion molecule L1 is associated with metastasis in cutaneous malignant melanoma. Eur J Cancer 2002;38:1708-16.

29. Ouellet V, Provencher DM, Maugard CM, et al. Discrimination between serous low malignant potential and invasive epithelial ovarian tumors using molecular profiling. Oncogene 2005;24:4672-87.

30. Andersen, C. L., J. L. Jensen, et al. (2004). "Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets." Cancer Res 64(15): 5245-50.

31. Barsyte-Lovejoy, D., S. K. Lau, et al. (2006). "The c-Myc Oncogene Directly Induces the H19 Noncoding RNA by Allele-Specific Binding to Potentiate Tumorigenesis." Cancer Res 66(10): 5330-5337.

32. Der, S. D., M. Pintilie, et al. (2010). Validation of a prognostic gene signature for early stage non-small cell lung cancer in an independent cohort. AACR- IASLC Joint Conference on Molecular Origins of Lung Cancer. San Diego, IASLC.

33. Drury, S., H. Anderson, et al. (2009). "Selection of REFERENCE genes for normalization of qRT-PCR data derived from FFPE breast tumors." Diagn Mol Pathol 18(2): 103-7.

34. Jemal, A., R. Siegel, et al. (2010). "Cancer Statistics, 2010." CA Cancer J Clin.

35. Landis, J. R. and G. G. Koch (1977). "The measurement of observer agreement for categorical data." Biometrics 33(1): 159-74.

36. Mountain, C. F. (1997). "Revisions in the International System for Staging Lung Cancer." Chest 111(6): 1710-7.

37. Nguewa, P. A., J. Agorreta, et al. (2008). "Identification of importin 8 (IPO8) as the most accurate reference gene for the clinicopathological analysis of lung specimens." BMC Mol Biol 9: 103.

38. Perez-Novo, C. A., C. Claeys, et al. (2005). "Impact of RNA quality on reference gene expression stability." Biotechniques 39(1): 52, 54, 56.

39. Saviozzi, S., F. Cordero, et al. (2006). "Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer." BMC Cancer 6: 200.

40. Shrout, P. E. and J. L. Fleiss (1979). "Intraclass correlations: uses in assessing rater reliability." Psychol Bull 86(2): 420-8.

41. Vandesompele, J., K. De Preter, et al. (2002). "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes." Genome Biol 3(7): RESEARCH0034.

42. Zhu, C. Q., K. Ding, et al. (2010). "Prognostic and Predictive Gene Signature for Adjuvant Chemotherapy in Resected Non-Small-Cell Lung cancer." J Clin Oncol.

Claims:

1. A method for predicting benefit of adjuvant chemotherapy for a non-small cell lung cancer (NSCLC) patient, comprising:

   determining a gene expression profile from a sample of the patient's lung tumor, the gene expression profile comprising the level of expression of from 5 to 14 genes each correlative with survival in NSCLC, and being listed in Table 3; and

   classifying the gene expression profile as being predictive of a positive response to adjuvant chemotherapy or as being predictive of non-responsiveness to adjuvant chemotherapy.

2. The method of claim 1, wherein the gene expression profile comprises the expression level of from 5 to 13 genes that are correlative with survival in NSCLC, the genes being selected from Table 4A.

3. The method of claim 1, wherein the gene expression profile does not include expression level of one or both of MLANA and MYT1L.

4. The method of any one of claims 1 to 3, wherein the gene expression profile further comprises the level of expression of one or more normalization genes in the sample.

5. The method of claim 4, wherein the normalization gene(s) include one or more of BAT1, TBP, PP1A, and GUSB.

6. The method of claim 5, wherein a normalization gene is GUSB.

7. The method of any one of claims 1 to 6, wherein the gene expression profile is determined by quantitative RT-PCR.

8. The method of any one of claims 1 to 7, wherein the sample is a frozen tumor tissue specimen, or is a formalin-fixed paraffin-embedded tumor tissue specimen.

9. The method of claim 7, wherein the RT-PCR amplifies at least one target sequence listed in Table 11.

10. The method of claim 7, wherein the quantitative PCR amplifies at least one target sequence listed in Table 9.

11. The method of claim 7, wherein the quantitative PCR employs at least one primer or primer set of Table 7 or Table 16.

12. The method of any one of claims 1 to 11, wherein a product of amplification is detected with a probe selected from Table 16.

13. A method for preparing a gene expression profile indicative of response to adjuvant chemotherapy for non-small cell lung cancer (NSCLC), comprising:

determining the level of expression of at least 5 genes from Table 4A in a formalin-fixed paraffin-embedded tumor tissue sample from a lung cancer patient;

wherein the expression levels are determined by quantitative RT-PCR, and wherein the gene expression profile includes the expression level of fewer than 15 genes that are correlative with survival in NSCLC.

14. The method of claim 13, wherein the gene expression profile contains the expression level of from 5 to 13 genes selected from Table 4A.

15. The method of claim 13, wherein the gene expression profile does not include the expression level of one or both of MLANA and MYT1L.

16. The method of any one of claims 13 to 15, wherein the gene expression profile further comprises the level of expression of one or more normalization genes in the sample.

17. The method of claim 16, wherein the normalization gene(s) include one or more of BAT1, TBP, PP1A, and GUSB.

18. The method of claim 13, wherein a normalization gene is GUSB.

19. The method of any one of claims 13 to 18, wherein the quantitative RT-PCR amplifies a sequence that comprises at least one target sequence listed in Table 11.

20. The method of claim 19, wherein the quantitative RT-PCR amplifies at least one target sequence listed in Table 9.

21. The method of claim 20, wherein the quantitative RT-PCR employs at least one primer or primer set of Table 7 or Table 16.

22. The method of any one of claims 13 to 21, wherein the amplification product is detected with a probe selected from Table 16.

23. The method of any one of claims 1 to 22, wherein the tumor tissue sample is from a stage I or stage II NSCLC patient.

24. A method of prognosing or classifying a subject with non-small cell lung cancer (NSCLC) comprising the steps:

    a. determining the expression of thirteen biomarkers in a test tumor tissue sample from the subject, wherein the biomarkers correspond to genes in Table 4A, and

    b. comparing the expression of the thirteen biomarkers in the test sample with expression of the thirteen biomarkers in a control sample,

wherein a difference or a similarity in the expression of the thirteen biomarkers between the control and the test sample is used to prognose or classify the subject with NSCLC into a poor survival group or a good survival group.

25. A method of predicting prognosis in a subject with non-small cell lung cancer (NSCLC) comprising the steps:

a. obtaining a subject biomarker expression profile in a tumor tissue sample of the subject;

b. obtaining a biomarker reference expression profile associated with a prognosis, wherein the subject biomarker expression profile and the biomarker reference expression profile each have thirteen values, each value representing the expression level of a biomarker, wherein each biomarker corresponds to one gene in Table 4A; and

c. selecting the biomarker reference expression profile most similar to the subject biomarker expression profile, to thereby predict a prognosis for the subject.

26. The method of claim 25, wherein the biomarker reference expression profile comprises a poor survival group or a good survival group.

27. The method of any one of claims 24 to 26 wherein the NSCLC is stage I or stage II.

28. The method of any one of claims 24 to 27, wherein determining the biomarker expression level comprises use of quantitative PCR or an array.

29. The method of claim 28, wherein the use of quantitative PCR comprises use of primers set out in Table 7 or Table 16.

30. The method of claim 29, wherein the array is a U133A chip.

31. The method of any one of claims 24 to 27, wherein determining the biomarker expression profile comprises use of an antibody to detect polypeptide products of the biomarker.

32. The method of any one of claims 24 to 31, wherein the sample comprises a tissue sample suitable for immunohistochemistry.

33. The method of claim 32, wherein the sample is a formalin-fixed paraffin-embedded (FFPE) tumor specimen.

34. A method of selecting a therapy for a subject with NSCLC, comprising the steps:

    a. classifying the subject with NSCLC into a poor survival group or a good survival group according to the method of any one of claims 22 to 31; and

    b. selecting adjuvant chemotherapy for the poor survival group or no adjuvant chemotherapy for the good survival group.

35. A method of selecting a therapy for a subject with NSCLC, comprising the steps:

    a. determining the expression of thirteen biomarkers in a test tumor tissue sample from the subject, wherein the thirteen biomarkers correspond to the thirteen genes in Table 4A;

    b. comparing the expression of the thirteen biomarkers in the test sample with the thirteen biomarkers in a control sample;

    c. classifying the subject in a poor survival group or a good survival group, wherein a difference or a similarity in the expression of the thirteen biomarkers between the control sample and the test sample is used to classify the subject into a poor survival group or a good survival group;

    d. selecting adjuvant chemotherapy if the subject is classified in the poor survival group and selecting no adjuvant chemotherapy if the subject is classified in the good survival group.

36. A composition comprising a plurality of isolated nucleic acid sequences, wherein each isolated nucleic acid sequence hybridizes to:

    a. a RNA product of 13 genes listed in Table 4A; and/or

    b. a nucleic acid complementary to a),

    wherein the composition is used to measure the level of RNA expression of the 13 genes.

37. The composition of claim 36, wherein the plurality of isolated nucleic acid sequences comprise isolated nucleic acid sequences hybridizable to 13 probe target sequences as set out in Table 9.

38. A composition comprising 26 primers for quantitative PCR as set out in Table 7 or Table 16.

39. An array comprising for each gene shown in Table 4A one or more polynucleotide probes complementary and hybridizable to an expression product of the gene.

40. A computer program product for use in conjunction with a computer having a processor and a memory connected to the processor, the computer program product comprising a computer readable storage medium having a computer mechanism encoded thereon, wherein the computer program mechanism may be loaded into the memory of the computer and cause the computer to carry out the method of any one of claims 25 to 34.

41. A computer implemented product for predicting a prognosis or classifying a subject with NSCLC comprising:

    a. a means for receiving values corresponding to a subject expression profile in a subject sample; and

b. a database comprising a reference expression profile associated with a prognosis, wherein the subject biomarker expression profile and the biomarker reference profile each has thirteen values, each value representing the expression level of a biomarker, wherein each biomarker corresponds to one gene in Table 4A;

wherein the computer implemented product selects the biomarker reference expression profile most similar to the subject biomarker expression profile, to thereby predict a prognosis or classify the subject.

42. A computer implemented product of claim 39 for use with the method of any one of claims 25 to 35.

43. A computer implemented product for determining therapy for a subject with NSCLC comprising:

a. a means for receiving values corresponding to a subject expression profile in a subject sample; and

b. a database comprising a reference expression profile associated with a therapy, wherein the subject biomarker expression profile and the biomarker reference profile each has thirteen values, each value representing the expression level of a biomarker, wherein each biomarker corresponds to one gene in Table 4A;

wherein the computer implemented product selects the biomarker reference expression profile most similar to the subject biomarker expression profile, to thereby predict the therapy.

44. The computer implemented product of claim 41 for use with the method of claim 25 or 26.

45. A computer readable medium having stored thereon a data structure for storing the computer implemented product of any one of claims 41 to 44.

46. The computer readable medium according to claim 45, wherein the data structure is capable of configuring a computer to respond to queries based on records belonging to the data structure, each of the records comprising:

    a. a value that identifies a biomarker reference expression profile of the 13 genes in Table 4A;

    b. a value that identifies the probability of a prognosis associated with the biomarker reference expression profile.

47. A computer system comprising

    a. a database including records comprising a biomarker reference expression profile of thirteen genes in Table 4A associated with a prognosis or therapy;

    b. a user interface capable of receiving a selection of gene expression levels of the 13 genes in Table 4A for use in comparing to the biomarker reference expression profile in the database;

    c. an output that displays a prediction of prognosis or therapy according to the biomarker reference expression profile most similar to the expression levels of the thirteen genes.

