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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 October 2015 (08.10.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/154018 A1

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

(21) International Application Number:
PCT/US2015/024331

(22) International Filing Date:
3 April 2015 (03.04.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/975,714 4 April 2014 (04.04.2014) US

(71) Applicant *(for all designated States except AL, AT, BA, BE, BG, CH, CN, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IN, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR)*: **GENENTECH, INC.** [US/US]; 1 Dna Way, South San Francisco, CA 94080-4990 (US).

(71) Applicant *(for AL, AT, BA, BE, BG, CH, CN, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IN, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR only)*: **F. HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventors: **MUNDT, Kirsten**; 1 Dna Way, South San Francisco, CA 94080-4990 (US). **BOURGON, Richard**; 1 Dna Way, South San Francisco, CA 94080-4990 (US). **WU, Qun**; 1 Dna Way, South San Francisco, CA 94080-4990 (US). **DU, Pan**; 1 Dna Way, South San Francisco, CA 94080-4990 (US).

(74) Agent: **ELBING, Karen, L.**; Clark & Elbing LLP, 101 Federal Street, 15th Floor, Boston, MA 02110 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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, ST, 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, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

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

— *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))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2015/154018 A1

(54) Title: PROGNOSTIC METHODS FOR DIFFUSE LARGE B-CELL LYMPHOMA

(57) Abstract: The present invention provides DNA methylation signatures in particular diffuse large B-cell lymphoma (DLBCL) subtypes and methods for using these DNA methylation signatures to provide diagnostic and prognostic information for patients who are at risk, diagnosed with, or have a particular subtype of DLBCL.

PROGNOSTIC METHODS FOR DIFFUSE LARGE B-CELL LYMPHOMA

Field of the Invention

The present invention is directed to methods for identifying patients with particular disease subtypes and predicting likely outcome to treatment based on the disease subtypes.

Background of the Invention

Diffuse large B-cell lymphoma (DLBCL) is the most common B-cell malignancy and is highly heterogenous in both clinical and molecular aspects. Gene expression profiling of primary DLBCL cases has identified biologically distinct subtypes of DLBCL. The two predominant molecular subtypes of DLBCL are germinal-center B-cell-like (GCB) DLBCL and activated B-cell-like (ABC) DLBCL. Though histologically indistinguishable, these subtypes are associated with distinct clinical outcomes, and may be predictive for response to, and progression-free survival after treatment with, for example, cyclophosphamide, hydroxydaunorubicin hydrochloride, vincristine, and prednisone; and rituximab with cyclophosphamide, hydroxydaunorubicin hydrochloride, vincristine, and prednisone (R-CHOP).

DNA methylation is involved in critical processes, such as normal cell development, cellular differentiation, genome imprinting, and X-chromosome inactivation. Dysregulation of gene expression involving epigenetic mechanisms, such as DNA methylation, is increasingly recognized as a hallmark of cancer.

There is a need for effective means to select patients, such as DLBCL patients, likely to respond to existing therapies and to guide development of novel therapies.

Summary of the Invention

The invention provides methods of determining whether a patient having diffuse large B-cell lymphoma (DLBCL) will likely respond to an anti-cancer therapy. The methods include: (a) detecting the DNA methylation level of one or more (e.g., 2-5) differentially methylated region (DMR) proximal to or within at least one (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15) gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient, (b) comparing the DNA methylation level of the DMR to a reference level, wherein a change in the level of methylation of the DMR in the patient sample relative to the reference level identifies a patient who is likely to respond to the therapy; and optionally (c) informing the patient that they have an increased likelihood of being responsive to the therapy. In various examples, detection of a DNA methylation signature (see Tables 1-3) consistent with GCB is indicative of a subject who is likely to respond.

The invention also includes methods of optimizing therapeutic efficacy of an anti-cancer therapy for a patient having diffuse large B-cell lymphoma (DLBCL). The methods include: (a) detecting the DNA methylation level of one or more (e.g., 2-5) DMR proximal to or within at least one (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15) gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient, (b) comparing the DNA methylation level of the DMR to a reference level, wherein a change in the level of methylation of the DMR in the patient sample relative to the reference level identifies a patient who is likely to respond to the therapy; and (c) providing a recommendation to the patient for

particular anti-cancer therapy. In various examples, detection of a DNA methylation signature (see Tables 1-3) consistent with GCB is indicative of the potential for increased therapeutic efficacy of anti-cancer therapy as described herein.

Further, the invention includes methods of selecting a therapy for a particular patient having diffuse large B-cell lymphoma (DLBCL) in a population of patients being considered for therapy, the methods include: (a) detecting the DNA methylation level of one or more (e.g., 2-5) DMR proximal to or within at least one (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15) gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient, (b) comparing the DNA methylation level of the DMR to a reference level, wherein a change in the level of methylation of the DMR in the patient sample relative to the reference level identifies a patient who is likely to respond to the therapy; and (c) selecting a particular therapy if the patient is identified as likely to respond to the particular therapy and recommending to the patient the particular therapy; or (d) not selecting a particular therapy if the patient is identified as likely to not respond to the particular therapy and not recommending to the patient the particular therapy. In various examples, detection of a DNA methylation signature (see Tables 1-3) consistent with GCB is indicative of selection of various therapies indicated herein (e.g., cyclophosphamide, hydroxydaunorubicin hydrochloride, vincristine, and prednisone; and rituximab with cyclophosphamide, hydroxydaunorubicin hydrochloride, vincristine, and prednisone (R-CHOP)).

In addition, the invention includes methods of optimizing or modifying a treatment regimen for a patient having diffuse large B-cell lymphoma (DLBCL), the methods include: (a) detecting the DNA methylation level of one or more (e.g., 2-5) DMR proximal to or within at least one (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15) gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient, (b) comparing the DNA methylation level of the DMR to a reference level, wherein a change in the level of methylation of the DMR in the patient sample relative to the reference level identifies a patient who may benefit from modification of their treatment regimen; and (c) providing a recommendation to the patient for the modification. In various examples, detection of a DNA methylation signature (see Tables 1-3) consistent with GCB is indicative of selection of various therapies indicated herein (e.g., cyclophosphamide, hydroxydaunorubicin hydrochloride, vincristine, and prednisone; and rituximab with cyclophosphamide, hydroxydaunorubicin hydrochloride, vincristine, and prednisone (R-CHOP)).

In various embodiments of the invention, the patient is in a population of patients being tested for responsiveness to at least one treatment for cancer and the reference level is the median level of DNA methylation of the DMR in the population of patients. In other embodiments, the patient is in a population of patients being considered for therapy and the reference level is the median level of DNA methylation of the DMR of the at least one gene in the population of patients. In other embodiments, the reference level is the median level in a population known to have DLBCL of the GCB or ABC subtype.

The methods of invention can also include, optionally, detecting the expression level of the at least one gene. The expression level can be detected by, e.g., measuring mRNA or plasma protein levels.

In various embodiments, the DNA methylation level is inversely correlated to gene expression level of the at least one gene.

The therapy in the methods of the invention can be, for example, an agent selected from the group consisting of: an anti-neoplastic agent, a chemotherapeutic agent, a growth inhibitory agent, a cytotoxic agent, and combinations thereof.

The methods of the invention can be used to determine whether the DLBCL is germinal-center B-cell-like (GCB) DLBCL, activated B-cell-like (ABC), DLBCL, or another form of DLBCL.

In various embodiments of the methods of the invention, the change in level of DNA methylation of the DMR in the patient sample is an increase relative to the reference level, or is a decrease relative to the reference level. See Tables 1-3 for additional information in this regard.

The methods of the invention can further optionally include administering an anti-cancer therapy to the patient if it is determined that the patient may benefit from the anti-cancer therapy.

The invention also provides methods for prognosing disease outcome in a patient having DLBCL. These methods can include the steps of: (a) detecting the DNA methylation level of a DMR of at least one gene (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15) listed in Table 1, 2, or 3 in a biological sample obtained from the patient; and (b) comparing the DNA methylation level of the DMR to a reference level, wherein a change in the level of methylation of the DMR in the patient sample relative to the reference level is prognostic of disease outcome. (See, e.g., Tables 1-3.)

In further embodiments, these methods include administering an anti-cancer therapy to the patient if the patient is identified as having DLBCL (in particular, GCB, or ABC).

Further, the invention provides kits for determining whether a patient may benefit from an anti-cancer therapy, the kit including: (a) reagents for determining the DNA methylation level of a DMR of at least one of the genes listed in Table 1, 2, or 3; and optionally (b) instructions for use of the reagents to determine the DNA methylation level of a DMR of at least one of the genes, wherein a change in the level of the DNA methylation relative to a reference level indicates that the patient may benefit from an anti-cancer therapy.

Other features and advantages of the invention will be apparent from the following detailed description, drawings, and claims.

Brief Description of the Drawings

The application file contains at least one drawing executed in color. Copies of this patent or patent application with color drawings will be provided by the Office upon request and payment of the necessary fee.

FIGURE 1 shows the performance of the DLBCL subgroup predictor using gene expression measurements from spotted cDNA microarrays as detailed in Wright et al. *Proc. Natl. Acad. Sci.*. 100(17):9991, 2003. The expression levels for the 27 genes in the subgroup predictor in 274 DLBCL samples are depicted according to the scale shown at the left. The 14 genes that were used to predict the DLBCL subgroups within the Affymetrix data set are indicated with asterisks. The probabilities that the DLBCL samples belong to the ABC or GCB subgroups are graphed at the top, and the DLBCL cases are arranged accordingly. The cases that belong to either the ABC or GCB DLBCL subgroups with ≥ 90% likelihood are indicated.

FIGURE 2 is a graph showing the classification of the GCB and ABC subtypes of DLBCL in 14 different cell lines based on gene expression assayed by qRT-PCR (LPS score).

FIGURE 3 is a schematic showing genome-wide DNA methylation profiling using Illumina Infinium BeadChip.

FIGURE 4 is a flowchart showing a process for identifying differentially methylated regions (DMRs) and their association to nearby genes.

FIGURE 5 shows the inverse correlation in DNA methylation and gene expression patterns as determined by two separate platforms in an eight-gene classifier that differentiates between ABC and GCB subtypes.

FIGURE 6 contains graphs showing the methylation patterns of two known classifier genes, Bcl-6 and IRF4, and the difference in methylation pattern in the ABC and GCB subtypes.

FIGURE 7 shows the results of RNAseq experiments performed for each of the 14 indicated DLBCL cell lines and the correlation of methylation profiles with expression of individual transcripts as evidence for transcript-specific regulation by DNA methylation.

FIGURE 8 shows hierarchical clustering of mean methylation for 160 differentially methylated DMRs (columns) and cell lines (rows). 45 of the genes show strong epigenetic signatures associated with ABC or GCB subtypes.

FIGURE 9 is a schematic showing Bcl-6 and IRF4 as the central nodes of an epigenetically regulated network of ABC and GCB subtypes. The Venn diagram shows the relationship between the 45 differentially methylated genes identified in the study and the 15 classifier genes originally defined by Wright et al., *Proc. Natl. Acad. Sci. USA.* 100(17):9991, 2003, and the overlap in the genes identified in both studies.

FIGURE 10 provides plots showing a comparison of methylation data of all selected CpG sites, and promoter and CpG island-associated sites. Red dots indicate samples in which methylation array QC metrics suggested low quality. In both cases, samples are largely on trend and were not excluded.

Detailed Description of the Preferred Embodiments

I. Introduction

The present invention provides DNA methylation signatures associated with particular DLBCL subtypes (e.g., GCB or ABC), and methods of using these DNA methylation signatures to provide diagnostic and prognostic information for patients who are at risk of, diagnosed with, or have a particular subtype of DLBCL.

The invention is based on the discovery that determination of DNA methylation signatures associated with at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 of the genes listed below in Tables 1-3 can define DLBCL prognostic subpopulations (GCB and ABC). This information can be used to inform clinical decision-making by, e.g., enabling the selection of patients likely to respond to existing and/or novel therapeutic treatment of DLBCL. In Tables 1 and 2, below, the DLBCL subtype (GCB or ABC) stated is the subtype in which methylation of the indicated gene is increased, as compared to the other subtype.

Table 1. Eight-gene signature classifier		
B-cell lymphoma 6 protein (BCL6) **(ABC)**	Inositol-trisphosphate 3-kinase B (ITPKB) **(ABC)**	Tyrosine-protein phosphatase non-receptor type 1 (PTPN1) **(GCB)**
Ets variant 6 (ETV6) **(GCB)**	Serine/threonine-protein kinase Nek6 (NEK6) **(ABC)**	SH3 domain-binding protein 5 (SH3BP5) **(GCB)**
Interferon regulatory factor 4 (IRF4) **(GCB)**	Proto-oncogene serine/threonine-protein kinase Pim-1 (PIM1) **(GCB)**	

Table 2. Signature Genes associated with ABC or GCB subtypes				
absent in melanoma 1 (AIM 1) **(ABC)**	CTD (carboxyl-terminal domain, RNA polymerase II, polypeptide A) phosphatase, subunit 1 (CTDP1) **(GCB)**	heparin sulfate (glucosamine) 3-O-sulfotransferase 1 (HS3ST1) **(GCB)**	phosphodiesterase 3B, cGMP-inhibited (PDE3B) **(GCB)**	synaptogyrin 1 (SYNGR1) **(ABC)**
angiomotin like 1 (AMOTL1) **(ABC)**	CWC27 spliceosome-associated protein homolog (S. cerevisiae) (CWC27) **(GCB)**	immediate early response 5-like (IER5L) **(GCB)**	phosphodiesterase 7B (PDE7B) **(ABC)**	synergin, gamma (SYNGR) **(GCB)**
ankyrin repeat and FYVE domain containing 1 (ANKFY1) **(GCB)**	deafness, autosomal recessive 31 (DFNB31) **(ABC)**	inter-alpha-trypsin inhibitor heavy chain 1 (ITIH1) **(GCB)**	phosphodiesterase 9A (PDE9A) **(GCB)**	synaptophysin-like 1 (SYPL1) **(ABC)**
ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 (ARAP3) **(GCB)**	DLG5 antisense RNA 1 (DLG5-AS1) **(GCB)**	potassium voltage-gated channel, subfamily G, member 2 (KCNG2) **(GCB)**	pleckstrin homology domain containing, family A member 7 (PLEKHA7) **(ABC)**	toll-like receptor 9 (TLR9) **(ABC)**
Rho GTPase activating protein 23 (ARHGAP23) **(ABC)**	dynein, axonemal, heavy chain 6 (DNAH6) **(GCB)**	lysine (K)-specific demethylase 4C (KDM4C) **(ABC)**	PR domain containing 15 (PRDM15) **(GCB)**	tumor necrosis factor receptor superfamily, member 17 (TNFRSF17) **(GCB)**

B-cell CLL/lymphoma 11A (zinc finger protein) (BCL11A) **(ABC)**	early B-cell factor 1 (EBF1) **(ABC)**	KIAA0907 (KIAA0907) **(GCB)**	PR domain containing 2, with ZNF domain (PRDM2) **(GCB)**	troponin C type 2 (fast) (TNNC2) **(GCB)**
BEN domain containing 3 (BEND3) **(ABC)**	enoyl-CoA delta isomerase 1 (ECI1) **(GCB)**	long intergenic non-protein coding RNA 161 (LINC00161) **(GCB)**	patched domain containing 2 (PTCHD2) **(GCB)**	torsin family 4, member A (TOR4A) **(ABC)**
bestrophin 4 (BEST4) **(ABC)**	EGF-like and EMI domain containing 1, pseudogene (EGFEM1P) **(GCB)**	long intergenic non-protein coding RNA 271 (LINC00271) **(ABC)**	prothymosin, alpha (PTMA) **(GCB)**	tyrosylprotein sulfotransferase 1 (TPST1) **(ABC)**
basic helix-loop-helix family, member a15 (BHLHA15) **(GCB)**	ectopic P-granules autophagy protein 5 homolog (C. elegans) (EPG5) **(GCB)**	long intergenic non-protein coding RNA 544 (LINC00544) **(ABC)**	protein tyrosine phosphatase, non-receptor type 18 (brain-derived) (PTPN18) **(ABC)**	tRNA aspartic acid methyltransferase 1 (TRDMT1) **(GCB)**
chromosome 12 open reading frame 42 (C12orf42) **(GCB)**	family with sequence similarity 69, member B (FAM69B) **(GCB)**	Gem (nuclear organelle) associated protein 2 psuedogene 2 (GEMIN2P2; LOC100287063) **(ABC)**	glutaminyl-tRNA synthase (glutamine-hydrolyzing)-like 1 (QRSL1) **(ABC)**	utrophin (UTRN) **(ABC)**
coiled-coil domain containing 42 (CCDC42) **(GCB)**	guanylate binding protein 3 (GBP3) **(GCB)**	lysophosphatidic acid receptor 5 (LPAR5) **(GCB)**	RAP2A, member of RAS oncogene family (RAP2A) **(GCB)**	zinc finger protein 217 (ZNF217) **(GCB)**
cadherin 6, type 2, K-cadherin (fetal kidney) (CDH6) **(GCB)**	golgin A3 (GOLGA3) **(GCB)**	low density lipoprotein receptor-related protein 5 (LRP5) **(ABC)**	rabphilin 3A-like (without C2 domains) (RPH3AL) **(GCB)**	DiGeorge Syndrome Critical Region Gene 6 (DGCR6) **(GCB)**
carnitine palmitoyltransferase 1A (liver) (CPT1A) **(GCB)**	G protein-coupled receptor 25 (GPR25) **(ABC)**	matrix metallopeptidase 15 (membrane-inserted) (MMP15) **(ABC)**	short stature homeobox 2 (SHOX2) **(ABC)**	Microtubule-associated protein 18 (MAP18) **(GCB)**
carboxypeptidase Z (CPZ) **(GCB)**	guanylate cyclase 2D, membrane (retina-specific) (GUCY2D) **(GCB)**	neuraminidase (NA) **(GCB)**	spleen focus forming virus (SFFV) proviral integration oncogene (SPI1) **(ABC)**	Prickle homolog 1 (PRICKLE1) **(GCB)**
casein kinase 1, alpha 1 (CSNK1A1) **(GCB)**	hippocalcin-like 1 (HPCAL1) **(ABC)**	oculocutaneous albinism II (OCA2) **(GCB)**	suppression of tumorigenicity 5 (ST5) **(ABC)**	EPH receptor A8 (EPHA8) **(GCB)**

