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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 December 2018 (20.12.2018)

W!P O | PCT

(10) International Publication Number
WO 2018/231589 Al

(51) International Patent Classification:
C12Q 1/6886 (2018.01)

(21) International Application Number:
PCT/US20 18/036084

(22) International Filing Date:
05 June 2018 (05.06.2018)

(25) Filing Language: English

(26) Publication Langiuage: English

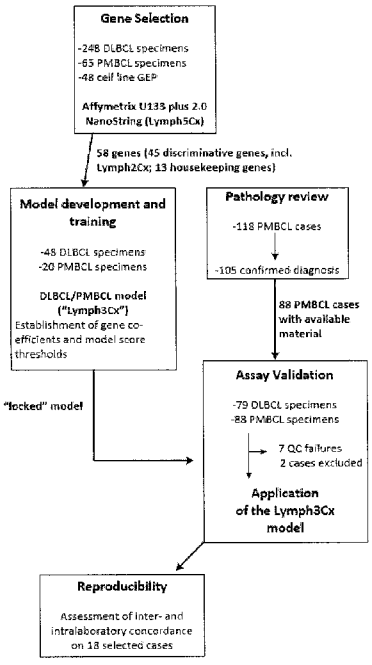
(30) Priority Data:
62/5 19,728 14 June 2017 (14.06.2017) US

(71) Applicants *(for all designated States except US):* **THE UNITED STATES OF AMERICA, AS REPRESENT¬ ED BY THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES** [US/US]; Of- fice of Technology Transfer, National Institutes of Health, 601 1 Executive Boulevard, Suite 325, MSC 7660, Bethesda, Maryland 20892-7660 (US). **BRITISH COLUMBIA CANCER AGENCY BRANCH** [CA/CA]; 600 West 10th Avenue, Vancouver, British Co- lumbia V5Z 4E6 (CA). **MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH** [US/US]; 20 First Street SW, Rochester, Minneso- ta 55905 (US). **JULIUS-MAXIMILIANS-UNIVERSI¬ TY OF WÜRZBURG** [DE/DE]; Sanderring 2, 97070 Wurzburg (DE). **BOARD OF REGENTS OF THE UNIVERSITY OF NEBRASKA** [US/US]; Varner Hall, 3835 Holdege Street, Lincoln, Nebraska 68583-0745 (US). **OSLO UNIVERSITY HOSPITAL HF** [NO/NO]; Tarn- bygget, Kirkeveien 166, 0450 Oslo (NO). **UNIVERSI¬ TY HEALTH NETWORK** [CA/CA]; 190 Elizabeth Street, R. Fraser Elliott Building - Room 1S-417, Toron- to, Ontario M5G 2C4 (CA). **HOSPITAL CLINIC DE BARCELONA** [ES/ES]; Villarroel 170, 08036 Barcelona (ES). **UNIVERSITAT DE BARCELONA** [ES/ES]; Center de Patents - Centre de Patents - UB, Baldiri Reixac, 4, 08028 Barcelona (ES). **INSTITUT D'INVESTIGACIONS BIOMÈDIQUES AUGUST PI I SUNYER (IDIBAPS)** [ES/ES]; C. Rossello, 149-153, 08036 Barcelona (ES). **ROBERT BOSCH HEALTH¬ CARE SYSTEMS, INC.** [US/US]; 2400 Geng Road, Suite 200, Palo Alto, California 94303 (US). **OREGON HEALTH & SCIENCE UNIVERSITY** [US/US]; 3181 SW Sam Jackson Park Rd., Portland, Oregon 97239 (US). **CITY OF HOPE** [US/US]; 1500 East Duarte Road, Duarte, California 91010-3000 (US). **THE CLEVELAND CLINIC FOUNDATION** [US/US]; 9500 Euclid Avenue, Cleveland, Ohio 44195 (US).

WO 2018/231589 A1

(54) Title: METHOD FOR DETERMINING LYMPHOMA TYPE

Figure 1



(57) Abstract: In embodiments of the invention, the invention provides a method for distinguishing between lymphoma types based on gene expression measurements. In em- bodiments, the invention distinguishes between PMBCL and DLBCL based on gene ex- pression signatures, and can further distinguish between DLBCL subtypes.

[Continued on nextpage]

(72) Inventors; and

(71) Applicants *(for US only):* **STAUDT, Louis M.** [US/US]; 9407 Spruce Tree Circle, Bethesda, Maryland 20814 (US). **STEEDL, Christian** [CA/CA]; 1038 Lillooet Road, North Vancouver, British Columbia V7J 2H5 (CA). **MOTTOK, Anja** [CA/CA]; 3869 Ontario Street, Vancouver, British Columbia V5V 0C8 (CA). **WRIGHT, George W.** [US/US]; 4008 London Terrace, Rockville, Maryland 20853 (US). **SCOTT, David William** [CA/CA]; 2355 Larch Street, Vancouver, British Columbia V6K 0A2 (CA). **RIMSZA, Lisa M.** [US/US]; 6818 E. Montgomery Road, Scottsdale, Arizona 85266 (US).

