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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 February 2013 (14.02.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/020690 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 33/574* (2006.01)

(21) International Application Number:
PCT/EP2012/003339

(22) International Filing Date:
4 August 2012 (04.08.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/205,469 8 August 2011 (08.08.2011) US

(71) Applicant *(for DE only)*: **ROCHE DIAGNOSTICS GMBH** [DE/DE]; Sandhofer Strasse 116, 68305 Mannheim (DE).

(71) Applicant *(for all designated States except DE, US)*: **F. HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventors: **LI, Yan**; 301 Whitclem Drive, Palo Alto, CA 94306 (US). **LIH, Chih-Jian**; 151 Monroe Drive, Palo Alto, CA 94306 (US). **LIU, Wei-Min**; 4929 Shelton Street, Dublin, CA 94568 (US). **TAI, Yu Chuan**; 3660 Andrews Drive #310, Pleasanton, CA 94588 (US).

(74) Agent: **ROCHE DIAGNOSTICS GMBH**; Patent Department, 68298 Mannheim (DE).

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

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

Declarations under Rule 4.17:

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

WO 2013/020690 A1

(54) Title: PREDICTING RESPONSE TO ANTI-CD20 THERAPY IN DLBCL PATIENTS

(57) Abstract: This invention provides methods, compositions, and kits relating to biomarkers whose expression levels are correlated with diffuse large B-cell lymphoma (DLCBL) patients' response to treatment with a CD20 antagonist, such as a CD20 antibody, exemplified by rituximab. The methods, compositions, and kits of the invention can be used to identify DLBCL patients who are likely or not likely, to respond to anti-CD20 treatments.

PREDICTING RESPONSE TO ANTI-CD20 THERAPY IN DLBCL PATIENTS

BACKGROUND OF THE INVENTION

Lymphoma is the fifth most common cancer in women and the sixth most common cancer in men in the Western world (*see* Murawski et al. (2010) *Unresolved issues in diffuse large B-cell lymphomas*, Expert Rev. Anticancer Ther. 10(3):387). 90% of aggressive lymphomas originate from B-cells and are classified as diffuse large B-cell lymphomas (DLBCL). Until recently, the accepted form of therapy for DLBCL was CHOP: a combination of cyclophosphamide, hydroxydaunorubicine (doxorubicin), Oncovin® (vincristine) and prednisone. In 1997, the FDA approved rituximab (Rituxan®) for treatment of aggressive Non-Hodgkin lymphomas. Rituximab is a chimeric mouse-human monoclonal antibody against a protein CD20 found primarily on the surface of B-cells. Rituximab has been shown to be effective as a single agent in DLBCL (*see* Coiffier et al. (1998) *Rituximab (anti-CD20 monoclonal antibody) for the treatment of patients with relapsing or refractory aggressive lymphoma: a multicenter Phase II study.* Blood 92(6):1927). Subsequent studies of a combination of rituximab with CHOP demonstrated high response rate with high overall and progression-free survival, even with fewer rounds of CHOP therapy (*see* Murawski et al. (2010), *supra*). However, it appears that patients receiving rituximab have different response rates, which result in different survival times. Thus, there is a need to identify patients who would benefit from the anti-CD 20 therapy such as rituximab treatment and patients who would likely not respond well to such therapy, and would need a different therapy instead.

SUMMARY OF THE INVENTION

In one aspect, the invention relates to a method of predicting whether a patient with diffuse large B-cell lymphoma (DLBCL) is likely to exhibit response to anti-CD20 therapy,

comprising: determining in a sample obtained from a patient the level of expression of one or more genes listed in Table 1 and Table 2; predicting whether the patient is likely to exhibit response to anti-CD20 therapy, based on the comparison between the measured level of expression of each gene in the patient's sample and a set of control samples. In some variations of this embodiment, the anti-CD20 therapy comprises an anti-CD20 antibody, for example rituximab. In some embodiments predicting whether the patient is likely to exhibit response to anti-CD20 therapy is based on the similarity between the measured level of expression of each gene in the patient's sample and a set of control samples. In certain embodiments of the method the level of expression of all the genes listed in Table 1 is determined and/or the level of expression of all the genes listed in Table 2 is determined. In certain embodiments the expression of the genes is measured by measuring the level of RNA transcribed from each gene. In other embodiments the expression of the genes is measured by measuring the level of protein corresponding to each gene. In another embodiment the expression of the genes is measured by measuring the level of RNA or protein of a gene situated downstream in a biological pathway from a gene listed in Table 1 or Table 2. In yet another embodiment the expression of the genes is measured by measuring the product of activity of a protein corresponding to a gene listed in Table 1 or Table 2. In certain embodiments response to anti-CD20 therapy constitutes survival for three or more years following the start of the therapy. In certain embodiments a set of control samples is used, where the expression of the corresponding genes has been determined. In some embodiments the control set of samples comprises a representative number of patients that exhibit response to anti-CD20 therapy and a representative number of patients that exhibit no response or poor response to anti-CD20 therapy. In certain embodiments predicting whether the patient is likely to exhibit response to anti-CD20 therapy is made using statistical methods. In certain embodiments the statistical method is the Support Vector Machine method. In certain embodiments the statistical method is the k-nearest-neighbor method.

In another aspect, the invention relates to a set of diagnostic probes for predicting whether a patient with diffuse large B-cell lymphoma (DLBCL) is likely to exhibit response to anti-CD20 therapy, comprising nucleic acid probes or antibodies for detecting expression of one or more genes listed in Table 1 and Table 2. In certain embodiments the diagnostic probes are nucleic acid probes. In other embodiments the diagnostic probes are antibodies.

In another aspect, the invention relates to the *in vitro* use of a set of diagnostic probes comprising nucleic acid probes or antibodies for detecting expression of one or more genes listed in Table 1 and Table 2 in a sample obtained from a patient for predicting whether the patient is likely to exhibit response to anti-CD20 therapy, based on the comparison between the expression of each gene measured in the patient's sample and a set of control samples. In certain embodiments the patient is a diffuse large B-cell lymphoma (DLBCL) patient and the prediction that the patient is likely to exhibit response to anti-CD20 therapy results in the administration of an anti-CD20 antibody. In another embodiment the prediction that the patient is not likely to exhibit response to anti-CD20 therapy results in no administration of an anti-CD20 antibody. In certain embodiments the anti-CD20 antibody is rituximab. In certain embodiments the expression of all the genes listed in Table 1 is detected and/or the expression of all the genes listed in Table 2 is detected. In some embodiments predicting whether the patient is likely to exhibit response to anti-CD20 therapy is based on the similarity between the measured level of expression of each gene in the patient's sample and a set of control samples. In certain embodiments a set of control samples is used, where the expression of the corresponding genes has been determined.

In another aspect, the invention relates to a kit for predicting whether a patient with diffuse large B-cell lymphoma (DLBCL) is likely to exhibit response to anti-CD20 therapy comprising a set of diagnostic probes comprising nucleic acid probes or antibodies for detecting expression of one or more genes for one or more genes listed in Table 1 and Table 2; and reagents necessary for detecting hybridization and/or binding of the diagnostic probes. In

certain embodiments the kit comprises nucleic acid probes for the genes listed in Table 1 and/or nucleic acid probes for the genes listed in Table 2. In variations of this embodiment, the kit comprises antibodies for proteins expressed from the genes listed in Table 1 and Table 2. In certain embodiments the kit further includes a control set of samples, comprising samples from a representative number of patients that exhibit response to anti-CD20 therapy and samples from a representative number of patients that exhibit no response or poor response to anti-CD20 therapy. In certain embodiments a set of control samples is used, where the expression of the corresponding genes has been determined. In certain embodiments the anti-CD20 therapy is treatment with rituximab.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a Kaplan-Meier plot showing differences in survival for the model built in Example I validated with re-substitution.

Figure 2 is a Kaplan-Meier plot showing differences in survival for leave-one-out cross validation of the model built in Example 1.