Tumor necrosis factor, alpha-induced protein 8-like 2 (TNFAIP8L2) **(ABC)**	Ribonuclease T2 (RNASET2) **(ABC)**	Transcription factor CP2 (TFCP2) **(ABC)**	Synaptotagmin-like 1 (SYTL1) **(ABC)**	ATP-binding cassette, subfamily B (MDR/TAP), member 6 (ABCB6) **(ABC)**
A disintegrin and metalloproteinase 19 (ADAM19) **(ABC)**	Paired-like homeodomain 1 (PITX1) **(ABC)**	Ring finger protein 14 (RNF14) **(ABC)**	Spleen focus forming virus (SFFV) proviral integration ocogene (SPI1) **(ABC)**	Carbonic anhydrase VII (CA7) **(GCB)**
Disks large-associated protein 2 (DLGAP2) **(GCB)**	Small Nucleolar RNA, C/D Box 115-43 (SNORD115-43) **(GCB)**	Kallikrein-related peptidase 14 (KLK14) **(GCB)**	Thyroid hormone receptor interactor 10 (TRIP10) **(GCB)**	Olfactory receptor, family 6, subfamily X, member 1 (OR6X1) **(GCB)**
Prepronociceptin (PNOC) **(GCB)**	Sprouty homolog 1 (SPRY1) **(GCB)**	RNA gene affiliated with lncRNA class (LOC727710) **(GCB)**	Zinc finger, AN1-type domain 2A (ZFAND2A) **(GCB)**	Src homology 2 domain containing adaptor protein B (SHB) **(GCB)**
Spermidine synthase (SRM) **(GCB)**	Solute carrier family 25 (SLC25A3) **(GCB)**	Zinc ribbon domain containing 1 (ZNRD1) **(GCB)**	Histon cluster 3, H2bb (HIST3H2BB) **(GCB)**	Zinc finger protein 77 (ZNF77) **(GCB)**

Table 3. DLBCL biomarker network

B-cell lymphoma 6 protein (Bcl-6)	LIM domain only 2 (rhombotin-like 1) (LMO2)	Phosphodiesterase 3 (PDE3B)	B-cell linker protein (BLNK)
Interferon regulatory factor 4 (IRF4)	COX Assembly mitochondrial protein 2 homolog (C16orf61)	Histone H2A type 3 (HIST3H2A)	Neprilysin (MME)
Lymphoid-restricted membrane protein (Lrmp)	Interleukin 16 (IL-16)	Transcription factor PU.1 (PU.1)	
G1/S-specific cyclin-D2 (Cyclin D2)	Human inositol 1,4,6-trisphosphate 3-kinase isoform B (IP3KB)	Transcription factor COE1 (EBF)	

II. Definitions

The term "DNA methylation signature" is used herein to refer to a distinct DNA methylation pattern characteristic or prognostic of a disorder. DNA methylation signatures may be associated with a set of genes (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 genes) that are differentially or aberrantly methylated and, as a result, silenced or activated in cancerous cells versus normal cells. To determine signatures of differential methylation that are characteristic of a particular cancer or cancer subtype, DNA methylation patterns of a sample from a subject diagnosed with or having the cancer is compared to the DNA methylation patterns of a sample from a subject who does not have the cancer or who has the cancer, e.g., in a known subtype. DNA methylation patterns of samples from subjects at different stages of the cancer, in remission, or in relapse can also be used to determine signatures of differential methylation. The sets of genes associated with DNA methylation signatures can serve as diagnostic and prognostic biomarkers. In particular examples of the invention, the DNA methylation signature of a test sample is compared to the DNA methylation signature associated with the same set of genes from a reference sample or samples, which can include samples from GCB and/or ABC subtypes of DLBCL.

The term "differentially methylated region" or "DMR" is used herein to refer to a region within genomic DNA that includes one or more CpG sites that exhibit patterns or degrees of methylation that vary in different conditions (e.g., diseases, such as cancers, or subtypes thereof (e.g., DLBCL subtypes GCB and ABC). DMRs can occur within a gene (e.g., with an exon, such as the first exon, or in an intron), or be proximal to a gene. A DMR is "proximal" to a gene if it is found within the promoter of a gene (e.g., a gene described herein), near the transcription start site (TSS) of a gene (e.g., within 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, or 7 kb), or within an exon upstream of a gene.

The terms "DNA methylation level" or "methylation level" are used interchangeably herein and generally refer to the amount of methylated DNA in a biological sample or at a particular site (e.g., a DMR or CpG site within a DMR). "Methylation" generally refers to the process by which a methyl group is added to a nucleotide, e.g., cytosine. DNA methylation at the 5 position of cytosine generally has the effect of reducing gene expression. Hypomethylation refers to a decrease in epigenetic methylation of a nucleotide, such as cytosine, in DNA. In general, hypomethylation arises earlier and is linked to chromosomal instability and loss of imprinting. Hypermethylation refers to an increase in the epigenetic methylation of a nucleotide, e.g., cytosine, in DNA and is associated and can arise secondary to gene (e.g., oncogene suppressor) silencing. According to the invention "methylation" of a gene or proximal DMR can, but is not required to, affect expression of the gene.

The terms "biomarker" and "marker" are used interchangeably herein to refer to a DNA, RNA, protein, carbohydrate, or glycolipid-based molecular marker, the methylation, expression, or presence of which in a subject's or patient's sample can be detected by standard methods (or methods disclosed herein) and is useful for diagnosing a cancer as described herein, predicting and monitoring the responsiveness or sensitivity of a mammalian subject to treatment of a cancer. Such biomarkers include, but are not limited to, the genes listed in Tables 1-3. Methylation or expression of such a biomarker may be determined to be higher or lower in a sample obtained from a subject predisposed, diagnosed, or having a cancer described herein (including, *e.g.*, the median methylation or expression level of the

biomarker in samples from a group/population of subjects being tested for a cancer; the level in a sample previously obtained from the individual at a prior time; the level in a sample from a subject who received prior treatment for a cancer, and who now may be experiencing relapse of the cancer; or the median level in samples of a population of subjects having a particular cancer subtype (e.g., DLBCL of the GCB or ABC subtype). Individuals having a methylation or expression level that is greater than or less than the reference methylation or expression level of at least one gene, such as those noted above, can be identified as subjects/patients likely to respond to treatment of a cancer. For example, such subjects/patients who exhibit gene methylation or expression levels at, e.g., the most extreme 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, or 5% relative to (i.e., higher or lower than) the reference level (such as the median level, noted above), can be identified as subjects/patients likely to respond to treatment of a cancer.

The terms "sample" and "biological sample" are used interchangeably to refer to any biological sample obtained from an individual including body fluids, body tissue (*e.g.*, tumor tissue), cells, or other sources. Body fluids are, *e.g.*, lymph, sera, whole fresh blood, peripheral blood mononuclear cells, frozen whole blood, plasma (including fresh or frozen), urine, saliva, semen, synovial fluid and spinal fluid. Samples also include breast tissue, renal tissue, colonic tissue, brain tissue, muscle tissue, synovial tissue, skin, hair follicle, bone marrow, and tumor tissue. Methods for obtaining tissue biopsies and body fluids from mammals are well known in the art.

An "effective response" of a subject or a subject's "responsiveness" or "sensitivity" to treatment of a cancer refers to the clinical or therapeutic benefit imparted to a subject at risk for or suffering from a cancer described herein (e.g., DLBCL) from or as a result of the treatment with existing therapies or novel therapies. Such benefit includes cellular or biological responses, a complete response, a partial response, a stable disease (without progression or relapse), or a response with a later relapse of the patient from or as a result of the treatment with existing therapies or novel therapies. For example, an effective response can be reduced tumor size or progression-free survival in a patient diagnosed as having a DNA methylation signature or expressing one or more of the biomarkers noted above, in a manner described herein, versus a patient not having the DNA methylation signature or expressing one or more of the biomarkers in such a manner. The detection of DNA methylation signatures, as described herein, and optionally expression of genetic biomarker(s), effectively predicts, or predicts with high sensitivity, such effective response. In one example, detection of a DNA methylation signature consistent with GCB is indicative of the potential for a more effective response.

A "disorder" or "disease" is any condition that would benefit from treatment with a substance/molecule or method of the invention. This includes chronic and acute disorders or diseases including those pathological conditions which predispose the mammal to the disorder in question. Non-limiting examples of disorders to be treated herein include malignant and benign tumors; non-leukemias and lymphoid malignancies; neuronal, glial, astrocytal, hypothalamic and other glandular, macrophagal, epithelial, stromal and blastocoelic disorders; and inflammatory, immunologic and other angiogenic disorders.

The terms "cell proliferative disorder" and "proliferative disorder" refer to disorders that are associated with some degree of abnormal cell proliferation. In one embodiment, the cell proliferative disorder is cancer.

"Tumor," as used herein, refers to all neoplastic cell growth and proliferation, whether malignant or benign, and all pre-cancerous and cancerous cells and tissues. The terms "cancer," "cancerous," "cell proliferative disorder," "proliferative disorder," and "tumor" are not mutually exclusive as referred to herein.

The terms "cancer" and "cancerous" refer to or describe the physiological condition in mammals that is typically characterized by unregulated cell proliferation. Examples of cancer include but are not limited to, carcinoma, lymphoma, blastoma, sarcoma, and leukemia. More particular examples of such cancers include squamous cell cancer, lung cancer (including small-cell lung cancer, non-small cell lung cancer, adenocarcinoma of the lung, and squamous carcinoma of the lung), cancer of the peritoneum, hepatocellular cancer, gastric or stomach cancer (including gastrointestinal cancer), pancreatic cancer, glioblastoma, cervical cancer, ovarian cancer, liver cancer, bladder cancer, hepatoma, breast cancer, colon cancer, colorectal cancer, endometrial or uterine carcinoma, salivary gland carcinoma, kidney or renal cancer, liver cancer, prostate cancer, vulval cancer, thyroid cancer, hepatic carcinoma and various types of head and neck cancer, as well as B-cell lymphoma (including low grade/follicular non-Hodgkin's lymphoma (NHL); small lymphocytic (SL) NHL; intermediate grade/follicular NHL; intermediate grade diffuse NHL; high grade immunoblastic NHL; high grade lymphoblastic NHL; high grade small non-cleaved cell NHL; bulky disease NHL; mantle cell lymphoma; AIDS-related lymphoma; and Waldenstrom's Macroglobulinemia); chronic lymphocytic leukemia (CLL); acute lymphoblastic leukemia (ALL); Hairy cell leukemia; chronic myeloblastic leukemia; and post-transplant lymphoproliferative disorder (PTLD), as well as abnormal vascular proliferation associated with phakomatoses, edema (such as that associated with brain tumors), and Meigs' syndrome.

The term "anti-neoplastic composition" or "anti-cancer composition" or "anti-cancer agent" refers to a composition useful in treating cancer comprising at least one active therapeutic agent, *e.g.*, "anti-cancer agent." Examples of therapeutic agents (anti-cancer agents) include, but are limited to, *e.g.*, chemotherapeutic agents, growth inhibitory agents, cytotoxic agents, agents used in radiation therapy, anti-angiogenesis agents, apoptotic agents, anti-tubulin agents, and other-agents to treat cancer, such as anti-HER-2 antibodies, anti-CD20 antibodies (e.g., rituximab), an epidermal growth factor receptor (EGFR) antagonist (*e.g.*, a tyrosine kinase inhibitor), HER1/EGFR inhibitor (*e.g.*, erlotinib (Tarceva™), platelet derived growth factor inhibitors (*e.g.*, Gleevec™ (Imatinib Mesylate)), a COX-2 inhibitor (*e.g.*, celecoxib), interferons, cytokines, antagonists (*e.g.*, neutralizing antibodies) that bind to one or more of the following targets ErbB2, ErbB3, ErbB4, PDGFR-beta, BlyS, APRIL, BCMA VEGF, or VEGF receptor(s), TRAIL/Apo2, and other bioactive and organic chemical agents, etc. Combinations thereof are also included in the invention.

Non-neoplastic and neoplastic conditions include, *e.g.*, cancer, especially vascularized solid tumors and metastatic tumors (including colon cancer, breast cancer, lung cancer (especially small-cell lung cancer), brain cancer (especially glioblastoma) or prostate cancer), undesired or aberrant hypertrophy, arthritis, rheumatoid arthritis (RA), inflammatory bowel disease or IBD (Crohn's disease and ulcerative colitis), psoriasis, psoriatic plaques, sarcoidosis, atherosclerosis, atherosclerotic plaques,

diabetic and other proliferative retinopathies including retinopathy of prematurity, retrolental fibroplasia, neovascular glaucoma, age-related macular degeneration, diabetic macular edema, corneal neovascularization, corneal graft neovascularization, corneal graft rejection, retinal/choroidal neovascularization, neovascularization of the anterior surface of the iris (rubeosis), ocular neovascular disease, vascular restenosis, arteriovenous malformations (AVM), meningioma, hemangioma, angiofibroma, thyroid hyperplasias (including Grave's disease), chronic inflammation, lung inflammation, acute lung injury/ARDS, sepsis, primary pulmonary hypertension, malignant pulmonary effusions, cerebral edema (*e.g.*, associated with acute stroke/ closed head injury/ trauma), synovial inflammation, myositis ossificans, hypertropic bone formation, osteoarthritis (OA), refractory ascites, polycystic ovarian disease, endometriosis, 3rd spacing of fluid diseases (pancreatitis, compartment syndrome, burns, bowel disease), uterine fibroids, premature labor, chronic inflammation such as IBD, renal allograft rejection, inflammatory bowel disease, nephrotic syndrome, undesired or aberrant tissue mass growth (non-cancer), hemophilic joints, hypertrophic scars, inhibition of hair growth, Osler-Weber syndrome, pyogenic granuloma retrolental fibroplasias, scleroderma, trachoma, vascular adhesions, synovitis, dermatitis, preeclampsia, ascites, pericardial effusion (such as that associated with pericarditis), and pleural effusion.

As used herein, "treatment" refers to clinical intervention in an attempt to alter the natural course of the individual or cell being treated, and can be performed either for prophylaxis or during the course of clinical pathology. Desirable effects of treatment include preventing occurrence or recurrence of disease, alleviation of symptoms, diminishment of any direct or indirect pathological consequences of the disease, preventing metastasis, decreasing the rate of disease progression, amelioration or palliation of the disease state, and remission or improved prognosis. In some embodiments, antibodies of the invention are used to delay development of a disease or disorder.

An "effective amount" refers to an amount effective, at dosages and for periods of time necessary, to achieve the desired therapeutic or prophylactic result.

A "therapeutically effective amount" of a substance/molecule of the invention, agonist or antagonist may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of the substance/molecule, agonist or antagonist to elicit a desired response in the individual. A therapeutically effective amount is also one in which any toxic or detrimental effects of the substance/molecule, agonist or antagonist are outweighed by the therapeutically beneficial effects. The term "therapeutically effective amount" refers to an amount of an antibody, polypeptide or antagonist of this invention effective to "treat" a disease or disorder in a mammal (e.g., patient). In the case of cancer, the therapeutically effective amount of the drug can reduce the number of cancer cells; reduce the tumor size or weight; inhibit (i.e., slow to some extent and preferably stop) cancer cell infiltration into peripheral organs; inhibit (i.e., slow to some extent and preferably stop) tumor metastasis; inhibit, to some extent, tumor growth; and/or relieve to some extent one or more of the symptoms associated with the cancer. To the extent the drug can prevent growth and/or kill existing cancer cells, it can be cytostatic and/or cytotoxic. In one embodiment, the therapeutically effective amount is a growth inhibitory amount. In another embodiment, the therapeutically effective amount is an amount that extends the survival of a patient. In another embodiment, the therapeutically effective amount is an amount that improves progression free survival of a patient.

