MOLECULAR DIAGNOSTIC TEST FOR LUNG CANCER

FIELD OF THE INVENTION

The present invention relates to a molecular diagnostic test useful for predicting responsiveness of lung cancers to particular treatments that includes the use of a DNA damage repair deficiency subtype. The invention includes the generation and use of various classifiers derived from identification of this subtype in NSCLC patients, such as use of a 44-gene classification model that is used to identify this DNA damage repair deficiency molecular subtype. One application is the stratification of response to, and selection of patients for Non Small Cell Lung cancer (NSCLC) therapeutic drug classes, including DNA damage causing agents and DNA repair targeted therapies. The present invention provides a test that can guide conventional therapy selection as well as selecting patient groups for enrichment strategies during clinical trial evaluation of novel therapeutics. DNA repair deficient subtypes can be identified, for example, from fresh/frozen (FF) or formalin fixed paraffin embedded (FFPE) patient samples.

BACKGROUND

The pharmaceutical industry continuously pursues new drug treatment options that are more effective, more specific or have fewer adverse side effects than currently administered drugs. Drug therapy alternatives are constantly being developed because genetic variability within the human population results in substantial differences in the effectiveness of many drugs. Therefore, although a wide variety of drug therapy options are currently available, more therapies are always needed in the event that a patient fails to respond.

Traditionally, the treatment paradigm used by physicians has been to prescribe a first-line drug therapy that results in the highest success rate possible for treating a disease. Alternative drug therapies are then prescribed if the first is ineffective. This paradigm is clearly not the best treatment method for certain diseases. For example, in diseases such as cancer, the first treatment is often the most important and offers the best opportunity for successful therapy, so there exists a heightened need to choose an initial drug that will be the most effective against that particular patient's disease.

Lung cancer is the most prevalent cancer globally, responsible for 1.37 million of the 7.6 million deaths due to cancer in 2008 (WHO Fact sheet N°297). In 2010, 42,026 people in the UK were diagnosed with lung cancer and there were 34,859 deaths from lung cancer, correlating to 6% of all deaths in the UK (CRUK stats). The advent of microarrays and molecular genomics has the potential for a significant impact on the diagnostic capability and prognostic classification of disease, which may aid in the prediction of the response of an individual patient to a defined therapeutic regimen. Microarrays provide for the analysis of large amounts of genetic information, thereby providing a genetic fingerprint of an individual. There is much enthusiasm that this technology will ultimately provide the necessary tools for custom-made drug treatment regimens.

Currently, healthcare professionals have few mechanisms to help them identify cancer patients who will benefit from chemotherapeutic agents. Identification of the optimal first-line drug has been

difficult because methods are not available for accurately predicting which drug treatment would be the most effective for a particular cancer's physiology. This deficiency results in relatively poor single agent response rates and increased cancer morbidity and death. Furthermore, patients often needlessly undergo ineffective, toxic drug therapy.

Molecular markers have been used to select appropriate treatments, for example, in breast cancer. Breast tumors that do not express the estrogen and progesterone hormone receptors as well as the HER2 growth factor receptor, called "triple negative", appear to be responsive to PARP-1 inhibitor therapy (Linn, S. C., and Van 't Veer, L., J. Eur J Cancer *45 Suppl 1*, 11-26 (2009); O'Shaughnessy, J., et al. N Engl J Med *364*, 205-214 (2011). Recent studies indicate that the triple negative status of a breast tumor may indicate responsiveness to combination therapy including PARP-1 inhibitors, but may not be sufficient to indicate responsiveness to individual PARP-1 inhibitors (O'Shaughnessy et al., 2011).

Furthermore, there have been other studies that have attempted to identify gene classifiers associated with molecular subtypes to indicate responsiveness of chemotherapeutic agents (Farmer *et al.* Nat Med *15*, 68-74 (2009); Konstantinopoulos, P. A., *et al.*, J Clin Oncol *28*, 3555-3561 (2010)).

WO 2012/037378 describes a 44-gene DNA microarray assay, the DNA damage repair deficient (DDRD) assay. This assay identifies a molecular subgroup of cancers that have lost the DNA damage response FA/BRCA pathway, resulting in sensitivity to DNA damaging chemotherapeutic agents (Kennedy & D'Andrea Journal of Clinical Oncology (2006) 24:3799, Turner *et al* Nature Reviews Cancer (2004) 4:814).

In breast cancer the DDRD assay has been shown to predict response to neoadjuvant DNA-damaging chemotherapy (5-fluorouracil, anthracycline andcyclophosphamide) in 203 breast cancer patients (odd ratio 4.01) (95% CI:1.69-9.54). In a cohort of 191 early breast cancer patients treated with adjuvant 5-fluorouracil, epirubicin and cyclophosphamide treatment, the assay predicted 5-year relapse free survival with a hazard ratio of 0.37 (95% CI:0.15-0.88).

SUMMARY OF THE INVENTION

Non-small cell lung cancer (NSCLC) is the second most common malignancy among men and third among women in the UK. Loss of the FA/BRCA pathway has been reported in up to 44% of NSCLC (Lee *et al* Clinical Cancer Research (2007) 26:2048). The NICE guidelines for the treatment of early stage-NSCLC were updated in 2011 and are outlined in the CG121 guidelines. Currently adjuvant Cisplatin/Carboplatin based therapy (ACT) should be offered to patients with high risk early NSCLC. However this only confers a 4-15% 5-year survival advantage suggesting that not all patients benefit. Furthermore, patients diagnosed with NSCLC can be poor candidates for chemotherapy as they are generally older and many are smokers with significant cardio-vascular and renal co-morbities. The risk of severe toxicity from ACT therefore outweighs the benefit for many patients, especially when the majority gain no survival advantage. The ability to determine which patients are not going to benefit from ACT could prevent over-treatment with unnecessary toxicities and may guide the use of alternative, non-DNA damaging therapies, such as taxanes or vincavina-alkaloids.

The present invention is based upon application of methods that identify deficiencies in DNA damage repair to determine which patients will benefit from certain therapies, such as ACT in order to treat lung cancer. The invention is directed to methods of using a collection of gene product markers

expressed in lung cancer such that when some or all of the transcripts are over or under-expressed, they identify a subtype of lung cancer that has a deficiency in DNA damage repair. The invention also provides methods for indicating responsiveness or resistance to DNA-damaging therapeutic agents. In different aspects, this gene or gene product list may form the basis of a single parameter or a multiparametric predictive test that could be delivered using methods known in the art such as microarray, Q-PCR, immunohistochemistry, ELISA or other technologies that can quantify mRNA or protein expression.

Thus, according to one aspect of the invention there is provided a method of predicting responsiveness of an individual having lung cancer such as (in particular) non-small cell lung cancer (NSCLC) to treatment with a DNA-damaging therapeutic agent comprising:

a. measuring expression levels of one or more biomarkers in a test sample obtained from the individual, wherein the one or more biomarkers are selected from Table 1A, 1B, 1C, 2A, 2B, 3A, 3B and/or 3C, such as from the group consisting of CXCL10, MX1, IDO1, IF144L, CD2, GBP5, PRAME, ITGAL, LRP4, and APOL3;

b. deriving a test score that captures the expression levels;

c. providing a threshold score comprising information correlating the test score and responsiveness;

d. and comparing the test score to the threshold score; wherein responsiveness is predicted when the test score exceeds the threshold score and/or wherein a lack of responsiveness is predicted when the test score does not exceed the threshold score.

The methods may be performed as a method for selecting a suitable treatment for an individual. Thus, in certain embodiments if the test score exceeds the threshold score (responsiveness is predicted) the individual is treated with the DNA-damaging therapeutic agent. Similarly, if the test score does not exceed the threshold score (responsiveness is not predicted) the individual is not treated with the DNA-damaging therapeutic agent. In those circumstances, alternative treatments may be contemplated. For NSCLC, the alternative treatments may comprise administration of a mitotic inhibitor, such as a vinca alkaloid or a taxane. Example vinca alkaloids include vinorelbine. Example taxanes include paclitaxel or docetaxel. Alternatively, the treatment may exclude chemotherapy altogether. The methods can, in some embodiments, also involve the subsequent treatment of the individual identified as responsive. Corresponding kits are also contemplated. The method is typically performed in vitro. The method is, therefore, performed using an isolated, or pre-isolated, sample. In some embodiments, the methods may encompass the step of obtaining a test sample from the individual. In certain embodiments, the method comprises measuring an expression level of at least 10 of the biomarkers from Table 1A in the test sample. More specifically, the method may comprise measuring the expression level of all 58 different biomarkers listed in Table 1A. In certain embodiments, expression levels are measured using primers or probes which bind to at least one of the target sequences set forth as SEQ ID NO: 1-80 (Table 1A), 81-260 (Table 3A), 261-313 (Table 3B), 314-337 (Table 1B) or 338-363 (Table 1C).

In some embodiments, the method further comprises measuring an expression level of one or more biomarkers in the test sample, wherein the one or more biomarkers are selected from the group consisting of CDR1, FYB, TSPAN7, RAC2, KLHDC7B, GRB14, AC138128.1, KIF26A, CD274, CD109, ETV7, MFAP5, OLFM4, PI15, FOSB, FAM19A5, NLRC5, PRICKLE1, EGR1, CLDN10, ADAMTS4,

SP140L, ANXA1, RSAD2, ESR1, IKZF3, OR2I1P, EGFR, NAT1, LATS2, CYP2B6, PTPRC, PPP1R1A, and AL137218.1. In certain embodiments, the test score captures the expression levels of all of the biomarkers (CXCL10, MX1, IDO1, IF144L, CD2, GBP5, PRAME, ITGAL, LRP4, and APOL3, and CDR1, FYB, TSPAN7, RAC2, KLHDC7B, GRB14, AC138128.1, KIF26A, CD274, CD109, ETV7, MFAP5, OLFM4, PI15, FOSB, FAM19A5, NLRC5, PRICKLE1, EGR1, CLDN10, ADAMTS4, SP140L, ANXA1, RSAD2, ESR1, IKZF3, OR2I1P, EGFR, NAT1, LATS2, CYP2B6, PTPRC, PPP1R1A, and AL137218.1; see Table 2B. In some embodiments, responsiveness may be predicted when the test score exceeds a threshold score at a value of between approximately 0.1 and 0.5 such as 0.1, 0.2, 0.3, 0.4 or 0.5. for example approximately 0.3681.

The lung cancer is typically non-small cell lung cancer (NSCLC) and may be early stage. Alternatively, the NSCLC may be late stage or metastatic disease. The NSCLC may be selected from one or more of adenocarcinoma, large-cell lung carcinoma and squamous cell carcinoma.

The treatment, for which responsiveness is predicted is typically adjuvant treatment. However, it may comprise neoadjuvant treatment additionally or alternatively.

The invention described herein is not limited to any one DNA-damaging therapeutic agent; it can be used to identify responders and non responders to any of a range of DNA-damaging therapeutic agent, for example those that directly or indirectly affect DNA damage and/or DNA damage repair. In some embodiments, the DNA-damaging therapeutic agent comprises one or more substances selected from the group consisting of: a DNA damaging agent, a DNA repair targeted therapy, an inhibitor of DNA damage signalling, an inhibitor of DNA damage induced cell cycle arrest, a histone deacetylase inhibitor, a heat shock protein inhibitor and an inhibitor of DNA synthesis. More specifically, the DNA-damaging therapeutic agent may be selected from one or more of a platinum-containing agent, a nucleoside analogue such as gemcitabine or 5-fluorouracil or a prodrug thereof such as capecitabine, an anthracycline such as epirubicin or doxorubicin, an alkylating agent such as cyclophosphamide, an ionising radiation or a combination of radiation and chemotherapy (chemoradiation). In particular embodiments, the DNA-damaging therapeutic agent comprises a platinum-containing agent, such as a platinum based agent selected from cisplatin, carboplatin and oxaliplatin. The methods may predict responsiveness to treatment with the DNA-damaging therapeutic agent together with a further drug. Thus, the methods may predict responsiveness to a combination therapy. For example, it is shown experimentally herein that the methods of the invention can identify a subpopulation of NSCLC patients who are more likely to benefit to adjuvant cisplatin based therapy, in combination with vinorelbine. Thus, in some embodiments, the further drug is a mitotic inhibitor. The mitotic inhibitor may be a vinca alkaloid or a taxane. In specific embodiments,the vinca alkaloid is vinorelbine In certain embodiments, responders to the following treatments are identified: cisplatin/carboplatin, Cisplatin/carboplatin and 5-fluorouracil (5-FU) (CF), cisplatin/carboplatin and capecitabine (CX), epirubicin/doxyrubicin,cisplatin/carboplatin and fluorouracil (ECF), epirubicin, oxaliplatin and capecitabine (EOX), gemcitabine, cyclophosphamide, radiation and chemoradiation. In specific aspects this invention, it is useful for evaluating cisplatin/ carboplatin (Paraplatin), cisplatin/carboplatin and etoposide (CP), gemcitabine and cisplatin/carboplatin (GemCarbo) cyclophosphamide epirubicin/doxorubicin and vincristine (CEV/CAV), CEV/CAV plus etoposide (CEVE/CAVE), epirubicin/doxorubicin, cyclophosphamide and etoposide (ECE/ACE) a combination of DNA damaging agents with topotecan, or cisplatin or carboplatin (Paraplatin) with at least one other drug such as Vinorelbine, Gemcitabine,

Paclitaxel (Taxol), Docetaxel (Taxotere), epirubicin/Doxorubicin, Etoposide, Pemetrexed or radiation in treatment of NSCLC.

The present invention relates to prediction of response to drugs (DNA-damaging therapeutic agents) using different classifications of response, such as overall survival, progression free survival, disease free survival, radiological response, as defined by RECIST, complete response, partial response, stable disease and serological markers such as, but not limited to, PSA, CEA, CA125, CA15-3 and CA19-9. In specific embodiments this invention can be used to evaluate standard chest roentgenography, computed tomography (CT), perfusion CT, dynamic contrast material–enhanced magnetic resonance (MR) diffusion-weighted (DW) MR or positron emission tomography (PET) with the glucose analog fluorine 18 fluorodeoxyglucose (FDG) (FDG-PET) response in NSCLC treated with DNA damaging therapeutic agents, including combination therapies, alone or in the context of standard treatment.

The present invention relies upon a DNA damage response deficiency (DDRD) molecular subtype, originally identified in breast and ovarian cancer (WO2012/037378; incorporated herein by reference). This molecular subtype can, in some embodiments, be detected by the use of two different gene classifiers – one being 40 genes in length and one being 44 genes in length. The DDRD classifier was first defined by a classifier consisting of 53 probesets on the Almac Breast Disease Specific Array (DSA™). So as to validate the functional relevance of this classifier in the context of its ability to predict response to DNA-damaging containing chemotherapy regimens, the classifier needed to be re-defined at a gene level. This facilitated evaluation of the DDRD classifier using microarray data from independent datasets that were profiled on microarray platforms other than the Almac Breast DSA®. In order to facilitate defining the classifier at a gene level, the genes to which the Almac Breast DSA® probesets map needed to be defined. This involved the utilization of publicly available genome browser databases such as Ensembl and NCBI Reference Sequence. The 44-gene DDRD classifier model supersedes that of the 40-gene DDRD classifier model. The results presented herein demonstrate that the probe sets can be mapped to NSCLC and used to generate a suitable classifier (see Table 1A). Results are also presented herein confirming that the 44 gene classifier is effective in predicting responsiveness to DNA-damaging therapeutic agents (cisplatin) in a range of NSC lung cancers (see Example 2). The 44 and 40 gene classifier models and related classifier models derived from the markers in Table 1A are effective and significant predictors of response to chemotherapy regimens that contain DNA damaging therapeutics in the context of NSCLC.

The identification of the DDRD subtype using classifier models based upon genes taken from Table 1A, such as using up to all 58 of the genes, and also from Tables 1B and 1C, such as by both the 40-gene classifier model and the 44-gene classifier model, can be used to predict response to, and select patients for, standard NSCLC cancer therapeutic drug classes, including DNA damage causing agents and DNA repair targeted therapies.