48. A kit comprising a primer set for amplifying a target sequence in each of from 5 to 14 genes from Table 3, and including at least 5 genes from Table 4A.

49. The kit of claim 48, wherein the primer set contains primer pairs for amplifying between 5 and 13 genes from Table 4A.

50. The kit of claim 48 or 49, further comprising a primer set for amplifying at least one normalization gene.

51. The kit of any one of claims 48 to 50, further comprising, at least one probe for detecting each target sequence.

52. A kit to prognose or classify a subject with early stage NSCLC, comprising detection agents that can detect the expression products of 13 biomarkers, wherein the 13 biomarkers comprise 13 genes in Table 4A and instructions for use.

53. A kit to select a therapy for a subject with NSCLC, comprising detection agents that can detect the expression products of 13 biomarkers, wherein the 13 biomarkers comprise 13 genes in Table 4A and instructions for use.

54. The kit of claim 52 or 53, wherein the detection agents are probes hybridizable to the probe target sequences as set out in Table 9.

55. The kit of claims 52 or 53, wherein the detection agents are primers for quantitative PCR as set out in Table 7 or Table 16.

56. A method for prognosing or classifying a subject with NSCLC comprising:

a. calculating a combined score from relative expression levels of at least 13 different biomarkers in a tumor tissue sample from the subject, wherein the at least 15 biomarkers comprise FAM64A, MB, EDN3, ZNF236, FOSL2, L1CAM, TRIM14, STMN2, UMPS, ATP1B1, HEXIM1, IKBKAP, and MDM2, and

b. classifying the subject into a high or low risk group based on the combined score.

57. The method of claim 56 wherein the combined score is calculated from the relative expression levels of FAM64A, MB, EDN3, ZNF236, FOSL2,

L1CAM, TRIM14, STMN2, UMPS, ATP1B1, HEXIM1, IKBKAP, and MDM2.

58. The method of claim 56, wherein the combined score is calculated from the relative expression levels of 14 different biomarkers, wherein the one additional biomarker is selected from the genes listed in Table 3.

59. The method of claim 58, wherein the additional one biomarker is selected from the group consisting of RGS4, UGT2B4, and MCF2.

60. A method for prognosing or classifying a subject with NSCLC comprising:

a. determining relative expression levels of at least 13 different biomarkers, wherein the biomarkers comprise FAM64A, MB, EDN3, ZNF236, FOSL2, L1CAM, TRIM14, STMN2, UMPS, ATP1B1, HEXIM1, IKBKAP, and MDM2,

b. calculating a combined score from the relative expression levels of at least 13 different biomarkers in the subject, and

c. classifying the subject into a high or low risk group based on the combined score.

61. The method of claim 56 to 60, wherein the relative expression levels of thirteen or fourteen different biomarkers selected from the group consisting of FAM64A, MB, EDN3, ZNF236, FOSL2, L1CAM, TRIM14, STMN2, UMPS, ATP1B1, HEXIM1, IKBKAP, MDM2, RGS4, UGT2B4, and MCF2 are determined.

62. The method according to any one of claims 56 to 61, wherein the combined score is calculated according to Formula I.

63. The method of any one of claims 56 to 61 wherein the subject is a human.

64. A method for selecting therapy comprising the steps of claim 56 or 60, and further comprising selecting adjuvant chemotherapy for a subject in the high risk group or no adjuvant chemotherapy for a subject in the low risk group, wherein the subject is a human.

65. A kit to prognose or classify a subject with NSCLC comprising detection agents capable of detecting the expression product of at least 13 different biomarkers wherein the at least 13 different biomarkers comprise FAM64A, MB, EDN3, ZNF236, FOSL2, L1CAM, TRIM14, STMN2, UMPS, ATP1B1, HEXIM1, IKBKAP, and MDM2.

66. The kit of claim 65, comprising detection agents capable of detecting the expression product of 14 different biomarkers, wherein the additional one biomarker is selected from the genes listed in Table 3.

67. The kit of claim 65, comprising detection agents capable of detecting the expression products of 13 or 14 different biomarkers, selected from the group consisting of FAM64A, MB, EDN3, ZNF236, FOSL2, L1CAM, TRIM14, STMN2, UMPS, ATP1B1, HEXIM1, IKBKAP, MDM2, RGS4, UGT2B4, and MCF2.

68. The kit of claim 65, further comprising an addressable array comprising probes for the expression products of the at least 13 biomarkers.

69. The kit of claim 65, wherein the detection agents comprise primers capable of hybridizing to the expression products of at least 13 biomarkers.

70. The kit of claim 65, wherein the detection agents comprise primers capable of hybridizing to the expression products of 14 biomarkers.

71. A kit according to any one of claims 65 to 70, further comprising a computer implemented product for calculating a combined score for a subject.

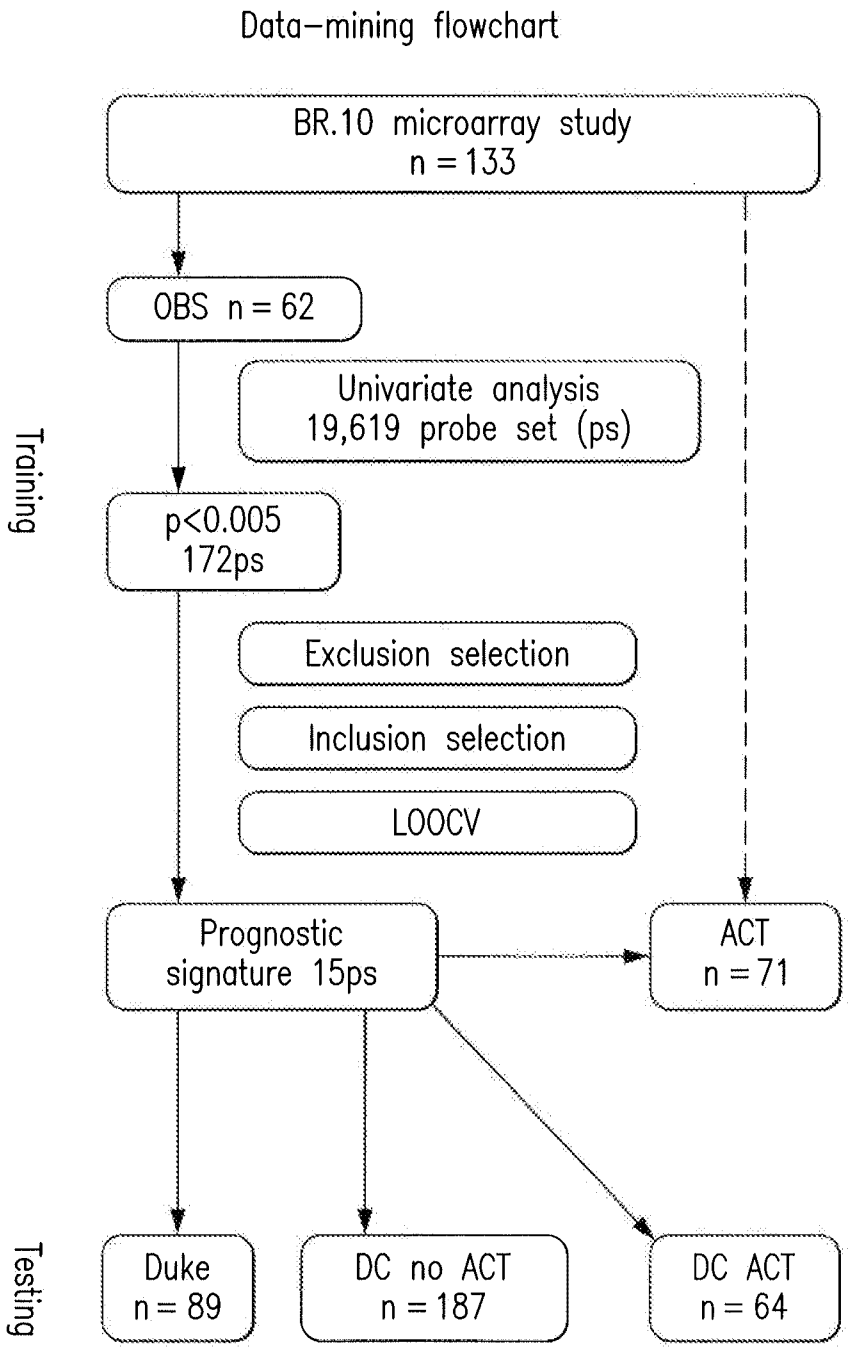

Data-mining flowchart
BR.10 microarray study
n = 133
OBS n = 62
Univariate analysis
19,619 probe set (ps)
p<0.005
172ps
Exclusion selection
Inclusion selection
LOOCV
Prognostic
signature 15ps
ACT
n = 71
Duke
n = 89
DC no ACT
n = 187
DC ACT
n = 64
Training
Testing

FIG. 1A

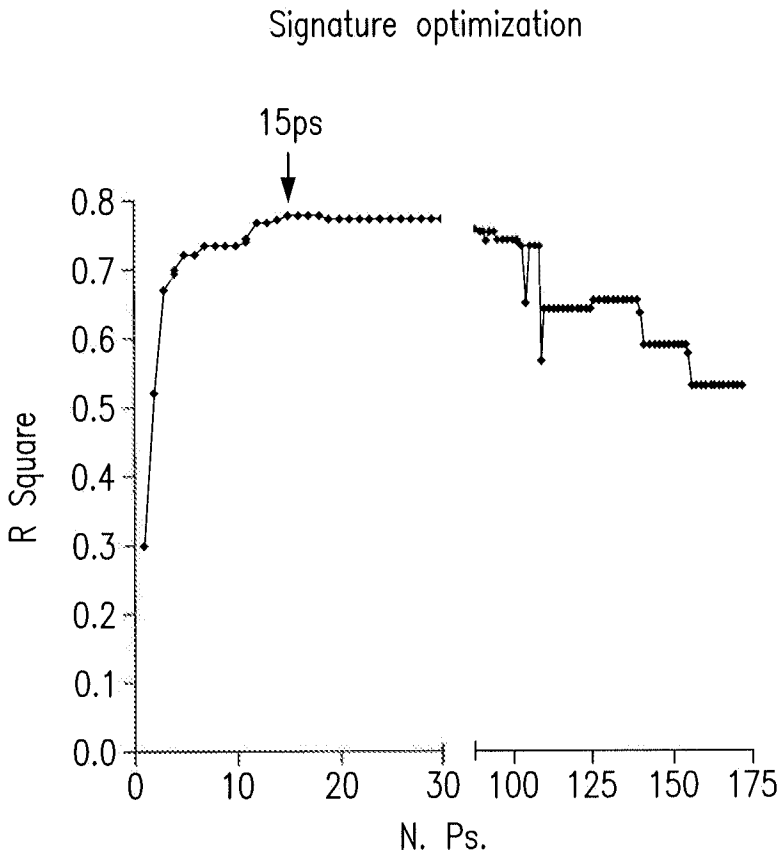
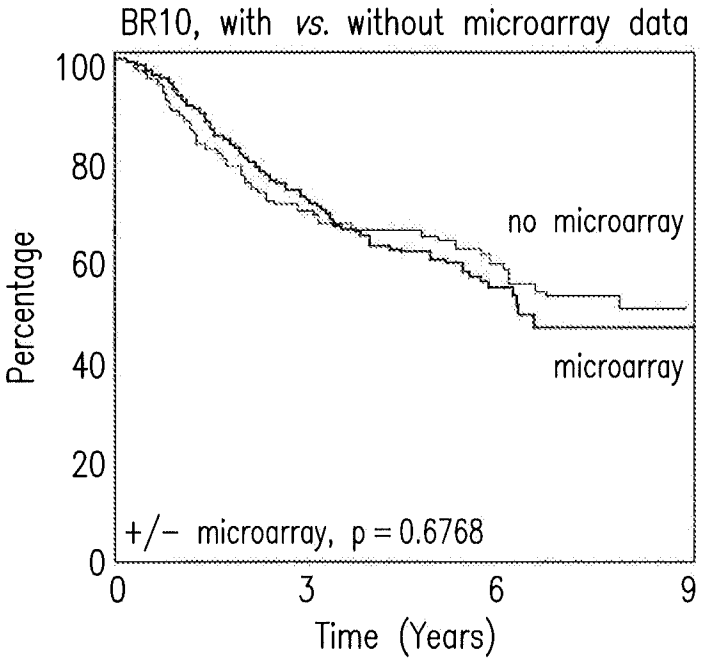
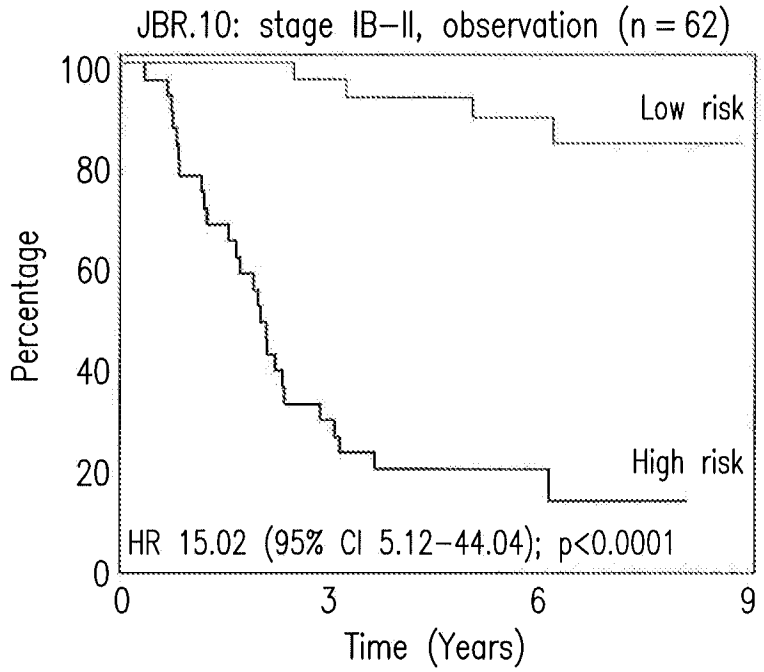