A "chemotherapeutic agent" is a chemical compound useful in the treatment of cancer. Examples of chemotherapeutic agents include alkylating agents such as thiotepa and CYTOXAN® cyclosphosphamide; alkyl sulfonates such as busulfan, improsulfan and piposulfan; aziridines such as benzodopa, carboquone, meturedopa, and uredopa; ethylenimines and methylamelamines including altretamine, triethylenemelamine, trietylenephosphoramide, triethiylenethiophosphoramide and trimethylolomelamine; acetogenins (especially bullatacin and bullatacinone); delta-9-tetrahydrocannabinol (dronabinol, MARINOL®); beta-lapachone; lapachol; colchicines; betulinic acid; a camptothecin (including the synthetic analogue topotecan (HYCAMTIN®), CPT-11 (irinotecan, CAMPTOSAR®), acetylcamptothecin, scopolectin, and 9-aminocamptothecin); bryostatin; callystatin; CC-1065 (including its adozelesin, carzelesin and bizelesin synthetic analogues); podophyllotoxin; podophyllinic acid; teniposide; cryptophycins (particularly cryptophycin 1 and cryptophycin 8); dolastatin; duocarmycin (including the synthetic analogues, KW-2189 and CB1-TM1); eleutherobin; pancratistatin; a sarcodictyin; spongistatin; nitrogen mustards such as chlorambucil, chlornaphazine, cholophosphamide, estramustine, ifosfamide, mechlorethamine, mechlorethamine oxide hydrochloride, melphalan, novembichin, phenesterine, prednimustine, trofosfamide, uracil mustard; nitrosureas such as carmustine, chlorozotocin, fotemustine, lomustine, nimustine, and ranimnustine; antibiotics such as the enediyne antibiotics (*e.g.*, calicheamicin, especially calicheamicin gamma1I and calicheamicin omegaI1 (see, *e.g.*, Agnew, *Chem Intl. Ed. Engl.* 33:183-186 (1994)); dynemicin, including dynemicin A; an esperamicin; as well as neocarzinostatin chromophore and related chromoprotein enediyne antiobiotic chromophores), aclacinomysins, actinomycin, authramycin, azaserine, bleomycins, cactinomycin, carabicin, carminomycin, carzinophilin, chromomycinis, dactinomycin, daunorubicin, detorubicin, 6-diazo-5-oxo-L-norleucine, ADRIAMYCIN® doxorubicin (including morpholino-doxorubicin, cyanomorpholino-doxorubicin, 2-pyrrolino-doxorubicin and deoxydoxorubicin), epirubicin, esorubicin, idarubicin, marcellomycin, mitomycins such as mitomycin C, mycophenolic acid, nogalamycin, olivomycins, peplomycin, potfiromycin, puromycin, quelamycin, rodorubicin, streptonigrin, streptozocin, tubercidin, ubenimex, zinostatin, zorubicin; anti-metabolites such as methotrexate and 5-fluorouracil (5-FU); folic acid analogues such as denopterin, methotrexate, pteropterin, trimetrexate; purine analogs such as fludarabine, 6-mercaptopurine, thiamiprine, thioguanine; pyrimidine analogs such as ancitabine, azacitidine, 6-azauridine, carmofur, cytarabine, dideoxyuridine, doxifluridine, enocitabine, floxuridine; androgens such as calusterone, dromostanolone propionate, epitiostanol, mepitiostane, testolactone; anti- adrenals such as aminoglutethimide, mitotane, trilostane; folic acid replenisher such as frolinic acid; aceglatone; aldophosphamide glycoside; aminolevulinic acid; eniluracil; amsacrine; bestrabucil; bisantrene; edatraxate; defofamine; demecolcine; diaziquone; elfornithine; elliptinium acetate; an epothilone; etoglucid; gallium nitrate; hydroxyurea; lentinan; lonidainine; maytansinoids such as maytansine and ansamitocins; mitoguazone; mitoxantrone; mopidanmol; nitraerine; pentostatin; phenamet; pirarubicin; losoxantrone; 2-ethylhydrazide; procarbazine; PSK® polysaccharide complex (JHS Natural Products, Eugene, OR); razoxane; rhizoxin; sizofiran; spirogermanium; tenuazonic acid; triaziquone; 2,2',2"-trichlorotriethylamine; trichothecenes (especially T-2 toxin, verracurin A, roridin A and anguidine); urethan; vindesine (ELDISINE®, FILDESIN®); dacarbazine; mannomustine; mitobronitol; mitolactol; pipobroman; gacytosine; arabinoside ("Ara-C"); thiotepa; taxoids, *e.g.*, TAXOL® paclitaxel (Bristol-Myers Squibb Oncology, Princeton, N.J.),

ABRAXANE™ Cremophor-free, albumin-engineered nanoparticle formulation of paclitaxel (American Pharmaceutical Partners, Schaumberg, Illinois), and TAXOTERE® doxetaxel (Rhône-Poulenc Rorer, Antony, France); chloranbucil; gemcitabine (GEMZAR®); 6-thioguanine; mercaptopurine; methotrexate; platinum analogs such as cisplatin and carboplatin; vinblastine (VELBAN®); platinum; etoposide (VP-16); ifosfamide; mitoxantrone; vincristine (ONCOVIN®); oxaliplatin; leucovovin; vinorelbine (NAVELBINE®); novantrone; edatrexate; daunomycin; aminopterin; ibandronate; topoisomerase inhibitor RFS 2000; difluorometlhylornithine (DMFO); retinoids such as retinoic acid; capecitabine (XELODA®); pharmaceutically acceptable salts, acids or derivatives of any of the above; as well as combinations of two or more of the above such as CHOP, an abbreviation for a combined therapy of cyclophosphamide, doxorubicin, vincristine, and prednisolone, and FOLFOX, an abbreviation for a treatment regimen with oxaliplatin (ELOXATIN™) combined with 5-FU and leucovovin. Additional chemotherapeutic agents include the cytotoxic agents useful as antibody drug conjugates, such as maytansinoids (DM1, for example) and the auristatins MMAE and MMAF, for example.

A "subject" herein is any single human subject, including a patient, eligible for treatment who is experiencing or has experienced one or more signs, symptoms, or other indicators of cancer disorder. Intended to be included as a subject are any subjects involved in clinical research trials not showing any clinical sign of disease, or subjects involved in epidemiological studies, or subjects once used as controls.

A "kit" is any manufacture (*e.g.*, a package or container) comprising at least one reagent, *e.g.*, a probe for specifically detecting a biomarker gene or protein of the invention. The manufacture is preferably promoted, distributed, or sold as a unit for performing the methods of the present invention.

By "correlate" or "correlating" is meant comparing, in any way, the performance and/or results of a first analysis or protocol with the performance and/or results of a second analysis or protocol. For example, one may use the results of a first analysis or protocol in carrying out a second protocols and/or one may use the results of a first analysis or protocol to determine whether a second analysis or protocol should be performed.

The terms "level of expression" or "expression level" are used interchangeably and generally refer to the amount of a polynucleotide or an amino acid product or protein in a biological sample. "Expression" generally refers to the process by which gene-encoded information is converted into the structures present and operating in the cell. Therefore, according to the invention "expression" of a gene may refer to transcription into a polynucleotide, translation into a protein, or even posttranslational modification of the protein. Fragments of the transcribed polynucleotide, the translated protein, or the post-translationally modified protein shall also be regarded as expressed whether they originate from a transcript generated by alternative splicing or a degraded transcript, or from a post-translational processing of the protein, *e.g.*, by proteolysis. "Expressed genes" include those that are transcribed into a polynucleotide as mRNA and then translated into a protein, and also those that are transcribed into RNA but not translated into a protein (for example, transfer and ribosomal RNAs).

III. Methods for Identifying Patients and Predicting Outcome of Treatment in Patients with Particular DLBCL Subtypes

As noted above, DLBCL is the most common B-cell malignancy and is highly heterogenous in both clinical and molecular features. Gene expression profiling has identified two predominant molecular subtypes of DLBCL, germinal-center B-cell-like (GCB) DLBCL and activated B-cell-like (ABC) DLBCL. Though histologically indistinguishable, these subtypes are associated with distinct clinical outcomes and may be predictive for response to and progression-free survival in patients treated with cyclophosphamide, hydroxydaunorubicin hydrochloride, vincristine, and prednisone and rituximab with cyclophosphamide, hydroxydaunorubicin hydrochloride, vincristine, and prednisone (R-CHOP) or novel therapies. In particular, GCB is, in general, associated with a better therapeutic outcome than ABC.

The present invention provides DNA methylation signatures associated with DLBCL subtypes, which can be used as biomarkers for these subtypes. In particular, distinct DNA methylation patterns associated with the two DLBCL molecular subtypes, GCB and ABC, revealed a correlation pattern of high mRNA expression associated with low methylation in the vicinity of the transcriptional start site (TSS), and vice versa. Comparison of differential methylation profiles with corresponding RNAseq data show that DNA methylation profiles underlie the gene expression patterns that are currently used to classify ABC and GCB DLBCL subtypes and show that differential methylation profiles can be used in sensitive, specific, and non-invasive approaches to screen for the presence of a cancer and, in particular, cancer sub-type (e.g., ABC or GCB DLBCL) in at-risk individuals.

A. Detection of DNA Methylation

Differentially methylated regions (DMRs) are typically located near or within the regulatory regions of genes, and correlation of methylation signature profiles with expression of individual transcripts can identify biomarkers that are regulated by methylation.

DNA methylation can be detected using any method known in the art. For example, high-throughput approaches to identify DNA methylation markers or signatures described herein such as restriction landmark genomic scanning, microarray gene expression profiling after 5-aza-2'-deoxycytidine (5-aza) treatment, and ChIP-on-chip approaches can be used (see, e.g., Costello et al., *Nat. Genet.* 24(2):132-138 (2000); Weber et al., *Nat. Genet.* 37(8):853-862 (2005); Suzuki et al., *Nat. Genet.* 31(2):141-149 (2002); Karpf et al., *Proc. Natl. Acad. Sci. USA.* 96(24):14007-14012 (1999)). Higher resolution methods, such as sodium bisulfate pyrosequencing and quantitative methylation-specific PCT (QMSP) can also be used in the methods described herein (see, e.g., Belinsky, *Nat. Rev. Cancer* 4(9):707-717 (2004); Belinksy et al., *Cancer Res.* 66(6):3338-3344 (2006)). The methylation assay described in Walter et al. *Clin Cancer Res.* 18:2360-2373 (2012), described in detail in the Examples, can also be used to detect DNA methylation. A summary of various methods for detecting DNA methylation is provided in Table 4, below.

Table 4. Methods for detecting DNA methylation

Methods	Throughput	Material	Application
Methylation specific PCR	Relatively high for samples; low for genes	Patient DNA, cell line DNA, bodily fluids (sputum, blood, urine)	Screening for new methylation markers, analyzing relatively large numbers of samples for known markers
Quantitative MSP (Taqman, sybr-green)	Relatively high for samples, difficult to set up, every gene requires empirical standardization	Patient DNA, cell line DNA, bodily fluids (sputum, blood, urine)	Quantitative analysis of patient DNA, correlations with clinical parameters
Pyrosequencing	Moderate for samples and genes, high throughput for sequencing	Patient DNA, cell line DNA, bodily fluids (sputum, blood, urine)	Quantitative analysis of patient DNA, correlations with clinical parameters, analysis of more CpG sites than MSP
Microarray expression profiling after reversal of DNA methylation	Low for samples, high for genes	Cell lines	Novel marker discovery
Restriction landmark genomic scanning	Low for samples, high for genes	Cell lines	Novel marker discovery
ChIP-on-chip	Low for samples, high for genes	Cell lines, potentially tumor material	Novel marker discovery

For additional detailed descriptions of methods for detecting DNA methylation, see, e.g., Herman et al., *Proc. Natl. Acad. Sci. USA.* 93(18):9821-9826 (1996); Zochbauer-Muller et al., *Cancer Res.* 61(1):249-255 (2001); Shivapurkar et al., *Int. J. Cancer* 116(4):656-660 (2005); Collela et al., *Biotechniques* 35 (1):146-150 (2003); 63 Keshet et al., Nat. Genet. 38(2):149-153 (2006); Fackler et al., *Cancer Res.* 64(13):4442-4452 (2004), Curtis & Coggins, *Methods Mol. Med.* 103:123-136 (2005); Fraga & Esteller, *Biotechniques.* 33(3):632,634,636-649 (2002); Costello et al., *Methods* 27(2):144-149 (2002).

B. Detection of Gene Expression

The genetic biomarkers described herein can be detected using any method known in the art. For example, tissue or cell samples from mammals can be conveniently assayed for, *e.g.*, mRNAs or DNAs from a genetic biomarker of interest using Northern, dot-blot, or polymerase chain reaction (PCR) analysis, array hybridization, RNase protection assay, or using DNA SNP chip microarrays, which are commercially available, including DNA microarray snapshots. For example, real-time PCR (RT-PCR) assays such as quantitative PCR assays are well known in the art. In an illustrative embodiment of the invention, a method for detecting mRNA from a genetic biomarker of interest in a biological sample comprises producing cDNA from the sample by reverse transcription using at least one primer; amplifying the cDNA so produced; and detecting the presence of the amplified cDNA. In addition, such methods can

include one or more steps that allow one to determine the levels of mRNA in a biological sample (*e.g.*, by simultaneously examining the levels a comparative control mRNA sequence of a "housekeeping" gene such as an actin family member). Optionally, the sequence of the amplified cDNA can be determined.

1. Detection of Nucleic Acids

Expression of the biomarker genes as described herein can be performed by RT-PCR technology. Probes used for PCR may be labeled with a detectable marker, such as, for example, a radioisotope, fluorescent compound, bioluminescent compound, a chemiluminescent compound, metal chelator, or enzyme. Such probes and primers can be used to detect the presence of expressed genes set forth in Tables 1-3 in a sample. As will be understood by the skilled artisan, a great many different primers and probes may be prepared and used effectively to amplify, clone and/or determine the presence and/or levels expressed of one or more of the genes listed in Tables 1-3.

Other methods include protocols that examine or detect mRNAs from at least one of the genes listed in Tables 1-3 in a tissue or cell sample by microarray technologies. Using nucleic acid microarrays, test and control mRNA samples from test and control tissue samples are reverse transcribed and labeled to generate cDNA probes. The probes are then hybridized to an array of nucleic acids immobilized on a solid support. The array is configured such that the sequence and position of each member of the array is known. For example, a selection of genes that have potential to be expressed in certain disease states may be arrayed on a solid support. Hybridization of a labeled probe with a particular array member indicates that the sample from which the probe was derived expresses that gene. Differential gene expression analysis of disease tissue can provide valuable information. Microarray technology utilizes nucleic acid hybridization techniques and computing technology to evaluate the mRNA expression profile of thousands of genes within a single experiment (see, *e.g.*, WO 2001/75166). See, for example, U.S. Patent No. 5,700,637, U.S. Patent No. 5,445,934, and U.S. Patent No. 5,807,522, Lockart, *Nature Biotechnology* 14:1675-1680 (1996); and Cheung et al., *Nature Genetics* 21(Suppl):15-19 (1999) for a discussion of array fabrication.

In addition, the DNA profiling and detection method utilizing microarrays described in EP 1753878 may be employed. This method rapidly identifies and distinguishes between different DNA sequences utilizing short tandem repeat (STR) analysis and DNA microarrays. In an embodiment, a labeled STR target sequence is hybridized to a DNA microarray carrying complementary probes. These probes vary in length to cover the range of possible STRs. The labeled single-stranded regions of the DNA hybrids are selectively removed from the microarray surface utilizing a post-hybridization enzymatic digestion. The number of repeats in the unknown target is deduced based on the pattern of target DNA that remains hybridized to the microarray.

One example of a microarray processor is the Affymetrix GENECHIP® system, which is commercially available and comprises arrays fabricated by direct synthesis of oligonucleotides on a glass surface. Other systems may be used as known to one skilled in the art.

Other methods for determining the level of the biomarker besides RT-PCR or another PCR-based method include proteomics techniques, as well as individualized genetic profiles that are necessary to treat cancer based on patient response at a molecular level. Other methods that can be used to detect

nucleic acids, for use in the invention, involve high throughput RNA sequence expression analysis, including RNA-based genomic analysis, such as, for example, RNASeq.

Many references are available to provide guidance in applying the above techniques (Kohler *et al., Hybridoma Techniques* (Cold Spring Harbor Laboratory, New York, 1980); Tijssen, *Practice and Theory of Enzyme Inimunoassays* (Elsevier, Amsterdam, 1985); Campbell, *Monoclonal Antibody Technology* (Elsevier, Amsterdam, 1984); Hurrell, *Monoclonal Hybridoma Antibodies: Techniques and Applications* (CRC Press, Boca Raton, FL, 1982); and Zola, *Monoclonal Antibodies: A Manual of Techniques*, pp. 147-1 58 (CRC Press, Inc., 1987)). Northern blot analysis is a conventional technique well known in the art and is described, for example, in *Molecular Cloning, a Laboratory Manual,* second edition, 1989, Sambrook, Fritch, Maniatis, Cold Spring Harbor Press, 10 Skyline Drive, Plainview, NY 11803-2500. Typical protocols for evaluating the status of genes and gene products are found, for example in Ausubel *et al.*, eds., 1995, *Current Protocols In Molecular Biology*, Units 2 (Northern Blotting), 4 (Southern Blotting), 15 (Immunoblotting) and 18 (PCR Analysis).

2. Detection of Proteins

As to detection of protein biomarkers such as a protein biomarker corresponding to at least one of the genes listed in Tables 1-3, for example, various protein assays are available including, for example, antibody-based methods as well as mass spectroscopy and other similar means known in the art. In the case of antibody-based methods, for example, the sample may be contacted with an antibody specific for said biomarker under conditions sufficient for an antibody-biomarker complex to form, and then detecting said complex. Detection of the presence of the protein biomarker may be accomplished in a number of ways, such as by Western blotting (with or without immunoprecipitation), 2-dimensional SDS-PAGE, immunoprecipitation, fluorescence activated cell sorting (FACS), flow cytometry, and ELISA procedures for assaying a wide variety of tissues and samples, including plasma or serum. A wide range of immunoassay techniques using such an assay format are available, see, *e.g.*, U.S. Patent Nos. 4,016,043, 4,424,279, and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as in the traditional competitive binding assays. These assays also include direct binding of a labeled antibody to a target biomarker.

Sandwich assays are among the most useful and commonly used assays. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate, and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labeled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labeled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of biomarker.

Variations on the forward assay include a simultaneous assay, in which both sample and labeled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In a typical forward sandwich assay, a first antibody having specificity for the biomarker is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride, or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (*e.g.*, 2-40 minutes or overnight if more convenient) and under suitable conditions (*e.g.*, from room temperature to 40°C such as between 25°C and 32°C inclusive) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the biomarker. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the molecular marker.