(72) Inventors; and

(71) Applicants: ROSENWALD, Andreas [DE/DE]; Kronbergstrae 14, 97078 Wurzburg (DE). **GASCOYNE, Randy** [CA/CA]; # 113-245 W. 15th Street, North Vancouver, British Columbia V7M 1S3 (CA). **GREINER, Timothy** [US/US]; 7345 Stafford Drive, Council Bluffs, Iowa 51503 (US). **WEISENBURGER, Dennis** [US/US]; 130 Martindale Way, Glendora, California 91741 (US). **SMELAND, Erlend B.** [NO/NO]; Lochenveien 5A, 0286 Oslo (NO). **DELABIE, Jan** [BE/CA]; 32 Gothic Avenue, Unit 4, Toronto, Ontario M6P 2V9 (CA). **CAMPO GUERRI, Elias** [ES/ES]; Avenida Roma 10, 13-2, 08036 Barcelona (ES). **OTT, German** [DE/DE]; Grnwiesenstrae 47, 74321 Bietigheim-Bissingen (DE). **BRAZIEL, Rita** [US/US]; 541 Ashdown Circle, West Linn, Oregon 97068 (US). **JAFFE, Elaine S.** [US/US]; 424 River Bend Road, Great Falls, Virginia 22066 (US). **FU, Kai** [US/US]; 3878 S. 175th Ave., Omaha, Nebraska 68130 (US). **CHAN, Wing C.** [US/US]; 3336 Villa Mesa Road, Pasadena, California 91107 (US). **SONG, Joo** [US/US]; 1500 East Duarte Road, Duarte, California 91010 (US). **COOK, James R.** [US/US]; 2860 Attleboro Road, Shaker Heights, Ohio 44120 (US).

(74) Agent: SPENNER, Jonathan M. et al; Leydig, Voit & Mayer, Ltd., Two Prudential Plaza, Suite 4900, 180 North Stetson Avenue, Chicago, Illinois 60601 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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).

Published:

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

— *with sequence listing part of description (Rule 5.2(a))*

METHOD FOR DETERMINING LYMPHOMA TYPE

CROSS-REFERENCE TO A RELATED APPLICATION

[0001] This patent application claims the benefit of U.S. Provisional Patent Application No. 62/519,728, filed June 14, 2017, which is incorporated by reference herein in its entirely.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] This invention was made with Government support under Grant Number U01CA157581 awarded by the Strategic Partnering to Evaluate Cancer Signatures (SPECS II). The Government has certain rights in this invention.

INCORPORATION-BY-REFERENCE OF MATERIAL SUBMITTED ELECTRONICALLY

[0003] Incorporated by reference in its entirety herein is a computer-readable nucleotide/amino acid sequence listing submitted concurrently herewith and identified as follows: one 16,028 Byte ASCII (Text) file named "739259_ST25.txt" dated June 5, 2018.

BACKGROUND OF THE INVENTION

[0004] Primary mediastinal large B cell lymphoma (PMBCL), known as the "third subtype" of diffuse large B cell lymphoma (DLBCL), represents 2-4% of all B cell non-Hodgkin Lymphoma. PMBCL is recognized as a distinct clinico-pathological entity in the current World Health Organization classification. The pathogenic hallmarks of PMBCL include (1) activation of the JAK-STAT pathway, (2) activation of the NF-κB pathway, and (3) overexpression of specific immune checkpoint molecules, such as programmed death ligands (PDL) 1 and 2. The current classification of PMBCL from DLBCL is based on clinico-pathologic consensus. Presently the diagnosis of PMBCL relies on the integration of clinical characteristics and clinical/pathological presentation because a reliable distinction from DLBCL solely based on morphological or immunophenotypic features can be challenging for pathologists for various reasons. Reasons for the difficulty in diagnosing PMBCL include, among others, (1) that clinico-pathologic consensus is not always well

instituted, and (2) gene expression profiling developed in fresh-frozen tissue is not routinely available in clinical practice.

[0005] Gene expression profiling studies provide evidence that PMBCL can be distinguished from DLBCL on a molecular level and supported a strong relationship between PMBCL and classical Hodgkin lymphoma. However, because these studies were performed using snap-frozen tissue, the molecular classification of PMBCL has not penetrated into clinical practice.

[0006] Therefore, there is an unmet need for a gene expression-based molecular classifier using formalin-fixed, paraffin-embedded (FFPE) samples, with the ability to distinguish PMBCL from DLBCL. There is also an unmet need for additional methods of classifying DLBCL subtype tumors. The present invention provides such methods.