Signature optimization
15ps
R Square
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
100
125
150
175
N. Ps.

FIG. 1B



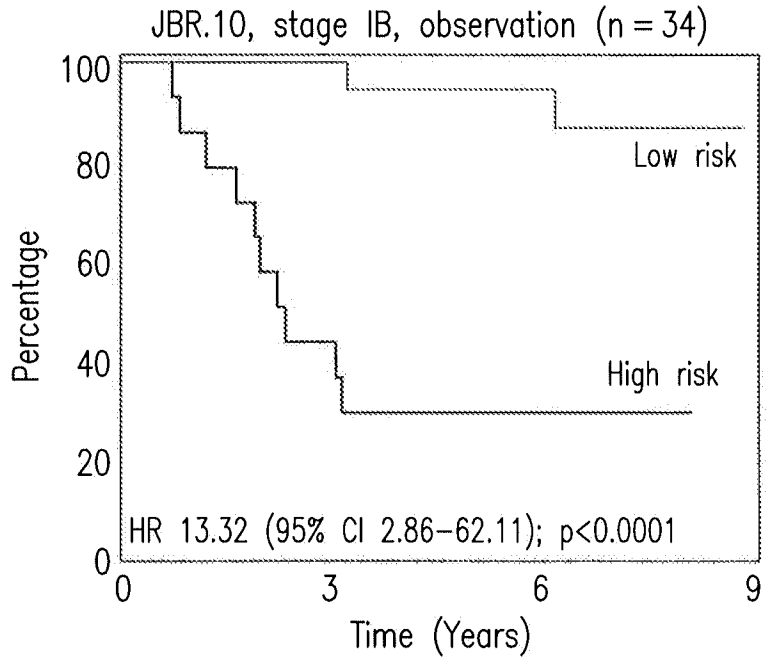
No microarray	349	219	41	0
Microarray	133	90	57	5

FIG. 2A



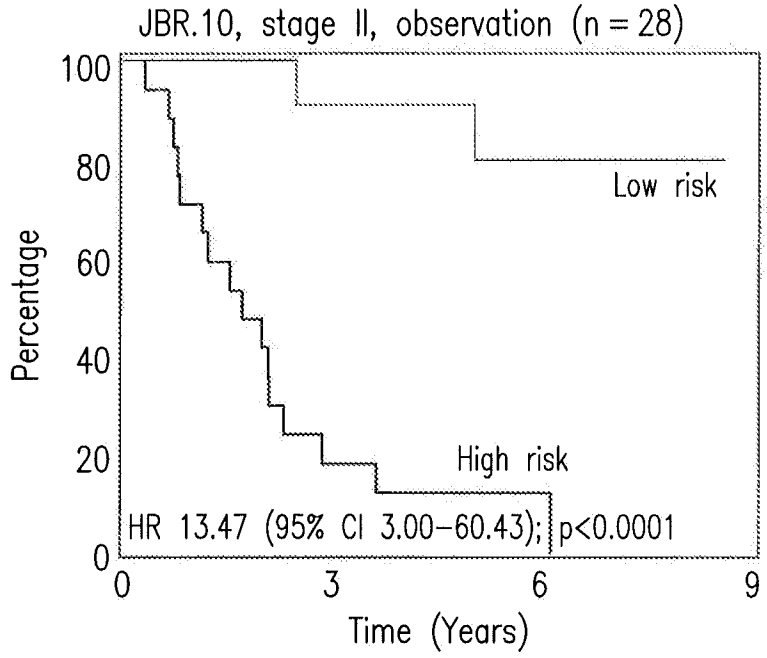
Low Risk	31	28	20	1
High Risk	31	9	3	0

FIG. 2B



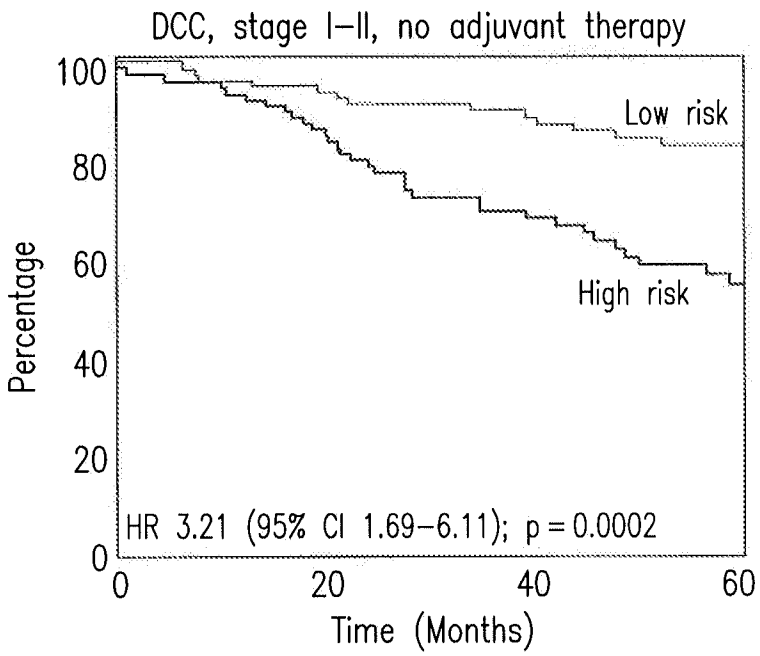
Low Risk	20	19	13	1
High Risk	14	6	2	0

FIG. 2C



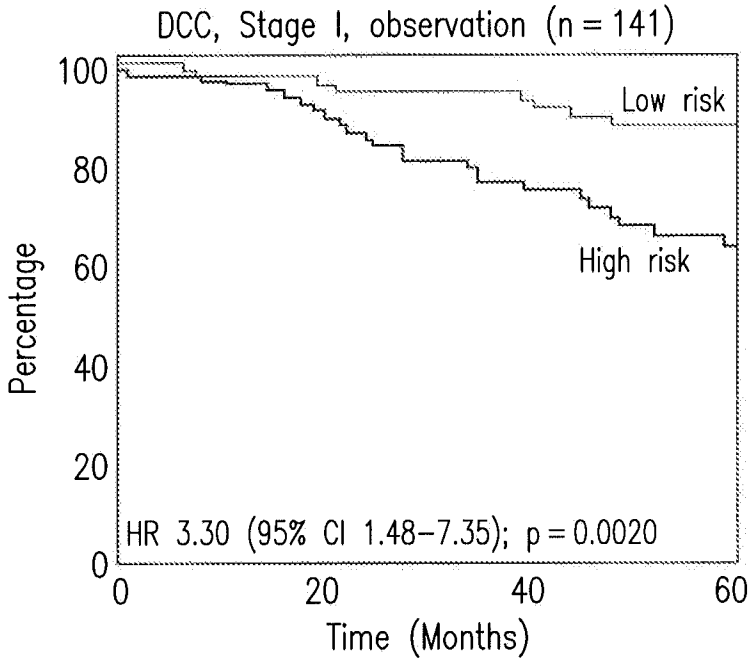
Low Risk	11	9	7	0
High Risk	17	3	1	0

FIG. 2D



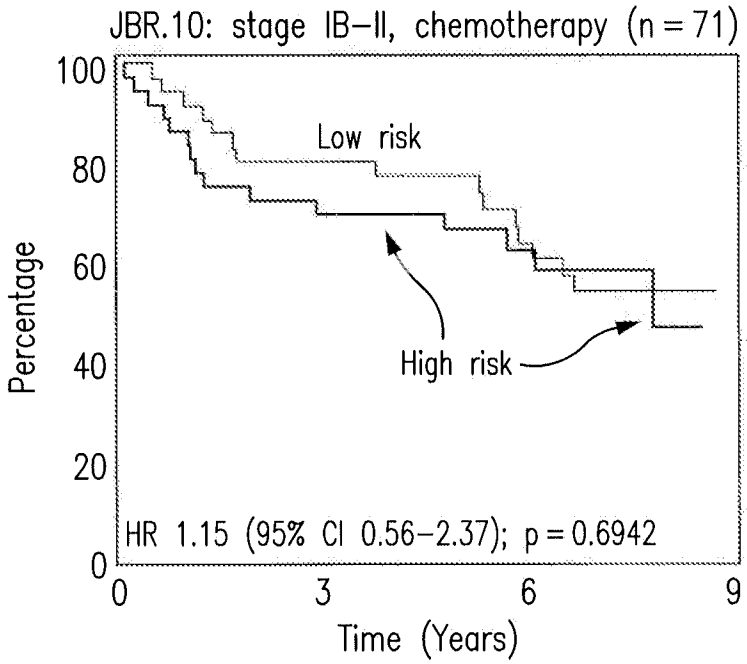
Low Risk	87	81	65	49
High Risk	82	69	48	28