In another aspect, the present invention relates to kits for conventional diagnostic uses listed above such as nucleic acid amplification, including PCR and all variants thereof such as real-time and end point methods and qPCR, Next generation Sequencing (NGS), microarray, and immunoassays such as immunohistochemistry, ELISA, Western blot and the like. Such kits include appropriate reagents and directions to assay the expression of the genes or gene products and quantify mRNA or protein expression. The kits may include suitable primers and/or probes to detect the expression levels of at

least one of the genes in Table 1A, 1B and/or 1C. The kits may contain primers and/or probes that bind to target sequences comprising, consisting essentially of or consisting of SEQ ID NO: 1-80, SEQ ID NO: 81-260 or SEQ ID NO: 261-363 (or SEQ ID NO: 1-80 (Table 1A), 81-260 (Table 3A), 261-313 (Table 3B), 314-337 (Table 1B), 338-363 (Table 1C)). The kits may contain primers and/or probes to determine expression levels of any one or more up to all of the 40, 44 or 58 (respectively) gene classifiers described herein. The kits may comprise primer and/or probes comprising, consisting essentially of or consisting of the nucleotide sequences set forth in Table 3C (SEQ ID NOs 364-455).

In some embodiments, the kits may also contain the specific DNA-damaging therapeutic agent to be administered in the event that the test predicts responsiveness. This agent may be provided in a form, such as a dosage form, that is tailored to NSCLC treatment specifically. The kit may be provided with suitable instructions for administration according to NSCLC treatment regimens.

The invention also provides methods for identifying DNA damage response-deficient (DDRD) human NSCLC tumors. It is likely that this invention can be used to identify patients that are sensitive to and respond, or are resistant to and do not respond, to DNA-damaging therapeutic agents, such as drugs that damage DNA directly, damage DNA indirectly or inhibit normal DNA damage signaling and/or repair processes.

The invention also relates to guiding conventional treatment of patients. The invention also relates to selecting patients for clinical trials where novel DNA-damaging therapeutic agents, such as drugs of the classes that directly or indirectly affect DNA damage and/or DNA damage repair are to be tested.

The present invention and methods accommodate the use of archived formalin fixed paraffin-embedded (FFPE) biopsy material, including fine needle aspiration (FNA) as well as fresh/frozen (FF) tissue, for assay of all transcripts in the invention, and are therefore compatible with the most widely available type of biopsy material. The expression level may be determined using RNA obtained from FFPE tissue, fresh frozen tissue or fresh tissue that has been stored in solutions such as RNAlater®.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 provides a diagram representing the semi-supervised hierarchical clustering of the NSCL samples (columns) by the most variable genes (rows) defined in the DDRD discovery data set. Sample clinical information is represented as coloured bars above the cluster and described in the legend box. The right hand side table represents the overlap of the genes in each cluster with the DDRD genes from the Breast DDRD discovery data set. See Example 1.

FIG. 2 Is a Kaplan Meier (KM) plot showing the survival of treated (red) and non-treated (blue) patients in the DDRD cohort. See Example 1.

FIG. 3 Is a Kaplan Meier (KM) plot showing the survival of treated (red) and non-treated (blue) patients in the non DDRD cohort. See Example 1.

FIG. 4 is a Kaplan-Meier plot of overall survival following cisplatin based adjuvant chemotherapy when the 44 gene DDRD signature was applied to 60 non small cell lung cancer samples. See Example 2.

DETAILED DESCRIPTION OF THE INVENTION

Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods, devices, and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

All publications, published patent documents, and patent applications cited in this application are indicative of the level of skill in the art(s) to which the application pertains. All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element, unless explicitly indicated to the contrary.

A major goal of current research efforts in cancer is to increase the efficacy of perioperative systemic therapy in patients by incorporating molecular parameters into clinical therapeutic decisions. Pharmacogenetics/genomics is the study of genetic/genomic factors involved in an individual's response to a foreign compound or drug. Agents or modulators which have a stimulatory or inhibitory effect on expression of a marker of the invention can be administered to individuals to treat (prophylactically or therapeutically) lung cancer in a patient. It is ideal to also consider the pharmacogenomics of the individual in conjunction with such treatment. Differences in metabolism of therapeutics may possibly lead to severe toxicity or therapeutic failure by altering the relationship between dose and blood concentration of the pharmacologically active drug. Thus, understanding the pharmacogenomics of an individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the level of expression of a marker of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

The invention is directed to the application of a collection of gene or gene product markers (hereinafter referred to as "biomarkers") expressed in certain lung cancer tissue for predicting responsiveness to treatment using DNA-damaging therapeutic agents. In different aspects, this biomarker list may form the basis of a single parameter or multiparametric predictive test that could be delivered using methods known in the art such as microarray, Q-PCR, NGS, immunohistochemistry, ELISA or other technologies that can quantify mRNA or protein expression.

The present invention also relates to kits and methods that are useful for prognosis following cytotoxic chemotherapy or selection of specific treatments for lung cancer (particularly NSCLC). Methods are provided such that when some or all of the transcripts are over or under-expressed, the expression profile indicates responsiveness or resistance to DNA-damaging therapeutic agents. These kits and methods employ gene or gene product markers that are differentially expressed in tumors of patients with NSCLC. In one embodiment of the invention, the expression profiles of these biomarkers are correlated

with clinical outcome (response or survival) in archival tissue samples under a statistical method or a correlation model to create a database or model correlating expression profile with responsiveness to one or more DNA-damaging therapeutic agents. The predictive model may then be used to predict the responsiveness in a patient whose responsiveness to the DNA-damaging therapeutic agent(s) is unknown. In many other embodiments, a patient population can be divided into at least two classes based on patients' clinical outcome, prognosis, or responsiveness to DNA-damaging therapeutic agents, and the biomarkers are substantially correlated with a class distinction between these classes of patients. The biological pathways described herein have been shown to be predictive of responsiveness to treatment of NSCLC using DNA-damaging therapeutic agents.

Predictive Marker Panels/Expression Classifiers

A unique collection of biomarkers as a genetic classifier expressed in lung cancer/NSCLC tissue is provided that is useful in determining responsiveness or resistance to therapeutic agents, such as DNA-damaging therapeutic agents, used to treat lung cancer/NSCLC. Such a collection may be termed a "marker panel", "expression classifier", or "classifier". The collection is shown in Table 1A. This collection was derived from an original collection of biomarkers as shown in Tables 1B and 1C (see WO 2012/037378) which were then mapped to an NSCLC platform (see Example 1 herein). A hierarchical clustering analysis identified a DDRD cluster that defines those individuals likely to respond to certain treatments of NSCLC. This cluster, or collection, of biomarkers makes up Table 1A. This represents 58 different genes and 80 different target sequences within those 58 genes. The invention may involve determining expression levels of any one or more of these genes or target sequences. Evidence is also presented herein (example 2) that the 44 gene classifier (Table 2B and 3C) is effective in predicting responsiveness to DNA-damaging therapeutic agents (cisplatin) in various NSC lung cancers, including adenocarcinoma, squamous cell carcinoma and large cell carcinoma.

The biomarkers useful in the present methods are thus identified in the tables herein, such as Tables 1A, 1B and 1C. These biomarkers are identified as having predictive value to determine a patient (having NSCLC) response to a therapeutic agent, or lack thereof. Their expression correlates with the response to an agent, and more specifically, a DNA-damaging therapeutic agent. By examining the expression of a collection of the identified biomarkers in a lung tumor, in particular an adenocarcinoma, large-cell lung carcinoma or squamous cell carcinoma, it is possible to determine which therapeutic agent or combination of agents will be most likely to reduce the growth rate of the cancer, and in some embodiments, NSCLC cells. By examining a collection of identified transcript gene or gene product markers, it is also possible to determine which therapeutic agent or combination of agents will be the least likely to reduce the growth rate of the cancer. By examining the expression of a collection of biomarkers, it is therefore possible to eliminate ineffective or inappropriate therapeutic agents. Importantly, in certain embodiments, these determinations can be made on a patient-by-patient basis or on an agent-by-agent basis. Thus, one can determine whether or not a particular therapeutic regimen is likely to benefit a particular patient or type of patient, and/or whether a particular regimen should be continued.

Table 1A – Genes (biomarkers) and target sequences therein relevant for defining DDRD status in NSCLC patients

Probe Set ID	Patent ID	Gene symbol	SEQ ID NO of target sequence
204205_at	DDRD_Lung_SSA-1	APOBEC3G	1
204416_x_at	DDRD_Lung_SSA-2	APOC1	2
213553_x_at	DDRD_Lung_SSA-3	APOC1	3
209846_s_at	DDRD_Lung_SSA-4	BTN3A2	4
212613_at	DDRD_Lung_SSA-5	BTN3A2	5
218232_at	DDRD_Lung_SSA-6	C1QA	6
212886_at	DDRD_Lung_SSA-7	CCDC69	7
204606_at	DDRD_Lung_SSA-8	CCL21	8
1405_i_at	DDRD_Lung_SSA-9	CCL5	9
204655_at	DDRD_Lung_SSA-10	CCL5	10
206337_at	DDRD_Lung_SSA-11	CCR7	11
203645_s_at	DDRD_Lung_SSA-12	CD163	12
215049_x_at	DDRD_Lung_SSA-13	CD163	13
205831_at	DDRD_Lung_SSA-14	CD2	14
207277_at	DDRD_Lung_SSA-15	CD209	15
213539_at	DDRD_Lung_SSA-16	CD3D	16
205456_at	DDRD_Lung_SSA-17	CD3E	17
204661_at	DDRD_Lung_SSA-18	CD52	18
34210_at	DDRD_Lung_SSA-19	CD52	19
203416_at	DDRD_Lung_SSA-20	CD53	20
219505_at	DDRD_Lung_SSA-21	CECR1	21
202357_s_at	DDRD_Lung_SSA-22	CFB	22
209395_at	DDRD_Lung_SSA-23	CHI3L1	23
209396_s_at	DDRD_Lung_SSA-24	CHI3L1	24
213060_s_at	DDRD_Lung_SSA-25	CHI3L2	25
212865_s_at	DDRD_Lung_SSA-26	COL14A1	26
200838_at	DDRD_Lung_SSA-27	CTSB	27
200839_s_at	DDRD_Lung_SSA-28	CTSB	28
213274_s_at	DDRD_Lung_SSA-29	CTSB	29
213275_x_at	DDRD_Lung_SSA-30	CTSB	30
203922_s_at	DDRD_Lung_SSA-31	CYBB	31

203923_s_at	DDRD_Lung_SSA-32	CYBB	32
217838_s_at	DDRD_Lung_SSA-33	EVL	33
220306_at	DDRD_Lung_SSA-34	FAM46C	34
205285_s_at	DDRD_Lung_SSA-35	FYB	35
211795_s_at	DDRD_Lung_SSA-36	FYB	36
219243_at	DDRD_Lung_SSA-37	GIMAP4	37
211990_at	DDRD_Lung_SSA-38	HLA-DPA1	38
211991_s_at	DDRD_Lung_SSA-39	HLA-DPA1	39
213537_at	DDRD_Lung_SSA-40	HLA-DPA1	40
209540_at	DDRD_Lung_SSA-41	IGF1	41
209541_at	DDRD_Lung_SSA-42	IGF1	42
209542_x_at	DDRD_Lung_SSA-43	IGF1	43
211577_s_at	DDRD_Lung_SSA-44	IGF1	44
205038_at	DDRD_Lung_SSA-45	IKZF1	45
204912_at	DDRD_Lung_SSA-46	IL10RA	46
204116_at	DDRD_Lung_SSA-47	IL2RG	47
203828_s_at	DDRD_Lung_SSA-48	IL32	48
205798_at	DDRD_Lung_SSA-49	IL7R	49
202531_at	DDRD_Lung_SSA-50	IRF1	50
213475_s_at	DDRD_Lung_SSA-51	ITGAL	51
202746_at	DDRD_Lung_SSA-52	ITM2A	52
202747_s_at	DDRD_Lung_SSA-53	ITM2A	53
205821_at	DDRD_Lung_SSA-54	KLRK1	54
208071_s_at	DDRD_Lung_SSA-55	LAIR1	55
210644_s_at	DDRD_Lung_SSA-56	LAIR1	56
208885_at	DDRD_Lung_SSA-57	LCP1	57
213975_s_at	DDRD_Lung_SSA-58	LYZ	58
210356_x_at	DDRD_Lung_SSA-59	MS4A1	59
217418_x_at	DDRD_Lung_SSA-60	MS4A1	60
209734_at	DDRD_Lung_SSA-61	NCKAP1L	61

206370_at	DDRD_Lung_SSA-62	PIK3CG	62
204269_at	DDRD_Lung_SSA-63	PIM2	63
203471_s_at	DDRD_Lung_SSA-64	PLEK	64
205267_at	DDRD_Lung_SSA-65	POU2AF1	65
204279_at	DDRD_Lung_SSA-66	PSMB9	66
207419_s_at	DDRD_Lung_SSA-67	RAC2	67
213603_s_at	DDRD_Lung_SSA-68	RAC2	68
204070_at	DDRD_Lung_SSA-69	RARRES3	69
203485_at	DDRD_Lung_SSA-70	RTN1	70
210222_s_at	DDRD_Lung_SSA-71	RTN1	71
204923_at	DDRD_Lung_SSA-72	SASH3	72
204563_at	DDRD_Lung_SSA-73	SELL	73
219159_s_at	DDRD_Lung_SSA-74	SLAMF7	74
219993_at	DDRD_Lung_SSA-75	SOX17	75
202524_s_at	DDRD_Lung_SSA-76	SPOCK2	76
202307_s_at	DDRD_Lung_SSA-77	TAP1	77
205922_at	DDRD_Lung_SSA-78	VNN2	78
202663_at	DDRD_Lung_SSA-79	WIPF1	79
202665_s_at	DDRD_Lung_SSA-80	WIPF1	80

Table 1B – Original list of genes tested in breast cancer and mapped to NSCLC

Sense genes (166)		Antisense of known genes (24)		
Gene Symbol	EntrezGene ID	Almac Gene ID	Almac Gene symbol	SEQ ID NO:
ABCA12	26154		N/A	
ALDH3B2	222		N/A	
APOBEC3G	60489		N/A	
APOC1	341		N/A	
APOL6	80830		N/A	
ARHGAP9	64333		N/A	
BAMBI	25805		N/A	
BIK	638		N/A	
BIRC3	330	AS1_BIRC3	Hs127799.0C7n9_at	314
BTN3A3	10384		N/A	
C12orf48	55010		N/A	
C17orf28	283987		N/A	
C1orf162	128346		N/A	
C1orf64	149563		N/A	
C1QA	712		N/A	
C21orf70	85395		N/A	
C22orf32	91689		N/A	

C6orf211	79624		N/A	
CACNG4	27092		N/A	
CCDC69	26112		N/A	
CCL5	6352		N/A	
CCNB2	9133		N/A	
CCND1	595		N/A	
CCR7	1236		N/A	
CD163	9332		N/A	
CD2	914		N/A	
CD22	933		N/A	
CD24	100133941		N/A	
CD274	29126		N/A	
CD3D	915		N/A	
CD3E	916		N/A	
CD52	1043		N/A	
CD53	963		N/A	
CD79A	973		N/A	
CDH1	999		N/A	
CDKN3	1033		N/A	
CECR1	51816		N/A	
CHEK1	1111		N/A	
CKMT1B	1159		N/A	
CMPK2	129607		N/A	
CNTNAP2	26047		N/A	
COX16	51241		N/A	
CRIP1	1396		N/A	
CXCL10	3627		N/A	
CXCL9	4283		N/A	
CYBB	1536		N/A	
CYP2B6	1555		N/A	
DDX58	23586		N/A	
DDX60L	91351		N/A	
ERBB2	2064		N/A	
ETV7	51513		N/A	
FADS2	9415		N/A	
FAM26F	441168		N/A	
FAM46C	54855		N/A	
FASN	2194		N/A	
FBP1	2203		N/A	
FBXO2	26232		N/A	
FKBP4	2288		N/A	
FLJ40330	645784		N/A	
FYB	2533		N/A	
GBP1	2633		N/A	
GBP4	115361		N/A	
GBP5	115362	AS1_GBP5	BRMX.5143C1n2_at	315
GIMAP4	55303		N/A	
GLRX	2745		N/A	
GLUL	2752		N/A	
GVIN1	387751		N/A	
H2AFJ	55766		N/A	
HGD	3081		N/A	
HIST1H2BK	85236		N/A	
HIST3H2A	92815		N/A	
HLA-DOA	3111		N/A	
HLA-DPB1	3115		N/A	
HMGB2	3148		N/A	
HMGB3	3149		N/A	
HSP90AA1	3320		N/A	
IDO1	3620		N/A	