BRIEF SUMMARY OF THE INVENTION

[0007] In an embodiment, the present invention provides a method for classifying the lymphoma type of a sample, which method comprises providing a formalin-fixed and paraffin-embedded (FFPE) lymphoma sample from the subject, isolating RNA from the sample, obtaining gene expression data from the RNA, wherein the gene expression data comprises signal values that represent expression levels for each gene of Table 1, and determining a predictor score from the gene expression data, wherein the tumor predictors score is calculated by

$$S = \sum_{k=0}^{58} a_i x_i$$

wherein a_i is the model coefficient value for gene i, as listed in Table 1, column D for determining whether the sample is PMBCL or DLBCL and as listed in Table 1 column E for determining whether a sample is ABC DLBCL or GCB DLBCL, and x_i is the log_2 transformed expression signal value for gene i; and when the coefficient values in column D of Table 1 are used, classifying the lymphoma as DLBCL when S is less than -57.95, PMBCL when S is greater than -23.57, or uncertain DLBCL/PMBCL when S is between -57.95 and -23.57; and when the coefficient values in column E are used, classifying

the lymphoma as GCB DLBCL when S is less than 798.5, ABC DLBCL when S is greater than 1324.5, or uncertain ABC/GCB DLBCL when S is between 798.5 and 1324.5.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Figure 1 presents a schematic overview of the studies described in the Example.

[0009] Figure 2 is a dot plot of the genes used in selecting Nanostring probe set for the novel Lymph3Cx assay. The large circles represent the higher expressing genes in PMBCL (bottom) and DLBCL (top). The Y axis represents the Lymph5Cx standardized mean difference and the X axis represents the Affymetrix standardized mean difference. The genes with higher expression in DLBCL are CARDl 1, BTG2, PRKCBB, FAM159A, PRDX2, and BANK1. The genes that are higher expressed in PMBLC are MAL, HOMER2, CCL17, FSCN1, TFP12, TMOD1, SLAMF1, RCL1, PRR6, IL13RA1, MST1R, LIMA1, TRAF1, PDCD1LG2, PTGIR, NFIL3, BATF3, RHOF, SNX11, QSOX1, AUH, IFIH1, MOBKL2C, andNECAP2.

[0010] Figure 3 shows a dot plot of the pathology review (Y axis) against the PMBCL model score (X axis) from the Lymph3Cx assay.

[0011] Figure 4 is a graph showing comparison of the linear predictor scores (LPS) for 66 cases which were run using the published Lymph2Cx assay and the Lymph3Cx assay. Dotted lines represent the thresholds previously defined for COO-assignment.

[0012] Figure 5 is a graph showing comparison of Lymph3Cx scores for selected cases of the validation cohort from two independent laboratories (BC Cancer Agency (BCCA) and Mayo Clinic). Dotted lines represent the defined thresholds to discriminate PMBCL from DLBCL using the Lymph3Cx assay. Of note, no case changed subtype assignment between the different laboratories.

DETAILED DESCRIPTION OF THE INVENTION

[0013] In an embodiment, the present invention provides a method for classifying the lymphoma type of a sample, which method comprises providing a formalin-fixed and paraffin-embedded (FFPE) lymphoma sample from the subject, isolating RNA from the sample, obtaining gene expression data from the RNA, wherein the gene expression data comprises signal values that represent expression levels for each gene of Table 1, and determining a tumor predictor score from the gene expression data, wherein the tumor predictors score is calculated by

$$S = \sum_{k=0}^{58} a_i x_i$$

wherein a_i is the model coefficient value for gene i, as listed in Table 1, column D for determining whether the sample is PMBCL or DLBCL and as listed in Table 1 column E for determining whether a sample is ABC DLBCL or GCB DLBCL, and x_i is the log_2 transformed expression signal value for gene i; and when the coefficient values in column D of Table 1 are used, classifying the lymphoma as DLBCL when S is less than -57.95, PMBCL when S is greater than -23.57, or uncertain DLBCL/PMBCL when S is between -57.95 and -23.57; and when the coefficient values in column E are used, classifying the lymphoma as GCB DLBCL when S is less than 798.5, ABC DLBCL when S is greater than 1324.5, or uncertain ABC/GCB DLBCL when S is between 798.5 and 1324.5.

[0014] In an embodiment, the method further comprises determining the probability that the sample is PMBCL or ABC DLBCL, wherein the probability is determined by

(a) determining the probability that the sample is PMBCL by calculating the probability score of

$$P(PMBCL) = \frac{\varphi(S_{PMBCL/DLBCL};\mu_{PMBCL},\sigma_{PMBCL})}{\varphi(S;\mu_{PMBCL},\sigma_{PMBCL}) + \phi(S;\mu_{DLBCL},\sigma_{DLBCL})},$$

wherein $S_{PMBCL/DLBCL}$ is the tumor predictor score; $\mu_{PMBCL} > \mu_{DLBCL}$ and σ_{PMBCL}, σ_{DLBCL} represent the mean and standard deviations of the PMBCL and DLBCL subtypes as indicated in Table 2 provided herein;