FIG. 2E



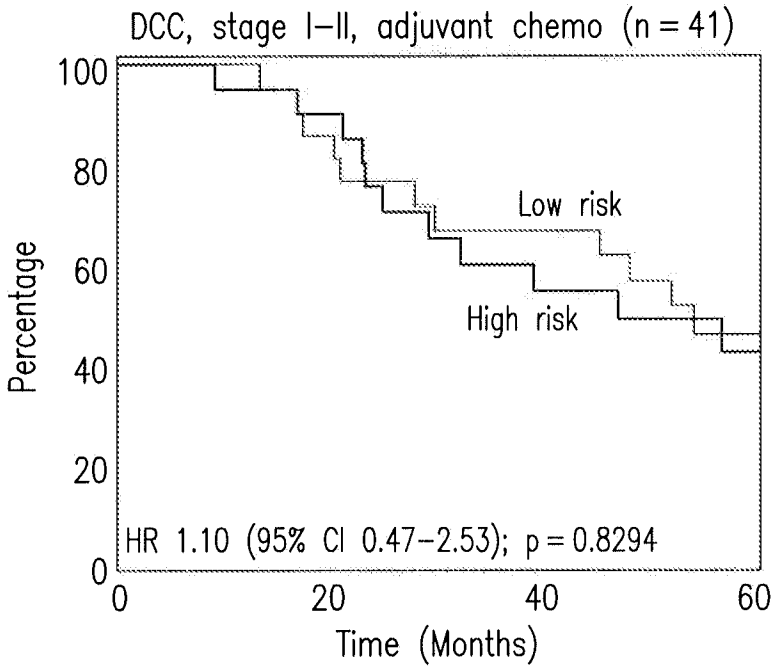
Low Risk	69	65	54	42
High Risk	72	63	46	28

FIG. 2F



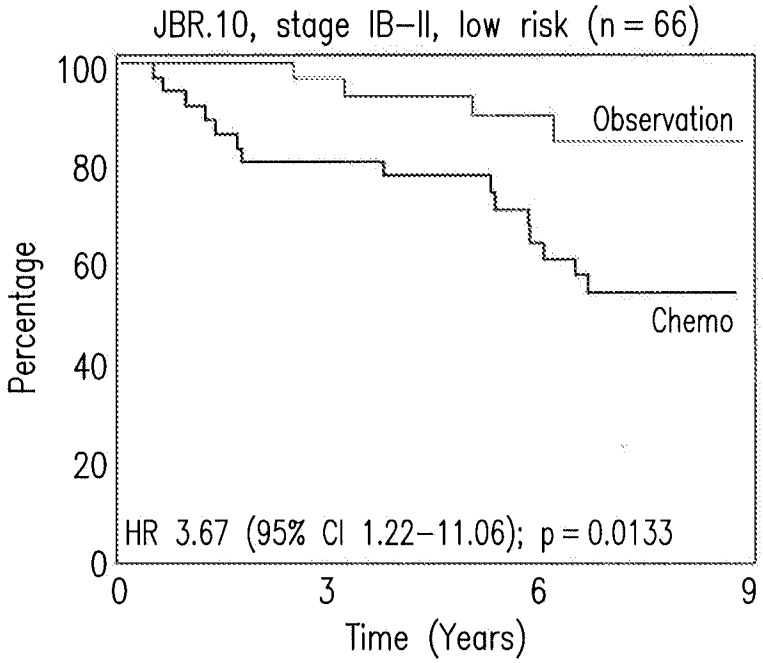
Low Risk	35	28	19	3
High Risk	36	25	15	1

FIG. 3A



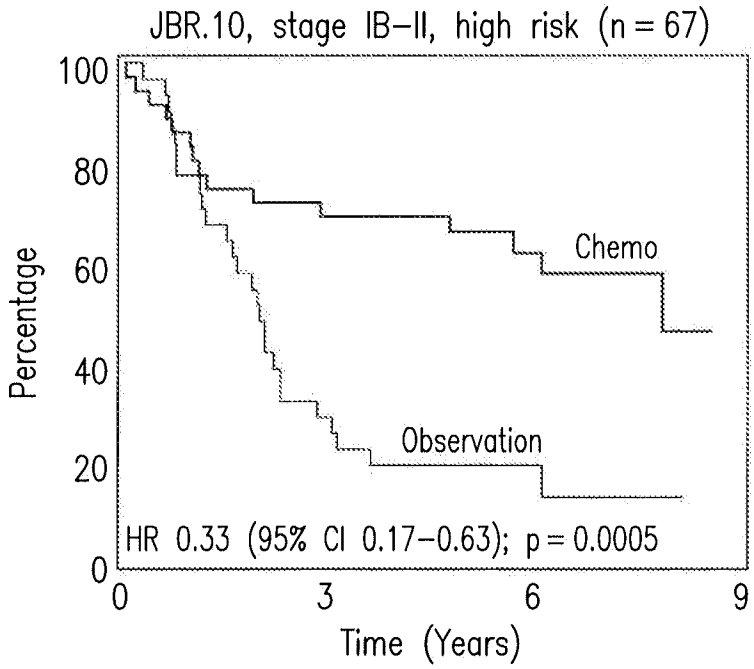
Low Risk	21	18	13	8
High Risk	20	18	10	6