IFI27	3429		N/A	
IFI44	10561		N/A	
IFI44L	10964	AS1_IFI44L	BRSA.1606C1n4_at	316
IFI6	2537		N/A	
IFIH1	64135		N/A	
IGJ	3512	AS1_IGJ	BRIH.1231C2n2_at	317
IKZF1	10320		N/A	
IL10RA	3587		N/A	
IL2RG	3561		N/A	
IL7R	3575		N/A	
IMPAD1	54928		N/A	
IQGAP3	128239	AS1_IQGAP3	BRAD.30779_s_at	318
IRF1	3659		N/A	
ISG15	9636		N/A	
ITGAL	3683		N/A	
KIAA1467	57613		N/A	
KIF20A	10112		N/A	
KITLG	4254		N/A	
KLRK1	22914		N/A	
KRT19	3880		N/A	
LAIR1	3903		N/A	
LCP1	3936		N/A	
LOC100289702	100289702		N/A	
LOC100294459	100294459	AS1_LOC100294459	BRSA.396C1n2_at	319
LOC150519	150519		N/A	
LOC439949	439949		N/A	
LYZ	4069		N/A	
MAL2	114569		N/A	
MGC29506	51237		N/A	
MIAT	440823		N/A	
MS4A1	931		N/A	
MX1	4599	AS1_MX1	BRMX.2948C3n7_at	320
NAPSB	256236		N/A	
NCKAP1L	3071		N/A	
NEK2	4751		N/A	
NLRC3	197358		N/A	
NLRC5	84166		N/A	
NPNT	255743		N/A	
NQO1	1728		N/A	
OAS2	4939		N/A	
OAS3	4940		N/A	
PAQR4	124222		N/A	
PARP14	54625		N/A	
PARP9	83666		N/A	
PIK3CG	5294		N/A	
PIM2	11040		N/A	
PLEK	5341		N/A	
POU2AF1	5450		N/A	
PP14571	100130449		N/A	
PPP2R2C	5522		N/A	
PSMB9	5698		N/A	
PTPRC	5788		N/A	
RAC2	5880		N/A	
RAMP1	10267		N/A	
RARA	5914		N/A	
RASSF7	8045		N/A	
RSAD2	91543		N/A	
RTP4	64108		N/A	
SAMD9	54809		N/A	
SAMD9L	219285		N/A	

SASH3	54440		N/A	
SCD	6319		N/A	
SELL	6402		N/A	
SIX1	6495	AS1_SIX1	Hs539969.0C4n3_at	321
SLAMF7	57823		N/A	
SLC12A2	6558		N/A	
SLC9A3R1	9368	AS1_SLC9A3R1	Hs396783.3C1n4_at	322
SPOCK2	9806		N/A	
SQLE	6713		N/A	
ST20	400410		N/A	
ST6GALNAC2	10610		N/A	
STAT1	6772	AS1_STAT1	BRMX.13670C1n2_at	323
STRA13	201254		N/A	
SUSD4	55061		N/A	
SYT12	91683		N/A	
TAP1	6890		N/A	
TBC1D10C	374403		N/A	
TNFRSF13B	23495		N/A	
TNFSF10	8743		N/A	
TOB1	10140	AS1_TOB1	BRAD.30243_at	324
TOM1L1	10040		N/A	
TRIM22	10346		N/A	
UBD	10537	AS1_UBD	BRMX.941C2n2_at	325
UBE2T	29089		N/A	
UCK2	7371		N/A	
USP18	11274		N/A	
VNN2	8875		N/A	
XAF1	54739		N/A	
ZWINT	11130		N/A	
		AS1_C1QC	BRMX.4154C1n3_s_at	326
		AS1_C2orf14	BRAD.39498_at	327
		AS1_EPSTI1	BRAD.34868_s_at	328
		AS1_GALNT6	5505575.0C1n42_at	329
		AS1_HIST1H4H	BREM.1442_at	330
		AS1_HIST2H4B	BRHP.827_s_at	331
		AS2_HIST2H4B	BRRS.18322_s_at	332
		AS3_HIST2H4B	BRRS.18792_s_at	333
		AS1_KIAA1244	Hs632609.0C1n37_at	334
		AS1_LOC100287927	Hs449575.0C1n22_at	335
		AS1_LOC100291682	BRAD.18827_s_at	336
		AS1_LOC100293679	BREM.2466_s_at	337

Table 1C – Original list of genes tested in breast cancer and mapped to NSCLC

Novel genes	
Gene symbol	**SEQ ID NO:**
BRAD.2605_at	338
BRAD.33618_at	339
BRAD.36579_s_at	340
BRAD1_5440961_s_at	341
BRAD1_66786229_s_at	342
BREM.2104_at	343
BRAG_AK097020.1_at	344
BRAD.20415_at	345
BRAD.29668_at	346
BRAD.30228_at	347
BRAD.34830_at	348
BRAD.37011_s_at	349
BRAD.37762_at	350

BRAD.40217_at	351
BRAD1_4307876_at	352
BREM.2505_at	353
Hs149363.0CB4n5_s_at	354
Hs172587.9C1n9_at	355
Hs271955.16C1n9_at	356
Hs368433.18C1n6_at	357
Hs435736.0C1n27_s_at	358
Hs493096.15C1n6_at	359
Hs493096.2C1n15_s_at	360
Hs592929.0CB2n8_at	361
Hs79953.0C1n23_at	362
BRMX.2377C1n3_at	363

All or a portion of the biomarkers recited in Tables 1A, 1B and/or 1C may be used in a predictive biomarker panel. For example, biomarker panels selected from the biomarkers in Tables 1A, 1B and 1C can be generated using the methods provided herein and can comprise between one, and all of the biomarkers set forth in Tables 1A, 1B and/or 1C and each and every combination in between (e.g., four selected biomarkers, 16 selected biomarkers, 74 selected biomarkers, etc.). In some embodiments, the predictive biomarker set comprises at least 5, 10, 20, 40, 60, 100, 150, 200, or 300 or more biomarkers. In other embodiments, the predictive biomarker set comprises no more than 5, 10, 20, 40, 60, 100, 150, 200, 300, 400, 500, 600 or 700 biomarkers. In some embodiments, the predictive biomarker set includes a plurality of biomarkers listed in Tables 1A, 1B and/or 1C. In some embodiments the predictive biomarker set includes at least about 1%, about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, about 96%, about 97%, about 98%, or about 99% of the biomarkers listed in Tables 1A, 1B and/or 1C. Selected predictive biomarker sets can be assembled from the predictive biomarkers provided using methods described herein and analogous methods known in the art. In one embodiment, the biomarker panel contains all 203 biomarkers in Table 1B and/or 1C. In another embodiment, the biomarker panel contains the 58 different genes/biomarkers or 80 different target sequences in Table 1A. In another embodiment, the biomarker panel corresponds to the 40 or 44 gene panel described in tables 2A and 2B.

Predictive biomarker sets may be defined in combination with corresponding scalar weights on the real scale with varying magnitude, which are further combined through linear or non-linear, algebraic, trigonometric or correlative means into a single scalar value via an algebraic, statistical learning, Bayesian, regression, or similar algorithms which together with a mathematically derived decision function on the scalar value provide a predictive model by which expression profiles from samples may be resolved into discrete classes of responder or non-responder, resistant or non-resistant, to a specified drug or drug class. Such predictive models, including biomarker membership, are developed by learning weights and the decision threshold, optimized for sensitivity, specificity, negative and positive predictive values, hazard ratio or any combination thereof, under cross-validation, bootstrapping or similar sampling techniques, from a set of representative expression profiles from historical patient samples with known drug response and/or resistance or with known molecular subtype (i.e. DDRD) classification.

In one embodiment, the biomarkers are used to form a weighted sum of their signals, where individual weights can be positive or negative. The resulting sum ("decisive function") is compared with a pre-determined reference point or value. The comparison with the reference point or value may be used to diagnose, or predict a clinical condition or outcome.

As described above, one of ordinary skill in the art will appreciate that the biomarkers included in the classifier or classifiers provided in Tables 1A, 1B and 1C will carry unequal weights in a classifier for responsiveness or resistance to a therapeutic agent. Therefore, while as few as one sequence may be used to diagnose or predict an outcome such as responsiveness to therapeutic agent, the specificity and sensitivity or diagnosis or prediction accuracy may increase using more sequences.

As used herein, the term "weight" refers to the relative importance of an item in a statistical calculation. The weight of each biomarker in a gene expression classifier may be determined on a data set of patient samples using analytical methods known in the art. Gene specific bias values may also be applied. Gene specific bias may be required to mean centre each gene in the classifier relative to a training data set, as would be understood by one skilled in the art.

In one embodiment the biomarker panel is directed to the 40 biomarkers detailed in Table 2A with corresponding ranks and weights detailed in the table or alternative rankings and weightings, depending, for example, on the disease setting. In another embodiment, the biomarker panel is directed to the 44 biomarkers detailed in Table 2B with corresponding ranks and weights detailed in the table or alternative rankings and weightings, depending, for example, on the disease setting. Tables 2A and 2B rank the biomarkers in order of decreasing weight in the classifier, defined as the rank of the average weight in the compound decision score function measured under cross-validation.

Table 2A
Gene IDs and EntrezGene IDs for 40-gene DDRD classifier model with associated ranking and weightings

DDRD classifier 40 gene model			
Rank	**Genes Symbol**	**EntrezGene ID**	**Weights**
1	GBP5	115362	0.022389581
2	CXCL10	3627	0.021941734
3	IDO1	3620	0.020991115
4	MX1	4599	0.020098675
5	IFI44L	10964	0.018204957
6	CD2	914	0.018080661
7	PRAME	23532	0.016850837
8	ITGAL	3683	0.016783359
9	LRP4	4038	-0.015129969
10	SP140L	93349	0.014646025
11	APOL3	80833	0.014407174
12	FOSB	2354	-0.014310521
13	CDR1	1038	-0.014209848
14	RSAD2	91543	0.014177132
15	TSPAN7	7102	-0.014111562
16	RAC2	5880	0.014093627
17	FYB	2533	0.01400475
18	KLHDC7B	113730	0.013298413
19	GRB14	2888	0.013031204
20	KIF26A	26153	-0.012942351
21	CD274	29126	0.012651964
22	CD109	135228	-0.012239425
23	ETV7	51513	0.011787297
24	MFAP5	8076	-0.011480443
25	OLFM4	10562	-0.011130113
26	PI15	51050	-0.010904326
27	FAM19A5	25817	-0.010500936
28	NLRC5	84166	0.009593449

29	EGR1	1958	-0.008947963
30	ANXA1	301	-0.008373991
31	CLDN10	9071	-0.008165127
32	ADAMTS4	9507	-0.008109892
33	ESR1	2099	0.007524594
34	PTPRC	5788	0.007258669
35	EGFR	1956	-0.007176203
36	NAT1	9	0.006165534
37	LATS2	26524	-0.005951091
38	CYP2B6	1555	0.005838391
39	PPP1R1A	5502	-0.003898835
40	TERF1P1	348567	0.002706847

Table 2B
Gene IDs and EntrezGene IDs for 44-gene DDRD classifier model with associated ranking and weightings

DDRD Classifier - 44 Gene Model (NA:genomic sequence)			
Rank	**Gene symbol**	**EntrezGene ID**	**Weight**
1	CXCL10	3627	0.023
2	MX1	4599	0.0226
3	IDO1	3620	0.0221
4	IFI44L	10964	0.0191
5	CD2	914	0.019
6	GBP5	115362	0.0181
7	PRAME	23532	0.0177
8	ITGAL	3683	0.0176
9	LRP4	4038	-0.0159
10	APOL3	80833	0.0151
11	CDR1	1038	-0.0149
12	FYB	2533	-0.0149
13	TSPAN7	7102	0.0148
14	RAC2	5880	-0.0148
15	KLHDC7B	113730	0.014
16	GRB14	2888	0.0137
17	AC138128.1	N/A	-0.0136
18	KIF26A	26153	-0.0136
19	CD274	29126	0.0133
20	CD109	135228	-0.0129
21	ETV7	51513	0.0124
22	MFAP5	8076	-0.0121
23	OLFM4	10562	-0.0117
24	PI15	51050	-0.0115

25	FOSB	2354	-0.0111
26	FAM19A5	25817	0.0101
27	NLRC5	84166	-0.011
28	PRICKLE1	144165	-0.0089
29	EGR1	1958	-0.0086
30	CLDN10	9071	-0.0086
31	ADAMTS4	9507	-0.0085
32	SP140L	93349	0.0084
33	ANXA1	301	-0.0082
34	RSAD2	91543	0.0081
35	ESR1	2099	0.0079
36	IKZF3	22806	0.0073
37	OR2I1P	442197	0.007
38	EGFR	1956	-0.0066
39	NAT1	9	0.0065
40	LATS2	26524	-0.0063
41	CYP2B6	1555	0.0061
42	PTPRC	5788	0.0051
43	PPP1R1A	5502	-0.0041
44	AL137218.1	N/A	-0.0017

Table 3A presents the probe sets from the Xcel Array (Almac) that represent the genes in Table 2A and 2B with reference to their sequence ID numbers. Table 3B presents the probe sets from the Human Genome U133A array (Affymetrix) that represent the genes in Table 2A and 2B with reference to their sequence ID numbers. Table 3C presents the probe sets from the Human Genome U133A plus 2.0 array (Affymetrix) that represent the genes in Table 2A and 2B.

Table 3A - Probe set IDs and SEQ Numbers for target sequences of genes contained in 44-gene signature as mapped to Xcel platform

Gene	Probeset ID	SEQ ID NO: of Target sequence
AC138128.1	NONMATCH	#N/A
ADAMTS4	ADXEC.29185.C1_at	81
ADAMTS4	ADXECAD.1557_at	82
ADAMTS4	ADXECAD.1557_x_at	83
ADAMTS4	ADXECNTDJ.9649_at	84
AL137218.1	ADXECADA.15298_x_at	85
ANXA1	ADXEC.961.C1_at	86
ANXA1	ADXEC.961.C2_s_at	87
ANXA1	ADXEC.961.C3_at	88
ANXA1	ADXECAD.8396_at	89
APOL3	ADXEC.11171.C1_s_at	90
CD109	ADXEC.11145.C1_s_at	91

CD109	ADXEC.11777.C1_at	92
CD109	ADXEC.12292.C1_at	93
CD2	ADXEC.7301.C1-a_s_at	94
CD2	ADXEC.7301.C1_at	95
CD2	ADXECEMUTR.6872_at	96
CD2	ADXECRS.12205_s_at	97
CD274	ADXEC.11136.C1_at	98
CD274	ADXEC.23232.C1_at	99
CD274	ADXECNTDJ.4196_s_at	100
CD274	ADXECNTDJ.4198_s_at	101
CDR1	ADXECRS.7695_s_at	102
CLDN10	ADXEC.19503.C1_s_at	103
CLDN10	ADXECEMUTR.6957_at	104
CLDN10	ADXECRS.17517_s_at	105
CXCL10	ADXEC.11676.C1_at	106
CYP2B6	ADXEC.20112.C1_s_at	107
CYP2B6	ADXECAD.18663_x_at	108
CYP2B6	ADXLCEC.9263.C1_at	109
EGFR	ADXEC.14093.C1_at	110
EGFR	ADXEC.1866.C1_at	111
EGFR	ADXEC.1866.C1_x_at	112
EGFR	ADXEC.21483.C1_at	113
EGFR	ADXEC.23775.C1_at	114
EGFR	ADXEC.31869.C1_at	115
EGFR	ADXEC.4451.C1_at	116
EGFR	ADXECAD.18126_at	117
EGFR	ADXECAD.19259_at	118
EGFR	ADXECADA.15206_at	119
EGFR	ADXECADA.21225_s_at	120
EGFR	ADXECADA.8307_at	121
EGFR	ADXECEMUTR.2965_at	122
EGFR	ADXECEMUTR.3575_at	123
EGFR	ADXECNTDJ.6255_at	124
EGFR	ADXECNTDJ.6256_at	125
EGFR	ADXECNTDJ.6256_x_at	126
EGFR	ADXECRS.19907_at	127
EGFR	ADXECRS.19907_s_at	128
EGFR	ADXECRS.24032_at	129
EGFR	ADXLCEC.7900.C1_at	130
EGFR	ADXPCEC.14538.C1_at	131
EGR1	ADXEC.2432.C2_s_at	132
EGR1	ADXEC.2432.C4_at	133
EGR1	ADXEC.2432.C6-a_s_at	134
ESR1	ADXEC.27541.C1_at	135
ESR1	ADXEC.29140.C1_s_at	136
ESR1	ADXEC.33997.C1_at	137