(b) determining the probability that the sample is ABC DLBCL by calculating the probability score of

$$P(ABC) = \frac{\varphi(S_{ABC/GCB};\mu_{ABC},\sigma_{ABC})}{\varphi(S;\mu_{ABC},\sigma_{ABC}) + \varphi(S;\mu_{GCB},\sigma_{GCB})},$$

wherein $S_{ABC/GCB}$ is the tumor predictor score; $\mu_{ABC>HGCB}$ and σ_{ABC}, σ_{GCB} represent the mean and standard deviations of the ABC and GCB subtypes as indicated in Table 2 herein; and

wherein $\phi(\chi;\mu,\sigma)$ is the standard normal density calculated by

$$\varphi(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\cdot\right).$$

[0015] In an embodiment, a score of P(PMBCL) ≥ 0.9 indicates that the sample is PMBCL, regardless of the P(ABC) score. In an embodiment, a score of P(PMBCL) ≤ 0.1 and a score of P(ABC) ≤ 0.1 indicates that the tumor is GCB DLBCL. In an embodiment, wherein a score of P(PMBCL) ≤ 0.1 and a score of P(ABC) ≥ 0.9 indicates that the tumor is ABC DLBCL. In an embodiment, a score of P(PMBCL) ≤ 0.1 and a score of P(ABC) greater than 0.1 and less than 0.9 indicates that the tumor is an unclassified DLBCL.

[0016] In an embodiment, the method encompasses genes which have a coefficient of zero. In an embodiment, the equation

$$S = \sum_{k=0}^{58} a_i x_i$$

may be utilized for determining a tumor predictor score for a sample, taking into account genes having a coefficient of zero when classifying PMBCL v. DLBCL and GCB DLBCL v. ABC DLBCL. In an embodiment the genes with a coefficient of zero need not be taken into account in the equation.

[0017] In an embodiment, the RNA gene expression data is obtained using a NanoString Technologies® nCounter® assay.

[0018] The inventive method comprises isolating sufficient RNA from a human subject, e.g., from a sample from a subject, such as from fresh tissue, a snap-frozen sample from a subject, or a formalin-fixed and paraffin-embedded (FFPE) sample from a subject. The sample may be a biopsy sample. As understood by one of ordinary skill in the art, the phrase "a snap-frozen sample from a subject" means that a sample is first taken from a subject and afterwards snap-frozen, and the phrase "obtaining or providing a formalin-fixed and paraffin-embedded (FFPE) sample from the subject" means that a sample is first taken from a subject and afterwards fixed with formalin and embedded in paraffin.

[0019] The gene expression product, e.g., the main mRNA species, is RNA, for example, total cellular mRNA. The RNA gene expression product may be obtained from the subject in any suitable manner. For example, one or more samples may be obtained from a patient that

has been diagnosed as having a non-Hodgkin lymphoma, and the samples can be formalin-fixed and paraffin-embedded using protocols that are known in the art or are commercially available (see, e.g., Keirnan, J. (ed.), *Histological and Histochemical Methods: Theory and Practice, 4th edition,* Cold Spring Harbor Laboratory Press (2008), incorporated herein by reference). The RNA can be extracted from an FFPE sample using methods that are known in the art or are commercially available (see, e.g., Huang et al., *Cancer Epidemiol Biomarkers Prev., 19:* 973-977 (2010), incorporated herein by reference; QIAGEN AllPREP DNA/RNA FFPE Kit (Qiagen, Venlo, Netherlands)). The digital gene expression profile may be obtained from archived FFPE tissue.

[0020] The inventive method further comprises obtaining gene expression data from the isolated RNA, wherein the gene expression data comprises data for genes in a gene expression signature. The phrase "gene expression data" as used herein refers to information regarding the relative or absolute level of expression of RNA species. "Gene expression data" may be acquired for an individual cell, or for a group of cells such as a tumor or biopsy sample.

[0021] Any effective method of quantifying the expression of at least one gene, gene set, or group of gene sets may be used to acquire gene expression data for use in the invention. For example, gene expression data may be measured or estimated using one or more microarrays, where, e.g., the microarrays produce a signal value for each gene and the signal values of all genes in a gene expression signature may comprise the gene expression data. See, for example, the methods as described in the Example below.

[0022] Nucleic acid microarrays generally comprise nucleic acid probes derived from individual genes and placed in an ordered array on a support. This support may be, for example, a glass slide, a nylon membrane, or a silicon wafer. Gene expression patterns in a sample are obtained by hybridizing the microarray with the RNA gene expression product from the sample. The RNA gene expression product from a sample is labeled with a radioactive, fluorescent, or other label to allow for detection. Following hybridization, the microarray is washed, and hybridization of RNA gene expression product to each nucleic acid probe on the microarray is detected and quantified using a detection device such as a phosphoimager or scanning confocal microscope.