Figure 3 is a Kaplan-Meier plot constructed with the classification according to the SVM method for the model built in Example 2.

Figure 4 is a Kaplan-Meier plot constructed with the probability assigned according to the SVM method for the model built in Example 2.

Figure 5 is a Kaplan-Meier plot showing differences in survival for the independent set of samples in Example 3.

DEFINITIONS

The terms "array", "microarray", and "DNA chip" are used herein interchangeably to refer to an array of distinct polynucleotides affixed to a substrate, such as glass, plastic, paper, nylon, or

other type of membrane, filter, chip, or any other suitable solid support. The polynucleotides can be synthesized directly on the substrate, or synthesized separate from the substrate and then affixed to the substrate. In another aspect, an "array" or "microarray" relates to an antibody array that provides for efficient profiling of protein expressions, e.g. from cell lysates, and of biomarkers, e.g. from serum or urine. Herein, an antibody microarray constitutes a collection of capture antibodies that are spotted and fixed on a solid surface such as glass, plastic or silicon chip, for the purpose of detecting antigens. Commercial arrays containing diagnostic probes (i.e., nucleic acid probes or antibodies) for all the targets potentially present in a particular type of sample are available. A commercial array may contain probes for all the nucleotide sequences present in the genome. Alternatively, an array may contain probes for only expressed sequences. The term "custom array" refers to an array that contains probes for only selected targets. In the context of the present invention, a custom array may contain probes for some or all of the genes in the expression profile predictive of response to anti-CD20 therapy.

The term "biomarker" refers to a gene or nucleic acid sequence that is of interest for a particular phenotype, such as a disease. For example, a biomarker could be a protein-coding gene whose activation or inactivation leads to a disease. In that case, the presence of the mRNA or the protein could be indicative or predictive of the disease. In another example, a biomarker could be a sequence polymorphism in linkage disequilibrium with an unknown gene causing disease. In that case, the presence of the polymorphism could be indicative or predictive of the disease. In yet another example, a biomarker is a somatic mutation in a gene, wherein the presence of the mutation is correlated with a disease. In that case, the presence of the mRNA mutation could be indicative or predictive of the disease. In the context of the present invention, biomarkers are the genes whose expression (measured by the presence of the mRNA, cDNA or protein) is in correlation with response or lack of response to anti-CD20 therapy.

The terms "gene expression profile" or "gene expression signature" refer to a collection of expression levels of a number of genes. The genes in the profile are markers that discriminate between individuals with different phenotypes. The discrimination is achieved because the expression levels of each gene exhibit a statistically significant difference among groups of individuals with different phenotypes. Thus each of the phenotypes is characterized by a lowered expression of some genes in the profile, and overexpression of other genes in the profile. By determining the "expression profile", i.e. expression of some or all the genes in the profile, one can assign a phenotype to an individual whose phenotype has not yet manifested itself. In the context of the present invention, by determining the gene expression profile disclosed herein, one can predict a DLBCL patient's response to anti-CD20 therapy.

The term "hybridization" refers to the formation of a duplex between two single-stranded nucleic acids due to complementary base pairing. Hybridization can occur between perfectly complementary nucleic acid strands or between substantially (but not perfectly) complementary nucleic acid strands that contain one or more mismatches. Conditions under which only perfectly complementary nucleic acid strands will hybridize are referred to as "stringent" hybridization conditions. Stable duplexes of substantially complementary nucleic acids can be achieved under less stringent hybridization conditions. Those skilled in the art of nucleic acid technology can determine duplex stability using for example, computer software such as Visual OMP® (DNA Software, Inc., Ann Arbor, Mich.). The primary interaction between the antiparallel polynucleotide molecules is typically base specific, *e.g.*, A/T and G/C, by Watson/Crick and/or Hoogsteen-type hydrogen bonding. In some aspects, a hybridization complex can form from intermolecular interactions, or alternatively, can form from intramolecular interactions.

The term "primer" refers to a nucleic acid, i.e. a polynucleotide or an oligonucleotide, which acts as a point of initiation of DNA synthesis under suitable conditions in the presence of nucleic acid precursors and an agent for polymerization. Further, a primer can initiate the

polymerization of nucleotides in a template-dependent manner to yield a polynucleotide that is complementary to the target polynucleotide. A primer does not need to have 100% complementarity with its template subsequence for primer elongation to occur; primers with less than 100% complementarity can be sufficient for hybridization and polymerase elongation to occur. A primer can either consist entirely of the target-hybridizing region or can contain additional features which allow for detection, immobilization, or manipulation of the amplified product.

In one aspect of the invention, the term "probe" refers to a nucleic acid, i.e. a polynucleotide or an oligonucleotide, which selectively hybridizes to a target nucleic acid under suitable conditions. A nucleic acid probe can either consist entirely of the target-hybridizing region or can contain additional features which allow for the detection, immobilization, or manipulation of the probe and/or the probe-target duplex. The nucleic acid probe may contain modifications to its primary structure by the addition of labels, linkers, peptides or any other groups necessary to perform the detection assay in the chosen format. Detection systems for labeled nucleic acid probes include, but are not limited to, the detection of fluorescence, fluorescence quenching (e.g., when using a FRET pair detection system), enzymatic activity, absorbance, molecular mass, radioactivity, luminescence or binding properties that permit specific binding of the reporter (*e.g.*, where the reporter is an antibody). Typically, a nucleic acid probe is sufficiently complementary to a specific target sequence contained in a nucleic acid to form a stable hybridization complex with the target sequence under a selected hybridization condition, such as, but not limited to, a stringent hybridization condition. In another aspect of the invention, a probe can be an antibody, rather than a nucleic acid, that has binding specificity for a nucleic acid nucleotide sequence of interest or for a protein of interest. Hence, according to the current invention a “diagnostic probe” may either be a nucleic acid probe or an antibody suitable for detecting expression of one or more target genes by either hybridizing a nucleic acid probe to a target nucleic acid under suitable conditions to form a

hybridization complex or by binding of an antibody probe to a target nucleic acid or a target protein under suitable conditions to form an antigen-antibody binding complex.

An "antibody" as referred to herein includes both polyclonal and monoclonal antibodies, as well as fragments thereof, such as Fv, Fab and F(ab)2 fragments that are capable of binding antigen or hapten. Herein, the antigen or hapten may either be a protein or a nucleic acid (i.e. RNA or DNA). Also included are single chain antibodies and humanized hybrid antibodies wherein amino acid sequences of a non-human donor antibody exhibiting a desired antigen-specificity are combined with sequences of a human acceptor antibody. The donor sequences will usually include at least the antigen-binding amino acid residues of the donor but may comprise other structurally and/or functionally relevant amino acid residues of the donor antibody as well. Such hybrids can be prepared by several methods well known in the art. Preferably, the ligand binds specifically to the antibody peptide or polypeptide. Specific binding means that the ligand or agent should not bind substantially to, i.e. cross-react with, another peptide, polypeptide, nucleic acid or substance present in the sample to be analyzed. Preferably, the specifically bound peptide or polypeptide should be bound with at least 3 times higher, more preferably at least 10 times higher and even more preferably at least 50 times higher affinity than any other relevant peptide or polypeptide. Binding of the ligand can be measured by any method known in the art. Preferably, said method is semi-quantitative or quantitative.

The term "probe set" refers to a unique set of nucleic acid probes or antibodies capable of detecting one or more markers. A set of nucleic acid probes may be used in amplification / detection reactions (such as e.g. PCR, LAMP, NASBA, RDC, PEAR) for determining the expression level of one or more genes. Alternatively, if antibodies are uses as the probe a set of antibodies may be used, e.g., in an immunoassay or ELISA for determining the expression of one or more genes. In some embodiments a probe set may be provided on an array or microarray. Typically, a microarray contains several probe sets for detecting each marker or

gene. In commercial microarrays, each probe set is associated with a sequence that is searchable in a public database by a unique accession number.