ESR1	ADXECAD.12370_s_at	138
ESR1	ADXECAD.18631_at	139
ESR1	ADXECAD.24092_s_at	140
ESR1	ADXECADA.11317_s_at	141
ESR1	ADXECADA.9299_at	142
ESR1	ADXECNTDJ.3778_at	143
ESR1	ADXECNTDJ.3779_at	144
ESR1	ADXOCEC.10271.C1_at	145
ESR1	ADXOCEC.10271.C1_x_at	146
ESR1	ADXOCEC.9813.C1_at	147
ETV7	ADXEC.745.C1_s_at	148
ETV7	ADXECEMUTR.534_s_at	149
FAM19A5	ADXEC.10689.C1_at	150
FAM19A5	ADXEC.13789.C1_at	151
FAM19A5	ADXEC.13789.C1_s_at	152
FAM19A5	ADXEC.13789.C1_x_at	153
FAM19A5	ADXECADA.11183_at	154
FAM19A5	ADXECADA.11183_s_at	155
FAM19A5	ADXECADA.11183_x_at	156
FAM19A5	ADXECNTDJ.10271_at	157
FOSB	ADXEC.34273.C1_at	158
FOSB	ADXEC.34273.C1_x_at	159
FOSB	ADXEC.9157.C1-a_s_at	160
FOSB	ADXEC.9157.C1_at	161
FOSB	ADXECNTDJ.4222_s_at	162
FOSB	ADXECNTDJ.4223_at	163
FOSB	ADXECNTDJ.4223_x_at	164
FOSB	ADXPCEC.11652.C1_x_at	165
FYB	ADXECAD.24300_s_at	166
FYB	ADXECADA.2898_at	167
FYB	ADXECNTDJ.82_s_at	168
GBP5	ADXEC.6891.C2_at	169
GBP5	ADXEC.6891.C2_s_at	170
GBP5	ADXEC.8878.C1_at	171
GRB14	ADXEC.13641.C1_s_at	172
IDO1	ADXEC.20415.C1-a_s_at	173
IFI44L	ADXEC.30980.C1_at	174
IFI44L	ADXEC.30980.C1_x_at	175
IFI44L	ADXEC.6079.C1_at	176
IFI44L	ADXEC.6079.C1_x_at	177
IFI44L	ADXOCEC.12110.C2_s_at	178
IFI44L	ADXOCEC.9547.C1_at	179
IFI44L	ADXOCEC.9547.C1_x_at	180
IKZF3	ADXEC.22688.C1_at	181
IKZF3	ADXEC.32096.C1_at	182
IKZF3	ADXEC.32096.C1_x_at	183

IKZF3	ADXECAD.25262_s_at	184
IKZF3	ADXECADA.10727_at	185
IKZF3	ADXECRS.658_s_at	186
ITGAL	ADXEC.7237.C1_s_at	187
ITGAL	ADXECADA.387_x_at	188
KIF26A	ADXEC.10112.C1_at	189
KIF26A	ADXEC.10112.C1_s_at	190
KLHDC7B	ADXEC.11833.C1_at	191
KLHDC7B	ADXECADA.94_at	192
LATS2	ADXEC.11588.C1_s_at	193
LATS2	ADXEC.8316.C2_s_at	194
LATS2	ADXECAD.19393_at	195
LRP4	ADXEC.13953.C1_at	196
LRP4	ADXEC.15783.C1_at	197
LRP4	ADXECADA.18233_at	198
MFAP5	ADXEC.18200.C1_at	199
MFAP5	ADXEC.8579.C1-a_s_at	200
MFAP5	ADXEC.8579.C1_at	201
MFAP5	ADXEC.8579.C2_s_at	202
MX1	ADXEC.6683.C1_at	203
MX1	ADXEC.6683.C1_s_at	204
MX1	ADXEC.6842.C2_at	205
MX1	ADXEC.6842.C2_x_at	206
NAT1	ADXEC.20034.C1-a_s_at	207
NAT1	ADXEC.20034.C1_at	208
NAT1	ADXEC.20034.C2_s_at	209
NAT1	ADXECEMUTR.4521_s_at	210
NAT1	ADXECNTDJ.5862_s_at	211
NAT1	ADXECNTDJ.5864_s_at	212
NAT1	ADXECNTDJ.5866_s_at	213
NAT1	ADXECNTDJ.5867_at	214
NAT1	ADXECNTDJ.5868_at	215
NLRC5	ADXEC.23051.C1_s_at	216
NLRC5	ADXEC.5068.C1_at	217
NLRC5	ADXECEMUTR.5074_at	218
NLRC5	ADXECEMUTR.5074_s_at	219
NLRC5	ADXECNTDJ.5048_s_at	220
OLFM4	ADXEC.8457.C1-a_s_at	221
OLFM4	ADXEC.8457.C1_s_at	222
OR2I1P	ADXECAD.16836_at	223
OR2I1P	ADXECAD.16836_s_at	224
PI15	ADXEC.29833.C1-a_s_at	225
PI15	ADXEC.29833.C1_at	226
PI15	ADXEC.29833.C1_s_at	227
PI15	ADXEC.7703.C1_at	228
PI15	ADXEC.7703.C1_x_at	229

PI15	ADXECAD.23062_at	230
PPP1R1A	ADXEC.14340.C1_at	231
PPP1R1A	ADXEC.15744.C1_at	232
PRAME	ADXEC.11333.C1_at	233
PRAME	ADXEC.11333.C1_x_at	234
PRICKLE1	ADXEC.9436.C1_at	235
PRICKLE1	ADXEC.9436.C1_x_at	236
PRICKLE1	ADXECAD.6243_s_at	237
PRICKLE1	ADXECAD.8320_at	238
PRICKLE1	ADXECRS.11172_s_at	239
PRICKLE1	ADXECRS.18104_s_at	240
PTPRC	ADXEC.8915.C1-a_s_at	241
PTPRC	ADXEC.8915.C1_at	242
PTPRC	ADXECAD.17697_at	243
PTPRC	ADXECADA.4026_at	244
PTPRC	ADXECADA.52_at	245
PTPRC	ADXECNTDJ.2722_s_at	246
PTPRC	ADXECNTDJ.2723_s_at	247
RAC2	ADXEC.15369.C1_s_at	248
RSAD2	ADXEC.8308.C1-a_s_at	249
RSAD2	ADXEC.8308.C1_at	250
RSAD2	ADXECAD.11200_at	251
RSAD2	ADXECADA.13258_s_at	252
RSAD2	ADXECNTDJ.5191_at	253
RSAD2	ADXECRS.4576_s_at	254
SP140L	ADXEC.31390.C1_at	255
SP140L	ADXECADA.3222_at	256
TSPAN7	ADXEC.12786.C1_at	257
TSPAN7	ADXECADA.9258_at	258
TSPAN7	ADXECADA.9258_x_at	259
TSPAN7	ADXECNTDJ.7964_at	260

Table 3B - Probe set IDs and SEQ Numbers for target sequences of genes contained in 44-gene signature as mapped to U133A platform

Gene	Probeset ID	SEQ ID NO of Target sequence
AC138128.1	NONMATCH	#N/A
ADAMTS4	NONMATCH	#N/A
AL137218.1	NONMATCH	#N/A
ANXA1	201012_at	261
APOL3	221087_s_at	262
CD109	NONMATCH	#N/A
CD2	205831_at	263
CD274	NONMATCH	#N/A
CDR1	207276_at	264
CLDN10	205328_at	265
CXCL10	204533_at	266
CYP2B6	206754_s_at	267
CYP2B6	206755_at	268
CYP2B6	217133_x_at	269
EGFR	201983_s_at	270
EGFR	201984_s_at	271
EGFR	210984_x_at	272
EGFR	211550_at	273
EGFR	211551_at	274
EGFR	211607_x_at	275
EGR1	201693_s_at	276
EGR1	201694_s_at	277
ESR1	205225_at	278
ESR1	211233_x_at	279
ESR1	211234_x_at	280
ESR1	211235_s_at	281
ESR1	211627_x_at	282
ESR1	215552_s_at	283
ESR1	217163_at	284
ESR1	217190_x_at	285
ETV7	221680_s_at	286
FAM19A5	NONMATCH	#N/A
FOSB	202768_at	287
FYB	205285_s_at	288
FYB	211794_at	289
FYB	211795_s_at	290
GBP5	NONMATCH	#N/A
GRB14	206204_at	291
IDO1	210029_at	292
IFI44L	204439_at	293
IKZF3	221092_at	294
ITGAL	213475_s_at	295

KIF26A	NONMATCH	#N/A
KLHDC7B	NONMATCH	#N/A
LATS2	NONMATCH	#N/A
LRP4	212850_s_at	296
MFAP5	209758_s_at	297
MFAP5	213764_s_at	298
MFAP5	213765_at	299
MX1	202086_at	300
NAT1	214440_at	301
NLRC5	NONMATCH	#N/A
OLFM4	212768_s_at	302
OR2I1P	NONMATCH	#N/A
PI15	207938_at	303
PPP1R1A	205478_at	304
PRAME	204086_at	305
PRICKLE1	NONMATCH	#N/A
PTPRC	207238_s_at	306
PTPRC	212587_s_at	307
PTPRC	212588_at	308
RAC2	207419_s_at	309
RAC2	213603_s_at	310
RSAD2	213797_at	311
SP140L	214791_at	312
TSPAN7	202242_at	313

Table 3C – Probe set IDs for target sequences of genes contained in 44-gene signature as mapped to Affymetrix GeneChip® human genome U133 plus 2.0 array, plus corresponding gene symbols and SEQ ID NOs for probe sequences

Probeset ID	**Gene symbol**	**SEQ ID NO**
NONMATCH	AC138128.1	
1555380_at	ADAMTS4	364
NONMATCH	AL137218.1	
201012_at	ANXA1	365
233011_at	ANXA1	366
221087_s_at	APOL3	367
226545_at	CD109	368
229900_at	CD109	369
239719_at	CD109	370
205831_at	CD2	371
223834_at	CD274	372
207276_at	CDR1	373
1556687_a_at	CLDN10	374
205328_at	CLDN10	375
204533_at	CXCL10	376

206754_s_at	CYP2B6	377
206755_at	CYP2B6	378
217133_x_at	CYP2B6	379
1565483_at	EGFR	380
1565484_x_at	EGFR	381
201983_s_at	EGFR	382
201984_s_at	EGFR	383
210984_x_at	EGFR	384
211550_at	EGFR	385
211551_at	EGFR	386
211607_x_at	EGFR	387
201693_s_at	EGR1	388
201694_s_at	EGR1	389
227404_s_at	EGR1	390
205225_at	ESR1	391
211233_x_at	ESR1	392
211234_x_at	ESR1	393
211235_s_at	ESR1	394
211627_x_at	ESR1	395
215551_at	ESR1	396
215552_s_at	ESR1	397
217163_at	ESR1	398
217190_x_at	ESR1	399
221680_s_at	ETV7	400
224225_s_at	ETV7	401
229459_at	FAM19A5	402
229655_at	FAM19A5	403
237094_at	FAM19A5	404
202768_at	FOSB	405
205285_s_at	FYB	406
211794_at	FYB	407
211795_s_at	FYB	408
224148_at	FYB	409
227266_s_at	FYB	410
229625_at	GBP5	411
238581_at	GBP5	412
206204_at	GRB14	413
210029_at	IDO1	414
204439_at	IFI44L	415
221092_at	IKZF3	416
1554240_a_at	ITGAL	417
213475_s_at	ITGAL	418
232069_at	KIF26A	419
234307_s_at	KIF26A	420
1552639_at	KLHDC7B	421
236285_at	KLHDC7B	422

223379_s_at	LATS2	423
223380_s_at	LATS2	424
227013_at	LATS2	425
230348_at	LATS2	426
212850_s_at	LRP4	427
209758_s_at	MFAP5	428
213764_s_at	MFAP5	429
213765_at	MFAP5	430
202086_at	MX1	431
214440_at	NAT1	432
226474_at	NLRC5	433
212768_s_at	OLFM4	434
NONMATCH	OR2I1P	
207938_at	PI15	435
229947_at	PI15	436
205478_at	PPP1R1A	437
235129_at	PPP1R1A	438
204086_at	PRAME	439
226065_at	PRICKLE1	440
226069_at	PRICKLE1	441
230708_at	PRICKLE1	442
232811_x_at	PRICKLE1	443
1552480_s_at	PTPRC	444
1569830_at	PTPRC	445
207238_s_at	PTPRC	446
212587_s_at	PTPRC	447
212588_at	PTPRC	448
207419_s_at	RAC2	449
213603_s_at	RAC2	450
213797_at	RSAD2	451
242625_at	RSAD2	452
214791_at	SP140L	453
223934_at	SP140L	454
202242_at	TSPAN7	455

In different embodiments, subsets of the biomarkers listed in Tables 1A, 1B and/or 1C, Table 2A and/or Table 2B and/or Tables 3A and/or 3B and/or 3C may be used in the methods described herein. These subsets include but are not limited to biomarkers ranked 1-2, 1-3, 1-4, 1-5, 1-10, 1-20, 1-30, 1-40, 1-44, 6-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 36-44, 11-20, 21-30, 31-40, and 31-44 in Table 2A or Table 2B. In one aspect, therapeutic responsiveness is predicted in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to at least one of the biomarkers from Table 1A and at least N additional biomarkers selected from the list of biomarkers in Table 1A, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46 ,47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57.

In one aspect, therapeutic responsiveness is predicted in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to at least one of the biomarkers GBP5, CXCL10, IDO1 and MX1 and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, or 36. As used herein, the term "biomarker" can refer to a gene, an mRNA, cDNA, an antisense transcript, a miRNA, a polypeptide, a protein, a protein fragment, or any other nucleic acid sequence or polypeptide sequence that indicates either gene expression levels or protein production levels. In some embodiments, when referring to a biomarker of CXCL10, IDO1, CD2, GBP5, PRAME, ITGAL, LRP4, APOL3, CDR1, FYB, TSPAN7, RAC2, KLHDC7B, GRB14, AC138128.1, KIF26A, CD274, ETV7, MFAP5, OLFM4, PI15, FOSB, FAM19A5, NLRC5, PRICKLE1, EGR1, CLDN10, ADAMTS4, SP140L, ANXA1, RSAD2, ESR1, IKZF3, OR2I1P, EGFR, NAT1, LATS2, CYP2B6, PTPRC, PPP1R1A, or AL137218.1, the biomarker comprises an mRNA of CXCL10, IDO1, CD2, GBP5, PRAME, ITGAL, LRP4, APOL3, CDR1, FYB, TSPAN7, RAC2, KLHDC7B, GRB14, AC138128.1, KIF26A, CD274, ETV7, MFAP5, OLFM4, PI15, FOSB, FAM19A5, NLRC5, PRICKLE1, EGR1, CLDN10, ADAMTS4, SP140L, ANXA1, RSAD2, ESR1, IKZF3, OR2I1P, EGFR, NAT1, LATS2, CYP2B6, PTPRC, PPP1R1A, or AL137218.1, respectively. In further or other embodiments, when referring to a biomarker of MX1, GBP5, IFI44L, BIRC3, IGJ, IQGAP3, LOC100294459, SIX1, SLC9A3R1, STAT1, TOB1, UBD, C1QC, C2orf14, EPSTI, GALNT6, HIST1H4H, HIST2H4B, KIAA1244, LOC100287927, LOC100291682, or LOC100293679, the biomarker comprises an antisense transcript of MX1, IFI44L, GBP5, BIRC3, IGJ, IQGAP3, LOC100294459, SIX1, SLC9A3R1, STAT1, TOB1, UBD, C1QC, C2orf14, EPSTI, GALNT6, HIST1H4H, HIST2H4B, KIAA1244, LOC100287927, LOC100291682, or LOC100293679, respectively.

In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarkers GBP5, CXCL10, IDO1 and MX1 and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, or 36. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarker GBP5 and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38 or 39. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarker CXCL10 and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38 or 39. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarker IDO1 and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38

or 39. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarker MX-1 and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38 or 39.

In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to at least two of the biomarkers CXCL10, MX1, IDO1 and IFI44L and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarkers CXCL10, MX1, IDO1 and IFI44L and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarker CXCL10 and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42 or 43. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarker MX1 and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42 or 43. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarker IDO1 and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42 or 43. In a further aspect, therapeutic responsiveness is predicted, or a cancer diagnosis is indicated, in an individual by conducting an assay on a test (biological) sample from the individual and detecting biomarker values that each correspond to the biomarker IFI44L and one of at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42 or 43.

In other embodiments, the target sequences/probes listed in Tables 1A, 3A, 3B and/or 3C, or subsets thereof, may be used in the methods described herein. The target sequences may be utilised for the purposes of designing primers and/or probes which hybridize to the target sequences. Design of suitable primers and/or probes is within the capability of one skilled in the art once the target sequence is

identified. Various primer design tools are freely available to assist in this process, such as the NCBI Primer-BLAST tool; see Ye *et al*, BMC Bioinformatics. 13:134 (2012). The primers and/or probes may be designed such that they hybridize to the target sequence under stringent conditions (as defined herein). Primers and/or probes may be at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 (or more) nucleotides in length. It should be understood that each subset can include multiple primers and/or probes directed to the same biomarker. The tables show in some cases multiple target sequences within the same overall gene. Such primers and/or probes may be included in kits useful for performing the methods of the invention. The kits may be array or PCR based kits for example and may include additional reagents, such as a polymerase and/or dNTPs for example.