An alternative method involves immobilizing the target biomarkers in the sample and then exposing the immobilized target to specific antibody which may or may not be labeled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labeling with the antibody. Alternatively, a second labeled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule. By "reporter molecule," as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (*i.e.*, radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase, and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable color change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labeled antibody is added to the first antibody-molecular marker complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of biomarker which was present in the sample. Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be

chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labeled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic color visually detectable with a light microscope. As in the EIA, the fluorescent labeled antibody is allowed to bind to the first antibody-molecular marker complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength, the fluorescence observed indicates the presence of the molecular marker of interest. Immunofluorescence and EIA techniques are both very well established in the art. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

C. Kits

For use in detection of the biomarkers of the present invention, kits or articles of manufacture are also provided by the invention. Such kits can be used to determine if a subject is at risk for or has a cancer as described herein (e.g., DLBCL), to determine the subtype of cancer (e.g., GCB or ABC), and the likelihood that the subject will be responsive to treatment of the cancer. These kits can comprise a carrier means being compartmentalized to receive in close confinement one or more container means such as vials, tubes, and the like, each of the container means comprising one of the separate compounds or elements to be used in the method. For example, one of the container means may comprise a probe that is or can be detectably labeled. Such a probe may be a polynucleotide specific for a message. Where the kit utilizes nucleic acid hybridization to detect the target nucleic acid, the kit may also have containers containing nucleotide(s) for amplification of the target nucleic acid sequence and/or a container comprising a reporter-means, such as a biotin-binding protein, *e.g.*, avidin or streptavidin, bound to a reporter molecule, such as an enzymatic, florescent, or radioisotope label.

Such a kit will typically comprise the container described above and one or more other containers comprising materials desirable from a commercial and user standpoint, including buffers, diluents, filters, needles, syringes, and package inserts with instructions for use. A label may be present on the container to indicate that the composition is used for a specific application, and may also indicate directions for either *in vivo* or *in vitro* use, such as those described above.

The kits of the invention have a number of embodiments. A typical embodiment is a kit comprising a container, a label on said container, and a composition contained within said container, wherein the composition includes polynucleotides that hybridize to the target nucleic acid, and the label on said container indicates that the composition can be used to evaluate the presence of such nucleic acid in a sample, and wherein the kit includes instructions for using the polynucleotides for evaluating the presence of methylated regions in a particular sample type. In case of DNA methylation, the kit may contain reagents such as sodium bisulfite and desulphonation columns, wash buffers, dilution buffers and binding buffers for purification of detected methylated DNA. The kit can further comprise a set of instructions and materials for preparing a sample and hybridizing the sample. The kit may include polynucleotides conjugated to a label for data read out.

Another embodiment is a kit comprising a container, a label on said container, and a composition contained within said container, wherein the composition includes one or more polynucleotides that

hybridize to a complement of a biomarker as described herein under stringent conditions, and the label on said container indicates that the composition can be used to evaluate the presence of a biomarker as described herein in a sample, and wherein the kit includes instructions for using the polynucleotide(s) for evaluating the presence of the biomarker RNA or DNA in a particular sample type.

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, *e.g.*, a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a biomarker protein or (2) a pair of primers useful for amplifying a biomarker nucleic acid molecule. The kit can also comprise, *e.g.*, a buffering agent, a preservative, or a protein stabilizing agent. The kit can further comprise components necessary for detecting the detectable label (*e.g.*, an enzyme or a substrate). The kit can also contain a control sample or a series of control samples that can be assayed and compared to the test sample. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit. Kits can also include instructions for interpreting the results obtained using the kit.

D. Statistics

As used herein, the general form of a prediction rule consists in the specification of a function of one or multiple biomarkers potentially including clinical covariates to predict response or non-response, or more generally, predict benefit or lack of benefit in terms of suitably defined clinical endpoints.

The simplest form of a prediction rule consists of a univariate model without covariates, wherein the prediction is determined by means of a cutoff or threshold. This can be phrased in terms of the Heaviside function for a specific cutoff c and a biomarker measurement x, where the binary prediction A or B is to be made, then if $H(x-c)=0$, then predict A, if $H(x-c)=1$, then predict *B*.

This is the simplest way of using univariate biomarker measurements in prediction rules. If such a simple rule is sufficient, it allows for a simple identification of the direction of the effect, *i.e.*, whether high or low methylation or expression levels are beneficial for the patient.

The situation can be more complicated if clinical covariates need to be considered and/or if multiple biomarkers are used in multivariate prediction rules. The two hypothetical examples below illustrate the issues involved:

1. Covariate Adjustment (Hypothetical Example):

For a biomarker X it is found in a clinical trial population that high methylation or expression levels are associated with a worse clinical response (univariate analysis). A closer analysis shows that there are two types of clinical response in the population, a first group which possesses a worse response than the second group and at the same time the biomarker methylation or expression for the first group is generally higher following administration of one or more anti-cancer therapies. An adjusted covariate analysis reveals that for each of the groups the relation of clinical benefit and clinical response is reversed, *i.e.*, within the groups, lower methylation or expression levels are associated with better clinical response. The overall opposite effect was masked by the covariate type--and the covariate adjusted analysis as part of the prediction rule reversed the direction.

2. Multivariate Prediction (Hypothetical Example):

For a biomarker X it is found in a clinical trial population that high methylation or expression levels are slightly associated with a worse clinical response (univariate analysis). For a second biomarker Y a similar observation was made by univariate analysis. The combination of X and Y revealed that a good clinical response is seen if both biomarkers are low. This makes the rule to predict benefit if both biomarkers are below some cutoffs (AND--connection of a Heaviside prediction function). For the combination rule, a simple rule no longer applies in a univariate sense; for example, having low methylation or expression levels in X will not automatically predict a better clinical response.

EXAMPLES

The following examples are provided to illustrate, but not to limit, the presently claimed invention.

Experimental Methods

The methods described in this section are used in Examples 1-5 set forth below. The Gene Expression Data and Formulation of the DLBCL Subgroup Predictor methods are described in Wright *et al.*, *Proc. Natl. Acad. Sci. USA*. 100(17):9991 (2003) and reproduced below.

Gene Expression Data

DLBCL gene expression data generated by using Lymphochip microarrays were obtained from supporting information of Rosenwald et al., *N. Engl. J. Med.* 346:1937-1947 (2002) at http://llmpp.nih.gov/DLBCL. DLBCL gene expression data generated by using Affymetrix HU6500 microarrays were obtained from supporting information of Shipp et al., Nat. Med. 8:68-74 (2002) at www.genome.wi.mit.edu/MPR/lymphoma and were normalized as follows. Genes that were listed as present on >50% of the samples were identified and then the signal values on each array were multiplied by a factor to make the median value of these genes equal to 1,000. After this normalization, all signal values that were <50 to a value of 50 were set and then a $\log_2$ transformation was applied. All gene expression data used in the present analysis can be obtained from http://llmpp.nih.gov/DLBCLpredictor.

Formulation of the DLBCL Subgroup Predictor

A linear predictor score (LPS) was calculated for each sample *X* of the form

$$\mathrm{LPS}(X) = \sum_j a_j X_j, \qquad [1]$$

where X_j represents the gene expression of gene *j*, and a_j is a scaling factor whose value depends on the degree to which each gene discriminates the subgroups. The scaling factors were chosen to be the *t* statistics generated by a *t* test for the difference in expression between the two subgroups. Only the *k* genes with the most significant *t* statistics were used to form the LPS, with the optimal *k* determined empirically (see below). For genes represented by multiple features on the microarray, the feature with the most significant *t* statistic was used.

Because the LPS is a linear combination of gene expression values, its distribution within each subgroup should be approximately normal, provided it includes a sufficient number of genes and the correlation structure of those genes is not extreme. The mean and variance of these normal distributions can then be estimated from the LPSs calculated for the samples in each subgroup. Given the LPS distribution of each subgroup, it is possible to estimate the likelihood that a new sample is in each of the two subgroups by applying Bayes' rule, so that

$$P(X \text{ in group } 1) = \frac{\phi(LPS(X); \hat{\mu}_1, \hat{\sigma}_1^2)}{\phi(LPS(X); \hat{\mu}_1, \hat{\sigma}_1^2) + \phi(LPS(X); \hat{\mu}_2, \hat{\sigma}_2^2)},$$

[2]

where $\phi(x; \mu, \sigma^2)$ represents the normal density function with mean μ, and variance σ^2, and $\hat{\mu}_1$, $\hat{\sigma}_1^2$, $\hat{\mu}_2$, and $\hat{\sigma}_2^2$ are the observed mean and variance of the LPSs within subgroup 1 and subgroup 2, respectively.

Because the samples that are used to estimate the distribution of the LPSs are also used to generate the model, there is a possibility of overfitting, resulting in a model that would indicate a larger separation between the subgroups' LPSs than would be found in independent data. Therefore, it is important to check the validity of the model on a separate validation data set. A training set consisting of 42 ABC DLBCL and 67 GCB DLBCL samples were constructed and a validation set consisting of 41 ABC DLBCL, 67 GCB DLBCL, and 57 type 3 DLBCL samples were also constructed. To choose the optimal number of genes (*k*) to include in the model, multiple models with different numbers of DLBCL subgroup discrimination genes were evaluated on the training set by using a leave-one-out cross-validation procedure; a model including 27 genes had the lowest average error rate. This model was applied to the validation set and observed that the distribution of the LPSs within each subgroup matched the corresponding distributions in the training set, thus demonstrating that model overfitting was not an issue. By using the probability estimates determined by Bayes' rule, a cutoff of 90% certainty was chosen to decide final subgroup membership. Those samples for which there was <90% likelihood of being in either subgroup were termed "unclassified."

To apply the DLBCL subgroup predictor to data obtained by using Affymetrix microarrays, Affymetrix microarray features with a median signal value of <200 across the samples were excluded and then multiple microarray features representing the same gene were averaged, if present. Of the 27 genes in the DLBCL subgroup predictor described above, only 14 were represented on the Affymetrix microarrays and passed this filtering process. These 14 genes were used to create a new DLBCL subgroup predictor in which the LPS scaling coefficients were again calculated based on the DLBCL subgroup distinction in the Lymphochip data set. For each gene, the expression values in the Affymetrix data set were shifted and scaled to match the mean and variance of the corresponding expression values in the Lymphochip data set, to account for systematic measurement differences between the two microarray platforms. The adjusted expression values for the 14 genes in the predictor were used to calculate LPSs for each sample in the Affymetrix data set, and DLBCL subgroup membership was assigned as above based on a cutoff of 90% certainty (Figure 1).

Fluidigm-based LPS scores and subtype assignment

ABC/GCB subtypes were assigned by implementing the linear predictor score (LPS) method published by Wright *et al.*, *Proc. Natl. Acad. Sci. USA*. 100(17):9991 (2003), as described above, on the Fluidigm real-time qPCR platform, which is suitable for FFPE samples, unlike the original microarrays. To obtain gene weights for the Fluidigm platform, expression of 21 classifier genes defined by Wright *et al.* (*supra*) was assayed in 166 unrelated DLBCL reference samples. For each sample, raw Ct measurements were first normalized to housekeeping genes (dCt), followed by gene-wise normalization to a universal reference sample (ddCt). Assignment of ABC/GCB label to the reference samples was exactly as in Wright *et al.*: through use of unsupervised clustering and subsequent selection of core clades whose gene expression patterns were highly consistent with previously established ABC and GCB prototypes. The following *t*-statistics were obtained for use as gene-specific weights when computing Fluidigm-specific LPS scores:

BCL6	10.13	**ITPKB**	11.17
BLNK	-3.18	**LMO2**	7.69
BMF	-1.57	**LRMP**	8.52
DDB1	1.64	**MME**	8.38
DENND3	9.24	**MYBL1**	16.48
ENTPD1	-5.21	**NEK6**	10.91
ETV6	-6.18	**PIM1**	-3.54
FUT8	-4.30	**PTPN1**	-4.52
IGHM	-7.75	**SH3BP5**	-6.19
IL16	-6.00	**TBC1D27**	-5.79
IRF4	-7.44		

LPS scores for new samples were calculated based on their normalized (ddCt) expression measurements. Classification probabilities were calculated based on the mean and standard deviation of LPS scores for ABC and GCB reference samples, as in Wright *et al.* Subtypes were assigned to samples with classification probability exceeding 90%; other sample were left unclassified (sometimes referred to as "Type III") (Figure 2).

Genome-wide DNA Methylation Profiling

Microarray data was collected at Expression Analysis, Inc. (Durham, NC; www.expressionanalysis.com) using the Illumina® Human Methylation450 BeadChip which is commercially available. These arrays contain probes for approximately 450,000 CpG loci sties. Target was prepared and hybridized according to the "Illumina Infinium HD Methylation Assay, Manual Protocol" (Illumina Part # 15019522 Rev. A) (Figure 3).

A bisulfite conversion reaction was employed using 500 ng of genomic DNA according to the manufacturer's protocol for the Zymo EZ DNA Methylation kit (Zymo Research). DNA was added to Zymo M-Dilution buffer and incubated for 15 min. at 37ºC. CT-conversion reagent was then added and the mixture was denatured by heating to 95ºC for 30 second followed by incubation for 1 hour at 50ºC. This denature/incubation cycle was repeated for a total of 16 hours. After bisulfite conversion, the DNA

was bound to a Zymo spin column and desulfonated on the column using desulfonation reagent per manufacturer's protocol. The bisulfite-converted DNA was eluted from the column in 10 µl of elution buffer (Figure 3).

4 µl of bisulfite converted product was transferred to a new plate with an equal amount of 0.1N NaOH and 20 µl of MA1 reagent (Illumina) then allowed to incubate at RT for 10 min. Immediately following incubation, 68 µl of MA2 reagent and 75 µl of MSM reagent (both Illumina) were added and the plate was incubated at 37ºC overnight for amplification. After amplification, the DNA was fragmented enzymatically, precipitated and resuspended in RA1 hybridization buffer.

Fragmented DNA was dispensed onto the multichannel HumanMethylation BeadChip and hybridization performed in an Illumina Hybridization oven for 20 hours. BeadChips were washed, primer extended, and stained per manufacturer protocols. BeadChips were coated and then imaged on an Illumina iScan Reader and images were processed with Genome Studio software methylation module (version 1.8 or later). Array data were analyzed and a methylation classifier was established using a "leave-one-out" cross validation strategy.

Preprocessing of microarray data using Bioconductor lumi package

Methylation data were processed using the Bioconductor lumi software package (Du, Kibbe et al. *Bioinformatics.* 24(13):1547-1548 (2008). The Infinium 450K platform includes Infinium I and II assays on the same array. The Infinium I assay employs two bead types per CpG locus, with the methylated state reported by the red dye in some cases and the green dye in others (identical to the previous Infinium 27K platform). The Infinium II assay uses one bead type and always reports the methylated state with the same dye, making dye bias a concern. A two-stage normalization procedure was applied to the arrays. First, for each array, a color-bias correction curve was estimated from Infinium I data using a smooth quantile normalization method; this correction curve was then applied to all data from that array. Second, arrays were normalized to one another by applying standard quantile normalization to all color-corrected signals. After pre-processing, both methylation M-values (log2 ratios of methylated to unmethylated probes) and β-values (a rescaling of the M-values to the 0 and 1 range via logistic transform) were computed for each sample (Du, Zhang et al. *BMC Bioinformatics* 11:587 (2010). For visualization, agglomerative hierarchical clustering of β-values was performed using complete linkage and Euclidean distance.

Detecting regions of significantly differentiated methylation

Supervised identification of CpG sites whose methylation signal is associated with a subtype required division of training lines into two groups. Two different divisions of varying stringency were considered:

1. **Standard**: classification of cell lines was based on application of the Fluidigm LPS classifier described below. This expression-based classifier assigns samples to one of ABC, GCB, or Unclassified. Standard classification contrasted ABC with GCB lines and omitted unclassifieds.
2. **Strict**: strict classification expanded the unclassified group and only retained ABC or GCB lines with the highest estimated probabilities of being correctly assigned to the respective class.

Other aspects of the procedure for identifying CpG sites which were differentially methylated between ABC and GCB cell lines were as described in Walter et al. *Clin Cancer Res.* 18:2360-2373 (2012), with the following minor modifications to significance and effect size thresholds: we required a difference of at least one unit for average *M*-values (i.e., a two-fold change between classes); and *t*-tests were considered significant if either the FDR-adjusted p-value was below 0.05 (loose) or below 0.01 (tight)..

After consideration of stringency levels 1 and 2 above, we found best performance on FFPE data (detailed below) when using (i) strict selection of cell lines and (ii) the tighter FDR cutoff for selection of differentially methylated CpG sites. Selected probes (one per CpG) site are detailed in the Appendix. The important columns from the Appendix are noted below.

- PROBEID: the Illumina identifier for the selected probe pair on the microarray. Genomic coordinates for the CpG site interrogated by this probe are given in the CHROMOSOME and POSITION columns, relative to human genome release hg19.
- GeneSymbol: selected probes/CpG sites are associated with the gene with the nearest transcription start site. The gene symbol is given here. Note that many genes are associated with more than one probe/CpG site; the *Probes by gene* tab provide unique gene identifier and probe count per gene.
- Promoter/CpG island: selected CpG sites which are close to transcription start sites or fall within annotated CpG islands are easier to interpret, although not necessarily better for prediction. This column identifies CpG sites within this category. The results considered below were considered for either all selected CpG sites in the spreadsheet, or only those which are promoter/CpG island proximal.

Hierarchical clustering used to identify distinct methylation signatures in disease subtypes

Hierarchical clustering was used to build a binary tree of the data that successively merges similar groups of points. Visualizing this tree provides a useful summary of the data. Softwares that include a function for creating hierarchical cluster analysis are commercially available and include softwares such as, MATLAB®, SAS®, and Mathematica®.