FIG. 3B



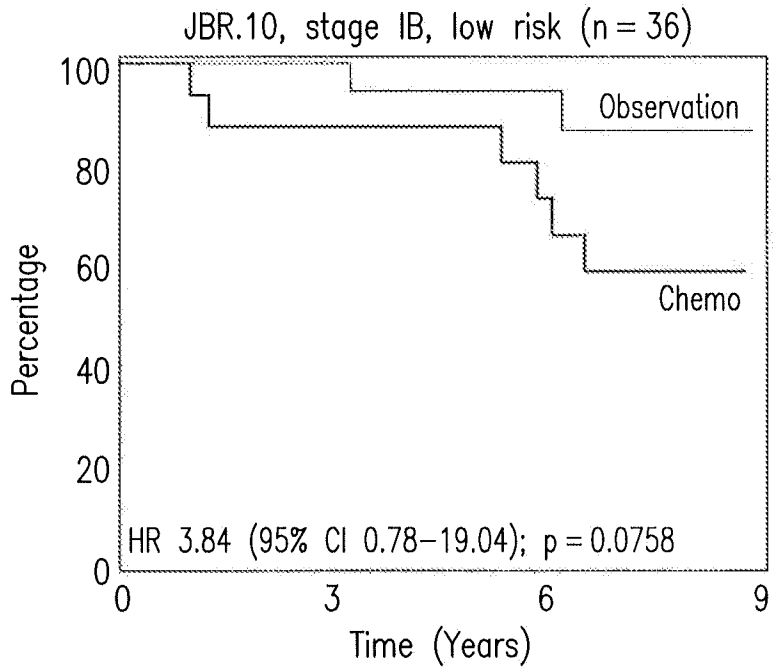
Observation	31	28	20	1
Chemo	35	28	19	3

FIG. 3C



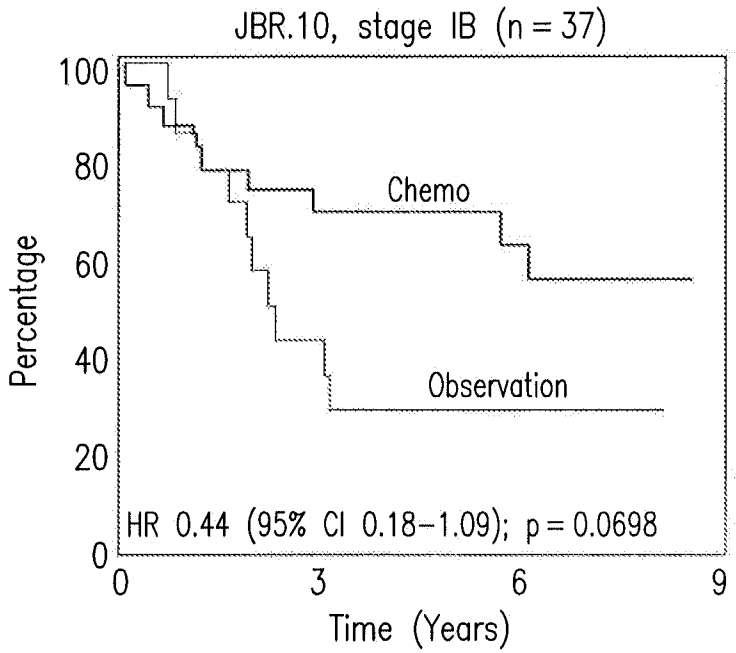
Observation	31	9	3	0
Chemo	36	25	15	1

FIG. 3D



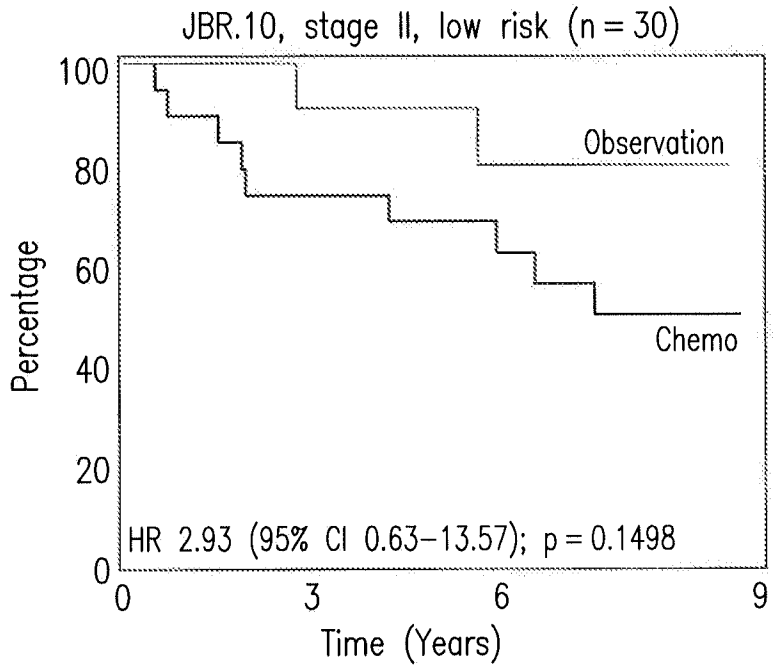
Observation	20	19	13	1
Chemo	16	14	10	1

FIG. 3E



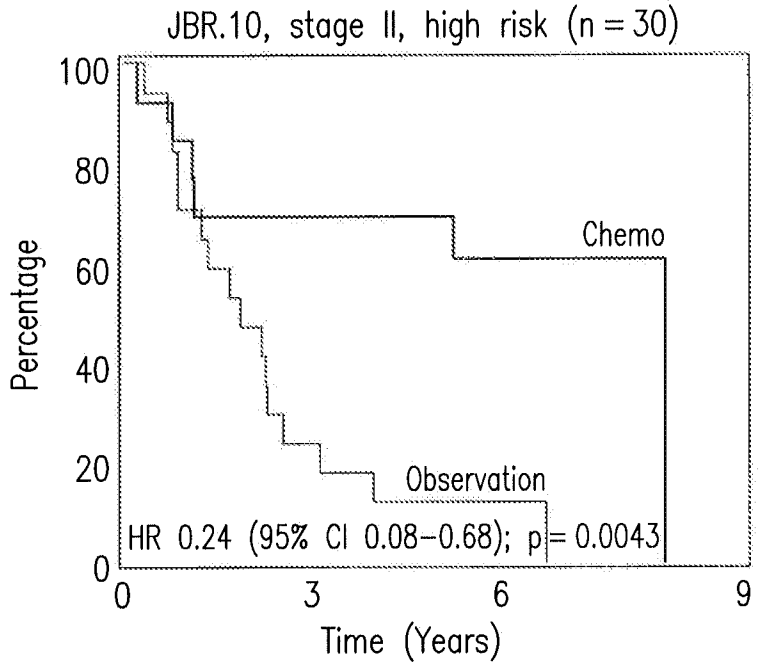
Observation	14	6	2	0
Chemo	23	16	9	1

FIG. 3F



Observation	11	9	7	0
Chemo	19	14	9	2

FIG. 3G



Observation	17	3	1	0
Chemo	13	9	6	0

FIG. 3H

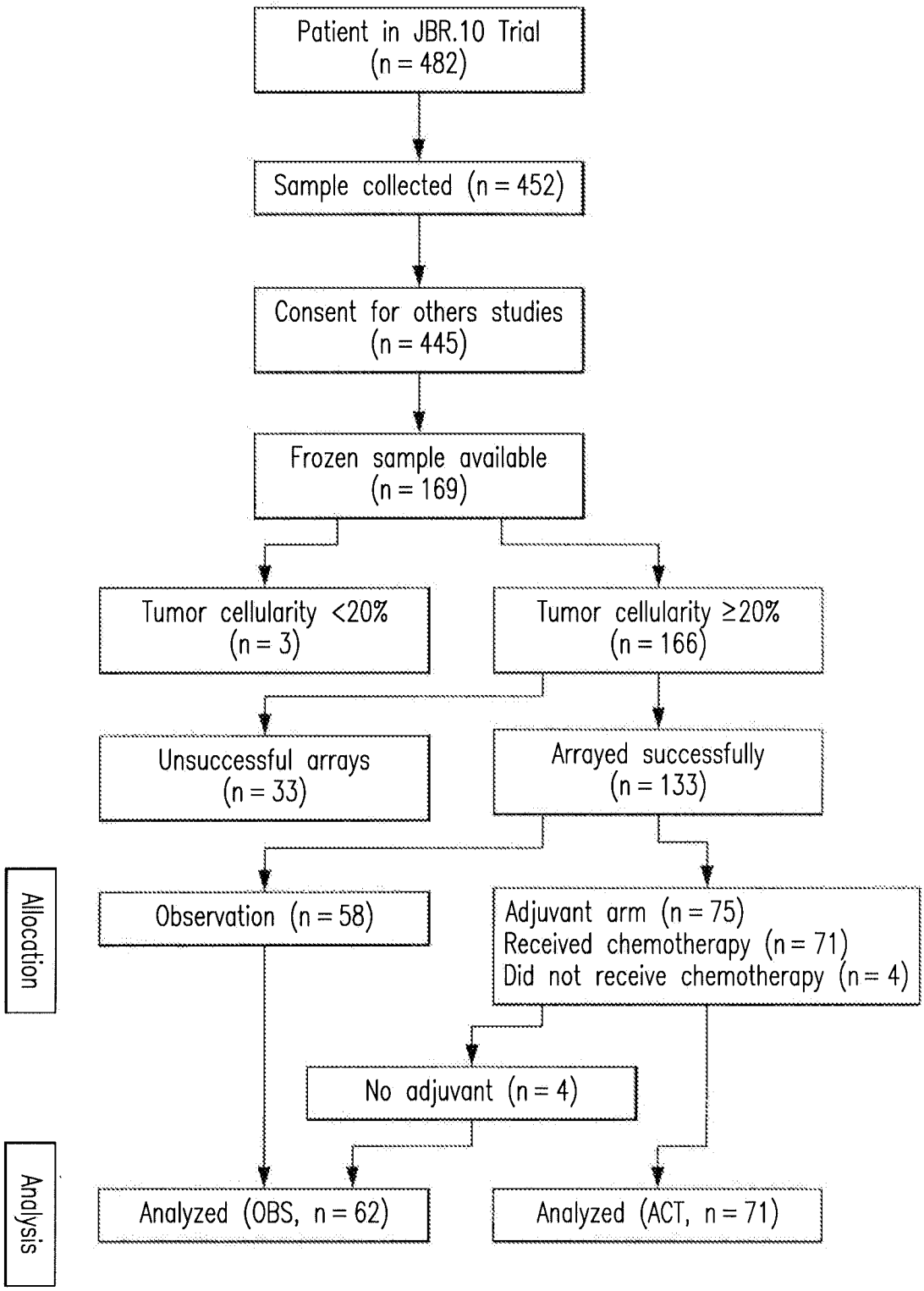

Patient in JBR.10 Trial (n = 482)
Sample collected (n = 452)
Consent for others studies (n = 445)
Frozen sample available (n = 169)
Tumor cellularity <20% (n = 3)
Tumor cellularity ≥20% (n = 166)
Unsuccessful arrays (n = 33)
Arrayed successfully (n = 133)
Allocation
Observation (n = 58)
Adjuvant arm (n = 75)
Received chemotherapy (n = 71)
Did not receive chemotherapy (n = 4)
No adjuvant (n = 4)
Analysis
Analyzed (OBS, n = 62)
Analyzed (ACT, n = 71)

FIG. 4

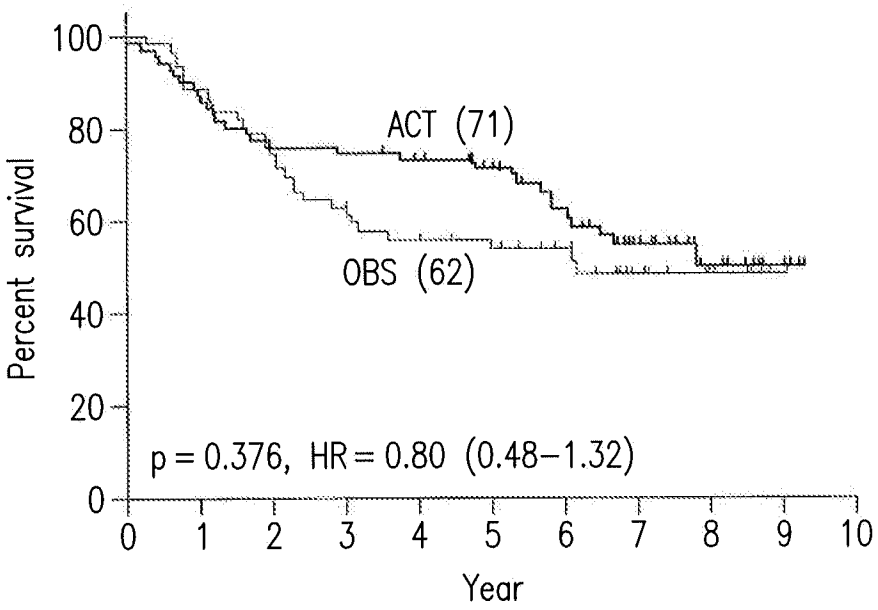


FIG. 5

Cluster	January 2004 (%)	June 2005 (%)
1	103 (93)	6 (27)
2	6 (5)	16 (73)
Outlier	2 (2)	
Total	111	22

FIG. 6

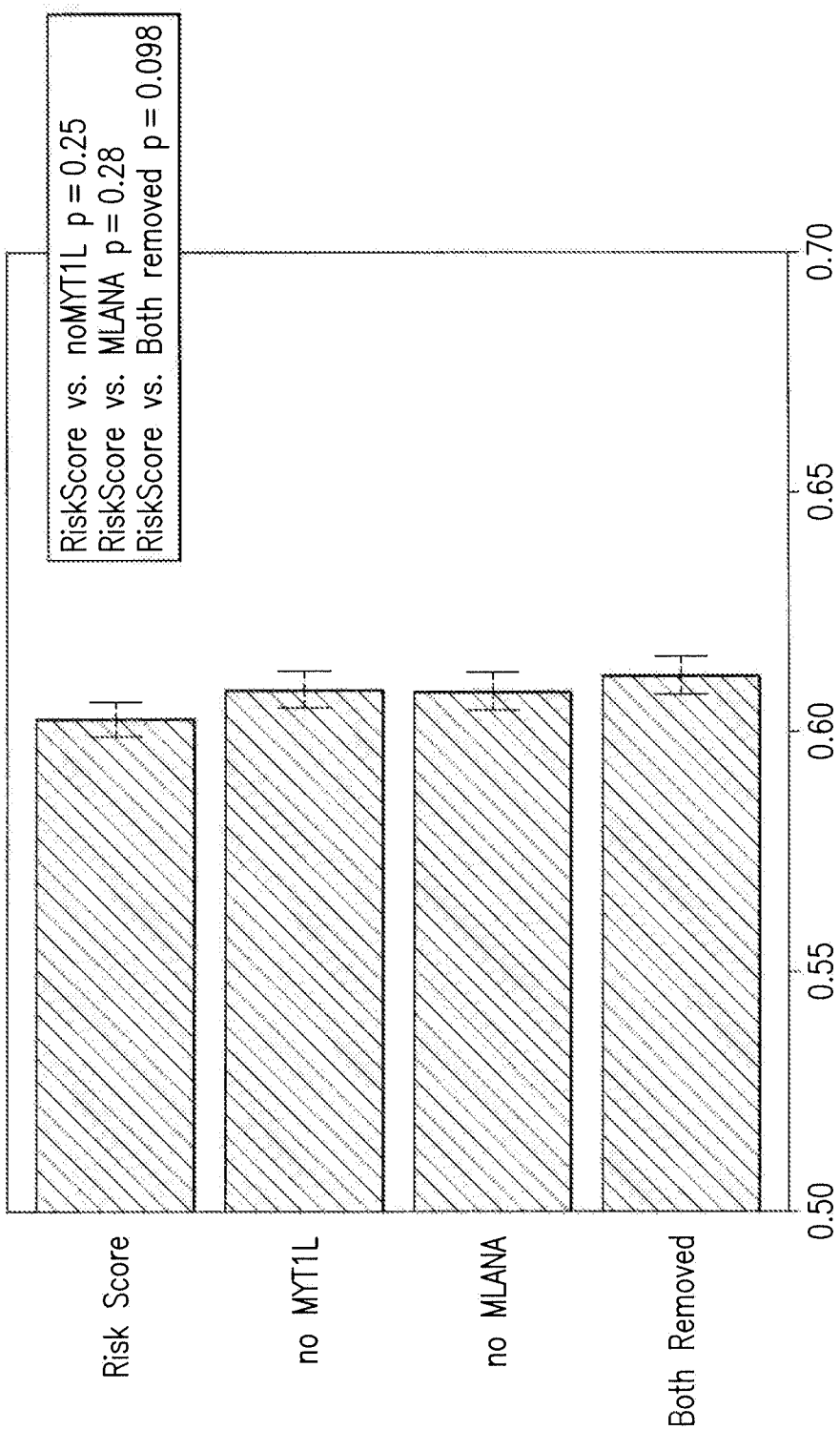

RiskScore vs. noMYT1L p = 0.25
RiskScore vs. MLANA p = 0.28
RiskScore vs. Both removed p = 0.098
0.50
0.55
0.60
0.65
0.70
C Index
Risk Score
no MYT1L
no MLANA
Both Removed