The terms "target sequence" or "target" refer to a region of a nucleic acid that is to be analyzed. In case an antibody is used as the diagnostic probe a "target" may either refer to a region of a nucleic acid or to a region of a protein that is to be analyzed and that provides the antigen for antibody binding.

The terms "nucleic acid," "polynucleotide" and "oligonucleotide" refer to target sequences, primers and probes. The terms are not limited by length and are generic to linear polymers of deoxyribonucleotides (single-stranded or double-stranded DNA), ribonucleotides (RNA), and any other N-glycoside of a purine or pyrimidine base, including adenosine, guanosine, cytidine, thymidine and uridine and modifications of these bases.

The term "response to therapy" refers to a benefit attributable to therapy. Response may be assessed by measuring a clinically relevant parameter, such as tumor shrinkage, length of overall survival or survival without disease progression. For example, the clinically relevant parameter may be the length of survival following the start of the therapy. The patient who has survived for a certain time, e.g. at least three years, may be considered to have responded to therapy. The patient, who has not survived past three years, is considered to not have responded or responded poorly to therapy.

The term "sample" refers to any composition containing or presumed to contain nucleic acids or proteins. This includes a sample of tissue or fluid isolated from an individual for example, skin, plasma, serum, spinal fluid, lymph fluid, synovial fluid, urine, tears, blood cells, organs and tumors, including the fresh-frozen tissue and formalin-fixed paraffin embedded tissue (FFPET), and also to samples of *in vitro* cultures established from cells taken from an individual, and nucleic acids or proteins isolated therefrom.

The term "training set" or "control set" refers to a set of samples used to establish a correlation between two variables. For example, a training set is a set of patient samples used to establish a correlation between the expression of a gene and the patient's condition. In the context of the present invention, the training set is a set of samples from DLBCL patients receiving anti-CD20 therapy, for which survival information is available. The training set is used to establish a correlation between the expression profile and survival.

The term "testing set" refers to a set of samples used to verify the correlation established using the training set. For example, a testing set is a set of samples from patients for whom both gene expression and the patient's condition are known. This set is used to verify whether the correlation established using the training set will correctly predict the patient's condition. In the context of the present invention, the testing set is an independent set of samples from DLBCL patients receiving anti-CD20 therapy, for which survival information is available. The testing set is used to measure expression of the genes in the profile and test whether the actual survival data matches the prediction made by measuring the gene expression.

The term "test sample" refers to a patient's sample that is to be tested for a particular parameter.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based on the discovery that expression of certain genes predicts response of diffuse large B-cell lymphoma (DLBCL) patients to anti-CD20 therapy. Specifically, it was discovered that some genes exhibit differential expression between a group of DLBCL patients with response to rituximab therapy and patients with poor response to rituximab therapy, where the response is measured by the length of survival following the therapy. The invention comprises such genes and gene expression profiles containing such genes. The invention further comprises the use of the genes and gene expression profiles to predict response of DLBCL patients to anti-CD20 therapy.

The genes identified as differentially expressed in responders and poor responders to anti-CD20 therapy among DLBCL patients are listed in Tables 1 and 2. In Tables 1 and 2, the column "PS ID" lists the identification code for the probe set in the Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array (Affymetrix, Santa Clara, Calif.) The column "NCBI" lists the corresponding public reference number (NCBI Accession No.). The column "Gene Title" lists the names of the genes, where available. The column "Gene Symbol" lists the gene symbols, where available. The column "p-value" in Table 1 lists the p-value statistic. The column "FDR_BH" in Table 2 lists the values of the statistic False Discovery Rate (FDR), calculated according to the method described in Benjamini Y. and Hochberg, Y., (1995) J Royal Stat. Soc. Ser. B 57:289.

Table 1

Predictive markers identified in Example 1

PS ID	NCBI	Symbol	Gene Title	p-value
202858_at	NM_006758	U2AF1	U2 small nuclear RNA auxiliary factor 1	8.40E-05
205215_at	NM_007212	RNF2	ring finger protein 2	4.10E-05
205877_s_at	NM_017590	ZC3H7B	zinc finger CCCH-type containing 7B	4.10E-05
208656_s_at	AF135162	CCNI	cyclin I	1.90E-05
208776_at	BF432873	PSMD11	proteasome (prosome, macropain) 26S subunit, non-ATPase, 11	6.00E-05
210461_s_at	BC002448	ABLIM1	actin binding LIM protein 1	3.50E-05
210964_s_at	U94364	GYG2	glycogenin 2	6.20E-05
212669_at	AI093569	CAMK2G	calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma	6.70E-06
212678_at	AW054826	NF1	neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)	3.40E-05
213658_at	BE858194	---	mRNA full length insert cDNA clone EUROIMAGE 826033	3.70E-08
213748_at	AW271713	TRIM66	tripartite motif-containing 66	3.30E-05

214891_at	U79257	FBXO21	F-box protein 21	4.90E-05
218205_s_at	NM_017572	MKNK2	MAP kinase interacting serine/threonine kinase 2	6.50E-05
224642_at	BG291550	FYTTD1	forty-two-three domain containing 1	7.70E-05
224808_s_at	AI090768	GET4	Golgi to ER traffic protein 4 homolog	7.80E-05
226004_at	AI910855	CABLES2	Cdk5 and Abl enzyme substrate 2	3.60E-05
227844_at	AI089932	FMNL3	formin-like 3	2.50E-05
230566_at	AI806805	C22orf27	chromosome 22 open reading frame 27	6.70E-05
231997_at	R69910	TBCEL	tubulin folding cofactor E-like	6.90E-05
232076_at	AW294133	ZNF707	zinc finger protein 707	3.60E-05
235640_at	AI763196	---	transcribed locus	7.80E-05
236449_at	AI885390	CSTB	cystatin B (stefin B)	2.70E-05
236604_at	BF195603	BAHCC1	BAH domain and coiled-coil containing 1	2.00E-05
239656_at	AA827176	LOC723809	hypothetical LOC723809	6.90E-05
240377_at	AI344289	NPIP	nuclear pore complex interacting protein	3.40E-05
241388_at	AL567118	---	cDNA FLJ40566 fis, clone THYMU2004733	1.40E-05
243518_at	BF195694	LOC730367	hypothetical protein LOC730367	1.10E-05
1553326_at	AF453828	RXFP2	relaxin/insulin-like family peptide receptor 2	3.70E-05
1554732_at	BC020886	MGC24125	hypothetical protein MGC24125	2.40E-05
1566337_x_at	AJ293390	---	mRNA, differentially expressed in malignant melanoma, clone MM A2	5.80E-05

Table 2

Predictive markers identified in Example 2

PS ID	NCBI	Gene Title	Symbol	FDR_BH
200730_s_at	BF576710	protein tyrosine phosphatase type IVA, member 1	PTP4A1	1.79E-07
201800_s_at	AF185696	oxysterol binding protein	OSBP	1.79E-07
213359_at	W74620	heterogeneous nuclear ribonucleo-protein D (AU-rich element RNA binding protein 1, 37kDa)	HNRNPD	1.79E-07