[0023] The microarray may be a cDNA microarray or an oligonucleotide microarray. cDNA arrays consist of hundreds or thousands of cDNA probes immobilized on a solid support, and are described in detail in, e.g., Southern et al., *Genomics, 13:* 1008-1017 (1992);

Southern et al., *Nucl. Acids. Res., 22:* 1368-1373 (1994); Gress et al., *Oncogene, 13:* 1819-1830 (1996); Pietu et al., *Genome Res., 6:* 492-503 (1996); Schena et al., *Science, 270:* 467-470 (1995); DeRisi et al., *Nat. Genet., 14:* 457-460 (1996); Schena et al, *Proc. Natl. Acad. Sci. USA, 93:* 10614-10619 (1996); Shalon et al., *Genome Res., 6:* 639-645 (1996); DeRisi et al., *Science, 278:* 680-686 (1997); Heller et al., *Proc. Natl. Acad. Sci. USA, 94:* 2150-2155 (1997); and Lashkari et al., *Proc. Natl. Acad. Sci. USA, 94:* 13057-13062 (1997), each incorporated herein by reference. Oligonucleotide arrays differ from cDNA arrays in that the probes are 20- to 25-mer oligonucleotides. Oligonucleotide arrays are generally produced by *in situ* oligonucleotide synthesis in conjunction with photolithographic masking techniques (see, e.g., Pease et al., *Proc. Natl. Acad. Sci. USA, 91:* 5022-5026 (1994); Lipshutz et al., *Biotechniques, 19:* 442-447 (1995); Chee et al., *Science, 274:* 610-14 (1996); Lockhart et al., *Nat. Biotechnol, 14:* 1675-1680 (1996); and Wodicka et al., *Nat. Biotechnol., 15:* 1359-1367 (1997), each incorporated herein by reference). The solid support for oligonucleotide arrays is typically a glass or silicon surface.

[0024] Methods and techniques applicable to array synthesis and use have been described in, for example, U.S. Patents 5,143,854, 5,242,974, 5,252,743, 5,324,633, 5,384,261, 5,424,186, 5,445,934, 5,451,683, 5,482,867, 5,491,074, 5,527,681, 5,550,215, 5,571,639, 5,578,832, 5,593,839, 5,599,695, 5,624,711, 5,631,734, 5,795,716, 5,831,070, 5,837,832, 5,856,101, 5,858,659, 5,936,324, 5,968,740, 5,974,164, 5,981,185, 5,981,956, 6,025,601, 6,033,860, 6,040,193, 6,090,555, and 6,410,229, and U.S. Patent Application Publication 2003/010441 1, each incorporated herein by reference. Techniques for the synthesis of microarrays using mechanical synthesis methods are described in, for example, U.S. Patents. 5,384,261 and 6,040,193, each incorporated herein by reference. Microarrays may be nucleic acids on beads, gels, polymeric surfaces, fibers such as fiber optics, glass or any other appropriate substrate (see, e.g., U.S. Patents 5,708,153, 5,770,358, 5,789,162, 5,800,992, and 6,040,193, each incorporated herein by reference).

[0025] Microarrays may be packaged in such a manner as to allow for diagnostic use, or they may be an all-inclusive device (see, e.g., U.S. Patents 5,856,174 and 5,922,591, each incorporated herein by reference). Microarrays directed to a variety of purposes are commercially available from, e.g., Affymetrix (Affymetrix, Santa Clara, CA, USA).

[0026] In an embodiment, the signal value comprises digital counts. Gene expression data can be obtained and analyzed using a variety of digital methods known in the art, such as, for example, serial analysis of gene expression (SAGE) (see, e.g., Velculescu et al.,

Science, 270(5235): 484-487 (1995)), SuperSAGE (see e.g., Matsumura et al., *Proc. Natl. Acad. Sci. USA, 100* (26): 15718-15723 (2003)), digital northern analysis (see, e.g., Cao et al., *Breast Cancer Research,* 10: R91 (2008)), and RNA-seq (see, e.g., Mortazavi et al. *Nat Methods,* 5(7):62 1-628 (2008)), each article incorporated herein by reference. In an embodiment, the RNA gene expression data is obtained using a NanoString Technologies® nCounter® assay available from NanoString Technologies®, Inc. (Seattle, WA, USA)

[0027] The NanoString platform is used for subtyping lymphomas (see e.g., Scott et al., *J. Clin. Oncol.* 31(6): 692-700 (2013), Scott et al., *Blood* 123(8): 1214-1217 (2014), Scott and Mottok et al., *J. Clin. Oncol.* 33(26): 2848-2856 (2015), Kridel and Mottok et al., *Blood* 126(18): 21118-2127 (2015), Scott and Abrisqueta et al., *J. Clin. Oncol.* 35(15): 1668-1677, Rosenwald et al., *J. Exper. Med.* 198(6): 851-862 (2003), and Savage et al., *Blood* 102(12): 3871-3879 (2003), each incorporated herein by reference.