Measuring Gene Expression Using Classifier Models

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

When a biomarker indicates or is a sign of an abnormal process, disease or other condition in an individual, that biomarker is generally described as being either over-expressed or under-expressed as compared to an expression level or value of the biomarker that indicates or is a sign of a normal process, an absence of a disease or other condition in an individual. "Up-regulation", "up-regulated", "over-expression", "over-expressed", and any variations thereof are used interchangeably to refer to a value or level of a biomarker in a biological sample that is greater than a value or level (or range of values or levels) of the biomarker that is typically detected in similar biological samples from healthy or normal individuals. The terms may also refer to a value or level of a biomarker in a biological sample that is greater than a value or level (or range of values or levels) of the biomarker that may be detected at a different stage of a particular disease.

"Down-regulation", "down-regulated", "under-expression", "under-expressed", and any variations thereof are used interchangeably to refer to a value or level of a biomarker in a biological sample that is less than a value or level (or range of values or levels) of the biomarker that is typically detected in similar biological samples from healthy or normal individuals. The terms may also refer to a value or level of a biomarker in a biological sample that is less than a value or level (or range of values or levels) of the biomarker that may be detected at a different stage of a particular disease.

Further, a biomarker that is either over-expressed or under-expressed can also be referred to as being "differentially expressed" or as having a "differential level" or "differential value" as compared to a "normal" expression level or value of the biomarker that indicates or is a sign of a normal process or an absence of a disease or other condition in an individual. Thus, "differential expression" of a biomarker can also be referred to as a variation from a "normal" expression level of the biomarker.

The terms "differential biomarker expression" and "differential expression" are used interchangeably to refer to a biomarker whose expression is activated to a higher or lower level in a subject suffering from a specific disease, relative to its expression in a normal subject, or relative to its expression in a patient that responds differently to a particular therapy or has a different prognosis. The

terms also include biomarkers whose expression is activated to a higher or lower level at different stages of the same disease. It is also understood that a differentially expressed biomarker may be either activated or inhibited at the nucleic acid level or protein level, or may be subject to alternative splicing to result in a different polypeptide product. Such differences may be evidenced by a variety of changes including mRNA levels, miRNA levels, antisense transcript levels, or protein surface expression, secretion or other partitioning of a polypeptide. Differential biomarker expression may include a comparison of expression between two or more genes or their gene products; or a comparison of the ratios of the expression between two or more genes or their gene products; or even a comparison of two differently processed products of the same gene, which differ between normal subjects and subjects suffering from a disease; or between various stages of the same disease. Differential expression includes both quantitative, as well as qualitative, differences in the temporal or cellular expression pattern in a biomarker among, for example, normal and diseased cells, or among cells which have undergone different disease events or disease stages.

In certain embodiments, the expression profile obtained is a genomic or nucleic acid expression profile, where the amount or level of one or more nucleic acids in the sample is determined. In these embodiments, the sample that is assayed to generate the expression profile (i.e. to measure the expression levels of the one or more biomarkers in the sample) employed in the diagnostic or prognostic methods comprises a nucleic acid sample. The nucleic acid sample includes a population of nucleic acids that includes the expression information of the phenotype determinative biomarkers of the cell or tissue being analyzed. In some embodiments, the nucleic acid may include RNA or DNA nucleic acids, e.g., mRNA, cRNA, cDNA etc., so long as the sample retains the expression information of the host cell or tissue from which it is obtained. The sample may be prepared in a number of different ways, as is known in the art, e.g., by mRNA isolation from a cell, where the isolated mRNA is used as isolated, amplified, or employed to prepare cDNA, cRNA, etc., as is known in the field of differential gene expression. Accordingly, determining the level of mRNA in a sample includes preparing cDNA or cRNA from the mRNA and subsequently measuring the cDNA or cRNA. The sample is typically prepared from a cell or tissue harvested from a subject in need of treatment, e.g., via biopsy of tissue, using standard protocols, where cell types or tissues from which such nucleic acids may be generated include any tissue in which the expression pattern of the to be determined phenotype exists, including, but not limited to, disease cells or tissue, body fluids, etc.

The expression profile, representing the measured expression levels of one or more biomarkers in the test sample may be generated from the initial nucleic acid sample using any convenient protocol. While a variety of different manners of generating expression profiles are known, such as those employed in the field of differential gene expression/biomarker analysis, one representative and convenient type of protocol for generating expression profiles is array-based gene expression profile generation protocols. Such applications are hybridization assays in which a surface such as a (glass) chip, on which several probes for each of several thousand genes are immobilized is employed. On these surfaces there are generally multiple target regions within each gene to be analysed, and multiple (usually from 11 to 100) probes per target region. In this way, expression of each gene is evaluated by hybridization to multiple (tens) of probes on the surface. In these assays, a sample of target nucleic acids is first prepared from the initial nucleic acid sample being assayed, where preparation may include labeling of the target nucleic acids with a label, e.g., a member of a signal producing system. Following

target nucleic acid sample preparation, the sample is contacted with the array under hybridization conditions, whereby complexes are formed between target nucleic acids that are complementary to probe sequences attached to the array surface. The presence of hybridized complexes is then detected, either qualitatively or quantitatively. Specific hybridization technology which may be practiced to generate the expression profiles employed in the subject methods includes the technology described in U.S. Pat. Nos. 5,143,854; 5,288,644; 5,324,633; 5,432,049; 5,470,710; 5,492,806; 5,503,980; 5,510,270; 5,525,464; 5,547,839; 5,580,732; 5,661,028; 5,800,992; the disclosures of which are herein incorporated by reference; as well as WO 95/21265; WO 96/31622; WO 97/10365; WO 97/27317; EP 373 203; and EP 785 280. In these methods, an array of "probe" nucleic acids that includes one or several probes for each of the biomarkers whose expression is being assayed is contacted with target nucleic acids as described above. Contact is carried out under hybridization conditions, e.g., stringent hybridization conditions as described above, and unbound nucleic acid is then removed. The resultant pattern of hybridized nucleic acids provides information regarding expression for each of the biomarkers that have been probed, where the expression information is in terms of whether or not the gene is expressed and, typically, at what level, where the expression data, i.e., expression profile, may be both qualitative and quantitative. The methods may include normalizing the hybridization pattern against a subset of or all other probes on the array.

Creating a Biomarker Expression Classifier

In one embodiment, the relative expression levels of biomarkers in a cancer tissue are measured to form a gene expression profile. The gene expression profile of a set of biomarkers from a patient tissue sample is summarized in the form of a compound decision score (or test score) and compared to a score threshold that may be mathematically derived from a training set of patient data. The score threshold separates a patient group based on different characteristics such as, but not limited to, responsiveness/non-responsiveness to treatment. The patient training set data is preferably derived from NSCLC tissue samples having been characterized by prognosis, likelihood of recurrence, long term survival, clinical outcome, treatment response, diagnosis, cancer classification, or personalized genomics profile. Alternatively it may represent a data set from a cohort of patients in which the molecular subtype (DDRD) is well defined and characterised. Expression profiles, and corresponding decision scores from patient samples (test scores) may be correlated with the characteristics of patient samples in the training set that are on the same side of the mathematically derived score decision threshold. The threshold of the linear classifier scalar output may be optimized to maximize the sum of sensitivity and specificity under cross-validation as observed within the training dataset. Alternatively the sensitivity and positive predictive value of the assay may be increased at the expense of the specificity and negative predictive value or vice versa depending on the proposed clinical utility of the test in different disease indications.

The overall expression data for a given sample is normalized using methods known to those skilled in the art in order to correct for differing amounts of starting material, varying efficiencies of the extraction and amplification reactions, etc. Using a linear classifier on the normalized data to make a diagnostic or prognostic call (e.g. responsiveness or resistance to therapeutic agent) effectively means to split the data space, i.e. all possible combinations of expression values for all genes in the classifier, into

two disjoint halves by means of a separating hyperplane. This split may be empirically derived on a large set of training examples, for example from patients showing responsiveness or resistance to a therapeutic agent. Without loss of generality, one can assume a certain fixed set of values for all but one biomarker, which would automatically define a threshold value for this remaining biomarker where the decision would change from, for example, responsiveness or resistance to a therapeutic agent. Expression values above this dynamic threshold would then either indicate resistance (for a biomarker with a negative weight) or responsiveness (for a biomarker with a positive weight) to a therapeutic agent. The precise value of this threshold depends on the actual measured expression profile of all other biomarkers within the classifier, but the general indication of certain biomarkers remains fixed, i.e. high values or "relative over-expression" always contributes to either a responsiveness (genes with a positive weight) or resistance (genes with a negative weights). Therefore, in the context of the overall gene expression classifier, relative expression can indicate if either up- or down-regulation of a certain biomarker is indicative of responsiveness or resistance to a therapeutic agent.

In one embodiment, the biomarker expression profile of a test sample, for example a patient tissue sample, is evaluated by a linear classifier. As used herein, a linear classifier refers to a weighted sum of the individual biomarker intensities into a compound decision score ("decision function"). The decision score is then compared to a pre-defined cut-off score threshold, corresponding to a certain set-point in terms of sensitivity and specificity which indicates if a sample is above the score threshold (decision function positive) or below (decision function negative).

Effectively, this means that the data space, i.e. the set of all possible combinations of biomarker expression values, is split into two mutually exclusive halves corresponding to different clinical classifications or predictions, e.g. one corresponding to responsiveness to a therapeutic agent and the other to resistance. In the context of the overall classifier, relative over-expression of a certain biomarker can either increase the decision score (positive weight) or reduce it (negative weight) and thus contribute to an overall decision of, for example, responsiveness or resistance to a therapeutic agent.

The term "area under the curve" or "AUC" refers to the area under the curve of a receiver operating characteristic (ROC) curve, both of which are well known in the art. AUC measures are useful for comparing the accuracy of a classifier across the complete data range. Classifiers with a greater AUC have a greater capacity to classify unknowns correctly between two groups of interest (e.g., NSCLC cancer samples and normal or control samples). ROC curves are useful for plotting the performance of a particular feature (e.g., any of the biomarkers described herein and/or any item of additional biomedical information) in distinguishing between two populations (e.g., individuals responding and not responding to a therapeutic agent). Typically, the feature data across the entire population (e.g., the cases and controls) are sorted in ascending order based on the value of a single feature. Then, for each value for that feature, the true positive and false positive rates for the data are calculated. The true positive rate is determined by counting the number of cases above the value for that feature and then dividing by the total number of cases. The false positive rate is determined by counting the number of controls above the value for that feature and then dividing by the total number of controls. Although this definition refers to scenarios in which a feature is elevated in cases compared to controls, this definition also applies to scenarios in which a feature is lower in cases compared to the controls (in such a scenario, samples below the value for that feature would be counted). ROC curves can be generated for a single feature as well as for other single outputs, for example, a combination of two or more features can be

mathematically combined (e.g., added, subtracted, multiplied, etc.) to provide a single sum value, and this single sum value can be plotted in a ROC curve. Additionally, any combination of multiple features, in which the combination derives a single output value, can be plotted in a ROC curve. These combinations of features may comprise a test. The ROC curve is the plot of the true positive rate (sensitivity) of a test against the false positive rate (1-specificity) of the test.

The interpretation of this quantity, i.e. the cut-off threshold responsiveness or resistance to a therapeutic agent, is derived in the development phase ("training") from a set of patients with known outcome. The corresponding weights and the responsiveness/resistance cut-off threshold for the decision score are fixed *a priori* from training data by methods known to those skilled in the art. In a preferred embodiment of the present method, Partial Least Squares Discriminant Analysis (PLS-DA) is used for determining the weights. (L. Ståhle, S. Wold, J. Chemom. 1 (1987) 185-196; D. V. Nguyen, D.M. Rocke, Bioinformatics 18 (2002) 39-50). Other methods for performing the classification, known to those skilled in the art, may also be used with the methods described herein, for example when applied to the transcripts of a lung cancer classifier.

Different methods can be used to convert quantitative data measured on these biomarkers into a prognosis or other predictive use. These methods include, but not limited to methods from the fields of pattern recognition (Duda et al. Pattern Classification, 2nd ed., John Wiley, New York 2001), machine learning (Schölkopf et al. Learning with Kernels, MIT Press, Cambridge 2002, Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford 1995), statistics (Hastie et al. The Elements of Statistical Learning, Springer, New York 2001), bioinformatics (Dudoit et al., 2002, J. Am. Statist. Assoc. 97:77-87, Tibshirani et al., 2002, Proc. Natl. Acad. Sci. USA 99:6567-6572) or chemometrics (Vandeginste, et al., Handbook of Chemometrics and Qualimetrics, Part B, Elsevier, Amsterdam 1998).

In a training step, a set of patient samples for both responsiveness/resistance cases are measured and the prediction method is optimised using the inherent information from this training data to optimally predict the training set or a future sample set. In this training step, the used method is trained or parameterised to predict from a specific intensity pattern to a specific predictive call. Suitable transformation or pre-processing steps might be performed with the measured data before it is subjected to the prognostic method or algorithm.

In a preferred embodiment of the invention, a weighted sum of the pre-processed intensity values for each transcript is formed and compared with a threshold value optimised on the training set (Duda et al. Pattern Classification, 2nd ed., John Wiley, New York 2001). The weights can be derived by a multitude of linear classification methods, including but not limited to Partial Least Squares (PLS, (Nguyen et al., 2002, Bioinformatics 18 (2002) 39-50)) or Support Vector Machines (SVM, (Schölkopf et al. Learning with Kernels, MIT Press, Cambridge 2002)).

In another embodiment of the invention, the data is transformed non-linearly before applying a weighted sum as described above. This non-linear transformation might include increasing the dimensionality of the data. The non-linear transformation and weighted summation might also be performed implicitly, e.g. through the use of a kernel function. (Schölkopf et al. Learning with Kernels, MIT Press, Cambridge 2002).

In another embodiment of the invention, a new data sample is compared with two or more class prototypes, being either real measured training samples or artificially created prototypes. This comparison is performed using suitable similarity measures, for example, but not limited to Euclidean

distance (Duda et al. Pattern Classification, 2nd ed., John Wiley, New York 2001), correlation coefficient (Van't Veer, et al. 2002, Nature 415:530) etc. A new sample is then assigned to the prognostic group with the closest prototype or the highest number of prototypes in the vicinity.

In another embodiment of the invention, decision trees (Hastie et al., The Elements of Statistical Learning, Springer, New York 2001) or random forests (Breiman, Random Forests, Machine Learning 45:5 2001) are used to make a prognostic call from the measured intensity data for the transcript set or their products.

In another embodiment of the invention neural networks (Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford 1995) are used to make a prognostic call from the measured intensity data for the transcript set or their products.

In another embodiment of the invention, discriminant analysis (Duda et al., Pattern Classification, 2nd ed., John Wiley, New York 2001), comprising but not limited to linear, diagonal linear, quadratic and logistic discriminant analysis, is used to make a prognostic call from the measured intensity data for the transcript set or their products.

In another embodiment of the invention, Prediction Analysis for Microarrays (PAM, (Tibshirani et al., 2002, Proc. Natl. Acad. Sci. USA 99:6567-6572)) is used to make a prognostic call from the measured intensity data for the transcript set or their products.

In another embodiment of the invention, Soft Independent Modelling of Class Analogy (SIMCA, (Wold, 1976, Pattern Recogn. 8:127-139)) is used to make a predictive call from the measured intensity data for the transcript set or their products.

In another embodiment of the invention, c-index is used to quantify predictive ability. This index applies biomarkers to a continuous response variable that can be censored. The c index is the proportion of all pairs of subjects whose survival times can be ordered such that the subject with the higher predicted survival is the one who survived longer. Two subject's survival times cannot be ordered if both subjects are censored or if one has failed and the follow up time of the other is less than the failure time of the first. The c index is the probability of concordance between predicted and observed survival, with c=0.5 for random prediction and c=1 for a perfectly discriminating model. (Frank E. Harrell, Jr. Regression Modeling Strategies, 2001).

Therapeutic agents

As described above, the methods described herein permit the classification of a patient suffering from NSCLC, including early stage NSCLC as responsive or non-responsive to a therapeutic agent that targets tumors with abnormal DNA repair (hereinafter referred to as a "DNA-damaging therapeutic agent"). As used herein "DNA-damaging therapeutic agent" includes agents known to damage DNA directly, agents that prevent DNA damage repair, agents that inhibit DNA damage signaling, agents that inhibit DNA damage induced cell cycle arrest, and agents that inhibit processes indirectly leading to DNA damage. Some current such therapeutics used to treat NSCLC include, but are not limited to, the following DNA-damaging therapeutic agents.