200848_at	AA479488	adenosylhomocysteinase-like 1	AHCYL1	2.45E-07
37170_at	AB015331	BMP2 inducible kinase	BMP2K	3.00E-07
202438_x_at	BF346014	iduronate 2-sulfatase	IDS	6.15E-07
242814_at	AI986192	serpin peptidase inhibitor, clade B (ovalbumin), member 9	SERPINB9	6.15E-07
233396_s_at	AK023759	CSRP2 binding protein	CSRP2BP	8.81E-07
211744_s_at	BC005930	CD58 molecule	CD58	8.98E-07
233509_at	AK021844	hect domain and RLD 4	HERC4	8.98E-07
210093_s_at	AF067173	mago-nashi homolog, proliferation-associated (Drosophila)	MAGOH	1.24E-06
223892_s_at	AF182414	transmembrane BAX inhibitor motif containing 4	TMBIM4	1.60E-06
243361_at	N51597	splicing factor, arginine/serine-rich 12	SFRS12	2.02E-06
216252_x_at	Z70519	Fas (TNF receptor superfamily, member 6)	FAS	2.68E-06
231369_at	BG149482	Zinc finger protein 333	ZNF333	2.68E-06
243267_x_at	AI127295	---	---	2.79E-06
1555063_at	BC029495	ubiquitin specific peptidase 6 (Tre-2 oncogene)	USP6	3.39E-06
225026_at	BF572029	chromodomain helicase DNA binding protein 6	CHD6	4.47E-06
236246_x_at	BF195670	hypothetical protein LOC653160	LOC653160	4.47E-06
243931_at	R64696	---	---	4.47E-06
200918_s_at	NM 003139	signal recognition particle receptor (docking protein)	SRPR	5.19E-06
215719_x_at	X83493	Fas (TNF receptor superfamily, member 6)	FAS	5.33E-06
224917_at	BF674052	microRNA 21	MIR21	5.33E-06
225173_at	BE501862	Rho GTPase activating protein 18	ARHGAP18	5.33E-06
239387_at	AW004885	---	---	5.33E-06
238549_at	AI420611	core-binding factor, runt domain, alpha	CBFA2T2	5.77E-06

		subunit 2; translocated to 2		
201150_s_at	NM 000362	TIMP metallopeptidase inhibitor 3	TIMP3	6.02E-06
218924_s_at	NM004388	di-N-acetyl-chitobiase	CTBS	6.92E-06
228822_s_at	AI435036	ubiquitin specific peptidase 16	USP16	7.26E-06
210621_s_at	M23612	RAS p21 protein activator (GTPase activating protein) 1	RASA1	7.69E-06
229961_x_at	AI871270	YjeF N-terminal domain containing -3	YJEFN3	8.19E-06

Measuring expression of one or more of the genes in the profile allows predicting whether a DLBCL patient is, or is not likely to respond to anti-CD20 therapy, such as for example, rituximab. In the context of the present invention, measuring the level of expression of a gene in a patient's sample may be accomplished by several means. Gene expression may be measured by measuring the level of RNA expressed by the gene. The amount of RNA may be determined for example, by reverse transcription followed by quantitative PCR (qPCR) amplification. Gene expression may also be detected by hybridizing a single probe to the RNA in a blot or hybridizing the RNA to multiple probes in an array and quantifying the hybridization signal. Alternatively, expression of protein-coding gene can be determined by measuring the level of expression of the protein encoded by the gene or measuring enzymatic activity of the expressed protein consisting in part or entirely of the protein encoded by the gene. As an additional alternative, it is possible to indirectly assess the expression of a gene listed in Table 1 or Table 2 by measuring the level of expression or activity of a gene or protein situated downstream in a biological pathway from any gene listed in Table 1 or Table 2. For such downstream pathway gene or protein, the expression may also be measured at various levels, including the nucleic acid level, protein level, and activity level.

To predict likelihood of a DLBCL patient's response to anti-CD20 therapy, expression of at least one, more than one or all of the genes listed in Tables 1 and 2 may be measured. A person

skilled in the art of statistics would recognize that a smaller subset of genes among those listed in Tables 1 and 2 may be sufficient to predict a patient's response. For example, a subset of genes may be selected based on the value of the T-statistic, the p-value or the FDR value. A person skilled in the art is aware of the statistical methods for selecting a subset of genes from an expression profile. For example, the smallest number of genes that is sufficient to predict response may be determined by calculating the area under the curve of the receiver operating characteristic ("AUC of RUC" method) as described in Green, D.M., Swets, J.A., (1966) *Signal detection theory and psychophysics*, Wiley, New York; Swets, J.A., Pickett, R.M. (1982) *Evaluation of diagnostic systems: Methods from signal detection theory*, Academic Press, New York; and Pepe, M.S. (2003), *The statistical evaluation of medical tests for classification and prediction*, Oxford Univ. Press.

Additional genes may be added to the profile predictive of a DLBCL patient's response to anti-CD20 therapy, as long as expression of the genes differs between the patients who respond and patients who do not respond to anti-CD20 therapy. For example, when measuring the difference in expression, the p-value or the FDR value may be used to select additional genes to be added to the profile.

In the examples of the expression profiles of the present invention, the expression of each gene was measured in a first set of patient samples whose response (or lack of response) to anti-CD20 therapy has been documented. This first set of samples (referred to as the training set) contained an adequate number of responder patients and an adequate number of non-responder patients.

To predict response to anti-CD20 therapy in a patient, the expression profile in the patient's sample was determined and compared to the expression profile in the samples of the training set. A statistical model was used to determine whether based on the expression profile, the patient is likely to belong to a responder group or a non-responder group for the anti-CD20 therapy. In the method of the present invention, the gene expression in the test sample and the

samples of the training set need not be measured simultaneously. For convenience, expression data from the training set samples may be stored on a computer readable medium and accessed each time a new patient sample is tested.

RESPONSE PREDICTION

In one embodiment, the invention comprises a method of predicting response of a DLBCL patient to anti-CD20 therapy, such as rituximab. The method comprises measuring expression level of one or more genes listed in Tables 1 and 2 in a sample obtained from a patient diagnosed with DLBCL prior to the patient receiving the therapy and comparing the expression to the expression of the same genes in the samples of the training set.

The expression levels of any gene listed in Tables 1 and 2 may be measured for example, by measuring the level of mRNA expressed from the gene in the patient's sample by reverse transcription and quantitative PCR (qPCR) amplification of the resulting cDNA, *see* U.S. Pat. Nos. 5,210,015 and 5,487,972 and Holland *et al.* (1991) PNAS USA 88: 7276. For multiple genes, the levels of mRNA may be measured in a microarray-based assay. For example, multiple mRNA molecules can be reverse-transcribed using common poly-A primers, converted into cDNA and then into labeled amplified RNA (aRNA) that is applied to an array of immobilized probes. (MAQC Consortium (2006) Nat Biotechnol. 24(9):1151-61). The expression may be detected using commercially available microarrays, for example from Roche NimbleGen, Inc. (Madison, Wisc.) or Affymetrix, Inc. (Santa Clara, Calif.). Alternatively, only selected probe sets for each of the genes listed in Tables 1 and 2 may be used. The sequences of probes corresponding to each gene are listed in the literature accompanying the Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array product. A person skilled in the art of nucleic acid hybridization may also design custom hybridization probes for the genes listed in Tables 1 and 2 using the gene sequences available through the public databases. A person skilled in the art of nucleic acid arrays may also design a custom

array containing previously known or custom designed probes for one or more, or all the genes listed in Tables 1 and 2 according to *DNA Microarrays: A Molecular Cloning Manual*, (2003), Eds. Bowtell and Sambrook, Cold Spring Harbor Laboratory Press.

Any alternative methods of amplifying and quantifying mRNA or cDNA may also be used in the context of the present invention. Furthermore, as an alternative to detecting gene transcripts on the nucleic acid level, the protein products of the genes in Tables 1 and 2 may be detected, where the protein products are present. The proteins may be detected, for example, by immunoassays. In another variation of this embodiment, the mRNA transcripts or protein products of genes that are linked to the genes in Tables 1 and 2 by a biological pathway, may be detected instead of (or in addition to) detecting the mRNA transcripts or proteins corresponding to the genes in Tables 1 and 2.

After the expression level of one or more genes from Tables 1 and 2 has been determined in the patient's sample, the expression is compared to the documented expression of the same gene or genes in the set of samples taken from patients whose response anti-CD20 therapy (e.g. rituximab) has been documented (training set). A statistical tool is then used to determine whether the test patient will respond to anti-CD20 therapy such as rituximab. For example, to determine whether the patient is classified in the responder group or a non-responder group, the Support Vector Machine (SVM) method (Cortes, C. and Vapnik, V. (1995) Machine Learning, 20:273-297) or the k-nearest neighbor method described e.g. in Hastie et al., (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, NY, may be used.