RNAseq Transcriptome Analysis

RNAseq was used to analyze the transcriptome of each of the 14 DLBCL cell lines. In general, a population of RNA (total or fractionated, such as poly(A)+) is converted to a library of cDNA fragments with adaptors attached to one or both ends. Each molecule, with or without amplification, is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). The reads are typically 30–400 bp, depending on the DNA-sequencing technology used. In principle, any high-throughput sequencing technology can be used for RNAseq, and the Illumina IG®, Applied Biosystems SOLiD®, and Roche 454 Life Science® systems have already been applied for this purpose. Following sequencing, the resulting reads are either aligned to a reference genome or reference transcripts, or assembled *de novo* without the genomic sequence to produce a genome-scale transcription map that consists of both the transcriptional structure and/or level

of expression for each gene. For an in depth description of the RNAseq technique, see, e.g., Wang et al., *Nat. Rev. Genet.* 10(1):57-63 (2009).

Example 1. Identification of differentially methylated regions/genes in ABC and GCB samples

Methylation at 485,512 sites was profiled for each of the 14 cell lines with the Illumina® Infinium BeadChip technology as described above. Subtype-specific methylation levels were evaluated by comparing the 7 cell lines classified as ABC to the 7 cell lines classified as GCB detailed in Table 5. The experimental schematic for narrowing the number of assayed probes to a smaller subset of significant probes is detailed in Figure 4. Significant probes were selected by t-test of locally smoothed methylation data. The selected probes in close proximity to another were merged to form differentially methylated regions (DMRs). This differential methylation analysis identified 215 significant DMRs. The 215 DMRs were then assigned to genes if the DMRs were within 2 kb from a transcription start site, within a promoter region, or within the first exon. This process narrowed the initial 215 DMRs down to 160 DMRs, assigned to 45 genes. This 160 significant DMRs associated with 45 genes were identified as differentiating DLBCL subtypes based on LPS classification.

Table 5. Cell lines classification as ABC or GCB subtype

Cell Line	*Classification*
A3/Kawakami (derived from malignant lymphoma of B(non-T)-Cell nature of the gastrointestinal tract)	ABC
A4/Fukada (derived from malignant lymphoma of B(non-T)-Cell nature of the gastrointestinal tract)	ABC
Ri-1 (Epstein-Barr virus-negative B-lymphoma cell line)	ABC
RC-K8 (human B-cell lymphoma)	ABC
SCC-3 (squamous cell carcinoma)	ABC
Toledo (derived from human peripheral blood B lymphocyte)	ABC
Pfeiffer (derived from human metastatic site lymphoblast B lymphocyte)	ABC
NU-DHL-1 (derived from human inguinal lymph node B lymphocyte)	GCB
DB (human ascites B lymphoblast in large cell lymphoma)	GCB
RL (human ascites B lymphoblast in non-Hodgkin's lymphoma)	GCB
HT (human ascites B lymphoblast in diffuse mixed lymphoma)	GCB
Karpas-422 (human B-cell Non-Hodgkin's Lymphoma)	GCB
SU-DHL-16 (human lymph node lymphoblast-like large cell lymphoma)	GCB
SU-DHL-10 (human lymph node lymphoblast-like large cell lymphoma)	GCB

Example 2. Analysis of multiple expression classifier genes show subtype-specific methylation

The 45 differentially methylated genes were then combined with the 15 DLBCL classifier genes originally defined by Wright et al., *Proc. Natl. Acad. Sci. USA* 100(17):9991 (2003) and mapped onto the GenGo MetaCore™ DLBCL disease network. Eight overlapping genes from the combined 45 differentially methylated genes and 15 DLBCL classifier genes defined by Wright et al. showed subtype-specific differential methylation. Furthermore, analysis of Illumina DNA methylation data and Affymetrix Gene Expression data of these eight genes revealed an inverse correlation between methylation and gene expression (Figure 5). The eight gene signature is shown in Table 1, above, which is reproduced below as Table 6.

Table 6. Eight-gene signature classifier

B-cell lymphoma 6 protein (BCL6) **(ABC)**	Inositol-trisphosphate 3-kinase B (ITPKB) **(ABC)**	Tyrosine-protein phosphatase non-receptor type 1 (PTPN1) **(GCB)**
Ets variant 6 (ETV6) **(GCB)**	Serine/threonine-protein kinase Nek6 (NEK6) **(ABC)**	SH3 domain-binding protein 5 (SH3BP5) **(GCB)**
Interferon regulatory factor 4 (IRF4) **(GCB)**	Proto-oncogene serine/threonine-protein kinase Pim-1 (PIM1) **(GCB)**	

The eight-gene signature shows a strong correlation between methylation and expression across different platforms and can be used to differentiate between ABC and GCB subtypes of DLBCL. For example, Figure 6 shows Bcl-6, found to be expressed in GCB subtype, is hypermethylated at 2 out of 3 loci in ABC cell lines. Conversely, IRF4, found to be highly expressed in ABC subtype, is methylated around the transcriptional start site in 3 out of 4 probes in GCB cell lines.

Example 3. Analysis of DNA methylation profiles shows transcript-specific regulation by DNA methylation

RNAseq was performed for each of the 14 DLBCL cell lines (Figure 7). Analysis of reads mapping to splice junction and isoform-specific exons allowed transcript-level quantitation. Correlation of methylation profiles with expression of individual transcripts pinpoints isoform-specific regulation by methylation. Comparison of methylation profiles with RNAseq data provided insights into methylation-based transcriptional regulation.

Example 4. ABC and GCB DLBCL subtypes have distinct methylation signatures

Hierarchical clustering of mean methylation for all 160 differentially methylated DMRs and cell lines are shown in Figure 8. The 45 gene signatures associated with ABC or GCB subtypes are shown in Table 2, above, which is reproduced below as Table 7. These results suggest that DNA methylation

contributes to the regulation of characteristic gene expression patterns in ABC and GCB subtypes of DLBCL.

Table 7. Signature Genes associated with ABC or GCB subtypes

absent in melanoma 1 (AIM 1) **(ABC)**	CTD (carboxyl-terminal domain, RNA polymerase II, polypeptide A) phosphatase, subunit 1 (CTDP1) **(GCB)**	heparin sulfate (glucosamine) 3-O-sulfotransferase 1 (HS3ST1) **(GCB)**	phosphodiesterase 3B, cGMP-inhibited (PDE3B) **(GCB)**	synaptogyrin 1 (SYNGR1) **(ABC)**
angiomotin like 1 (AMOTL1) **(ABC)**	CWC27 spliceosome-associated protein homolog (S. cerevisiae) (CWC27) **(GCB)**	immediate early response 5-like (IER5L) **(GCB)**	phosphodiesterase 7B (PDE7B) **(ABC)**	synergin, gamma (SYNGR) **(GCB)**
ankyrin repeat and FYVE domain containing 1 (ANKFY1) **(GCB)**	deafness, autosomal recessive 31 (DFNB31) **(ABC)**	inter-alpha-trypsin inhibitor heavy chain 1 (ITIH1) **(GCB)**	phosphodiesterase 9A (PDE9A) **(GCB)**	synaptophysin-like 1 (SYPL1) **(ABC)**
ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 (ARAP3) **(GCB)**	DLG5 antisense RNA 1 (DLG5-AS1) **(GCB)**	potassium voltage-gated channel, subfamily G, member 2 (KCNG2) **(GCB)**	pleckstrin homology domain containing, family A member 7 (PLEKHA7) **(ABC)**	toll-like receptor 9 (TLR9) **(ABC)**
Rho GTPase activating protein 23 (ARHGAP23) **(ABC)**	dynein, axonemal, heavy chain 6 (DNAH6) **(GCB)**	lysine (K)-specific demethylase 4C (KDM4C) **(ABC)**	PR domain containing 15 (PRDM15) **(GCB)**	tumor necrosis factor receptor superfamily, member 17 (TNFRSF17) **(GCB)**
B-cell CLL/lymphoma 11A (zinc finger protein) (BCL11A) **(ABC)**	early B-cell factor 1 (EBF1) **(ABC)**	KIAA0907 (KIAA0907) **(GCB)**	PR domain containing 2, with ZNF domain (PRDM2) **(GCB)**	troponin C type 2 (fast) (TNNC2) **(GCB)**
BEN domain containing 3 (BEND3) **(ABC)**	enoyl-CoA delta isomerase 1 (ECI1) **(GCB)**	long intergenic non-protein coding RNA 161 (LINC00161) **(GCB)**	patched domain containing 2 (PTCHD2) **(GCB)**	torsin family 4, member A (TOR4A) **(ABC)**
bestrophin 4 (BEST4) **(ABC)**	EGF-like and EMI domain containing 1, pseudogene (EGFEM1P) **(GCB)**	long intergenic non-protein coding RNA 271 (LINC00271) **(ABC)**	prothymosin, alpha (PTMA) **(GCB)**	tyrosylprotein sulfotransferase 1 (TPST1) **(ABC)**

basic helix-loop-helix family, member a15 (BHLHA15) **(GCB)**	ectopic P-granules autophagy protein 5 homolog (C. elegans) (EPG5) **(GCB)**	long intergenic non-protein coding RNA 544 (LINC00544) **(ABC)**	protein tyrosine phosphatase, non-receptor type 18 (brain-derived) (PTPN18) **(ABC)**	tRNA aspartic acid methyltransferase 1 (TRDMT1) **(GCB)**
chromosome 12 open reading frame 42 (C12orf42) **(GCB)**	family with sequence similarity 69, member B (FAM69B) **(GCB)**	Gem (nuclear organelle) associated protein 2 psuedogene 2 (GEMIN2P2; LOC100287063) **(ABC)**	glutaminyl-tRNA synthase (glutamine-hydrolyzing)-like 1 (QRSL1) **(ABC)**	utrophin (UTRN) **(ABC)**
coiled-coil domain containing 42 (CCDC42) **(GCB)**	guanylate binding protein 3 (GBP3) **(GCB)**	lysophosphatidic acid receptor 5 (LPAR5) **(GCB)**	RAP2A, member of RAS oncogene family (RAP2A) **(GCB)**	zinc finger protein 217 (ZNF217) **(GCB)**
cadherin 6, type 2, K-cadherin (fetal kidney) (CDH6) **(GCB)**	golgin A3 (GOLGA3) **(GCB)**	low density lipoprotein receptor-related protein 5 (LRP5) **(ABC)**	rabphilin 3A-like (without C2 domains) (RPH3AL) **(GCB)**	DiGeorge Syndrome Critical Region Gene 6 (DGCR6) **(GCB)**
carnitine palmitoyltransferase 1A (liver) (CPT1A) **(GCB)**	G protein-coupled receptor 25 (GPR25) **(ABC)**	matrix metallopeptidase 15 (membrane-inserted) (MMP15) **(ABC)**	short stature homeobox 2 (SHOX2) **(ABC)**	Microtubule-associated protein 18 (MAP18) **(GCB)**
carboxypeptidase Z (CPZ) **(GCB)**	guanylate cyclase 2D, membrane (retina-specific) (GUCY2D) **(GCB)**	neuraminidase (NA) **(GCB)**	spleen focus forming virus (SFFV) proviral integration oncogene (SPI1) **(ABC)**	Prickle homolog 1 (PRICKLE1) **(GCB)**
casein kinase 1, alpha 1 (CSNK1A1) **(GCB)**	hippocalcin-like 1 (HPCAL1) **(ABC)**	oculocutaneous albinism II (OCA2) **(GCB)**	suppression of tumorigenicity 5 (ST5) **(ABC)**	EPH receptor A8 (EPHA8) **(GCB)**
Tumor necrosis factor, alpha-induced protein 8-like 2 (TNFAIP8L2) **(ABC)**	Ribonuclease T2 (RNASET2) **(ABC)**	Transcription factor CP2 (TFCP2) **(ABC)**	Synaptotagmin-like 1 (SYTL1) **(ABC)**	ATP-binding cassette, subfamily B (MDR/TAP), member 6 (ABCB6) **(ABC)**
A disintegrin and metalloproteinase 19 (ADAM19) **(ABC)**	Paired-like homeodomain 1 (PITX1) **(ABC)**	Ring finger protein 14 (RNF14) **(ABC)**	Spleen focus forming virus (SFFV) proviral integration ocogene (SPI1) **(ABC)**	Carbonic anhydrase VII (CA7) **(GCB)**
Disks large-associated protein 2 (DLGAP2) **(GCB)**	Small Nucleolar RNA, C/D Box 115-43 (SNORD115-43) **(GCB)**	Kallikrein-related peptidase 14 (KLK14) **(GCB)**	Thyroid hormone receptor interactor 10 (TRIP10) **(GCB)**	Olfactory receptor, family 6, subfamily X, member 1 (OR6X1) **(GCB)**
Prepronociceptin (PNOC) **(GCB)**	Sprouty homolog 1 (SPRY1) **(GCB)**	RNA gene affiliated with	Zinc finger, AN1-type domain 2A	Src homology 2 domain containing

		lncRNA class (LOC727710) **(GCB)**	(ZFAND2A) **(GCB)**	adaptor protein B (SHB) **(GCB)**
Spermidine synthase (SRM) **(GCB)**	Solute carrier family 25 (SLC25A3) **(GCB)**	Zinc ribbon domain containing 1 (ZNRD1) **(GCB)**	Histon cluster 3, H2bb (HIST3H2BB) **(GCB)**	Zinc finger protein 77 (ZNF77) **(GCB)**

Example 5. Bcl-6 and IRF4 are identified as central nodes of the epigenetically regulated network of ABC and GCB DLBCL subtypes

This study identified a central Bcl-6/IRF4 network that is epigenetically, differentially regulated in ABC and GCB subtypes of DLBCL. 17 of the 60 unique genes (45 differentially methylated genes identified in the present study combined with the 15 DLBCL classifier genes defined by Wright *et al., supra*) connected via published interactions (Figure 9), including 12 differentially methylated genes are shown in Table 3, above, which is reproduced as Table 8, below (unconnected genes were filtered out). The DLBCL biomarker network identified in the present study features the transcription factors IRF4 and Bcl-6 as regulatory hubs.

Table 8. DLBCL biomarker network

B-cell lymphoma 6 protein (Bcl-6)	LIM domain only 2 (rhombotin-like 1) (LMO2)	Phosphodiesterase 3 (PDE3B)	B-cell linker protein (BLNK)
Interferon regulatory factor 4 (IRF4)	COX Assembly mitochondrial protein 2 homolog (C16orf61)	Histone H2A type 3 (HIST3H2A)	Neprilysin (MME)
Lymphoid-restricted membrane protein (Lrmp)	Interleukin 16 (IL-16)	Transcription factor PU.1 (PU.1)	
G1/S-specific cyclin-D2 (Cyclin D2)	Human inositol 1,4,6-trisphosphate 3-kinase isoform B (IP3KB)	Transcription factor COE1 (EBF)	

Example 6: Assessment of Performance in Formalin Fixed Paraffin Embedded (FFPE) tissue samples

CpG sites for use in a linear methylation signature were selected on the basis of differential methylation in cell lines, as described above. Given a set of such CpG sites, FFPE tissue samples were then scored on the basis of their methylation data, as follows:

- For selected sites, methylation *M*-values were centered and scaled across the tissue panel (to have mean 0, standard deviation 1 in this panel).
- Each subject was assigned a score based on the average of his or her centered and scaled *M-values*. Before averaging, sign of *M*-value was adjusted according to direction of differential methylation observed in cell lines.

All selected CpG sites as well as only those with 2kb of a transcript start site or within an annotated CpG island were both considered. Results are qualitatively similar, suggesting that the shorter set of sites and associated genes produced by the promoter/CpG island-proximal analysis is sufficient (Figure 10).

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, the descriptions and examples should not be construed as limiting the scope of the invention. The disclosures of all patents, patent applications, and scientific references, cited herein are expressly incorporated by reference in their entirety for all purposes as if each patent, patent application, and scientific reference were specifically and individually incorporated by reference.

WHAT IS CLAIMED IS:

1. A method of determining whether a patient having diffuse large B-cell lymphoma (DLBCL) will likely respond to an anti-cancer therapy, the method comprising:

(a) detecting the DNA methylation level of a differentially methylated region (DMR) proximal to at least one gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient,

(b) comparing the DNA methylation level of said DMR to a reference level, wherein a change in the level of methylation of said DMR in the patient sample relative to the reference level identifies a patient who is likely to respond to the therapy; and

(c) informing the patient that they have an increased likelihood of being responsive to the therapy.

2. A method of optimizing therapeutic efficacy of an anti-cancer therapy for a patient having diffuse large B-cell lymphoma (DLBCL), the method comprising:

(a) detecting the DNA methylation level of a DMR proximal to or within at least one gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient,

(b) comparing the DNA methylation level of said DMR to a reference level, wherein a change in the level of methylation of said DMR in the patient sample relative to the reference level identifies a patient who is likely to respond to the therapy; and

(c) providing a recommendation to the patient for a particular anti-cancer therapy.

3. A method of selecting a therapy for a particular patient having diffuse large B-cell lymphoma (DLBCL) in a population of patients being considered for therapy, the method comprising:

(a) detecting the DNA methylation level of a DMR proximal to or within at least one gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient,

(b) comparing the DNA methylation level of said DMR to a reference level, wherein a change in the level of methylation of said DMR in the patient sample relative to the reference level identifies a patient who is likely to respond to the therapy; and

(c) selecting a particular therapy if the patient is identified as likely to respond to the particular therapy and recommending to the patient the particular therapy; or

(d) not selecting a particular therapy if the patient is identified as likely to not respond to the particular therapy and not recommending to the patient the particular therapy.