1) DNA damaging agents:

a. Alkylating agents (platinum containing agents such as cisplatin, carboplatin, and oxaliplatin; cyclophosphamide; busulphan).
b. Topoisomerase I inhibitors (irinotecan;topotecan)
c. Topisomerase II inhibitors (etoposide;anthracyclines such as doxorubicin and epirubicin)
d. Ionising radiation

2) DNA repair targeted therapies
   a. Inhibitors of Non-homologous end-joining (DNA-PK inhibitors, Nu7441, NU7026)
   b. Inhibitors of homologous recombination
   c. Inhibitors of nucleotide excision repair
   d. Inhibitors of base excision repair (PARP inhibitors, AG014699, AZD2281, ABT-888, MK4827, BSI-201, INO-1001, TRC-102, APEX 1 inhibitors, APEX 2 inhibitors, Ligase III inhibitors
   e. Inhibitors of the Fanconianemia pathway

3) Inhibitors of DNA damage signalling
   a. ATM inhibitors (CP466722)
   b. CHK 1 inhibitors (XL-844,UCN-01, AZD7762, PF00477736)
   c. CHK 2 inhibitors (XL-844, AZD7762, PF00477736)
   d. ATR inhibitors (AZ20)

4) Inhibitors of DNA damage induced cell cycle arrest
   a. Wee1 kinase inhibitors
   b. CDC25a, b or c inhibitors

5) Inhibition of processes indirectly leading to DNA damage
   a. Histone deacetylase inhibitors
   b. Heat shock protein inhibitors (geldanamycin, AUY922),

6) Inhibitors of DNA synthesis:
   a. Pyrimidine analogues (5-FU, gemcitabine)
   b. Prodrugs (capecitabine)

As discussed above, the therapeutic agents, for which responsiveness is predicted may be applied in an adjuvant setting. However, they may be utilised in a neoadjuvant setting additionally or alternatively.

The invention described herein is not limited to any one DNA-damaging therapeutic agent; it can be used to identify responders and non-responders to any of a range of DNA-damaging therapeutic agent, for example those that directly or indirectly affect DNA damage and/or DNA damage repair. In some embodiments, the DNA-damaging therapeutic agent comprises one or more substances selected from the group consisting of: a DNA damaging agent, a DNA repair targeted therapy, an inhibitor of DNA damage signalling, an inhibitor of DNA damage induced cell cycle arrest, a histone deacetylase inhibitor,

a heat shock protein inhibitor and an inhibitor of DNA synthesis. More specifically, the DNA-damaging therapeutic agent may be selected from one or more of a platinum-containing agent, a nucleoside analogue such as gemcitabine or 5-fluorouracil or a prodrug thereof such as capecitabine, an anthracycline such as epirubicin or doxorubicin, an alkylating agent such as cyclophosphamide, an ionising radiation or a combination of radiation and chemotherapy (chemoradiation). In particular embodiments, the DNA-damaging therapeutic agent comprises a platinum-containing agent, such as a platinum based agent selected from cisplatin, carboplatin and oxaliplatin. The methods and kits may predict responsiveness to treatment with the DNA-damaging therapeutic agent together with a further drug. Thus, the methods and kits may predict responsiveness to a combination therapy. For example, it is shown experimentally herein that the methods of the invention can identify a subpopulation of NSCLC patients who are more likely to benefit to adjuvant cisplatin based therapy, in combination with vinorelbine. Thus, in some embodiments, the further drug is a mitotic inhibitor. The mitotic inhibitor may be a vinca alkaloid or a taxane. In specific embodiments, the vinca alkaloid is vinorelbine In certain embodiments, responders to the following treatments are identified: cisplatin/carboplatin, Cisplatin/carboplatin and 5-fluorouracil (5-FU) (CF), cisplatin/carboplatin and capecitabine (CX), epirubicin/doxyrubicin, cisplatin/carboplatin and fluorouracil (ECF), epirubicin, oxaliplatin and capecitabine (EOX), gemcitabine, cyclophosphamide, radiation and chemoradiation. In specific aspects this invention, it is useful for evaluating cisplatin/ carboplatin (Paraplatin), cisplatin/carboplatin and etoposide (CP), gemcitabine and cisplatin/carboplatin (GemCarbo) cyclophosphamide epirubicin/doxorubicin and vincristine (CEV/CAV), CEV/CAV plus etoposide (CEVE/CAVE), epirubicin/doxorubicin, cyclophosphamide and etoposide (ECE/ACE) a combination of DNA damaging agents with topotecan, or cisplatin or carboplatin (Paraplatin) with at least one other drug such as Vinorelbine, Gemcitabine, Paclitaxel (Taxol), Docetaxel (Taxotere), epirubicin/Doxorubicin, Etoposide, Pemetrexed or radiation in treatment of NSCLC.

Diseases and Tissue Sources

The predictive classifiers described herein are useful for determining responsiveness or resistance to a therapeutic agent for treating lung cancer, in particular NSCLC.

The lung cancer is typically non-small cell lung cancer (NSCLC) and may be early stage. The NSCLC may be selected from one or more of adenocarcinoma, large-cell lung carcinoma and squamous cell carcinoma.

In one embodiment, the methods described herein refer to NSCLCs that are treated with chemotherapeutic agents of the classes DNA damaging agents, DNA repair target therapies, inhibitors of DNA damage signalling, inhibitors of DNA damage induced cell cycle arrest, inhibition of processes indirectly leading to DNA damage and inhibition of DNA synthesis, but not limited to these classes. Each of these chemotherapeutic agents is considered a “DNA-damaging therapeutic agent” as the term is used herein.

"Biological sample", "sample", and "test sample" are used interchangeably herein to refer to any material, biological fluid, tissue, or cell obtained or otherwise derived from an individual. This includes blood (including whole blood, leukocytes, peripheral blood mononuclear cells, buffy coat, plasma, and serum), sputum, tears, mucus, nasal washes, nasal aspirate, breath, urine, semen, saliva, meningeal

fluid, amniotic fluid, glandular fluid, lymph fluid, nipple aspirate, bronchial aspirate, synovial fluid, joint aspirate, ascites, cells, a cellular extract, and cerebrospinal fluid. This also includes experimentally separated fractions of all of the preceding. For example, a blood sample can be fractionated into serum or into fractions containing particular types of blood cells, such as red blood cells or white blood cells (leukocytes). If desired, a sample can be a combination of samples from an individual, such as a combination of a tissue and fluid sample. The term "biological sample" also includes materials containing homogenized solid material, such as from a stool sample, a tissue sample, or a tissue biopsy, for example. The term "biological sample" also includes materials derived from a tissue culture or a cell culture. Any suitable methods for obtaining a biological sample can be employed; exemplary methods include, e.g., phlebotomy, swab (e.g., buccal swab), and a fine needle aspirate biopsy procedure. Samples may be obtained by bronchoscopy or by sputum cytology in some embodiments. A "biological sample" obtained or derived from an individual includes any such sample that has been processed in any suitable manner after being obtained from the individual.

In such cases, the target cells may be tumor cells, for example NSCLC cells. The target cells are derived from any tissue source, including human and animal tissue, such as, but not limited to, a newly obtained sample, a frozen sample, a biopsy sample, a sample of bodily fluid, a blood sample, preserved tissue such as a paraffin-embedded fixed tissue sample (i.e., a tissue block), or cell culture.

In some specific embodiments, the samples may or may not comprise vesicles.

Methods and Kits

Kits for Gene Expression Analysis

Reagents, tools, and/or instructions for performing the methods described herein can be provided in a kit. For example, the kit can contain reagents, tools, and instructions for determining an appropriate therapy for a lung cancer patient. Such a kit can include reagents for collecting a tissue sample from a patient, such as by biopsy, and reagents for processing the tissue. The kit can also include one or more reagents for performing a biomarker expression analysis, such as reagents for performing nucleic acid amplification, including RT-PCR and qPCR, NGS, northern blot, proteomic analysis, or immunohistochemistry to determine expression levels of biomarkers in a sample of a patient. For example, primers for performing RT-PCR, probes for performing northern blot analyses, and/or antibodies for performing proteomic analysis such as Western blot, immunohistochemistry and ELISA analyses can be included in such kits. Appropriate buffers for the assays can also be included. Detection reagents required for any of these assays can also be included. The appropriate reagents and methods are described in further detail below.

In certain embodiments, the target sequences listed in Tables 1A, 3A, 3B and 3C (and also 1B and 1C in some embodiments), or subsets thereof, may be used in the methods and kits described herein (such as SEQ ID NO: 1-80 (Table 1A), 81-260 (Table 3A), 261-313 (Table 3B), 314-337 (Table 1B), 338-363 (Table 1C), 364-455 (Table 3C)). The target sequences may be utilised for the purposes of designing primers and/or probes which hybridize to the target sequences. Design of suitable primers and/or probes is within the capability of one skilled in the art once the target sequence is identified. Various primer design tools are freely available to assist in this process such as the NCBI Primer-BLAST

tool. The primers and/or probes may be designed such that they hybridize to the target sequence under stringent conditions. Primers and/or probes may be at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 (or more) nucleotides in length. It should be understood that each subset can include multiple primers and/or probes directed to the same biomarker. The tables show in some cases multiple target sequences within the same overall gene. Such primers and/or probes may be included in kits useful for performing the methods of the invention. The kits may be array or PCR based kits for example and may include additional reagents, such as a polymerase and/or dNTPs for example. The kits featured herein can also include an instruction sheet describing how to perform the assays for measuring biomarker expression. The instruction sheet can also include instructions for how to determine a reference cohort, including how to determine expression levels of biomarkers in the reference cohort and how to assemble the expression data to establish a reference for comparison to a test patient. The instruction sheet can also include instructions for assaying biomarker expression in a test patient and for comparing the expression level with the expression in the reference cohort to subsequently determine the appropriate chemotherapy for the test patient. Methods for determining the appropriate chemotherapy are described above and can be described in detail in the instruction sheet.

Informational material included in the kits can be descriptive, instructional, marketing or other material that relates to the methods described herein and/or the use of the reagents for the methods described herein. For example, the informational material of the kit can contain contact information, e.g., a physical address, email address, website, or telephone number, where a user of the kit can obtain substantive information about performing a gene expression analysis and interpreting the results, particularly as they apply to a human's likelihood of having a positive response to a specific therapeutic agent.

The kits featured herein can also contain software necessary to infer a patient's likelihood of having a positive response to a specific therapeutic agent from the biomarker expression.

The kits may, in some embodiments, additionally contain the DNA-damaging therapeutic agent for administration in the event that the individual is predicted to be responsive. Any of the specific agents or combinations of agents described herein to treat NSCLC may be incorporated into the kits. The agent or combination of agents may be provided in a form, such as a dosage form, that is tailored to NSCLC treatment specifically. The kit may be provided with suitable instructions for administration according to NSCLC treatment regimens, for example in the context of adjuvant and/or neo-adjuvant treatment.

a) Gene expression profiling methods

Measuring mRNA in a biological sample may be used as a surrogate for detection of the level of the corresponding protein in the biological sample. Thus, any of the biomarkers or biomarker panels described herein can also be detected by detecting the appropriate RNA. Methods of gene expression profiling include, but are not limited to, microarray, RT-PCT, qPCR, NGS, northern blots, SAGE, mass spectrometry.

mRNA expression levels are measured by reverse transcription quantitative polymerase chain reaction (RT-PCR followed with qPCR). RT-PCR is used to create a cDNA from the mRNA. The cDNA may be used in a qPCR assay to produce fluorescence as the DNA amplification process progresses. By comparison to a standard curve, qPCR can produce an absolute measurement such as number of copies

of mRNA per cell. Northern blots, microarrays, Invader assays, and RT-PCR combined with capillary electrophoresis have all been used to measure expression levels of mRNA in a sample. See Gene Expression Profiling: Methods and Protocols, Richard A. Shimkets, editor, Humana Press, 2004.

miRNA molecules are small RNAs that are non-coding but may regulate gene expression. Any of the methods suited to the measurement of mRNA expression levels can also be used for the corresponding miRNA. Recently many laboratories have investigated the use of miRNAs as biomarkers for disease. Many diseases involve widespread transcriptional regulation, and it is not surprising that miRNAs might find a role as biomarkers. The connection between miRNA concentrations and disease is often even less clear than the connections between protein levels and disease, yet the value of miRNA biomarkers might be substantial. Of course, as with any RNA expressed differentially during disease, the problems facing the development of an in vitro diagnostic product will include the requirement that the miRNAs survive in the diseased cell and are easily extracted for analysis, or that the miRNAs are released into blood or other matrices where they must survive long enough to be measured. Protein biomarkers have similar requirements, although many potential protein biomarkers are secreted intentionally at the site of pathology and function, during disease, in a paracrine fashion. Many potential protein biomarkers are designed to function outside the cells within which those proteins are synthesized.

Gene expression may also be evaluated using mass spectrometry methods. A variety of configurations of mass spectrometers can be used to detect biomarker values. Several types of mass spectrometers are available or can be produced with various configurations. In general, a mass spectrometer has the following major components: a sample inlet, an ion source, a mass analyzer, a detector, a vacuum system, and instrument-control system, and a data system. Difference in the sample inlet, ion source, and mass analyzer generally define the type of instrument and its capabilities. For example, an inlet can be a capillary-column liquid chromatography source or can be a direct probe or stage such as used in matrix-assisted laser desorption. Common ion sources are, for example, electrospray, including nanospray and microspray or matrix-assisted laser desorption. Common mass analyzers include a quadrupole mass filter, ion trap mass analyzer and time-of-flight mass analyzer. Additional mass spectrometry methods are well known in the art (see Burlingame et al., Anal. Chem. 70:647 R-716R (1998); Kinter and Sherman, New York (2000)).

Protein biomarkers and biomarker values can be detected and measured by any of the following: electrospray ionization mass spectrometry (ESI-MS), ESI-MS/MS, ESI-MS/(MS)n, matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS), desorption/ionization on silicon (DIOS), secondary ion mass spectrometry (SIMS), quadrupole time-of-flight (Q-TOF), tandem time-of-flight (TOF/TOF) technology, called ultraflex III TOF/TOF, atmospheric pressure chemical ionization mass spectrometry (APCI-MS), APCI-MS/MS, APCI-(MS).sup.N, atmospheric pressure photoionization mass spectrometry (APPI-MS), APPI-MS/MS, and APPI-(MS).sup.N, quadrupole mass spectrometry, Fourier transform mass spectrometry (FTMS), quantitative mass spectrometry, and ion trap mass spectrometry.

Sample preparation strategies are used to label and enrich samples before mass spectroscopic characterization of protein biomarkers and determination biomarker values. Labeling methods include but are not limited to isobaric tag for relative and absolute quantitation (iTRAQ) and stable isotope labeling with amino acids in cell culture (SILAC). Capture reagents used to selectively enrich samples for

candidate biomarker proteins prior to mass spectroscopic analysis include but are not limited to aptamers, antibodies, nucleic acid probes, chimeras, small molecules, an $F(ab')_2$ fragment, a single chain antibody fragment, an Fv fragment, a single chain Fv fragment, a nucleic acid, a lectin, a ligand-binding receptor, affybodies, nanobodies, ankyrins, domain antibodies, alternative antibody scaffolds (e.g. diabodiesetc) imprinted polymers, avimers, peptidomimetics, peptoids, peptide nucleic acids, threose nucleic acid, a hormone receptor, a cytokine receptor, and synthetic receptors, and modifications and fragments of these.

The foregoing assays enable the detection of biomarker values that are useful in methods for predicting responsiveness of a cancer therapeutic agent, where the methods comprise detecting, in a biological sample from an individual suffering from NSCLC, at least N biomarker values that each correspond to a biomarker selected from the group consisting of the biomarkers provided in Tables 1 to 3, wherein a classification, as described in detail below, using the biomarker values indicates whether the individual will be responsive to a therapeutic agent. While certain of the described predictive biomarkers are useful alone for predicting responsiveness to a therapeutic agent, methods are also described herein for the grouping of multiple subsets of the biomarkers that are each useful as a panel of two or more biomarkers. Thus, various embodiments of the instant application provide combinations comprising N biomarkers, wherein N is at least three biomarkers. It will be appreciated that N can be selected to be any number from any of the above-described ranges, as well as similar, but higher order, ranges. In accordance with any of the methods described herein, biomarker values can be detected and classified individually or they can be detected and classified collectively, as for example in a multiplex assay format.

b) Microarray methods

In one embodiment, the present invention makes use of "oligonucleotide arrays" (also called herein "microarrays"). Microarrays can be employed for analyzing the expression of biomarkers in a cell, and especially for measuring the expression of biomarkers of cancer tissues.