METHOD OF TREATMENT

In another embodiment, the invention is a method of treatment of DLBCL. The method comprises collecting a sample from a patient diagnosed with DLBCL and measuring the level of expression of one or more genes from Tables 1 and 2 in the sample. In variations of this

embodiment, mRNA levels or levels of the protein products of the genes from Tables 1 and 2 may be detected. In another variation of this embodiment, the transcripts or protein products of genes that are linked to the genes in Tables 1 and 2 by a biological pathway may be detected instead of (or in addition to) detecting the transcripts or proteins corresponding to the genes in Tables 1 and 2.

In another embodiment, the invention is a method of treatment of a diffuse large B-cell lymphoma (DLBCL) patient, comprising: obtaining a sample from the patient; detecting in the sample the expression of one or more genes listed in Table 1 and Table 2; determining whether the patient is likely to exhibit response to anti-CD20 therapy, based on the similarity between the measured expression of each gene measured in the patient's sample and a set of control samples; and administering rituximab if the patient is determined to be likely to respond to anti-CD20 therapy.

After the expression level of one or more genes from Tables 1 and 2 has been determined in the patient's sample, the expression is compared to the expression of the same gene or genes in samples of patients whose response to anti-CD20 therapy has been documented. A statistical tool is then used to determine whether the test patient will likely respond to anti-CD20 therapy. If the patient is predicted to respond to anti-CD20 therapy, the therapy (e.g. an antibody such as rituximab) is administered alone or in combination with additional therapeutic agents.

GENE EXPRESSION PROFILE

In another embodiment, the invention comprises a set of detection probes for predicting response of DLBCL patients to anti-CD20 therapy (e.g. rituximab). The detection probes may comprise nucleic acid probes for detecting expression of the genes listed in Table 1, or Table 2, or both. The exact sequence of each nucleic acid probe is not critical. A person skilled in the art of nucleic acid hybridization will be able to select a probe or probes for detecting

expression of each gene based on the published sequence, and assemble the set of probes according to the present invention. For example, for each gene listed in Table 1 or Table 2, the probes listed in the product literature for the Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array may be used to assemble the set of probes according to the present invention.

In some embodiments, detection probes other than nucleic acid probes may be used. For example, antibodies may be used to detect the presence of proteins that are encoded by the genes listed in Table 1 or Table 2, where such proteins are available. In some embodiments, probes or antibodies for the genes linked to the genes listed in Table 1 or Table 2 in a biological pathway may also be included. Thus in some embodiments of the invention, the set of detection probes includes nucleic acid probes, antibodies, or a combination of both.

KITS

In another embodiment, the invention is a kit for predicting response to anti-CD20 therapy in a DLBCL patient. An exemplary kit comprises probes that detect expression of at least one, two or all of the genes listed in Tables 1 and 2. The kit may comprise additional reagents necessary to extract and amplify RNA from the patient's sample. In variations of this embodiment, the kit comprises a set of oligonucleotide probes that detect expression of at least one, two or all of the genes set forth in Tables 1 and 2, attached to a solid support, such as a microarray slide or chip. The kit may also contain reference material, e.g. samples from patients whose response to anti-CD20 therapy, for example, rituximab, has been documented.

In some embodiments of the invention, the kit may contain antibodies for detecting proteins that are encoded by the genes set forth in Tables 1 and 2. In some embodiments, probes or antibodies for the genes linked to the genes listed in Table 1 or Table 2 in a biological pathway may also be included. The kit may comprise additional reagents for antibody-based detection of proteins.

EXAMPLES

Example 1. Selecting predictive expression markers in DLBCL patients treated with rituximab using the Cox proportional hazards model and a k-nearest-neighbor method.

Samples were obtained from DLBCL patients undergoing therapy with CHOP-R (CHOP in combination with rituximab). For these patients, the three-year survival status has been documented. The samples underwent gene expression profiling through use of an Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array (Affymetrix, Inc. Santa Clara, Calif.). Applying a quality control, samples with low signals or less than 18% present calls (microarray signals) were eliminated. The resulting pool consisted of 85 patients falling into two groups: Group 1 – 21 long-term survivors, with survival greater than three years (1095 days) and Group 2 – 19 short term survivors, who died prior the end of three years and 45 patients who dropped out prior to the end of three years.

The microarray expression data was analyzed using the Cox proportional hazards model (Cox, D.R. (1972) *Regression models and life tables*, J R Statistical Soc. Ser., 34:187-220, Lachin, J. M. (2000) *Biostatistical Methods: The Assessment of Relative Risks*, Wiley, NY). For each gene, the Cox model was used to test association between the level of expression of the gene and survival past three years. Based on the value of the T-statistic, the Cox model selected 200 genes. Of the 200 genes, 30 genes with the lowest p-value (Table 1) were selected for the model.

Next, the 1-nearest neighbor method (a version of the k-nearest neighbor method with k=1 (T. Hastie et al., (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* Springer, NY)), was used to group patients according to the gene expression. The model was tested with two methods of cross-validation.

The first validation was performed using the re-substitution method. The accuracy of re-substitution validation was 90%. Results of the re-substitution validation are shown on Figure 1. Figure 1 is a Kaplan-Meier plot (Kaplan, E.L. and Meier, P. (1958) J. Amer. Stat. Assn.

53:457) demonstrating the difference in survival between the two groups selected according to the model. The upper curve represents the actual survival data for patients predicted to be respondents, while the lower curve represents the actual survival data for patients predicted to be non-respondents by the method of the present invention.

The second validation was performed using the leave-one-out cross validation method. The accuracy of leave-one-out cross-validation was 72%. Results of the leave-one-out cross validation are shown on Figure 2. Figure 2 is a Kaplan-Meier plot demonstrating the difference in survival between the two groups selected according to the model. The upper curve represents the actual survival data for patients predicted to be respondents, while the lower curve represents the actual survival data for patients predicted to be non-respondents by the method of the present invention.

Example 2. Selecting predictive expression markers in DLBCL patients treated with rituximab using the Cox proportional hazards model and Support Vector Machine method.

Samples were obtained from DLBCL patients undergoing therapy with CHOP and patients undergoing CHOP-R (CHOP in combination with rituximab) therapy. 159 patients for whom the three-year survival status has been documented were selected. Samples from patients who left the study prior to the end of the three year period were not used in this example.

The 159 samples were randomly split into a training and a testing set in three separate runs as follows: Run 1: 75 training, 84 testing; Run 2: 78 training, 81 testing; Run 3: 83 training, 76 testing. The samples underwent gene expression profiling through use of an Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array.

The microarray expression data was analyzed using the Cox proportional hazards model. For each gene, the Cox model was used to test association between the level of expression and survival past three years. The Cox model was applied to the three testing sets (Runs 1, 2 and 3). The genes with the lowest False Discovery Rate (FDR) were selected. The FDR cut-off of 0.01

yielded 31 genes (Table 2). These 31 genes were selected for model building and cross-validation.

To predict patient's survival past three years, the Support Vector Machine (SVM) method (Cortes, C. and Vapnik, V. (1995) Machine Learning, 20:273-297) was used. SVM is a binary classifier that classifies an input into one of the two classes based on the input data. SVM was applied to classify patient samples into Group 1 (survival more than three years) or Group 2 (survival less than three years) based on the expression levels of the 31 genes in Table 2.

After the samples were separated into Group 1 and Group 2 using the SVM method, a Kaplan-Meier plot was assembled (Figure 3). Group 1 contained 51 samples and Group 2 contained 33 samples. The upper curve represents the actual survival data for patients predicted to be respondents, while the lower curve represents the actual survival data for patients predicted to be non-respondents by the method of the present invention. The Log Rank test was used to determine whether the two curves (for Group 1 and Group 2) were significantly different. Log Rank test yielded a highly significant p- value of 6.06×10^{-10}.