FIG. 7

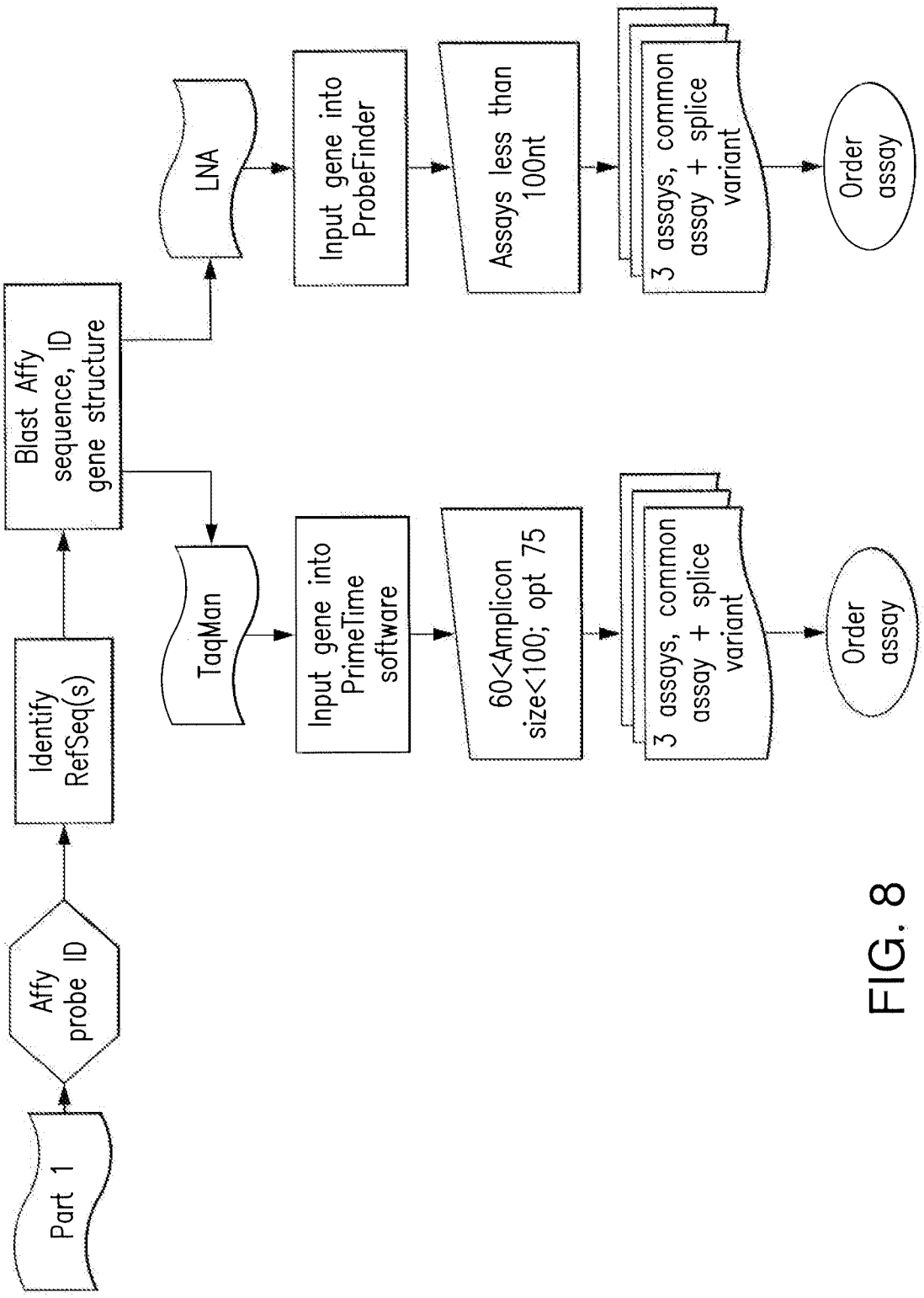

Part 1
Affy probe ID
Identify RefSeq(s)
Blast Affy sequence, ID gene structure
LNA
Input gene into ProbeFinder
Assays less than 100nt
3 assays, common assay + splice variant
Order assay
TaqMan
Input gene into PrimeTime software
60<Amplicon size<100; opt 75
3 assays, common assay + splice variant
Order assay

FIG. 8

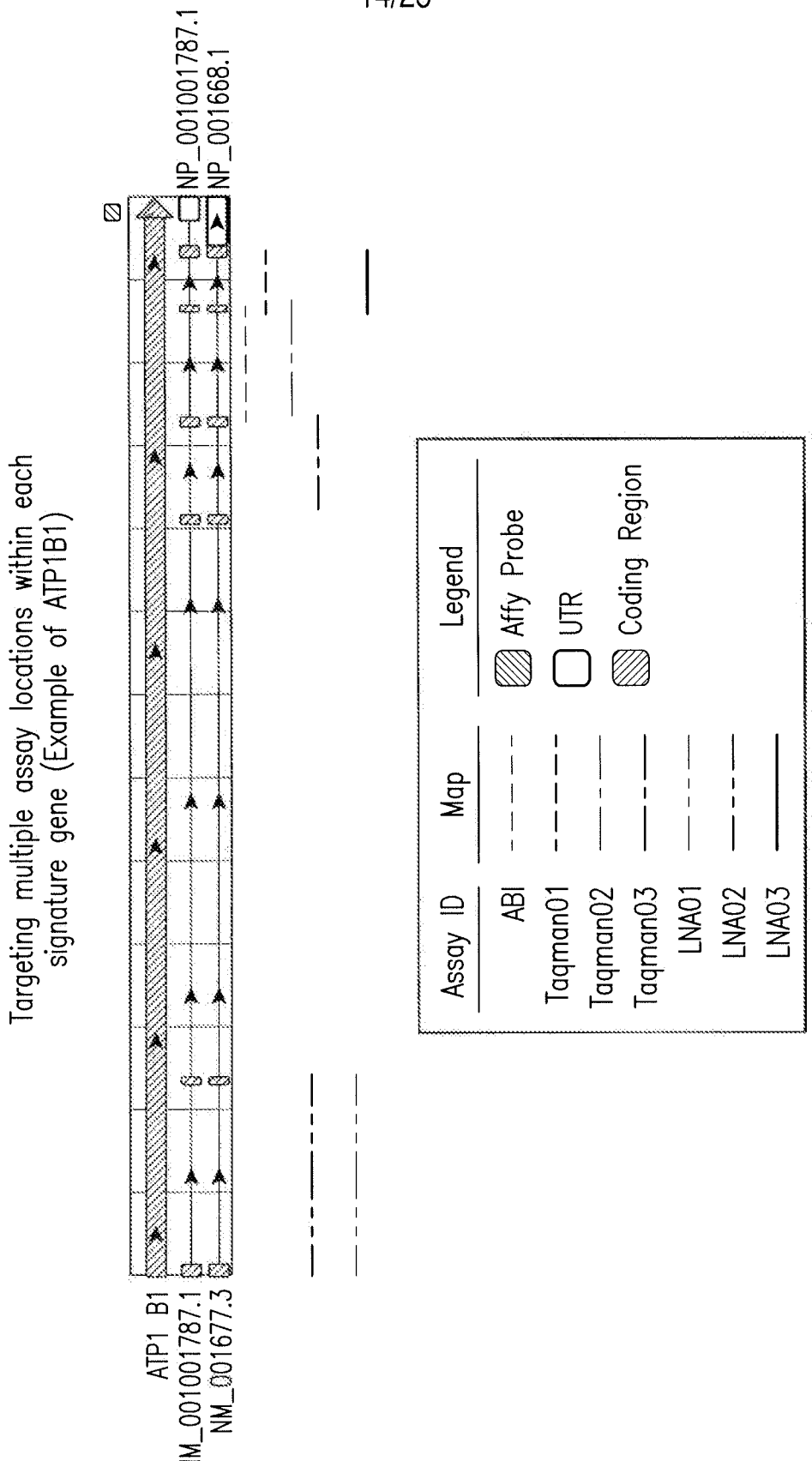
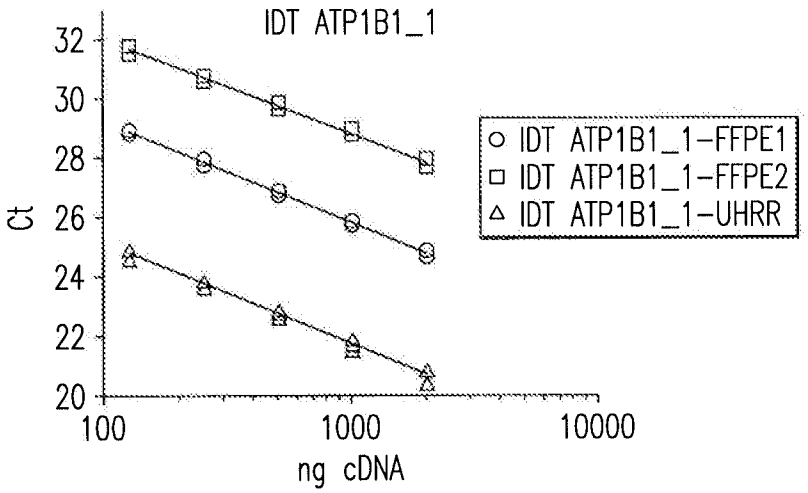

Targeting multiple assay locations within each signature gene (Example of ATP1B1)
ATP1 B1
NM_001001787.1
NM_001677.3
NP_001001787.1
NP_001668.1
Assay ID
Map
Legend
ABI
Taqman01
Taqman02
Taqman03
LNA01
LNA02
LNA03
Affy Probe
UTR
Coding Region

FIG. 9

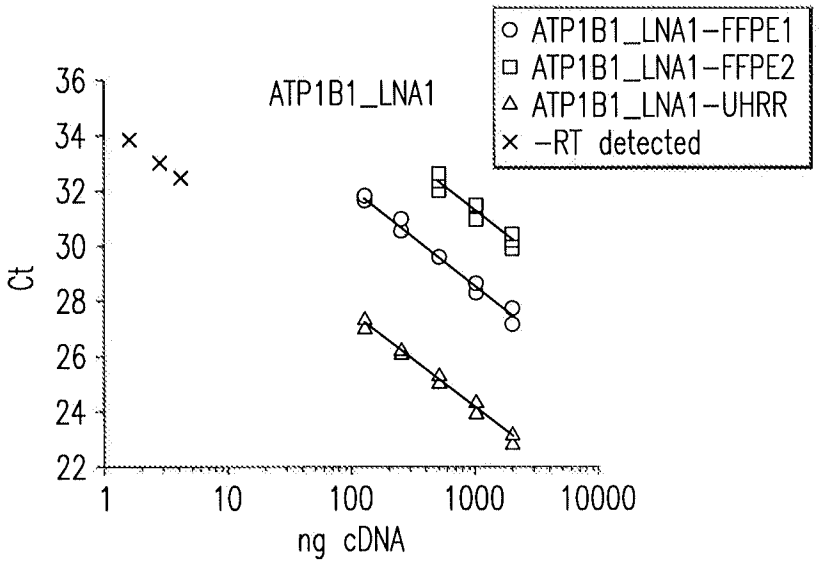
Assays compatible with both frozen and FFPE specimens
(Example of ATP1B1)



		Eff
IDT ATP1B1_1-FFPE1		*97%*
Slope	−3.40	
R^2	1.00	
IDT ATP1B1_1-FFPE2		*106%*
Slope	−3.18	
R^2	0.99	
IDT ATP1B1_1-UHRR		*98%*
Slope	−3.38	
R^2	0.99	

FIG. 10A

Assays compatible with both frozen and FFPE specimens
(Example of ATP1B1)