[0028] The nCounter® assay can detect the expression of up to 800 genes in a single reaction with high sensitivity and linearity across a broad range of expression levels. The nCounter ® assay is based on direct digital detection of mRNA molecules of interest using target-specific, color-coded probe pairs, and does not require the conversion of mRNA to cDNA by reverse transcription or the amplification of the resulting cDNA by PCR. Each target gene of interest is detected using a pair of reporter and capture probes carrying 35- to 50-nucleotide target-specific sequences. In addition, each reporter probe carries a unique color code at the 5' end that enables the molecular barcoding of the genes of interest, while the capture probes all carry a biotin label at the 3' end that provides a molecular handle for attachment of target genes to facilitate downstream digital detection. After solution-phase hybridization between target mRNA and reporter-capture probe pairs, excess probes are removed and the probe/target complexes are aligned and immobilized in an nCounter ® cartridge, which is then placed in a digital analyzer for image acquisition and data processing. Hundreds of thousands of color codes designating mRNA targets of interest are directly imaged on the surface of the cartridge. The expression level of a gene is measured by counting the number of times the color-coded barcode for that gene is detected, and the barcode counts are then tabulated. NanoString Technologies® technology and analysis of digital gene expression data is described in detail in, e.g., Kulkarni, M. M., "Digital Multiplexed Gene Expression Analysis Using the NanoString Technologies ® nCounter ® System," *Current Protocols in Molecular Biology. 94:* 25B.10.1-25B.10.17 (2011), incorporated herein by reference; Geiss et al., *Nature Biotechnology, 26:* 317-325 (2008),

incorporated herein by reference; and U.S. Patent 7,919,237, incorporated herein by reference.

[0029] The term "gene expression signature" as used herein refers to a group of coordinately expressed genes. The genes making up a particular signature may be expressed in a specific cell lineage, stage of differentiation, or during a particular biological response. The genes may reflect biological aspects of the tumors in which they are expressed, such as the cell of origin of the cancer, the nature of the non-malignant cells in the sample, and the oncogenic mechanisms responsible for the cancer (see, e.g., Shaffer et al., *Immunity, 15:* 375-385 (2001), incorporated herein by reference). Examples of gene expression signatures include lymph node (see Shaffer et al., *supra),* proliferation (see, e.g., Rosenwald et al., *New Engl. J. Med., 346:* 1937-1947 (2002), incorporated herein by reference), MHC class II, ABC DLBCL high, B-cell differentiation, T-cell, macrophage, immune response-1, immune response-2, and germinal center B cell.

[0030] The 58 genes of a gene expression signature of the present invention are shown in Table 1 with their respective coefficient values. When gene expression is detected using RNA, the sequences detected are the RNA sequences of the DNA target sequences, where the DNA sequences have thymine replaced with uracil.

Table 1

	A	B	C	D	E	
	Gene Symbol	Accession	Gene Type	PMBCL/ DLBCL Coefficient	ABC/GCB Coefficient	Gene Target Sequence
1	ASB13	NM_024701.3	GCB gene	0	-66.35	GGACACGTAGGCGGTACCACTAAGGTTTTGGTAATGAGC CATTCAAACCGACAGCAGTGTGAAGGTGTGTCAAGGTGT ATATTCTCGTGGCTCGGCATTC (SEQ ID NO: 1)
2	AUH	NM_001698.2	PMBCL gene	3.16	0	GGTGGTCTTGAACTGGCTTTAGCCTGTGATATACGAGTAG CAGCTTCCTCTGCAAAAATGGGCCTGGTTGAAACAAAAT TGGCGATTATTCCTGGTGGAG (SEQ ID NO: 2)
3	BANK1	NM_00 1083907. 1	DLBCL gene	-0.8	0	GGCAAATGAAATGGAAGGGGAAGGAAAACAGAATGGAT CAGGCATGGAGACCAAACACAGCCCACTAGAGGTTGGCA GTGAGAGTTCTGAAGACCAGTAT (SEQ ID NO: 3)
4	BATF3	NM_01 8664.2	PMBCL gene	2.44	0	CTGCTGTTATGCAGAGCCATTTCCTCTAGAATTTGGATAA TAAAGATGCTTATTGTCTCTCCCTTCTCCAGTTCTGGGAA TTTACAGGCACAATACACTT (SEQ ID NO: 4)
5	BTG2	NM_006763.2	DLBCL gene	-2.23	0	TGCTCTCCTTGGGATGATGGCTGGCTAGTCAGCCTTGCAT GTATTCCTTGGCTGAATGGGAGAGTGCCCCATGTTCTGCA AGACTACTTGGTATTCTTGT (SEQ ID NO: 5)
6	CARD 11	NM_03241 5.2	DLBCL gene	-3.12	0	TTGAAAATCGGCCCAAGAAGGAGCAGGTTCTGGAACTGG AGCGGGAGAATGAAATGCTGAAGACCAAAAACCAGGAG CTGCAGTCCATCATCCAGGCCGG (SEQ ID NO: 6)
7	CCDC50	NM_174908.3	ABC Gene	0	40.54	AAACACTTTCCAGAGTTCCCTGCAACCCGTGCTTATGCAG ATAGTTACTATTATGAAGATGGAGGAATGAAGCCAAGAG TGATGAAAGAAGCTGTATCTA (SEQ ID NO: 7)
8	CCL17	NM_002987.2	PMBCL gene	0.98	0	GCCTGGAGTACTTCAAGGGAGCCATTCCCCTTAGAAAGC TGAAGACGTGGTACCAGACATCTGAGGACTGCTCCAGGG ATGCCATCGTTTTTGTAACTGT (SEQ ID NO: 8)
9	CREB3L2	NM_19407 1.2	ABC Gene	0	65.79	ATGCCTGAGGGGATCAGGCTTTTCTACTCCAGGCAAACCT GCCCCATCTTGTCGCTTTTAGGACCTCCCACAACCTGGTT CCCCACACATCCATAGTTCT (SEQ ID NO: 9)