Alternatively, SVM was used to determine the probability that a sample belongs to one of the two groups. This approach yields three classifications: Group 1, Group 2 and unclassified samples for whom the probability of belonging to either group is less than 90%. After the samples were separated into Group 1 and Group 2 using the SVM method, a Kaplan-Meier plot was assembled (Figure 4). Group 1 contained 27 samples, Group 2 contained 31 samples and the remaining 26 samples were unclassified. Log Rank test yielded a highly significant p-value of 4.54×10^{-11}. Figure 4 shows the Kaplan-Meier plot for the 84 samples from the testing set in Run 1. The upper curve represents the actual survival data for patients predicted to be respondents, while the lower curve represents the actual survival data for patients predicted to be non-respondents by the method of the present invention. The middle curve represents the actual survival data for the unclassified group.

Example 3. Validating the model on DLBCL patients treated with rituximab

The method was applied to an independent set of samples: 233 samples from DLBCL patients on the CHOP-R therapy for which survival data were available. SVM was used to determine the probability of each patient belonging to Group 1 (survival more than three years) or Group 2 (survival less than three years). Samples with low probability of belonging to either group remained unclassified. After the samples were separated into Group 1 and Group 2 using the SVM method, a Kaplan-Meier plot was assembled using the survival information available for each sample (Figure 5). The middle curve represents the actual survival data for patients predicted to be respondents, while the lower curve represents the actual survival data for the unclassified patients. The upper curve represents the actual survival data for patients predicted to be non-respondents by the method of the present invention. The data does not contradict the teachings of the invention. The upper curve represents only three patients, all of whom were censored and two of them left the study prior to the three-year period (at 1.71 and 2.95 years). Unfortunately, although the curve shows a 100% survival, the censored patients may have passed away after censoring.

While the invention has been described in detail with reference to specific examples, it will be apparent to one skilled in the art that various modifications can be made within the scope of this invention. Thus the scope of the invention should not be limited by the examples described herein, but by the claims presented below.

CLAIMS:

1. A method of predicting whether a patient with diffuse large B-cell lymphoma (DLBCL) is likely to exhibit response to anti-CD20 therapy, comprising:

(a) determining in a sample obtained from the patient the level of expression of one or more genes selected from U2AF1, RNF2, ZC3H7B, CCNI, PSMD11, ABLIM1, GYG2, CAMK2G, NF1, gene coding for cDNA clone EUROIMAGE 826033 (NCBI Accession No. BE858194), TRIM66, FBXO21, MKNK2, FYTTD1, GET4, CABLES2, FMNL3, C22orf27, TBCEL, ZNF707, the gene corresponding to NCBI Accession No. AI763196, CSTB, BAHCC1, LOC723809, NPIP, the gene coding for CDNA FLJ40566 fis (NCBI Accession No. AL576118), LOC730367, RXFP2, MGC24125, the gene coding for mRNA, differentially expressed in malignant melanoma, clone MM A2 (NCBI Accession No. AJ293390), PTP4A1, OSBP, HNRNPD, AHCYL1, BMP2K, IDS, SERPINB9, CSRP2BP, CD58, HERC4, MAGOH, TMBIM4, SFRS12, FAS, ZNF333, the gene corresponding to NCBI Accession No. AI127295, USP6, CHD6, LOC653160, the gene corresponding to NCBI Accession No. R64696, SRPR, MIR21, ARHGAP18, the gene corresponding to NCBI Accession No. AW004885, CBFA2T2, TIMP3, CTBS, USP16, RASA1 and YJEFN3; and

(b) predicting whether the patient is likely to exhibit response to anti-CD20 therapy, based on the comparison between the level of expression of each gene measured in step (a) in the patient's sample and a set of control samples.

2. The method of claim 1, wherein anti-CD20 therapy comprises an anti-CD20 antibody.

3. The method of any one of claims 1 or 2, wherein the anti-CD20 antibody is rituximab.

4. The method of any one of claims 1 to 3, wherein in step (a), the level of expression of all the genes listed in Table 1 is determined and/or wherein in step (a), the level of expression of all the genes listed in Table 2 is determined.

5. The method of any one of claims 1 to 4, wherein the expression of the genes is measured by measuring the level of RNA transcribed from each gene.

6. The method of any one of claims 1 to 4, wherein the expression of the genes is measured by measuring the level of protein corresponding to each gene.

7. The method of any one of claims 1 to 4, wherein the expression of the genes is measured by measuring the level of RNA or protein of a gene situated downstream in a biological pathway from a gene listed in Table 1 or Table 2.

8. The method of any one of claims 1 to 4, wherein the expression of the genes is measured by measuring the product of activity of a protein corresponding to a gene listed in Table 1 or Table 2.

9. The method of any one of claims 1 to 8, wherein the control set of samples comprises a representative number of patients that exhibit response to anti-CD20 therapy and a representative number of patients that exhibit no response or poor response to anti-CD20 therapy.

10. A set of diagnostic probes for predicting whether a patient with diffuse large B-cell lymphoma (DLBCL) is likely to exhibit response to anti-CD20 therapy, comprising nucleic acid probes or antibodies for detecting expression of one or more genes selected from U2AF1, RNF2, ZC3H7B, CCNI, PSMD11, ABLIM1, GYG2, CAMK2G, NF1, gene coding for cDNA clone EUROIMAGE 826033 (NCBI Accession No. BE858194), TRIM66, FBXO21, MKNK2, FYTTD1, GET4, CABLES2, FMNL3, C22orf27, TBCEL, ZNF707, the gene corresponding to NCBI Accession No. AI763196, CSTB, BAHCC1, LOC723809, NPIP, the gene coding for CDNA FLJ40566 fis (NCBI Accession No.

AL576118), LOC730367, RXFP2, MGC24125, the gene coding for mRNA, differentially expressed in malignant melanoma, clone MM A2 (NCBI Accession No. AJ293390), PTP4A1, OSBP, HNRNPD, AHCYL1, BMP2K, IDS, SERPINB9, CSRP2BP, CD58, HERC4, MAGOH, TMBIM4, SFRS12, FAS, ZNF333, the gene corresponding to NCBI Accession No. AI127295, USP6, CHD6, LOC653160, the gene corresponding to NCBI Accession No. R64696, SRPR, MIR21, ARHGAP18, the gene corresponding to NCBI Accession No. AW004885, CBFA2T2, TIMP3, CTBS, USP16, RASA1 and YJEFN3.

11. In vitro use of a set of diagnostic probes comprising nucleic acid probes or antibodies for detecting expression of one or more genes in a sample obtained from a patient selected from U2AF1, RNF2, ZC3H7B, CCNI, PSMD11, ABLIM1, GYG2, CAMK2G, NF1, gene coding for cDNA clone EUROIMAGE 826033 (NCBI Accession No. BE858194), TRIM66, FBXO21, MKNK2, FYTTD1, GET4, CABLES2, FMNL3, C22orf27, TBCEL, ZNF707, the gene corresponding to NCBI Accession No. AI763196, CSTB, BAHCC1, LOC723809, NPIP, the gene coding for CDNA FLJ40566 fis (NCBI Accession No. AL576118), LOC730367, RXFP2, MGC24125, the gene coding for mRNA, differentially expressed in malignant melanoma, clone MM A2 (NCBI Accession No. AJ293390), PTP4A1, OSBP, HNRNPD, AHCYL1, BMP2K, IDS, SERPINB9, CSRP2BP, CD58, HERC4, MAGOH, TMBIM4, SFRS12, FAS, ZNF333, the gene corresponding to NCBI Accession No. AI127295, USP6, CHD6, LOC653160, the gene corresponding to NCBI Accession No. R64696, SRPR, MIR21, ARHGAP18, the gene corresponding to NCBI Accession No. AW004885, CBFA2T2, TIMP3, CTBS, USP16, RASA1 and YJEFN3 for predicting whether the patient is likely to exhibit response to anti-CD20 therapy, based on the comparison between the expression of each gene measured in the patient's sample and a set of control samples.