In one embodiment, biomarker arrays are produced by hybridizing detectably labeled polynucleotides representing the mRNA transcripts present in a cell (e.g., fluorescently-labeled cDNA synthesized from total cell mRNA or labeled cRNA) to a microarray. A microarray is a surface with an ordered array of binding (e.g., hybridization) sites for products of many of the genes in the genome of a cell or organism, preferably most or almost all of the genes. Microarrays can be made in a number of ways known in the art. However produced, microarrays share certain characteristics. The arrays are reproducible, allowing multiple copies of a given array to be produced and easily compared with each other. Preferably the microarrays are small, usually smaller than 5 cm^2, and they are made from materials that are stable under binding (e.g., nucleic acid hybridization) conditions. A given binding site or unique set of binding sites in the microarray will specifically bind the product of a single gene in the cell. In a specific embodiment, positionally addressable arrays containing affixed nucleic acids of known sequence at each location are used.

It will be appreciated that when cDNA complementary to the RNA of a cell is made and hybridized to a microarray under suitable hybridization conditions, the level of hybridization to the site in the array corresponding to any particular gene will reflect the prevalence in the cell of mRNA transcribed

from that gene/biomarker. For example, when detectably labeled (e.g., with a fluorophore) cDNA or cRNA complementary to the total cellular mRNA is hybridized to a microarray, the site on the array corresponding to a gene (i.e., capable of specifically binding the product of the gene) that is not transcribed in the cell will have little or no signal (e.g., fluorescent signal), and a gene for which the encoded mRNA is prevalent will have a relatively strong signal. Nucleic acid hybridization and wash conditions are chosen so that the probe "specifically binds" or "specifically hybridizes' to a specific array site, i.e., the probe hybridizes, duplexes or binds to a sequence array site with a complementary nucleic acid sequence but does not hybridize to a site with a non-complementary nucleic acid sequence. As used herein, one polynucleotide sequence is considered complementary to another when, if the shorter of the polynucleotides is less than or equal to 25 bases, there are no mismatches using standard base-pairing rules or, if the shorter of the polynucleotides is longer than 25 bases, there is no more than a 5% mismatch. Preferably, the polynucleotides are perfectly complementary (no mismatches). It can be demonstrated that specific hybridization conditions result in specific hybridization by carrying out a hybridization assay including negative controls using routine experimentation.

Optimal hybridization conditions will depend on the length (e.g., oligomer vs. polynucleotide greater than 200 bases) and type (e.g., RNA, DNA, PNA) of labeled probe and immobilized polynucleotide or oligonucleotide. General parameters for specific (i.e., stringent) hybridization conditions for nucleic acids are described in Sambrook et al., supra, and in Ausubel et al., "Current Protocols in Molecular Biology", Greene Publishing and Wiley-interscience, NY (1987), which is incorporated in its entirety for all purposes. When the cDNA microarrays are used, typical hybridization conditions are hybridization in 5xSSC plus 0.2% SDS at 65C for 4 hours followed by washes at 25℃ in low stringency wash buffer (1xSSC plus 0.2% SDS) followed by 10 minutes at 25℃ in high stringency wash buffer (0.1SSC plus 0.2% SDS) (see Shena *et al.*, Proc. Natl. Acad. Sci. USA, Vol. 93, p. 10614 (1996)). Useful hybridization conditions are also provided in, e.g., Tijessen, Hybridization With Nucleic Acid Probes", Elsevier Science Publishers B.V. (1993) and Kricka, "Nonisotopic DNA Probe Techniques", Academic Press, San Diego, Calif. (1992).

Microarray platforms include those manufactured by companies such as Affymetrix, Illumina and Agilent. Examples of microarray platforms manufactured by Affymetrix include the U133 Plus2 array, the Almac proprietary Xcel™ array and the Almac proprietary Cancer DSAs®, including the Breast Cancer DSA® and Lung Cancer DSA®.

c) Immunoassay methods

Immunoassay methods are based on the reaction of an antibody to its corresponding target or analyte and can detect the analyte in a sample depending on the specific assay format. To improve specificity and sensitivity of an assay method based on immunoreactivity, monoclonal antibodies are often used because of their specific epitope recognition. Polyclonal antibodies have also been successfully used in various immunoassays because of their increased affinity for the target as compared to monoclonal antibodies Immunoassays have been designed for use with a wide range of biological sample matrices Immunoassay formats have been designed to provide qualitative, semi-quantitative, and quantitative results.

Quantitative results may be generated through the use of a standard curve created with known

concentrations of the specific analyte to be detected. The response or signal from an unknown sample is plotted onto the standard curve, and a quantity or value corresponding to the target in the unknown sample is established.

Numerous immunoassay formats have been designed. ELISA or EIA can be quantitative for the detection of an analyte/biomarker. This method relies on attachment of a label to either the analyte or the antibody and the label component includes, either directly or indirectly, an enzyme. ELISA tests may be formatted for direct, indirect, competitive, or sandwich detection of the analyte. Other methods rely on labels such as, for example, radioisotopes (I^{125}) or fluorescence. Additional techniques include, for example, agglutination, nephelometry, turbidimetry, Western blot, immunoprecipitation, immunocytochemistry, immunohistochemistry, flow cytometry, Luminex assay, and others (see ImmunoAssay: A Practical Guide, edited by Brian Law, published by Taylor & Francis, Ltd., 2005 edition).

Exemplary assay formats include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay, fluorescent, chemiluminescence, and fluorescence resonance energy transfer (FRET) or time resolved-FRET (TR-FRET) immunoassays. Examples of procedures for detecting biomarkers include biomarker immunoprecipitation followed by quantitative methods that allow size and peptide level discrimination, such as gel electrophoresis, capillary electrophoresis, planar electrochromatography, and the like.

Methods of detecting and/or quantifying a detectable label or signal generating material depend on the nature of the label. The products of reactions catalyzed by appropriate enzymes (where the detectable label is an enzyme; see above) can be, without limitation, fluorescent, luminescent, or radioactive or they may absorb visible or ultraviolet light. Examples of detectors suitable for detecting such detectable labels include, without limitation, x-ray film, radioactivity counters, scintillation counters, spectrophotometers, colorimeters, fluorometers, luminometers, and densitometers.

Any of the methods for detection can be performed in any format that allows for any suitable preparation, processing, and analysis of the reactions. This can be, for example, in multi-well assay plates (e.g., 96 wells or 384 wells) or using any suitable array or microarray. Stock solutions for various agents can be made manually or robotically, and all subsequent pipetting, diluting, mixing, distribution, washing, incubating, sample readout, data collection and analysis can be done robotically using commercially available analysis software, robotics, and detection instrumentation capable of detecting a detectable label.

Clinical Uses

In some embodiments, methods are provided for identifying and/or selecting a NSCL cancer patient who is responsive to a therapeutic regimen. In particular, the methods are directed to identifying or selecting a cancer patient who is responsive to a therapeutic regimen that includes administering an agent that directly or indirectly damages DNA. Methods are also provided for identifying a patient who is non-responsive to a therapeutic regimen. These methods typically include determining the level of expression of a collection of predictive markers in a patient's tumor (primary, metastatic or other derivatives from the tumor such as, but not limited to, blood, or components in blood, urine, saliva and other bodily fluids)(e.g., a patient's cancer cells), comparing the level of expression to a reference expression level, and identifying whether expression in the sample includes a pattern or profile of

expression of a selected predictive biomarker or biomarker set which corresponds to response or non-response to therapeutic agent.

In some embodiments a method of predicting responsiveness of an individual having non-small cell lung cancer (NSCLC) to treatment with a DNA-damaging therapeutic agent comprises:

a. measuring expression levels of one or more biomarkers in a test sample obtained from the individual, wherein the one or more biomarkers are selected from Table 1A, 1B, 1C, 2A, 2B, 3A, 3B or 3C;
b. deriving a test score that captures the expression levels;
c. providing a threshold score comprising information correlating the test score and responsiveness;
d. and comparing the test score to the threshold score; wherein responsiveness is predicted when the test score exceeds the threshold score.

In specific embodiments, a method of predicting responsiveness of an individual having non-small cell lung cancer (NSCLC) to treatment with a DNA-damaging therapeutic agent comprises the following steps: obtaining a test sample from the individual; measuring expression levels of one or more biomarkers in the test sample, wherein the one or more biomarkers are selected from the group consisting of CXCL10, MX1, IDO1, IF144L, CD2, GBP5, PRAME, ITGAL, LRP4, and APOL3; deriving a test score that captures the expression levels; providing a threshold score comprising information correlating the test score and responsiveness; and comparing the test score to the threshold score; wherein responsiveness is predicted when the test score exceeds the threshold score. One of ordinary skill in the art can determine an appropriate threshold score, and appropriate biomarker weightings, using the teachings provided herein including the teachings of Example 1.

In other embodiments, the method of predicting responsiveness of an individual having non-small cell lung cancer (NSCLC) to treatment with to a DNA-damaging therapeutic agent comprises measuring the expression levels of one or more biomarkers in the test sample, wherein the one or more biomarkers are selected from the group consisting of CXCL10, MX1, IDO1, IF144L, CD2, GBP5, PRAME, ITGAL, LRP4, APOL3, CDR1, FYB, TSPAN7, RAC2, KLHDC7B, GRB14, AC138128.1, KIF26A, CD274, CD109, ETV7, MFAP5, OLFM4, PI15, FOSB, FAM19A5, NLRC5, PRICKLE1, EGR1, CLDN10, ADAMTS4, SP140L, ANXA1, RSAD2, ESR1, IKZF3, OR2I1P, EGFR, NAT1, LATS2, CYP2B6, PTPRC, PPP1R1A, and AL137218.1. Tables 2A and 2B provide exemplary gene signatures (or gene classifiers) wherein the biomarkers consist of 40 or 44 of the gene products listed therein, respectively, and wherein a threshold score is derived from the individual gene product weightings listed therein. In one of these embodiments wherein the biomarkers consist of the 44 gene products listed in Table 2B, and the biomarkers are associated with the weightings provided in Table 2B, a test score that exceeds a threshold score, such as a threshold score of 0.3681 indicates a likelihood that the individual will be responsive to a DNA-damaging therapeutic agent.

A cancer is "responsive" to a therapeutic agent if its rate of growth is inhibited as a result of contact with the therapeutic agent, compared to its growth in the absence of contact with the therapeutic agent. Growth of a cancer can be measured in a variety of ways, for instance, the size of a tumor or the expression of tumor markers appropriate for that tumor type may be measured.

A cancer is "non-responsive" to a therapeutic agent if its rate of growth is not inhibited, or inhibited to a very low degree, as a result of contact with the therapeutic agent when compared to its

growth in the absence of contact with the therapeutic agent. As stated above, growth of a cancer can be measured in a variety of ways, for instance, the size of a tumor or the expression of tumor markers appropriate for that tumor type may be measured. The quality of being non-responsive to a therapeutic agent is a highly variable one, with different cancers exhibiting different levels of "non-responsiveness" to a given therapeutic agent, under different conditions. Still further, measures of non-responsiveness can be assessed using additional criteria beyond growth size of a tumor, including patient quality of life, degree of metastases, etc.

An application of this test will predict end points including, but not limited to, overall survival, progression free survival, radiological response, as defined by RECIST, complete response, partial response, stable disease and serological markers such as, but not limited to, PSA, CEA, CA125, CA15-3 and CA19-9. In specific embodiments this invention can be used to evaluate standard chest roentgenography, computed tomography (CT), perfusion CT, dynamic contrast material–enhanced magnetic resonance (MR) diffusion-weighted (DW) MR or positron emission tomography (PET) with the glucose analog fluorine 18 fluorodeoxyglucose (FDG) (FDG-PET) response in NSCLC treated with DNA damaging combination therapies, alone or in the context of standard treatment.

Array or non-array based methods for detection, quantification and qualification of RNA, DNA or protein within a sample of one or more nucleic acids or their biological derivatives such as encoded proteins may be employed, including quantitative PCR (QPCR), enzyme-linked immunosorbent assay (ELISA) or immunohistochemistry (IHC) and the like.

After obtaining an expression profile from a sample being assayed, the expression profile is compared with a reference or control profile to make a diagnosis regarding the therapy responsive phenotype of the cell or tissue, and therefore host, from which the sample was obtained. The terms "reference" and "control" as used herein in relation to an expression profile mean a standardized pattern of gene or gene product expression or levels of expression of certain biomarkers to be used to interpret the expression classifier of a given patient and assign a prognostic or predictive class. The reference or control expression profile may be a profile that is obtained from a sample known to have the desired phenotype, e.g., responsive phenotype, and therefore may be a positive reference or control profile. In addition, the reference profile may be from a sample known to not have the desired phenotype, and therefore be a negative reference profile.

If quantitative PCR is employed as the method of quantitating the levels of one or more nucleic acids, this method may quantify the PCR product accumulation through measurement of fluorescence released by a dual-labeled fluorogenic probe (e.g. a TaqMan® probe or a molecular beacon or FRET/Light Cycler probes). Some methods may not require a separate probe, such as the Scorpion and Ampliflyor systems where the probes are built into the primers.

In certain embodiments, the obtained expression profile is compared to a single reference profile to obtain information regarding the phenotype of the sample being assayed. In yet other embodiments, the obtained expression profile is compared to two or more different reference profiles to obtain more in depth information regarding the phenotype of the assayed sample. For example, the obtained expression profile may be compared to a positive and negative reference profile to obtain confirmed information regarding whether the sample has the phenotype of interest.

The comparison of the obtained expression profile and the one or more reference profiles may be performed using any convenient methodology, where a variety of methodologies are known to those

of skill in the array art, e.g., by comparing digital images of the expression profiles, by comparing databases of expression data, etc. Patents describing ways of comparing expression profiles include, but are not limited to, U.S. Pat. Nos. 6,308,170 and 6,228,575, the disclosures of which are herein incorporated by reference. Methods of comparing expression profiles are also described above.

The comparison step results in information regarding how similar or dissimilar the obtained expression profile is to the one or more reference profiles, which similarity information is employed to determine the phenotype of the sample being assayed. For example, similarity with a positive control indicates that the assayed sample has a responsive phenotype similar to the responsive reference sample. Likewise, similarity with a negative control indicates that the assayed sample has a non-responsive phenotype to the non-responsive reference sample.

The level of expression of a biomarker can be further compared to different reference expression levels. For example, a reference expression level can be a predetermined standard reference level of expression in order to evaluate if expression of a biomarker or biomarker set is informative and make an assessment for determining whether the patient is responsive or non-responsive. Additionally, determining the level of expression of a biomarker can be compared to an internal reference marker level of expression which is measured at the same time as the biomarker in order to make an assessment for determining whether the patient is responsive or non-responsive. For example, expression of a distinct marker panel which is not comprised of biomarkers of the invention, but which is known to demonstrate a constant expression level can be assessed as an internal reference marker level, and the level of the biomarker expression is determined as compared to the reference. In an alternative example, expression of the selected biomarkers in a tissue sample which is a non-tumor sample can be assessed as an internal reference marker level. The level of expression of a biomarker may be determined as having increased expression in certain aspects. The level of expression of a biomarker may be determined as having decreased expression in other aspects. The level of expression may be determined as no informative change in expression as compared to a reference level. In still other aspects, the level of expression is determined against a pre-determined standard expression level as determined by the methods provided herein.

The invention is also related to guiding conventional treatment of patients. Patients in which the diagnostics test reveals that they are responders to the drugs, of the classes that directly or indirectly affect DNA damage and/or DNA damage repair, can be administered with that therapy and both patient and oncologist can be confident that the patient will benefit. Patients that are designated non-responders by the diagnostic test can be identified for alternative therapies which are more likely to offer benefit to them.

The invention further relates to selecting patients for clinical trials where novel drugs of the classes that directly or indirectly affect DNA damage and/or DNA damage repair in order to treat NSCLC. Enrichment of trial populations with potential responders will facilitate a more thorough evaluation of that drug under relevant criteria.