WO 2018/231589 PCT/US2018/036084

	A	B	C	D	E	
	Gene Symbol	Accession	Gene Type	PMBCL/ DLBCL Coefficient	ABC/GCB Coefficient	Gene Target Sequence
10	CYB5R2	NM_016229.3	ABC Gene	0	67.72	CCATGTCTTAGGGCTTCCTGTAGGTAACTATGTCCAGCTC TTGGCAAAAATCGATAATGAATTGGTGGTCAGGGCTTAC ACCCCTGTCTCCAGTGATGAT (SEQ ID NO: 10)
11	DNAJB12	NM_017626.4	Housekeeping	-2.97	-4.14	TTTCTTCCATGTTTTAGAAAATGAGGCCTGTTTGGGGAAG GTACCCTGGTGATGTTTTTGCTAGACATTAGCTGTAGCTG ACAGCATAAGGAGAGTCGCA (SEQ ID NO: 11)
12	FAM159A	NM_001042693.1	DLBCL gene	-1.78	0	ACAGCTACATGTGGTGGCTCAGCATTGGCGCTCTCATAGG CCTGTCCGTAGCAGCAGTGGTTCTTCTCGCCTTCATTGTT ACCGCCTGTGTGCTCTGCTA (SEQ ID NO: 12)
13	FSCN1	NM_003088.2	PMBCL gene	1.01	0	CCCTGCCCTCTTGTCTGCCACGGGGCGAGTCTGGCACCTC TTTCTTCTGACCTCAGACGGCTCTGAGCCTTATTTCTCTGG AAGCGGCTAAGGGACGGTT (SEQ ID NO: 13)
14	GIT2	NM_057169.2	Housekeeping	-2.97	-4.14	CAGATTTTACAGGCTGAATTATTGGCAGTATATGGAGCA GACCCAGGCACACAGGATTCTAGTGGGAAAACTCCCGTT GATTATGCAAGGCAAGGAGGGC (SEQ ID NO: 14)
15	GSK3B	NM_002093.2	Housekeeping	-2.97	-4.14	ACTGATTATACCTCTAGTATAGATGTATGGTCTGCTGGCT GTGTGTTGGCTGAGCTGTTACTAGGACAACCAATATTTCC AGGGGATAGTGGTGTGGATC (SEQ ID NO: 15)
16	HOMER2	NM_004839.2	PMBCL gene	0.89	0	TGGAAGACAAAGTGCGTTCCTTAAAGACAGACATTGAGG AGAGCAAATACCGACAGCGCCACCTGAAGGTGGAGTTGA AGAGCTTCCTGGAGGTGCTGGA (SEQ ID NO: 16)
17	IFIH1	NM_022168.2	PMBCL gene	3.35	0	GCTTGGGAGAACCCTCTCCCTTCTCTGAGAAAGAAAGAT GTCGAATGGGTATTCCACAGACGAGAATTTCCGCTATCTC ATCTCGTGCTTCAGGGCCAGG (SEQ ID NO: 17)
18	IK	NM_006083.3	Housekeeping	-2.97	-4.14	GTCCAAATTCTTGGGTGGTGACATGGAACACACCCATTTG GTGAAAGGCTTGGATTTTGCTCTGCTTCAAAAGGTACGAG CTGAGATTGCCAGCAAAGAG (SEQ ID NO: 18)
19	IL13RA1	NM_001560.2	PMBCL gene	1.63	0	TCTGCACTGGAAGAAGTACGACATCTATGAGAAGCAAAC CAAGGAGGAAACCGACTCTGTAGTGCTGATAGAAAACCT GAAGAAAGCCTCTCAGTGATGG (SEQ ID NO: 19)