12. Use according to claim 11, wherein the patient is a diffuse large B-cell lymphoma (DLBCL) patient and wherein the prediction that the patient is likely to exhibit response

to anti-CD20 therapy results in the administration of an anti-CD20 antibody, preferably rituximab.

13. A kit for predicting whether a patient with diffuse large B-cell lymphoma (DLBCL) is likely to exhibit response to anti-CD20 therapy comprising:

(a) a set of diagnostic probes comprising nucleic acid probes or antibodies for detecting expression of one or more genes selected from U2AF1, RNF2, ZC3H7B, CCNI, PSMD11, ABLIM1, GYG2, CAMK2G, NF1, gene coding for cDNA clone EUROIMAGE 826033 (NCBI Accession No. BE858194), TRIM66, FBXO21, MKNK2, FYTTD1, GET4, CABLES2, FMNL3, C22orf27, TBCEL, ZNF707, the gene corresponding to NCBI Accession No. AI763196, CSTB, BAHCC1, LOC723809, NPIP, the gene coding for CDNA FLJ40566 fis (NCBI Accession No. AL576118), LOC730367, RXFP2, MGC24125, the gene coding for mRNA, differentially expressed in malignant melanoma, clone MM A2 (NCBI Accession No. AJ293390), PTP4A1, OSBP, HNRNPD, AHCYL1, BMP2K, IDS, SERPINB9, CSRP2BP, CD58, HERC4, MAGOH, TMBIM4, SFRS12, FAS, ZNF333, the gene corresponding to NCBI Accession No. AI127295, USP6, CHD6, LOC653160, the gene corresponding to NCBI Accession No. R64696, SRPR, MIR21, ARHGAP18, the gene corresponding to NCBI Accession No. AW004885, CBFA2T2, TIMP3, CTBS, USP16, RASA1 and YJEFN3;

(b) reagents necessary for detecting hybridization and/or binding of the diagnostic probes.

14. The kit of claim 13, comprising nucleic acid probes for the genes listed in Table 1 and/or nucleic acid probes for the genes listed in Table 2.

15. The kit of claim 13, comprising antibodies for proteins expressed from the genes listed in Tables 1 and 2.

FIGURE 1

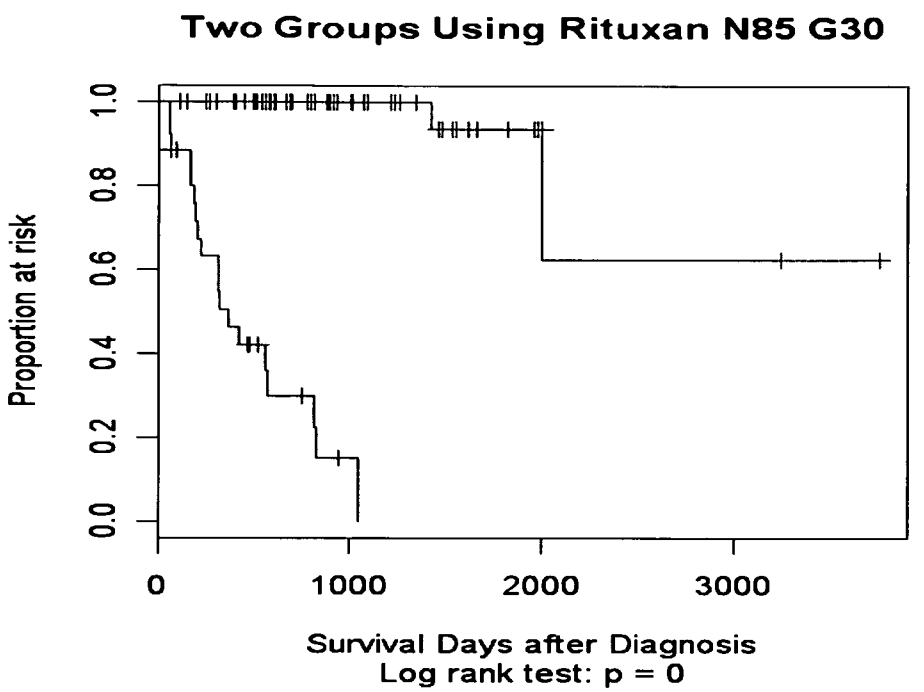

Two Groups Using Rituxan N85 G30
Proportion at risk
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
Survival Days after Diagnosis
Log rank test: p = 0

FIGURE 2

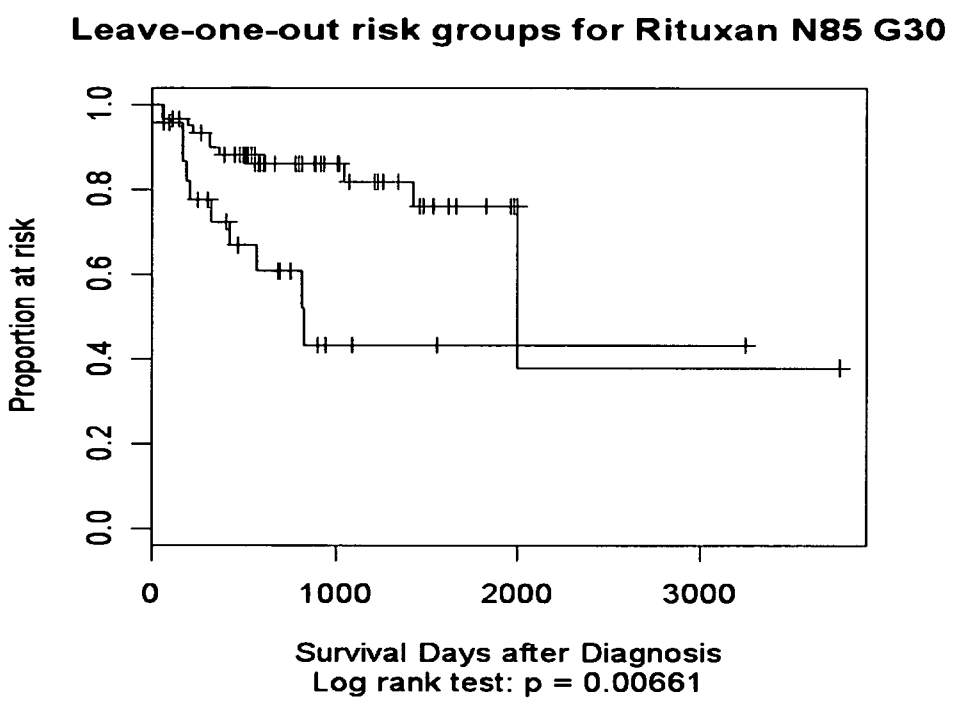

Leave-one-out risk groups for Rituxan N85 G30
Proportion at risk
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
Survival Days after Diagnosis
Log rank test: p = 0.00661

FIGURE 3

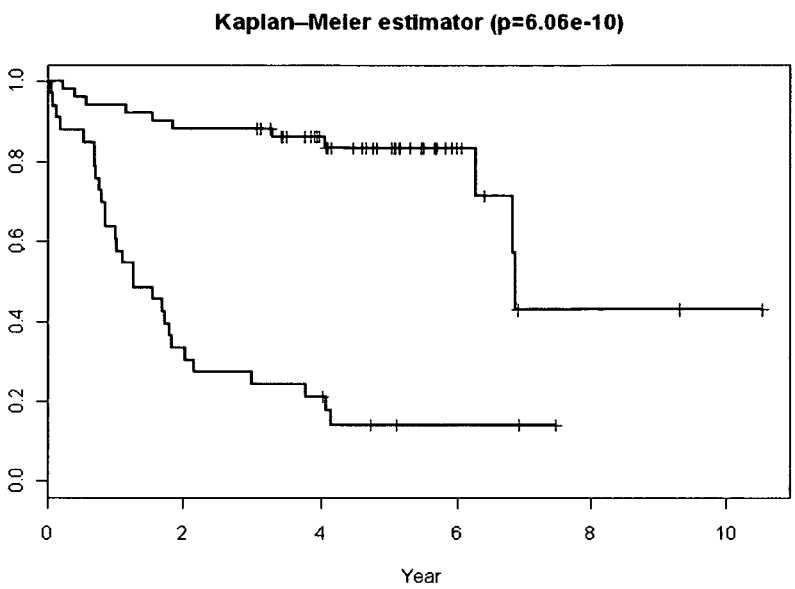

Kaplan–Meier estimator (p=6.06e-10)
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
Year