The invention still further relates to methods of diagnosing patients as having or being susceptible to developing NSCLC associated with a DNA damage response deficiency (DDRD). DDRD is defined herein as any condition wherein a cell or cells of the patient have a reduced ability to repair DNA damage, which reduced ability is a causative factor in the development or growth of a tumor. The DDRD diagnosis may be associated with a mutation in the Fanconi anemia/BRCA pathway. The DDRD

diagnosis may also be associated with adenocarcinoma, large-cell lung carcinoma or squamous cell carcinoma. The methods of diagnosing an individual having non-small cell lung cancer (NSCLC) may comprise:

a. measuring expression levels of one or more biomarkers in a test sample obtained from the individual, wherein the one or more biomarkers are selected from Table 1A, 1B, 1C, 2A, 2B, 3A, 3B or 3C;
b. deriving a test score that captures the expression levels;
c. providing a threshold score comprising information correlating the test score and diagnosis of NSCLC;
d. and comparing the test score to the threshold score; wherein the individual is determined to have NSCLC or be susceptible to developing NSCLC when the test score exceeds the threshold score.

The methods of diagnosis may comprise the steps of obtaining a test sample from the individual; measuring expression levels of one or more biomarkers in the test sample, wherein the one or more biomarkers are selected from the group consisting of CXCL10, MX1, IDO1, IF144L, CD2, GBP5, PRAME, ITGAL, LRP4, and APOL3; deriving a test score that captures the expression levels; providing a threshold score comprising information correlating the test score and a diagnosis of the NSCLC; and comparing the test score to the threshold score; wherein the individual is determined to have the cancer or is susceptible to developing the cancer when the test score exceeds the threshold score. One of ordinary skill in the art can determine an appropriate threshold score, and appropriate biomarker weightings, using the teachings provided herein including the teachings of Example 1.

In other embodiments, the methods of diagnosing patients as having or being susceptible to developing NSCLC associated with DDRD comprise measuring expression levels of one or more biomarkers in the test sample, wherein the one or more biomarkers are selected from the group consisting of CXCL10, MX1, IDO1, IF144L, CD2, GBP5, PRAME, ITGAL, LRP4, APOL3, CDR1, FYB, TSPAN7, RAC2, KLHDC7B, GRB14, AC138128.1, KIF26A, CD274, CD109, ETV7, MFAP5, OLFM4, PI15, FOSB, FAM19A5, NLRC5, PRICKLE1, EGR1, CLDN10, ADAMTS4, SP140L, ANXA1, RSAD2, ESR1, IKZF3, OR2I1P, EGFR, NAT1, LATS2, CYP2B6, PTPRC, PPP1R1A, and AL137218.1. Tables 2A and 2B provide exemplary gene signatures (or gene classifiers) wherein the biomarkers consist of 40 or 44 of the gene products listed therein, respectively, and wherein a threshold score is derived from the individual gene product weightings listed therein. In one of these embodiments wherein the biomarkers consist of the 44 gene products listed in Table 2B, and the biomarkers are associated with the weightings provided in Table 2B, a test score that exceeds a threshold score, such as a threshold score of 0.3681, indicates a diagnosis of NSCLC or of being susceptible to developing NSCLC.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Example 1

Application of DDRD Assay to NSCL cancer and Validation

Methods

Tumour Material

The gene expression analysis was conducted on a published cohort of 90 Non-Small Cell Lung (NSCL) frozen tumour tissue samples sourced from GEO (GSE14814). One sample was identified as outlier by Principal Component Analysis and was removed before further analysis was performed. This cohort of samples can be further described as follows:

- 39 samples were non treated while 50 samples received adjuvant platinum-based therapy (cisplatin, together with a mitotic inhibitor, vinorelbine) treatment
- Age: 63.2 [46.3-80.1]
- Sex: 66 Males and 23 females
- Stage: 45 stage I, 44 stage II

Data preparation

All samples were processed using RMA (Robust Multi-array Average) pre-processing.

Hierarchical clustering analysis

The probe sets from the original platform (Breast DSA®) were initially remapped to the probe sets on the NSCL platform (Affymetrix Human Genome U133A Array) to enable the transfer of information between platforms. The NSCL pre-processed data matrix was further filtered to remove all non-informative probe sets (PS) and retain the most variable genes identified in the original DDRD analysis. This gene set includes genes defining the DDRD samples and other genes biologically relevant to other functions
A Hierarchical agglomerative clustering analysis was performed using Euclidean as distance metrics and ward as linkage method.

Analysis of gene clusters

Genes were categorised as DDRD if they belong to a gene cluster defining the DDRD samples, in other words, the clusters enriched for DDRD and immune response functions. Other genes were defined as non DDRD.
The composition of each gene cluster in DDRD genes was calculated as a percentage of the size of each cluster size (number of DDRD genes/Number of genes in cluster).
A high expression of DDRD genes indicate a DDRD positive phenotype while a low expression of these genes represent a DDRD negative phenotype allowing the classification of samples as DDRD positive or DDRD negative.

Survival analyses

A Univariate survival analyses was performed within each DDRD sample group comparing treated samples versus non treated samples. The p-values and Hazards ratios were calculated using a cox proportional hazard ratio model.

Results

Identification of DDRD subtype

The clustering results are presented in figure 1.

Gene cluster #4 shows a high overlap with the DDRD genes showing supporting evidence of an active DDRD mechanism in Lung. These genes are listed in table 1A. It is composed of 65% of the original DDRD genes (see WO 2012/037378) while the other clusters including larger clusters only contain up to 12% of the DDRD genes. Strong expression pattern of these genes for the different sample clusters can be observed with a clear up-regulation of these genes for sample cluster 2. This expression pattern is similar to the original expression patters observed in the DDRD discovery set; namely a down regulated sample group, an up regulated sample group and a sample group with mixed expressions. All these observations suggest the existence of a DDRD subgroup in Lung.

Sample cluster 2 shows a strong up regulation for the DDRD gene cluster and was consequently labelled "DDRD positive", while the other two sample clusters (#1 and #3) were labelled "DDRD negative" for consistency with the discovery analysis of DDRD in Breast.

Survival analysis results

Differences in survival for treated patients versus non-treated patients were observed between the DDRD sample group and the non-DDRD sample groups. A significant difference in survival was found between treated and non-treated patients in the DDRD group: HR= 5.099 [0.9783-26.57], p-value= 0.032, figure 2 In comparison, no significant difference in survival was observed between treated and non-treated patients in the non-DDRD group: HR is 1.428 [0.6048-3.372], p-value= 0.414, figure 3

These observations suggest that our DDRD group is able to identify a subpopulation of patients which are more likely to benefit from adjuvant platinum-based (cisplatin based) therapy.

Conclusion

Evidence is provided demonstrating that the DDRD subtype is found in about 30% of NSCLC. These patients had a survival benefit following adjuvant cisplatin-based therapy (hazard ratio 5.01 p=0.032) compared to those outside the group (DDRD-) (hazard ratio 1.43 p=0.414). Therefore the DDRD Assay can predict benefit of chemotherapy in NSCL patients.

Example 2

Application of DDRD 44 gene signature to NSCL cancer

Methods

Tumour Material

The gene expression analysis was conducted on a published cohort of 60 Non-Small Cell Lung (NSCL) frozen tumour tissue samples sourced from Array Express and GEO (E-MTAB-923 and GSE37745). This cohort of samples can be further described as follows:

- All samples received adjuvant platinum-based therapy (cisplatin, together with a mitotic inhibitor, vinorelbine) treatment
- Histology: 46 Adenocarcinoma. 8 Squamous carcinoma and 6 large cell carcinoma
- Stage: 22 stage I, 14 stage II, 23 stage III and 1 stage IV

Data preparation

All samples were processed using RMA (Robust Multi-array Average) pre-processing.

DDRD classification

For each sample the intensities for each of the 44 signature genes was calculated using the median value of the probesets mapping to the gene on the Affymetrix GeneChip® human genome U133 plus 2.0 array (Table 3C). The DDRD score was calculated as a weighted sum of the intensities of the genes in the signature and a threshold of 0.65 was used to classify samples as DDRD positive and DDRD negative, where samples with a DDRD score greater than the threshold were classified as DDRD positive and samples with a DDRD score less than or equal to the threshold were classified as DDRD negative.

Survival analyses

A Univariate survival analyses was performed to determine the effect of DDRD status on overall survival following adjuvant chemotherapy. The p-values and Hazards ratios were calculated using a cox proportional hazard ratio model.

Results

DDRD classification

Application of the DDRD signature to this NSCL cancer cohort resulted in 30 samples (50%) being predicted as DDRD positive and 30 samples (50%) as DDRD negative

Survival analysis results

Significant differences in survival for DDRD positive patients versus DDRD negative patients were observed: HR= 0.4445 [0.2397-0.8241], p-value= 0.0098, Figure 4.

These observations suggest that our DDRD group is able to identify a subpopulation of patients which will benefit from adjuvant platinum-based (cisplatin based) therapy.

The various embodiments of the present invention are not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the various embodiments of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims. Moreover, all embodiments described herein are considered to be broadly applicable and combinable with any and all other consistent embodiments, as appropriate.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

CLAIMS

1. A method of predicting responsiveness of an individual having non-small cell lung cancer (NSCLC) to treatment with a DNA-damaging therapeutic agent comprising:
   a. measuring expression levels of one or more biomarkers in a test sample obtained from the individual, wherein the one or more biomarkers are selected from Table 2B, 1A, 1B, 1C, 2A, 3A, 3B and/or 3C;
   b. deriving a test score that captures the expression levels;
   c. providing a threshold score comprising information correlating the test score and responsiveness;
   d. and comparing the test score to the threshold score; wherein responsiveness is predicted when the test score exceeds the threshold score and/or wherein a lack of responsiveness is predicted when the test score does not exceed the threshold score.

2. The method of claim 1, wherein the one or more biomarkers are selected from the group consisting of CXCL10, MX1, IDO1, IF144L, CD2, GBP5, PRAME, ITGAL, LRP4, and APOL3 and/or the group consisting of CDR1, FYB, TSPAN7, RAC2, KLHDC7B, GRB14, AC138128.1, KIF26A, CD274, CD109, ETV7, MFAP5, OLFM4, PI15, FOSB, FAM19A5, NLRC5, PRICKLE1, EGR1, CLDN10, ADAMTS4, SP140L, ANXA1, RSAD2, ESR1, IKZF3, OR2I1P, EGFR, NAT1, LATS2, CYP2B6, PTPRC, PPP1R1A, and AL137218.1.

3. The method of claim 2, comprising measuring the expression level of all of the biomarkers.

4. The method of claim 1, comprising measuring the expression level of:
   a. at least 10 of the biomarkers from Table 1A in the test sample; and/or
   b. at least one or more up to all of CD2, FYB, ITGAL and RAC2

5. The method of claim 4, comprising measuring the expression level of all 58 different biomarkers listed in Table 1A.

6. The method of any of claims 1 to 5 where expression levels are measured using primers or probes which bind to at least one of the target sequences set forth as SEQ ID NO: 1-80 (Table 1A), 81-260 (Table 3A), 261-313 (Table 3B), 314-337 (Table 1B), 338-363 (Table 1C) or comprise at least 15 contiguous nucleotides of any one of SEQ ID NOs 364-455 (Table 3C).

7. The method of any one of claims 1 to 6, wherein the NSCLC is early stage, late stage or metastatic disease.

8. The method of any one of claims 1 to 7, wherein the NSCLC is selected from one or more of adenocarcinoma, large-cell lung carcinoma and squamous cell carcinoma.

9. The method of any one of claims 1 to 8, wherein the DNA-damaging therapeutic agent comprises one or more substances selected from the group consisting of: a DNA damaging

agent, a DNA repair targeted therapy, an inhibitor of DNA damage signalling, an inhibitor of DNA damage induced cell cycle arrest, a histone deacetylase inhibitor, a heat shock protein inhibitor and an inhibitor of DNA synthesis.

10. The method of claim 9, wherein the DNA-damaging therapeutic agent comprises one or more of a platinum-containing agent, a nucleoside analogue such as gemcitabine or 5-fluorouracil or a prodrug thereof such as capecitabine, an anthracycline such as epirubicin or doxorubicin, an alkylating agent such as cyclophosphamide, an ionising radiation or a combination of radiation and chemotherapy (chemoradiation).

11. The method of any one of claims 1 to 10, wherein the DNA-damaging therapeutic agent comprises a platinum-containing agent.

12. The method of claim 11, wherein the platinum based agent is selected from cisplatin, carboplatin and oxaliplatin.

13. The method of any one of claims 1 to 12, which predicts responsiveness to treatment with the DNA-damaging therapeutic agent together with a further therapy.

14. The method of claim 13 wherein the further therapy is (treatment with) a mitotic inhibitor.

15. The method of claim 14, wherein the mitotic inhibitor is a vinca alkaloid.

16. The method of claim 15, wherein the vinca alkaloid is vinorelbine.

17. The method of any one of claims 1 to 13 which predicts responsiveness to a combination therapy comprising a DNA-damaging therapeutic agent, wherein the combination therapy is selected from:
    a. cisplatin/carboplatin and 5-fluorouracil
    b. cisplatin/carboplatin and capecitabine
    c. epirubicin/doxorubicin, cisplatin/carboplatin and fluorouracil
    d. epirubicin/doxorubicin, oxaliplatin and capecitabine
    e. cisplatin/carboplatin and etoposide
    f. gemicitabine and cisplatin/carboplatin
    g. cyclophosphamide, epirubicin/doxorubicin and vincristine
    h. cyclophosphamide, epirubicin/doxorubicin, vincristine and etoposide
    i. epirubicin/doxorubicin, cyclophosphamide and etoposide

18. The method of any one of claims 1 to 17 wherein the treatment is adjuvant treatment and/or neoadjuvant treatment.

19. The method of any one of claims 1 to 18 wherein if responsiveness is predicted, the individual is treated with the DNA-damaging therapeutic agent.

20. The method of any one of claims 1 to 18 wherein if a lack of responsiveness is predicted, the individual is not treated with the DNA-damaging therapeutic agent.

21. The method of any one of claims 1 to 20 wherein the treatment is adjuvant cisplatin/vinorelbine treatment.

22. The method of claim 20 wherein if a lack of responsiveness is predicted, the individual is treated with a mitotic inhibitor.

23. The method of any one of claims 1 to 22 wherein responsiveness comprises or is increased overall survival, progression free survival and/or disease free survival.

24. (a) A method of treating NSCLC comprising administering a DNA-damaging therapeutic agent to a subject, wherein the subject is predicted to be responsive to the DNA-damaging therapeutic agent on the basis of a test score derived from expression levels of one or more biomarkers in a test sample obtained from the individual, wherein the one or more biomarkers are selected from those listed in Table 2B, 1A, 1B, 1C, 2A, 3A, 3B and/or 3C; or
(b) A DNA-damaging therapeutic agent for use in a method of treating NSCLC, wherein the subject is predicted to be responsive to the DNA-damaging therapeutic agent on the basis of a test score derived from expression levels of one or more biomarkers in a test sample obtained from the individual, wherein the one or more biomarkers are selected from those listed in Table 2B, 1A, 1B, 1C, 2A, 3A, 3B and/or 3C.

25. (a) A method of treating NSCLC comprising administering a mitotic inhibitor to a subject, wherein the subject is predicted to be non-responsive to a DNA-damaging therapeutic agent on the basis of a test score derived from expression levels of one or more biomarkers in a test sample obtained from the individual, wherein the one or more biomarkers are selected from those listed in Table 2B, 1A, 1B, 1C, 2A, 3A, 3B and/or 3C; or
(b) A mitotic inhibitor for use in a method of treating NSCLC, wherein the subject is predicted to be non-responsive to a DNA-damaging therapeutic agent on the basis of a test score derived from expression levels of one or more biomarkers in a test sample obtained from the individual, wherein the one or more biomarkers are selected from those listed in Table 2B, 1A, 1B, 1C, 2A, 3A, 3B and/or 3C.

26. The method or use of claim 24 or 25 wherein the test score has been derived according to a method as claimed in any one of claims 1 to 23.

27. A kit for predicting responsiveness of an individual having non-small cell lung cancer (NSCLC) to treatment with a DNA-damaging therapeutic agent comprising primers or probes which hybridize

to at least one of the target sequences set forth as SEQ ID NO: 1-80 (Table 1A), 81-260 (Table 3A), 261-313 (Table 3B), 314-337 (Table 1B), 338-363 (Table 1C) or comprise at least 15 contiguous nucleotides of any one of SEQ ID NOs 364-455 (Table 3C).

28. The kit of claim 27 wherein the primers or probes hybridize to at least 10 of the target sequences.

29. The kit of claim 27 or 28 further comprising a DNA-damaging therapeutic agent.

30. The kit of claim 29 wherein the DNA-damaging therapeutic agent is provided in a dosage form specifically for treatment of NSCLC.

31. The kit of claim 30 wherein the treatment is neo-adjuvant or adjuvant treatment.

32. The kit of any one of claims 27 to 31 wherein the DNA-damaging therapeutic agent comprises a platinum-based agent.