4. A method of optimizing or modifying a treatment regimen for a patient having diffuse large B-cell lymphoma (DLBCL), the method comprising:

(a) detecting the DNA methylation level of a DMR proximal to or within at least one gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient,

(b) comparing the DNA methylation level of said DMR to a reference level, wherein a change in the level of methylation of said DMR in the patient sample relative to the reference level identifies a patient who may benefit from modification of their treatment regimen; and

(c) providing a recommendation to the patient for said modification.

5. The method of claim 1 or 2, wherein the patient is in a population of patients being tested for responsiveness to at least one treatment for cancer and the reference level is the median level of DNA methylation of said DMR in the population of patients.

6. The method of claim 3, wherein the patient is in a population of patients being considered for therapy and the reference level is the median level of DNA methylation of said DMR of said at least one gene in the population of patients.

7. The method of any one of claims 1-6, wherein the method further comprises detecting the expression level of the at least one gene.

8. The method of any one of claims 1-6, wherein the DNA methylation level is inversely correlated to gene expression level of the at least one gene.

9. The method of any one of claims 1-4, wherein the therapy is an agent selected from the group consisting of: an anti-neoplastic agent, a chemotherapeutic agent, a growth inhibitory agent, a cytotoxic agent, and combinations thereof.

10. The method of any one of claims 1-4, comprising detection of methylation of two or more DMRs proximal to said at least one gene.

11. The method of claim 1, wherein said method is used to determine whether said DLBCL is germinal-center B-cell-like (GCB) or activated B-cell-like (ABC).

12. The method of any one of claims 1-4, wherein the change in level of DNA methylation of said DMR in the patient sample is an increase relative to the reference level.

13. The method of any one of claims 1-4, wherein the change in level of DNA methylation of said DMR in the patient sample is a decrease relative to the reference level.

14. The method of claim 6, wherein expression of the at least one gene in the biological sample obtained from the patient is detected by measuring mRNA.

15. The method of claim 6, wherein expression of the at least one gene in the biological sample obtained from the patient is detected by measuring plasma protein levels.

16. The method of any one of claims 1-4, further comprising detecting the DNA methylation level of a DMR of at least a second of said genes in the biological sample from the patient.

17. The method of claim 16, further comprising detecting the DNA methylation level of a DMR of at least a third of said genes in the biological sample from the patient.

18. The method of claim 17, further comprising detecting the DNA methylation level of a DMR of at least a fourth of said genes in the biological sample from the patient.

19. The method of any one of claims 1-18, further comprising administering an anti-cancer therapy to said patient if it is determined that the patient may benefit from said anti-cancer therapy.

20. A method for prognosing disease outcome in a patient having DLBCL, the method comprising the steps of:

(a) detecting the DNA methylation level of a DMR of at least one gene listed in Table 1, 2, or 3 in a biological sample obtained from the patient; and

(b) comparing the DNA methylation level of said DMR to a reference level, wherein a change in the level of methylation of said DMR in the patient sample relative to the reference level is prognostic of disease outcome.

21. The method of claim 20, wherein said method is used to determine whether said DLBCL is germinal-center B-cell-like (GCB) or activated B-cell-like (ABC).

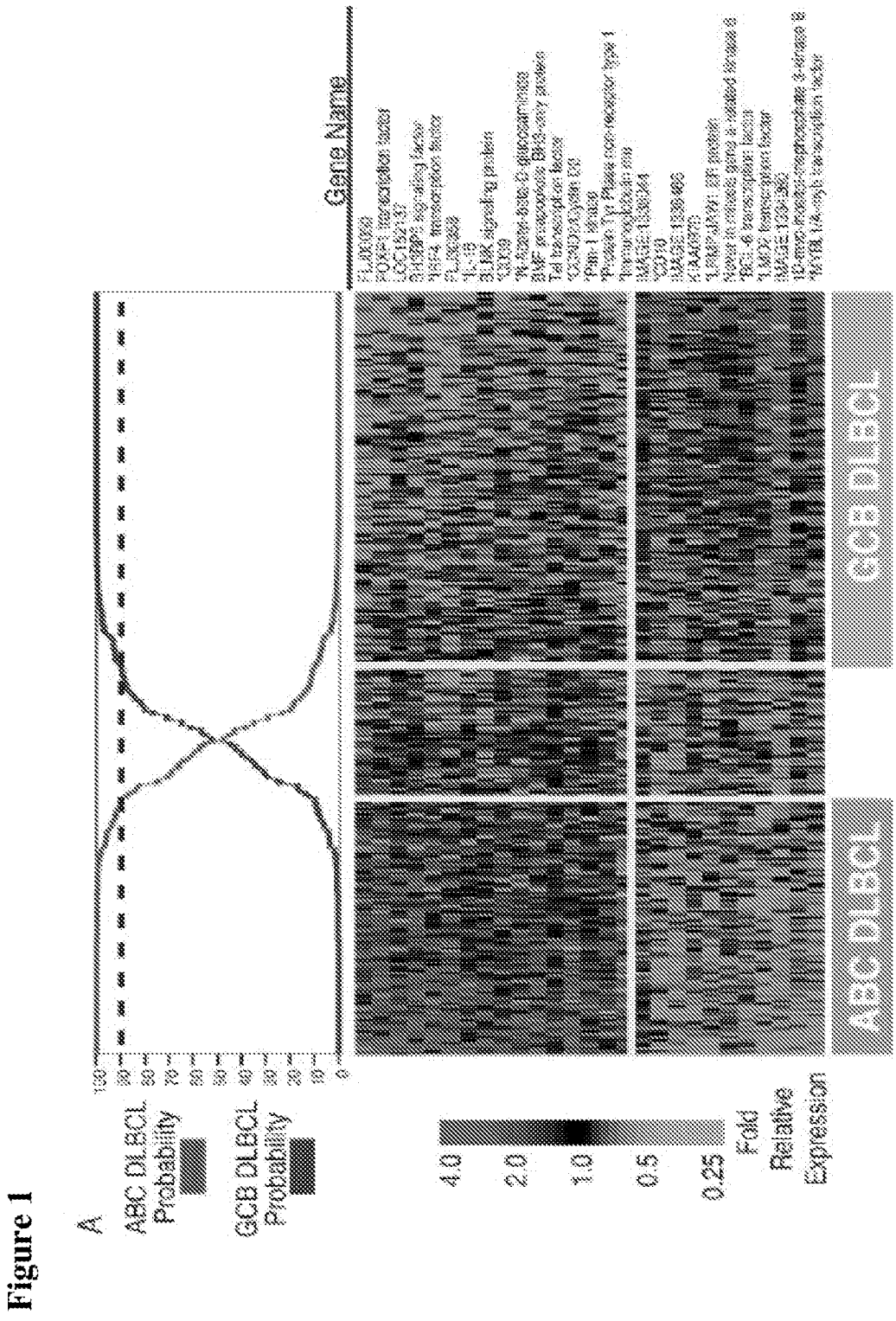
22. The method of claim 20, further comprising administering an anti-cancer therapy to the patient if the patient is identified as having DLBCL.

23. A kit for determining whether a patient may benefit from an anti-cancer therapy, the kit comprising:

(a) reagents for determining the DNA methylation level of a DMR of at least one of the genes listed in Table 1, 2, or 3; and optionally

(b) instructions for use of said reagents to determine the DNA methylation level of a DMR of at least one of said genes, wherein a change in the level of said DNA methylation relative to a reference level indicates that the patient may benefit from an anti-cancer therapy.

Figure 1


A
ABC DLBCL Probability
GCB DLBCL Probability
Gene Name
ABC DLBCL
GCB DLBCL
Fold Relative Expression
4.0
2.0
1.0
0.5
0.25

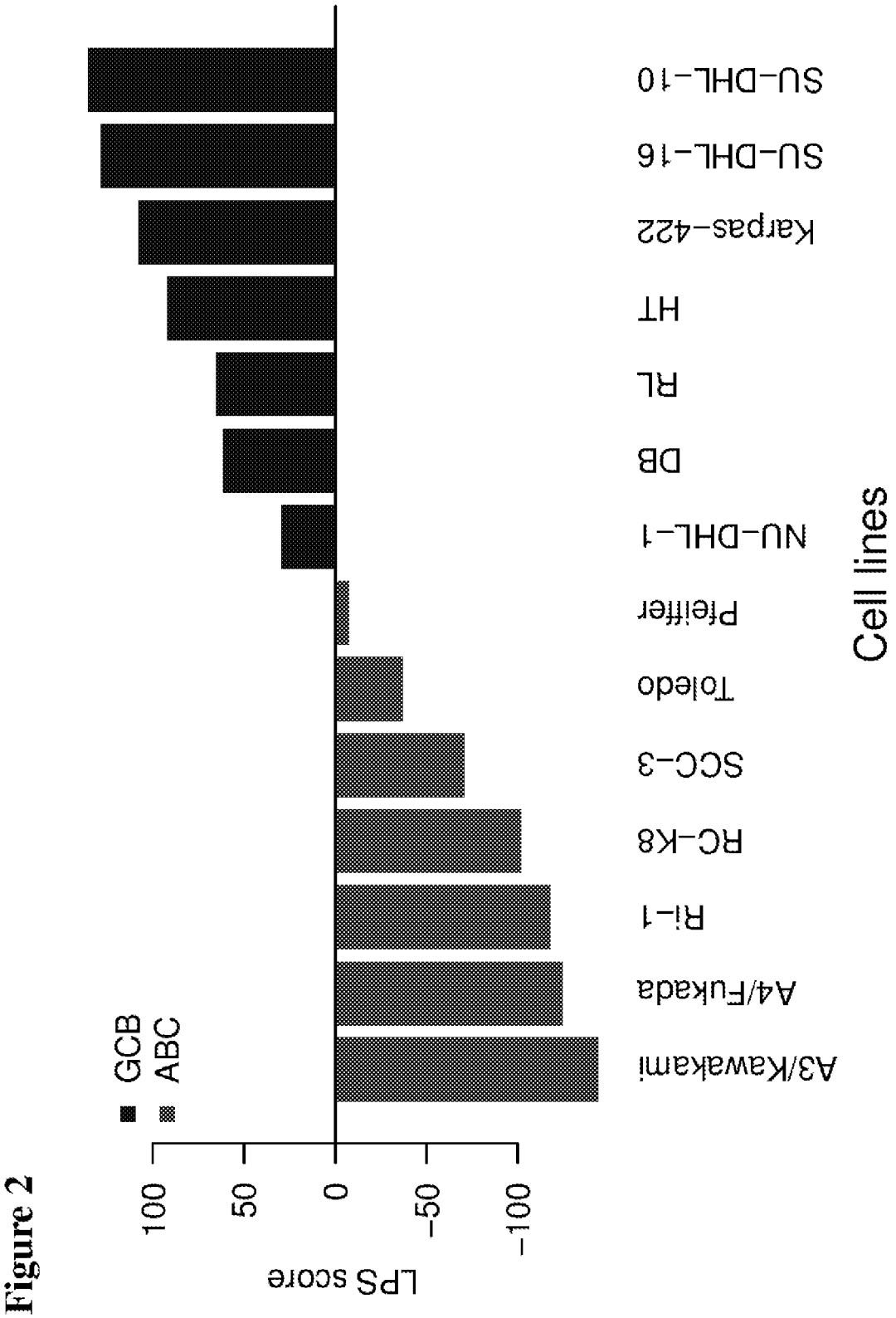

GCB
ABC
100
50
0
−50
−100
LPS score
A3/Kawakami
A4/Fukada
Ri−1
RC−K8
SCC−3
Toledo
Pfeiffer
NU−DHL−1
DB
RL
HT
Karpas−422
SU−DHL−16
SU−DHL−10
Cell lines