FIGURE 4

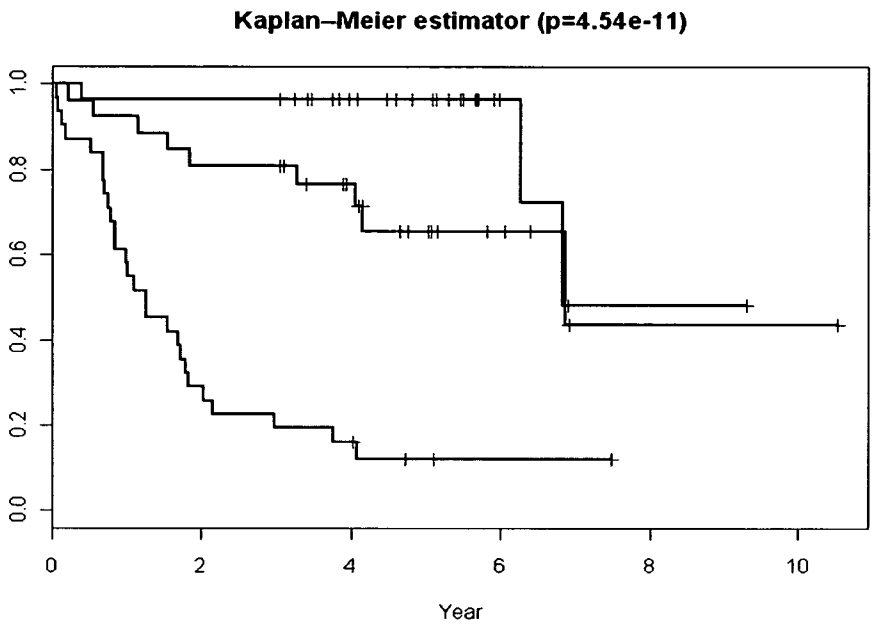
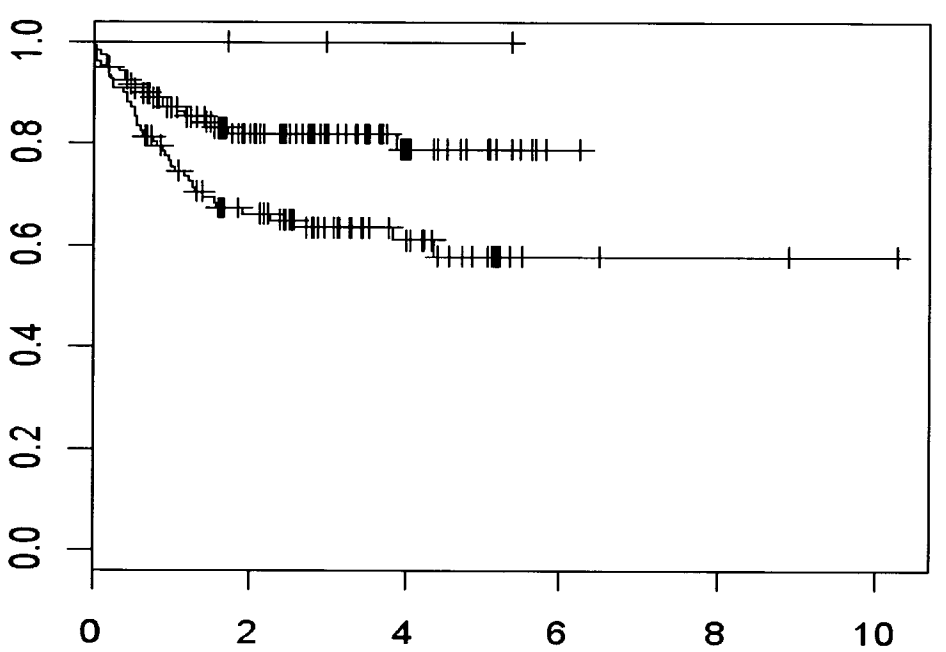

Kaplan–Meier estimator (p=4.54e-11)
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
Year

FIGURE 5



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/003339

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68 G01N33/574
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q G01N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	J-P JAIS ET AL: "The expression of 16 genes related to the cell of origin and immune response predicts survival in elderly patients with diffuse large B-cell lymphoma treated with CHOP and rituximab", LEUKEMIA, vol. 22, no. 10, 1 October 2008 (2008-10-01), pages 1917-1924, XP055036900, ISSN: 0887-6924, DOI: 10.1038/leu.2008.188 the whole document ----- -/--	1-15

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 September 2012	29/10/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Ulbrecht, Matthias

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2012/003339

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-15(partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	G. GUTIERREZ-GARCIA ET AL: "Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy", BLOOD, vol. 117, no. 18, 5 May 2011 (2011-05-05), pages 4836-4843, XP055036896, ISSN: 0006-4971, DOI: 10.1182/blood-2010-12-322362 the whole document	1-15
X	WO 2009/149297 A1 (ARIZONA BOARD REGENTS ON BEHAL [US]; HUTCHINSON FRED CANCER RES [US];) 10 December 2009 (2009-12-10) example 1 claims 1,3,4,5,11,1316,18,19,23,25,29,32,33,36,38	1-15
X	M. ZHANG ET AL: "Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit.", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 89, no. 18, 15 September 1992 (1992-09-15), pages 8769-8773, XP055037061, ISSN: 0027-8424, DOI: 10.1073/pnas.89.18.8769 page 8769, right-hand column, paragraph 2	10,13,15
X	SAMIT GHOSH ET AL: "Global Gene Expression Profiling of Endothelium Exposed to Heme Reveals an Organ-Specific Induction of Cytoprotective Enzymes in Sickle Cell Disease", PLOS ONE, vol. 6, no. 3, 1 January 2011 (2011-01-01) , pages e18399-e18399, XP055035872, ISSN: 1932-6203, DOI: 10.1371/journal.pone.0018399 abstract page 2, right-hand column, paragraph 2	10,13,14
X	-& AFFYMETRIX: "Gene Profiling Array cGMP U133 P2", 20120101, 1 January 2011 (2011-01-01), pages 1-2, XP007920950, the whole document	10,13,14
A	US 2007/172847 A1 (BONAVIDA BENJAMIN [US] ET AL) 26 July 2007 (2007-07-26) paragraph [0008] - paragraph [0010]	1-15

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2012/003339

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2009149297 A1	10-12-2009	CA 2726426 A1	10-12-2009
		EP 2297349 A1	23-03-2011
		JP 2011525106 A	15-09-2011
		US 2011159492 A1	30-06-2011
		WO 2009149297 A1	10-12-2009
US 2007172847 A1	26-07-2007	US 2007172847 A1	26-07-2007
		US 2009203050 A1	13-08-2009

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-15(partially)

A method of predicting whether a patient with DLBCL is likely to exhibit a response to anti-CD20 therapy as defined in claim 1, but limited to determining the level of expression of at least U2AF1; a set of diagnostic probes as defined in claim 10, but limited to nucleic acid probes or antibodies for detecting the expression of at least U2AF1; an in vitro use of a set of diagnostic probes as defined in claim 11, but limited nucleic acid probes or antibodies for detecting the expression of at least U2AF1.; a kit as defined in claim 13, but limited to nucleic acid probes or antibodies for detecting the expression of at least U2AF1.

2-60. claims: 1-15(partially)

Idem as invention 1, but each of inventions 2-60 limited to one of the further genes recited in claims 1, 10, 11 and 13, respectively.