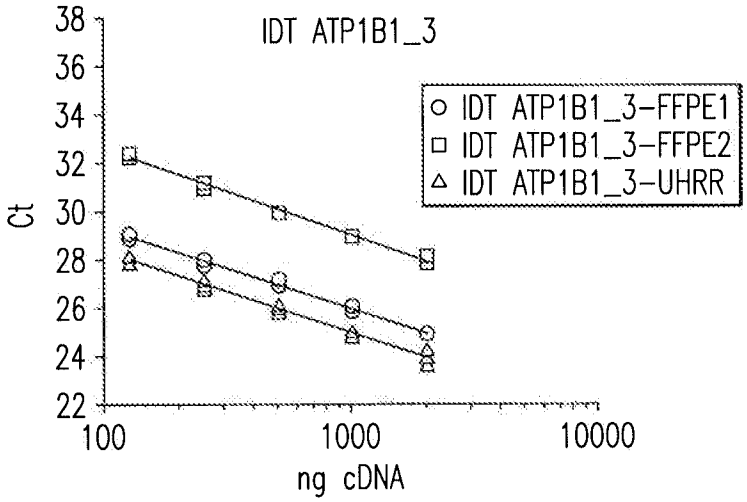
		Eff
ATP1B1_LNA1-FFPE2		*91%*
Slope	−3.55	
R^2	0.99	

		Eff
ATP1B1_LNA1-FFPE2		*92%*
Slope	−3.54	
R^2	0.92	

		Eff
ATP1B1_LNA1-UHRR		*96%*
Slope	−3.43	
R^2	0.99	

FIG. 10B

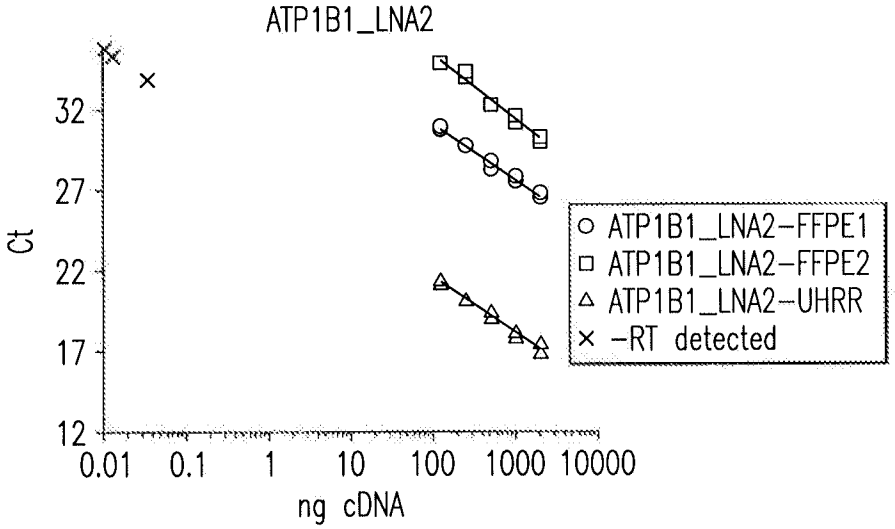
Assays compatible with both frozen and FFPE specimens
(Example of ATP1B1)



		Eff
IDT ATP1B1_3-FFPE1		
Slope	−3.34	*99%*
R^2	0.99	
IDT ATP1B1_3-FFPE2		
Slope	−3.51	*93%*
R^2	0.99	
IDT ATP1B1_3-UHRR		
Slope	−3.37	*98%*
R^2	0.99	

FIG. 10C

Assays compatible with both frozen and FFPE specimens
(Example of ATP1B1)



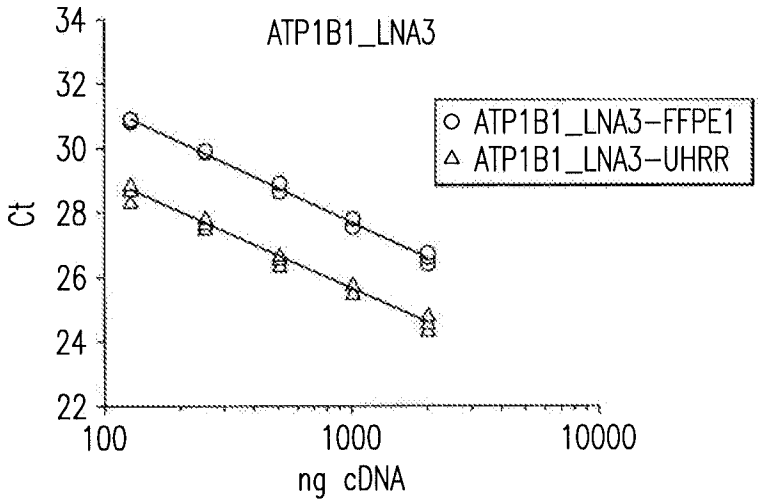
		Eff
ATP1B1_LNA2-FFPE1		*96%*
Slope	−3.43	
R^2	0.99	

		Eff
ATP1B1_LNA2-FFPE2		*76%*
Slope	−4.06	
R^2	0.98	

		Eff
ATP1B1_LNA2-UHRR		*92%*
Slope	−3.52	
R^2	0.99	

FIG. 10D

Assays compatible with both frozen and FFPE specimens
(Example of ATP1B1)



		Eff
ATP1B1_LNA3-FFPE1		
Slope	−3.56	*91%*
R^2	0.99	

		Eff
ATP1B1_LNA3-UHRR		
Slope	−3.38	*98%*
R^2	0.98	

FIG. 10E

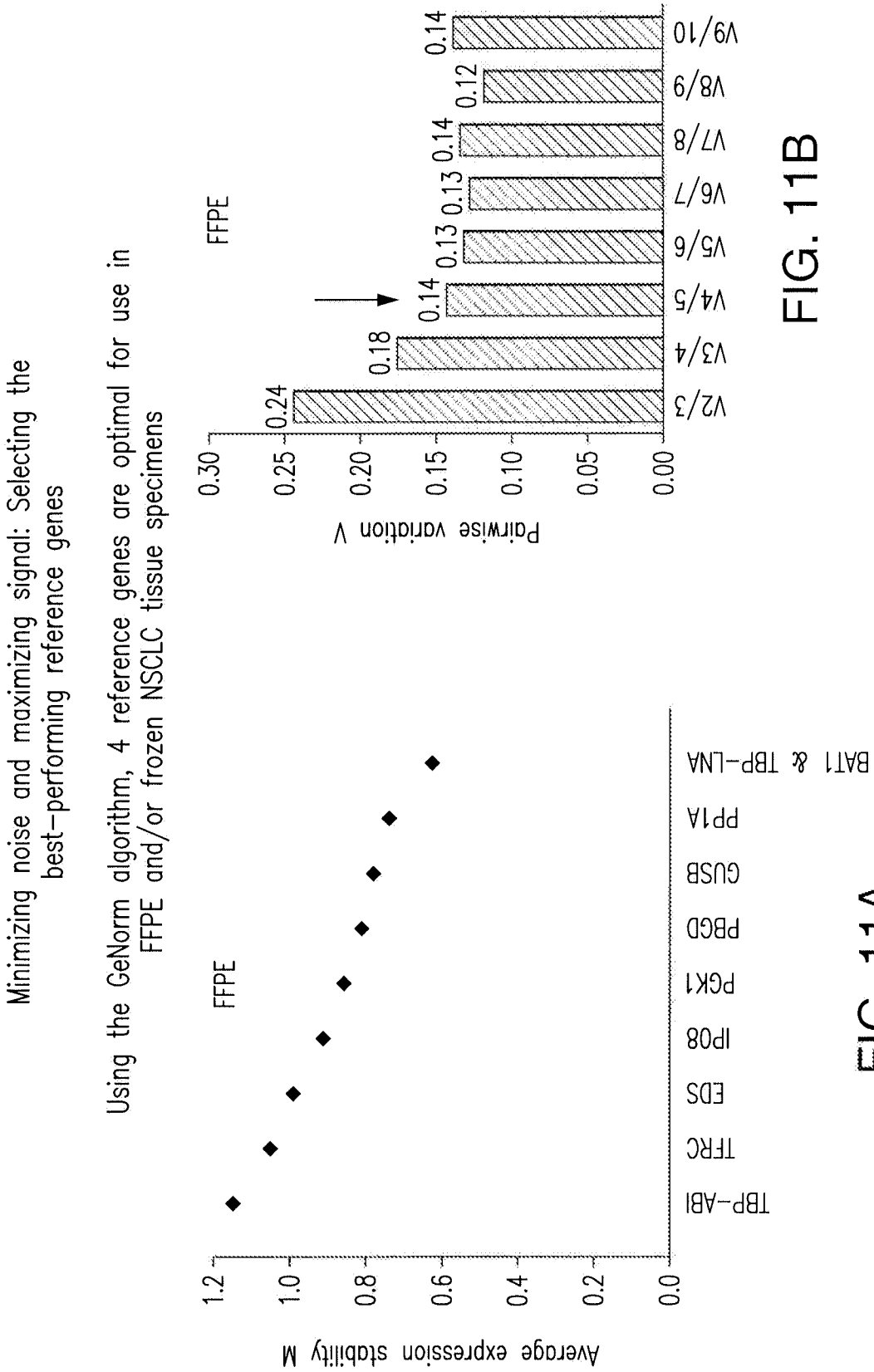


FIG. 11A

FIG. 11B

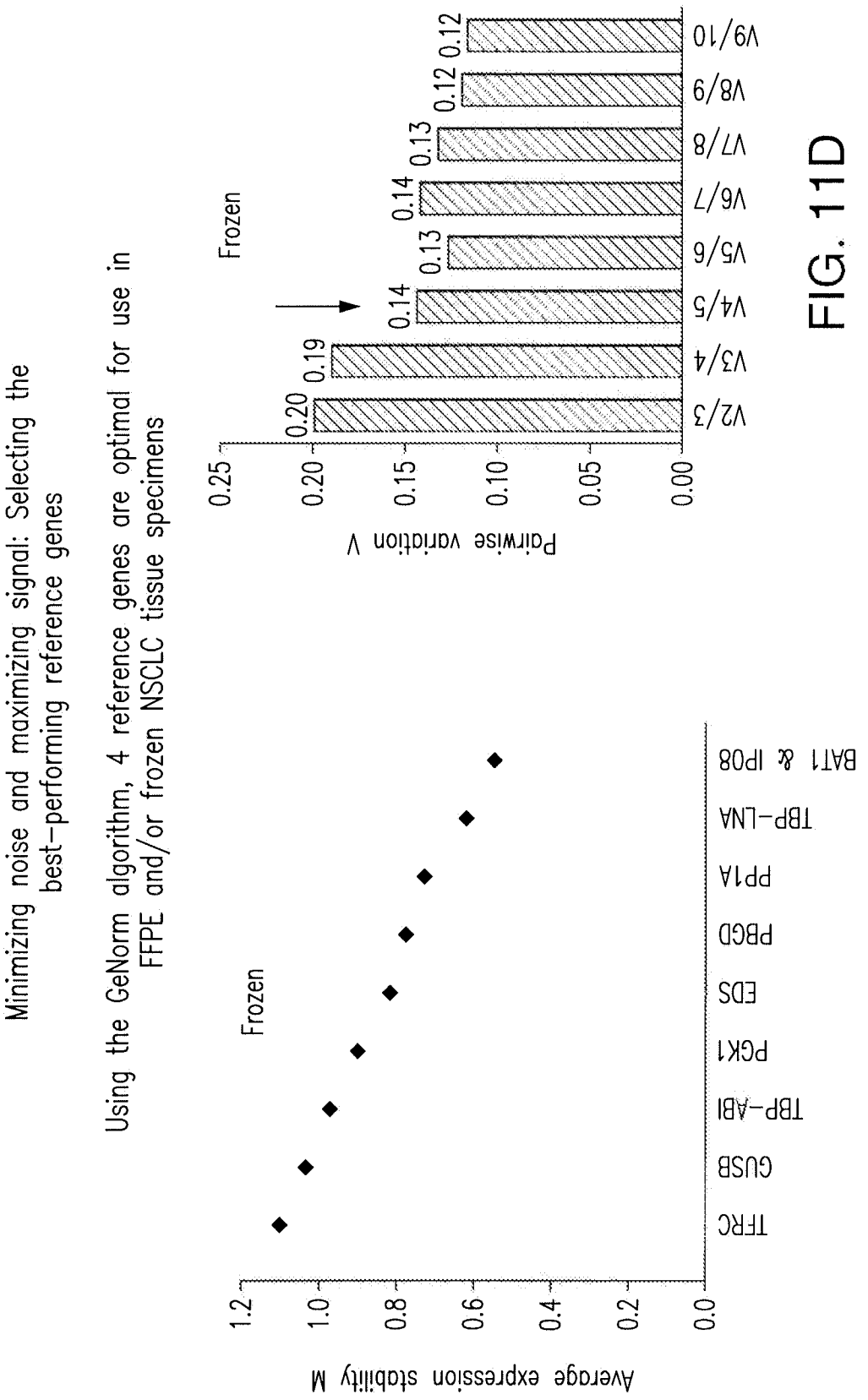


FIG. 11C

FIG. 11D

Minimizing noise and maximizing signal: Selecting the best-performing reference genes