Figure 2

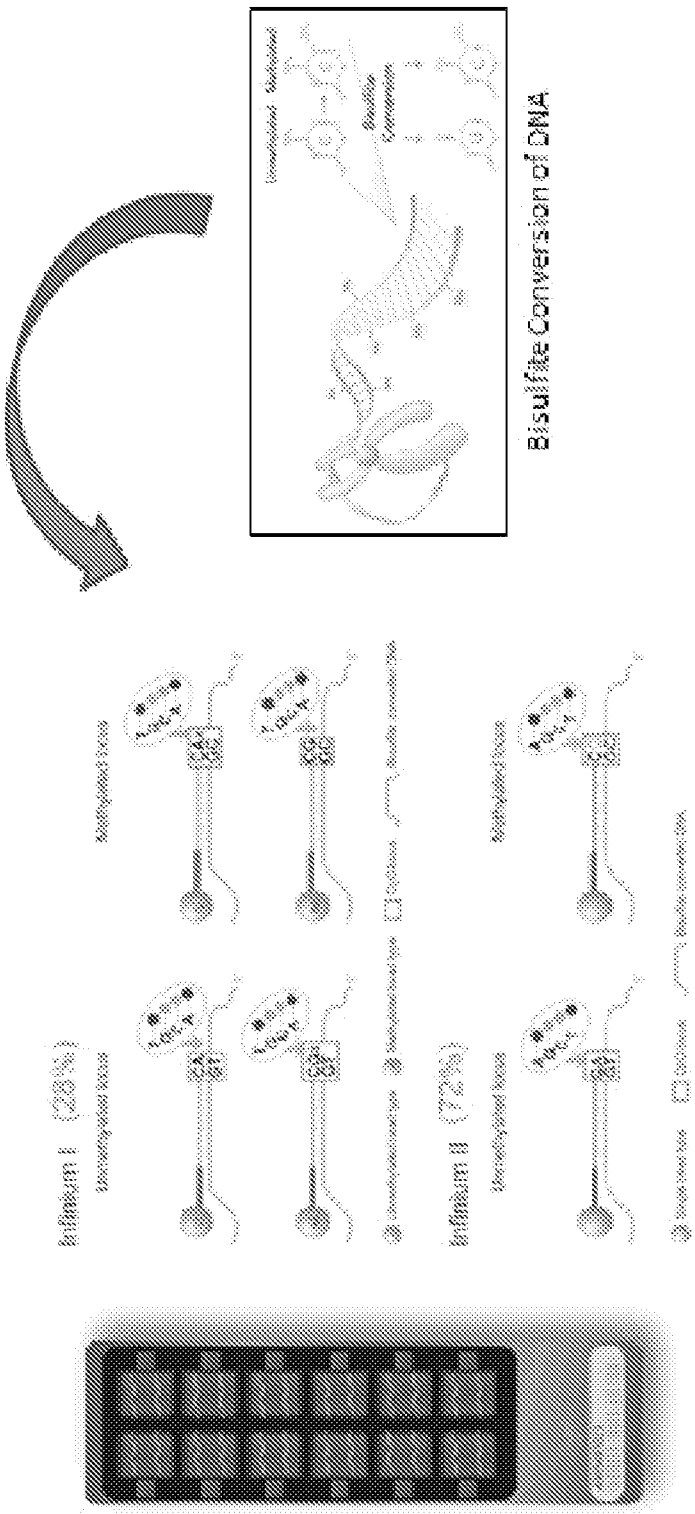


Figure 3

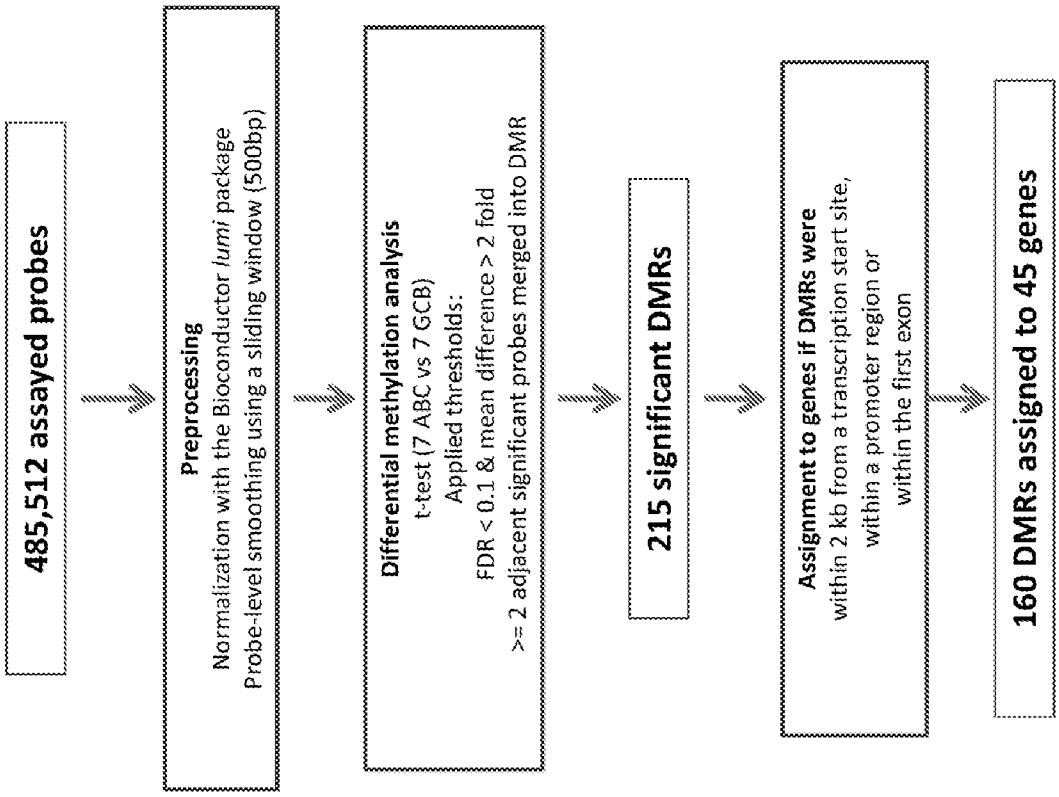
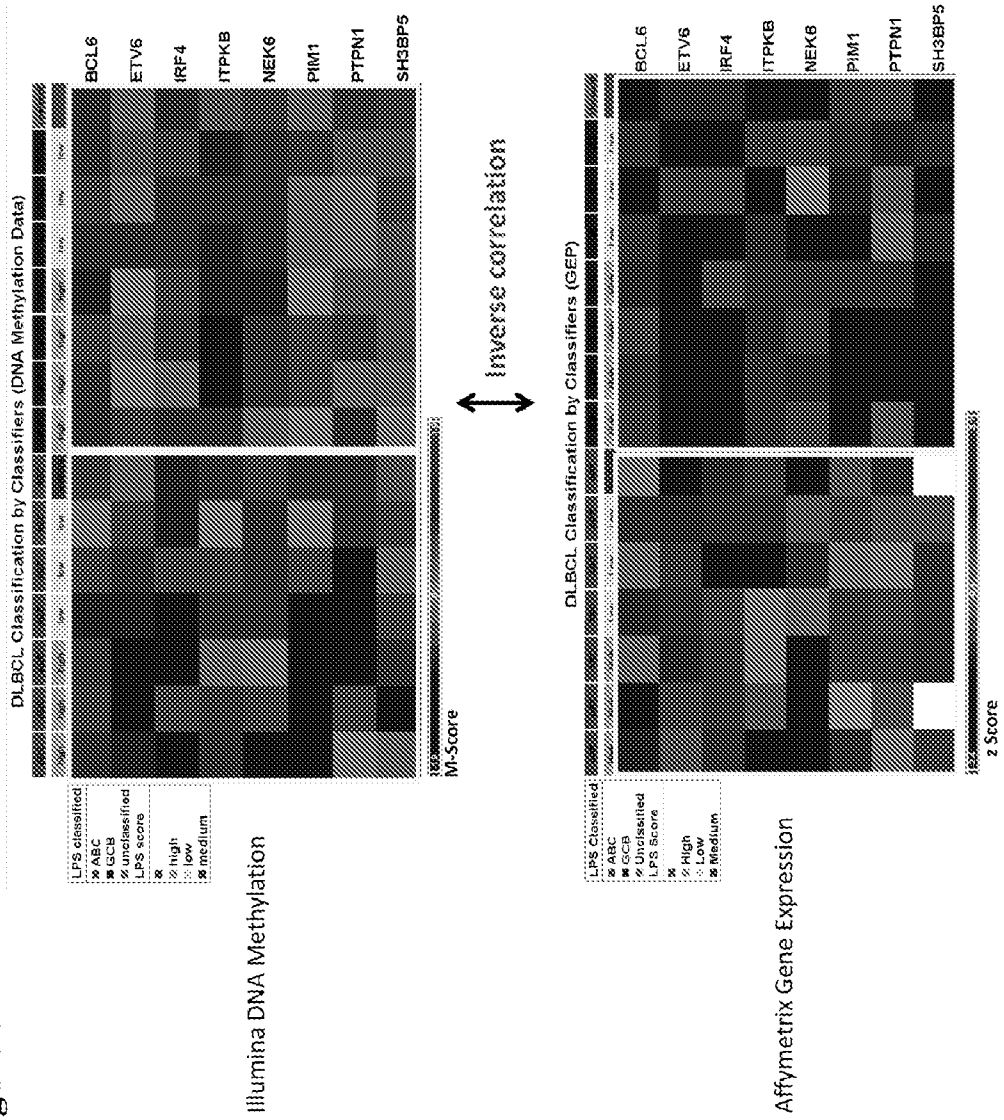

485,512 assayed probes
Preprocessing
Normalization with the Bioconductor lumi package
Probe-level smoothing using a sliding window (500bp)
Differential methylation analysis
t-test (7 ABC vs 7 GCB)
Applied thresholds:
FDR < 0.1 & mean difference > 2 fold
>= 2 adjacent significant probes merged into DMR
215 significant DMRs
Assignment to genes if DMRs were
within 2 kb from a transcription start site,
within a promoter region or
within the first exon
160 DMRs assigned to 45 genes

Figure 4

Figure 5


DLBCL Classification by Classifiers (DNA Methylation Data)
LPS classified
ABC
GCB
unclassified
LPS score
high
low
medium
BCL6
ETV6
IRF4
ITPKB
NEK6
PIM1
PTPN1
SH3BP5
M-Score
Illumina DNA Methylation
inverse correlation
DLBCL Classification by Classifiers (GEP)
LPS Classified
ABC
GCB
Unclassified
LPS Score
High
Low
Medium
BCL6
ETV6
IRF4
ITPKB
NEK6
PIM1
PTPN1
SH3BP5
Z Score
Affymetrix Gene Expression

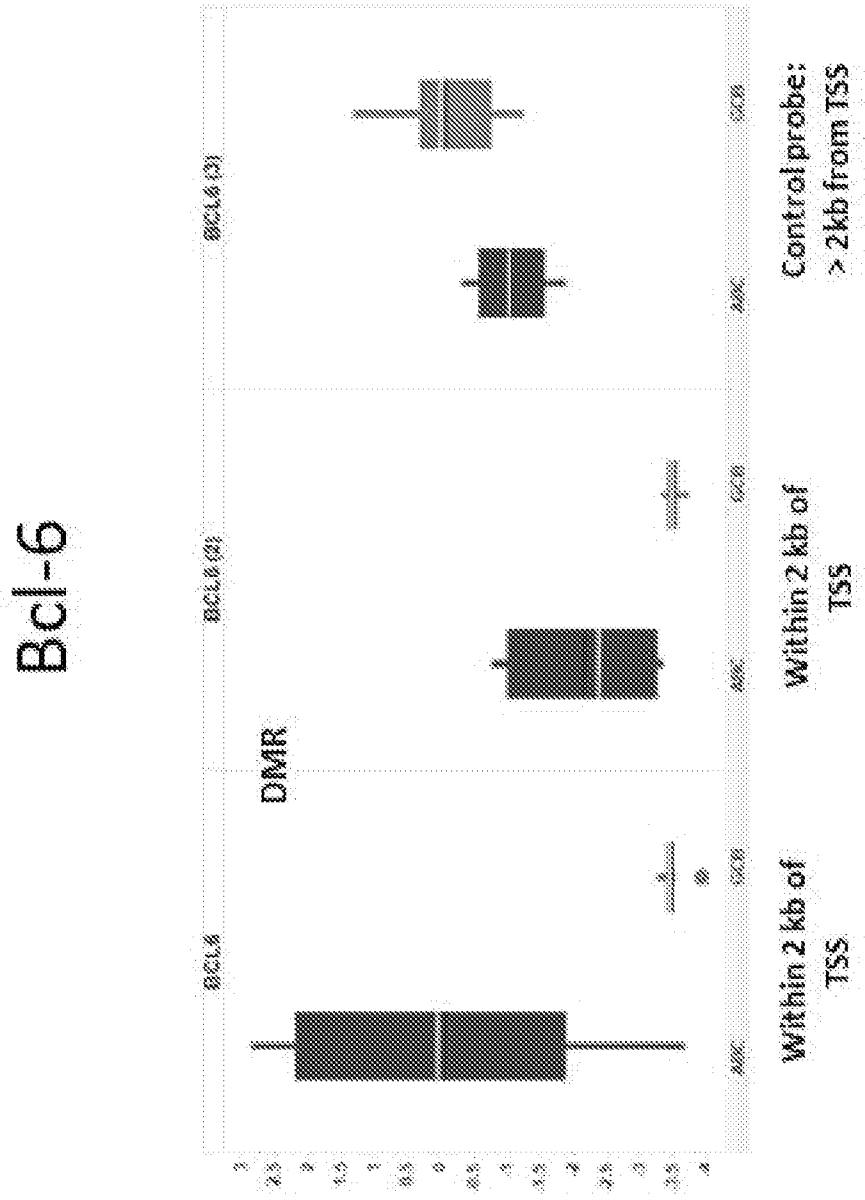

Bcl-6
DMR
Within 2 kb of TSS
Within 2 kb of TSS
Control probe: > 2kb from TSS

Figure 6

Figure 6 (continued)

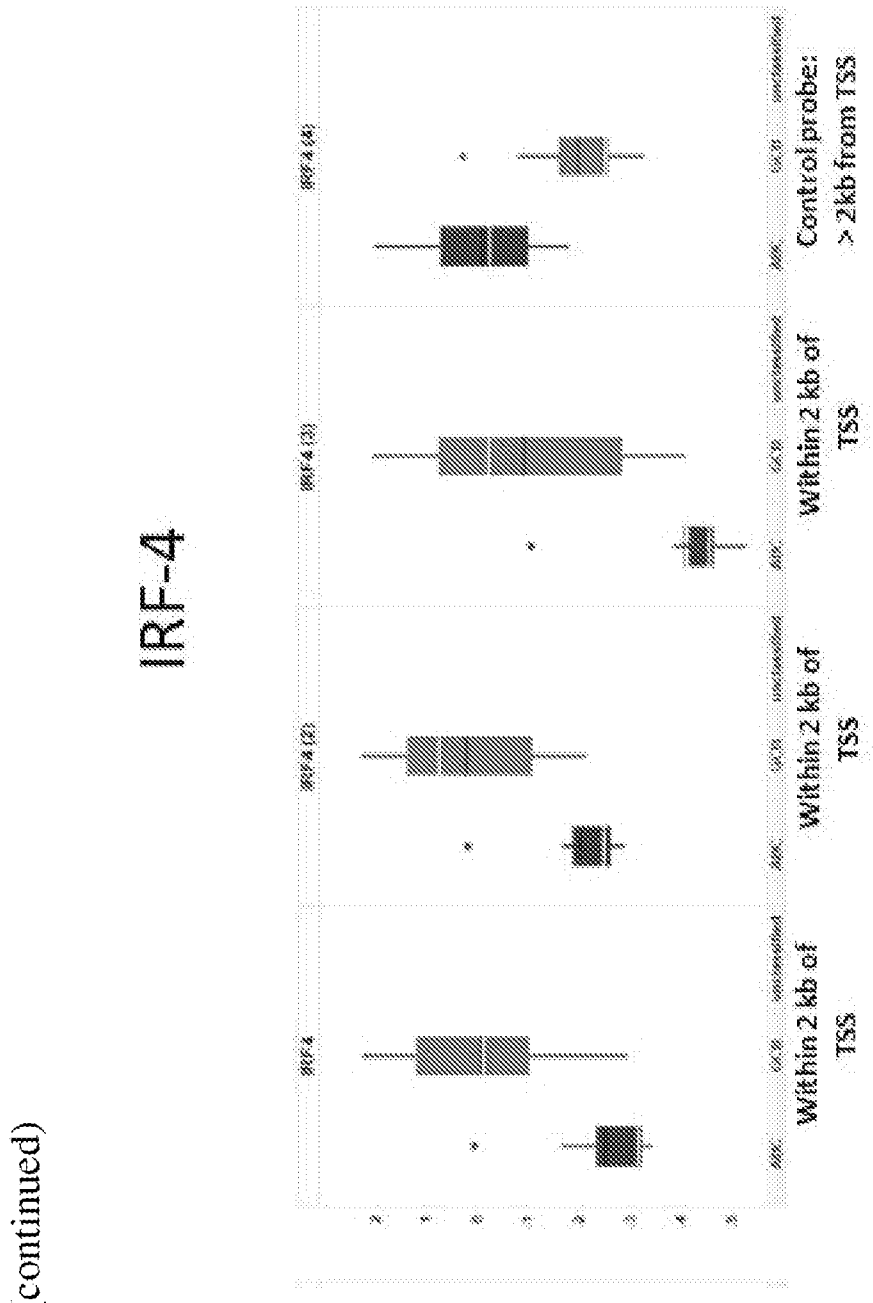
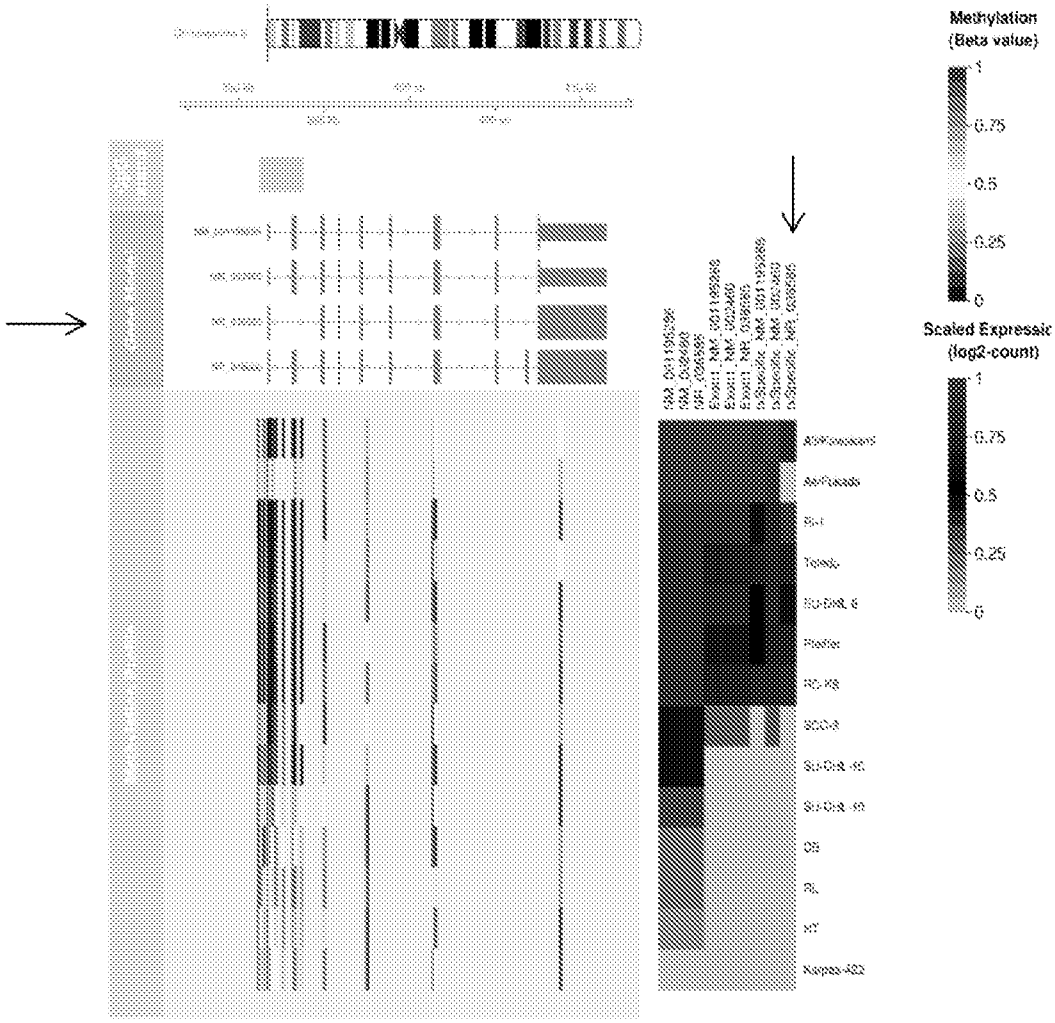
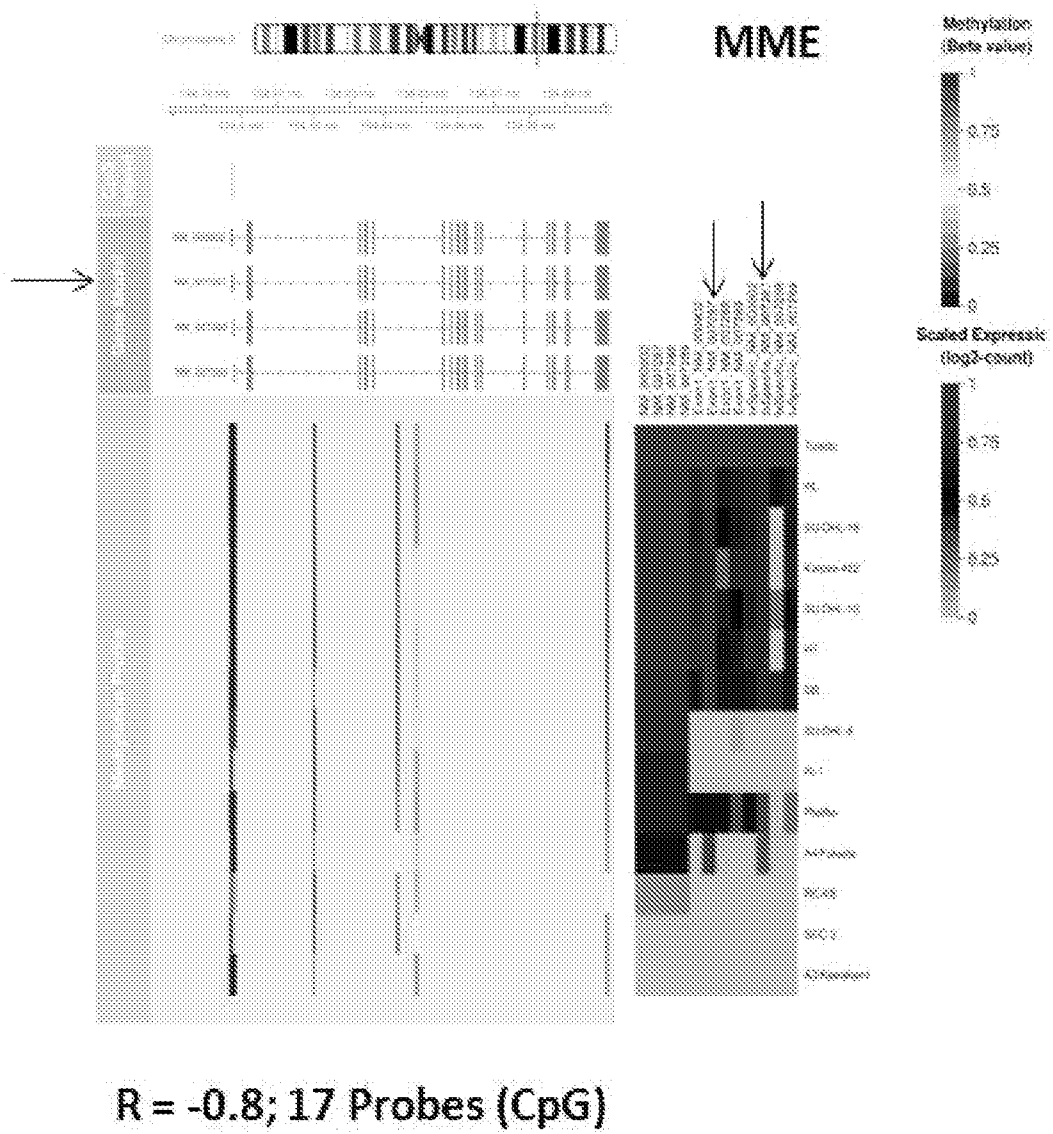


Figure 7



R = -0.8; 16 Probes (CpG)

Figure 7 (continued)



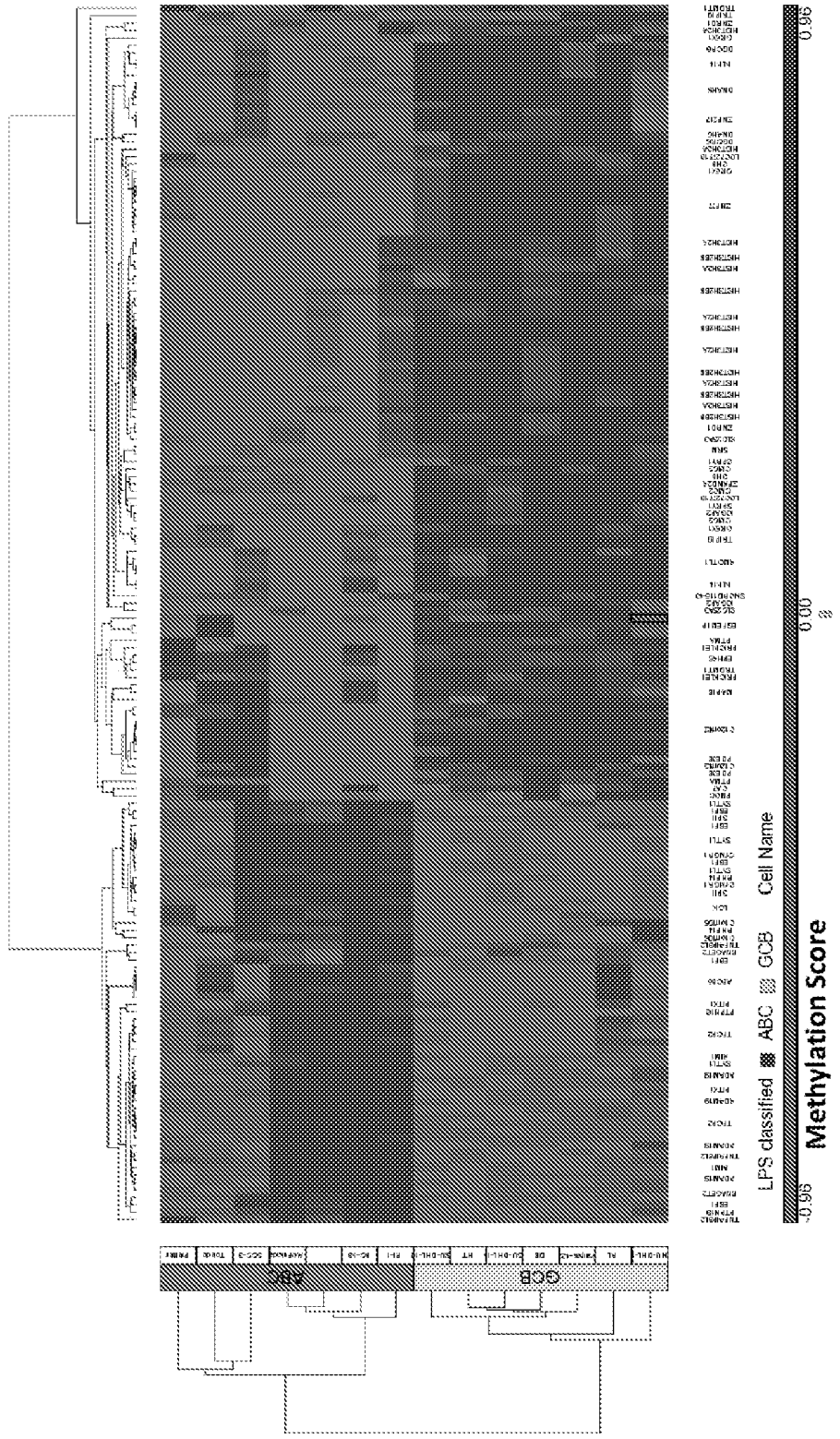

LPS classified
ABC
GCB
Cell Name
Methylation Score
0.96
0.00
-0.96
ABC
GCB

Figure 8

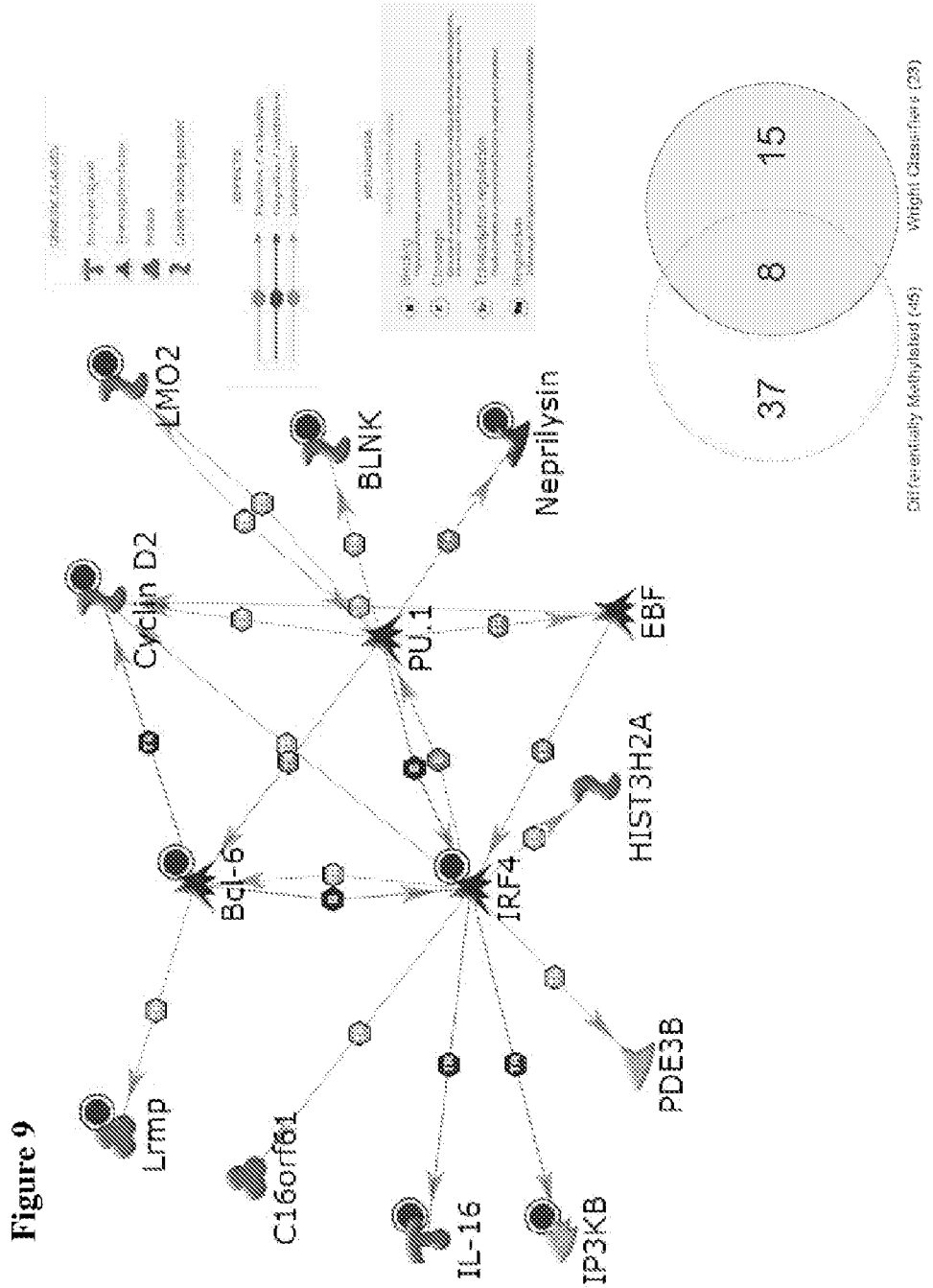

LMO2
BLNK
Neprilysin
Cyclin D2
PU.1
EBF
HIST3H2A
Bcl-6
IRF4
Lmp
C16orf61
IL-16
IP3KB
PDE3B
37
8
15
Differentially Methylated (45)
Wright Classifiers (23)

Figure 9

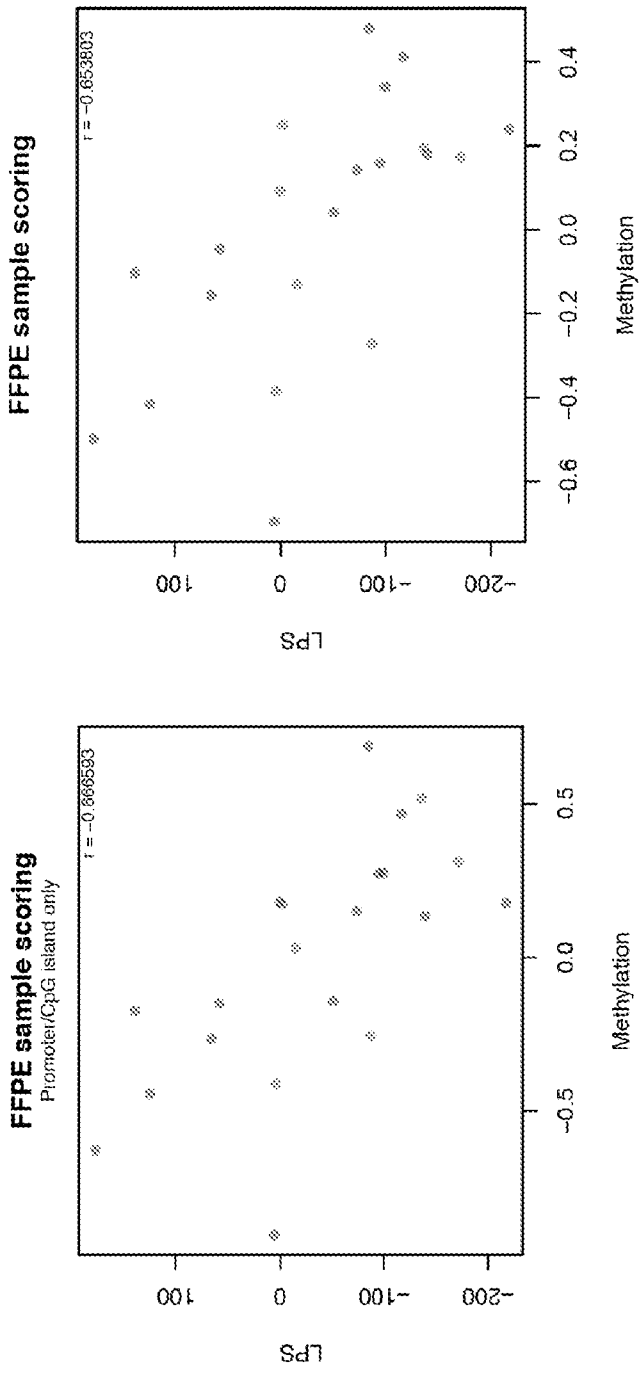

FFPE sample scoring
r = −0.653803
LPS
100
0
−100
−200
Methylation
−0.6
−0.4
−0.2
0.0
0.2
0.4
FFPE sample scoring
Promoter/CpG island only
r = −0.666593
LPS
100
0
−100
−200
Methylation
−0.5
0.0
0.5

Figure 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/024331

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
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

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.
Y	WO 02/27019 A1 (UNIV JOHNS HOPKINS MED [US]; BAYLIN STEPHEN B [US]; HERMAN JAMES G [US) 4 April 2002 (2002-04-04) claims 1, 10; figure 2; example 7 -----	1-23
Y	A. Y. LAI ET AL: "DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas", MOLECULAR CELL, vol. 13, no. 2, 30 August 2010 (2010-08-30), pages 291-1950, XP055198175, ISSN: 1097-2765, DOI: 10.1016/S1097-2765(04)00029-2 page 1940, paragraph 1; figure 2 -----	1-23

-/--

[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
26 June 2015	14/09/2015

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	Santagati, Fabio

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SUBHAJYOTI DE ET AL: "Aberration in DNA Methylation in B-Cell Lymphomas Has a Complex Origin and Increases with Disease Severity", PLOS GENETICS, vol. 9, no. 1, 10 January 2013 (2013-01-10), page e1003137, XP055198186, DOI: 10.1371/journal.pgen.1003137 figures 2, 7	1-23
X	MARINA BIBIKOVA ET AL: "High density DNA methylation array with single CpG site resolution", GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 98, no. 4, 26 July 2011 (2011-07-26), pages 288-295, XP028304083, ISSN: 0888-7543, DOI: 10.1016/J.YGENO.2011.07.007 [retrieved on 2011-08-02] the whole document	23
A	R. SHAKNOVICH ET AL: "DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma", BLOOD, vol. 116, no. 20, 7 July 2010 (2010-07-07) , pages e81-e89, XP055158085, ISSN: 0006-4971, DOI: 10.1182/blood-2010-05-285320 figures 4, 5	1-23
A	B L PIKE ET AL: "DNA methylation profiles in diffuse large B-cell lymphoma and their relationship to gene expression status", LEUKEMIA, vol. 22, no. 5, 21 February 2008 (2008-02-21), pages 1035-1043, XP055158206, ISSN: 0887-6924, DOI: 10.1038/leu.2008.18 table 1	1-23
A	XIAO-MING WANG ET AL: "Identification and functional relevance of de novo DNA methylation in cancerous B-cell populations", JOURNAL OF CELLULAR BIOCHEMISTRY, 1 March 2010 (2010-03-01), pages n/a-n/a, XP055197771, ISSN: 0730-2312, DOI: 10.1002/jcb.22461 table 1	1-23

-/--

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHAMBWE NYASHA ET AL: "Regular Article Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes", BLOOD, vol. 123, no. 11, 13 March 2014 (2014-03-13), pages 1699-1708, XP055197741, DOI: 10.1182/blood-2013-07- -----	1-23
A	WO 2013/082105 A1 (EINSTEIN COLL MED [US]; UNIV NEBRASKA [US]; YE BIHUI H [US]; CHAN WING) 6 June 2013 (2013-06-06) -----	1-23
A	I. S. LOSSOS: "Expression of a single gene, BCL-6, strongly predicts survival in patients with diffuse large B-cell lymphoma", BLOOD, vol. 98, no. 4, 15 August 2001 (2001-08-15), pages 945-951, XP055029462, ISSN: 0006-4971, DOI: 10.1182/blood.V98.4.945 -----	1-23

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/024331

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.:

1-23(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.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. claims: 1-23(partially)

A method of determining whether a patient having diffuse large B-cell lymphoma (DLBCL) will likely respond to an anti-cancer therapy or of f optimizing therapeutic efficacy of an anti-cancer therapy for a patient having diffuse large B-cell lymphoma (DLBCL) or of selecting a therapy for a particular patient having diffuse large B-cell lymphoma (DLBCL) in a population of patients being considered for therapy or of optimizing or modifying a treatment regimen for a patient having diffuse large B-cell lymphoma (DLBCL), the method comprising: detecting the DNA methylation level of a differentially methylated region (DMR) proximal to BCL6 in a biological sample obtained from the patient and comparing the DNA methylation level of said DMR to a reference level, wherein a change in the level of methylation of said DMR in the patient sample relative to the reference level identifies a patient who is likely to respond to the therapy. A method for prognosing disease outcome in a patient having DLBCL, the method comprising the steps of: (a) detecting the DNA methylation level of a DMR of BCL6 in a biological sample obtained from the patient; and (b) comparing the DNA methylation level of said DMR to a reference level, wherein a change in the level of methylation of said DMR in the patient sample relative to the reference level is prognostic of disease outcome. A kit for determining whether a patient may benefit from an anti-cancer therapy, the kit comprising:(a) reagents for determining the DNA methylation level of a DMR of BCL6.

2-120. claims: 1-23(partially)

Same as invention 1 but wherein the gene is
inv. 2 = ETV6
inv. 3 = IRF4
.....
inv. 9 = AIM1
.....
inv. 108 = ZNF77
inv. 109 = LRMP
.....
inv. 120 = MME
in the order of Tables 1, 2 and 3.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2015/024331

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0227019 A1	04-04-2002	AU 9651101 A	08-04-2002
		CA 2422890 A1	04-04-2002
		EP 1328656 A1	23-07-2003
		JP 2004509643 A	02-04-2004
		US 2002127572 A1	12-09-2002
		US 2004265887 A1	30-12-2004
		US 2010112593 A1	06-05-2010
		WO 0227019 A1	04-04-2002
WO 2013082105 A1	06-06-2013	US 2014329714 A1	06-11-2014
		WO 2013082105 A1	06-06-2013