Using the GeNorm algorithm, 4 reference genes are optimal for use in FFPE and/or frozen NSCLC tissue specimens

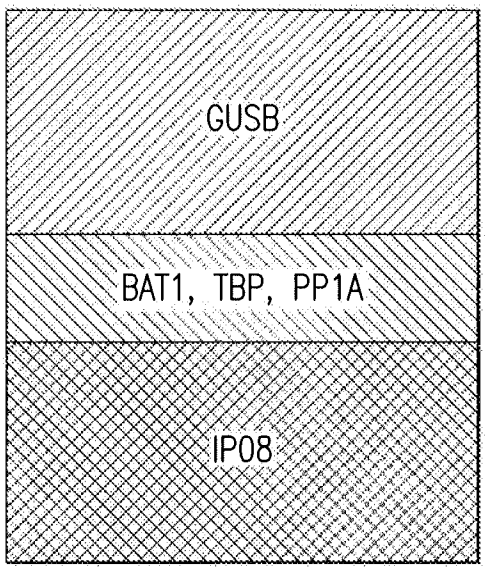


FIG. 11E

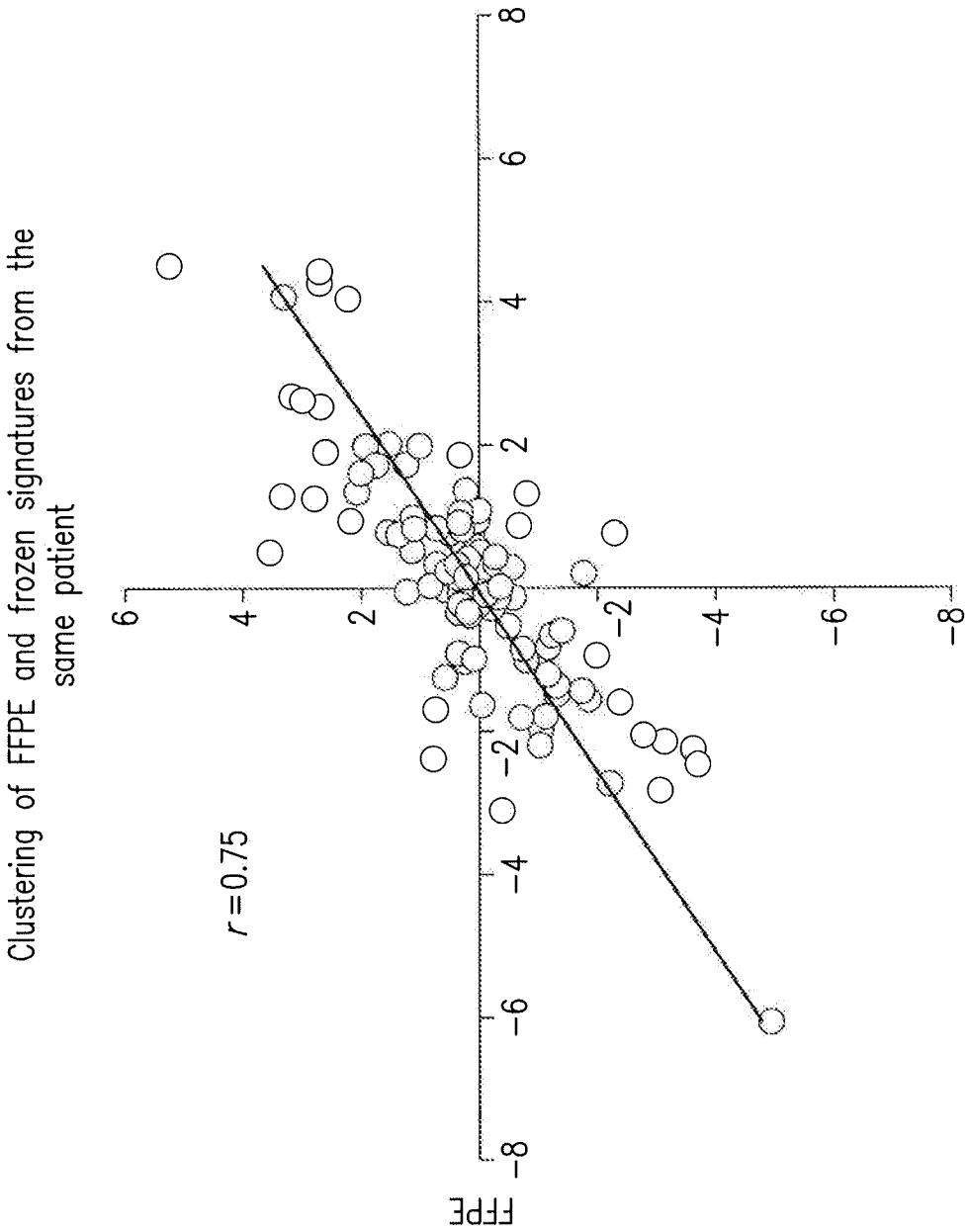

Clustering of FFPE and frozen signatures from the same patient
r=0.75
FFPE
Frozen
8
6
4
2
-2
-4
-6
-8

FIG. 12A

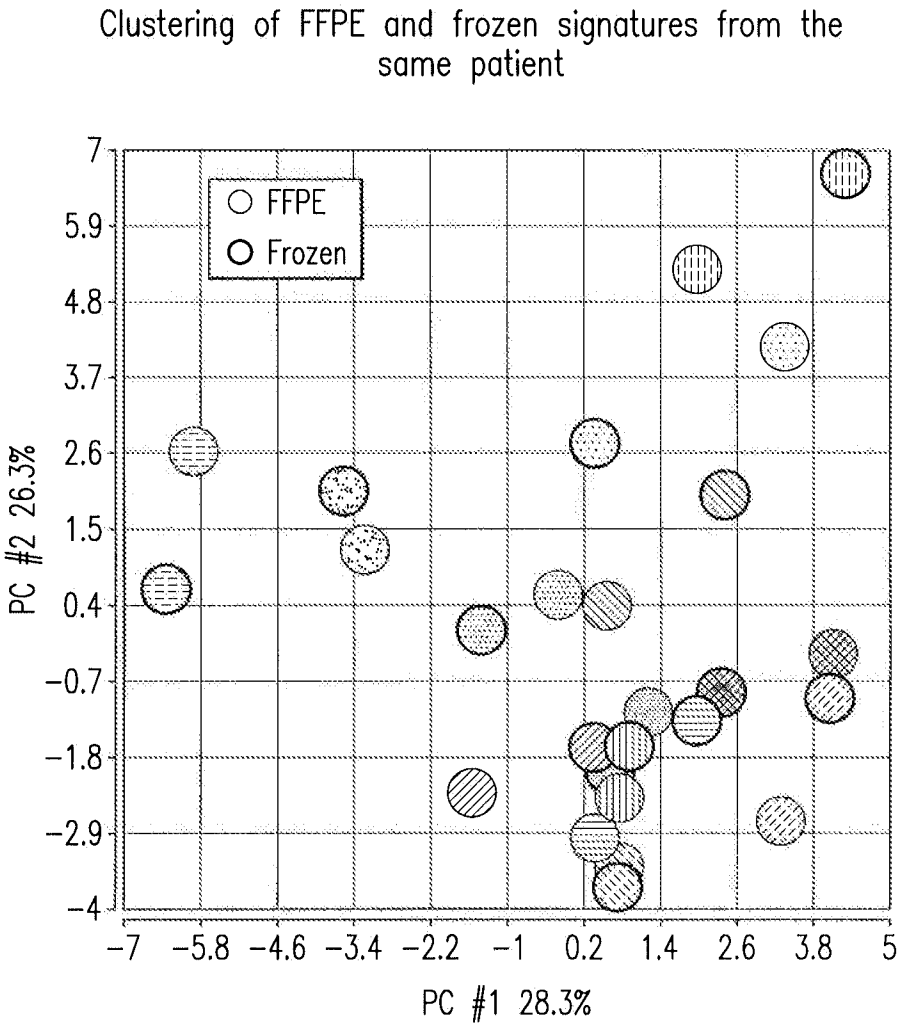

Clustering of FFPE and frozen signatures from the same patient
FFPE
Frozen
PC #2 26.3%
PC #1 28.3%
7
5.9
4.8
3.7
2.6
1.5
0.4
−0.7
−1.8
−2.9
−4
−7
−5.8
−4.6
−3.4
−2.2
−1
0.2
1.4
2.6
3.8
5

FIG. 12B

Substantial correlation between the qPCR and microarray–based risk scores

Purpose:
Correlation of microarray and qPCR platform using frozen NSCLC samples

Methodology:
Total RNA from 30 frozen NSCLC samples with available microarray data

Results:

Analysis (N 30) for UHN 30		
	HR	p value
Risk score (array)	2.1	0.16
Risk score (qPCR)	1.8	0.31

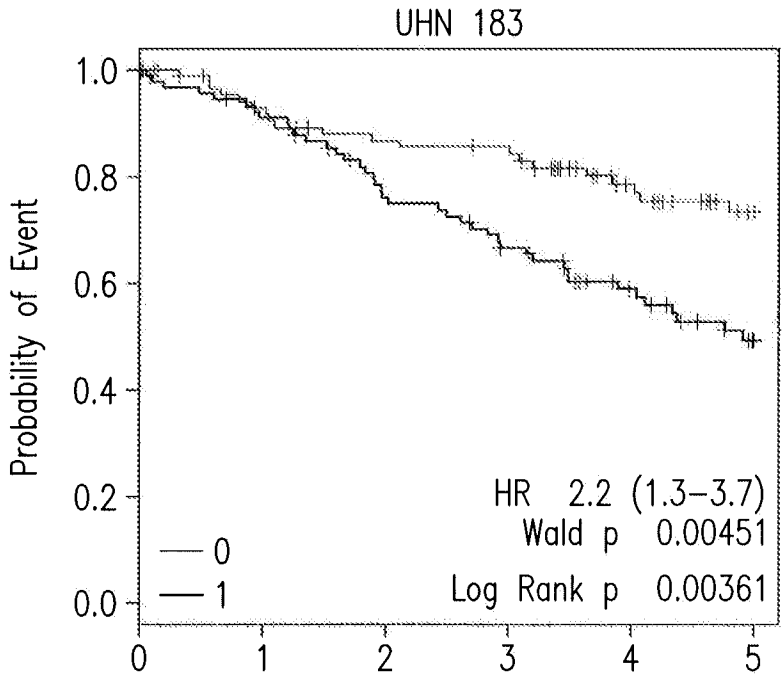


FIG. 13