	A	B	C	D	E	
	Gene Symbol	Accession	Gene Type	PMBCL/ DLBCL Coefficient	ABC/GCB Coefficient	Gene Target Sequence
20	IRF4	NM_002460. 1	ABC Gene	0	7 1.92	GGGCACTGTTTAAAGGAAAGTTCCGAGAAGGCATCGACA AGCCGGACCCTCCCACCTGGAAGACGCGCCTGCGGTGCG CTTTGAACAAGAGCAATGACTT (SEQ ID NO: 20)
2 1	ISY1	NM_02070 1.2	Housekeeping	-2.97	-4. 14	GGCAAAACATCAGTGTCTGTGGGTAGTTGGAATCTTCAGT TCCTGTGAGCGTCGGCGTCTTCTGGGCCTGTGGAGTTTCT TGGACAGGGGCCGCGGGGCT (SEQ ID NO: 2 1)
22	ITPKB	NM_00222 1.3	GCB gene	0	-67.78	GTGGCCTCCTGGCATCATTTGTTATTGCCTCTGAAACAAG CCTTACTGCCTGGAGGGCTTAGATTCCTGCTTCTCCAATG TAGTGTGGGTATCTTGTAGG (SEQ ID NO: 22)
23	LIMA1	NM_001 113547. 1	PMBCL gene	1.8	0	AACTACATCCTGAACTCGACGTCCTGAGGTATAATACAA CAGAGCACTTTTTGAGGCAATTGAAAAACCAACCTACAC TCTTCGGTGCTTAGAGAGATCT (SEQ ID NO: 23)
24	LIMD1	NM_01 4240.2	ABC Gene	0	6 1.92	AAGGCAAGTCTCAGGAACCCATGCAGGTACATCGCTTGC ACCTGTTTTTAGCTTATTTAATGACGGGCTTTTGGGAAGA GCTGCCCGCATACTGAGAGAC (SEQ ID NO: 24)
25	MAL	NM_002371.2	PMBCL gene	0.54	0	GCCTTCGCGTCCGGGTTGGGAGCTTGCTGTGTCTAACCTC CAACTGCTGTGCTGTCTGCTAGGGTCACCTCCTGTTTGTG AAAGGGGACCTTCTTGTTCG (SEQ ID NO: 25)
26	MAML3	NM_0 1871 7.4	GCB gene	0	-58.59	TGGAAGCCATCAACAATTTGCCCAGTAACATGCCACTGC CTTCAGCTTCTCCTCTTCACCAACTTGACCTGAAACCTTCT TTGCCCTTGCAGAACAGTGG (SEQ ID NO: 26)
27	MME	NM_000902.2	GCB gene	0	-56.55	GGATTGTAGGTGCAAGCTGTCCAGAGAAAAGAGTCCTTG TTCCAGCCCTATTCTGCCACTCCTGACAGGGTGACCTTGG GTATTTGCAATATTCCTTTGG (SEQ ID NO: 27)
28	MOBKL2C	NM_145279.4	PMBCL gene	3.37	0	TTCTCTTACCCAGAGATGCCCATGAGCTGACATTTTACTC ATCCCTCTGCCTCCAAGAAGGCCTGTATTATACGTGTCCT CCTGGGGGTTGGAGATGATC (SEQ ID NO: 28)
29	MST1R	NM_002447. 1	PMBCL gene	1.69	0	CCACTTTGGAGTTGTCTACCACGGAGAATACATAGACCA GGCCCAGAATCGAATCCAATGTGCCATCAAGTCACTAAG TCGCATCACAGAGATGCAGCAG (SEQ ID NO: 29)

	A	B	C	D	E	
	Gene Symbol	Accession	Gene Type	PMBCL/ DLBCL Coefficient	ABC/GCB Coefficient	Gene Target Sequence
30	MYBL1	XM_034274. 14	GCB gene	0	-72.92	GGCAAACGCTGTGTTATCCTCTTTGCAGACCATCCCAGAA TTTGCAGAGACTCTAGAACTTATTGAATCTGATCCTGTAG CATGGAGTGACGTTACCAGT (SEQ ID NO: 30)
31	NECAP2	NM_01 8090.4	PMBCL gene	6.6	0	CTCTCCTCTCCTCCTTGTCTGGCTCTGTTGACAAACCGGG CATGTTTGGCAGTAAATTGGCACCGTGTCACACTGTTTCC TGGGATTCAAGTATGCAACC (SEQ ID NO: 3 1)
32	NF1L3	NM_005384.2	PMBCL gene	2.06	0	CCTTTCTTTCTCCTCGCCGGCCCGAGAGCAGGAACACGAT AACGAAGGAGGCCCAACTTCATTCAATAAGGAGCCTGAC GGATTTATCCCAGACGGTAGA (SEQ ID NO: 32)
33	OPA1	NMJ30837. 1	Housekeeping	-2.97	-4. 14	CTGAGACCATATCCTTAAATGTAAAAGGCCCTGGACTAC AGAGGATGGTGCTTGTTGACTTACCAGGTGTGATTAATAC TGTGACATCAGGCATGGCTCC (SEQ ID NO: 33)
34	PDCD1 LG2	NM_025239.3	PMBCL gene	1.98	0	AGGAAAATAAACACTCACATCCTAAAGGTTCCAGAAACA GATGAGGTAGAGCTCACCTGCCAGGCTACAGGTTATCCT CTGGCAGAAGTATCCTGGCCAA (SEQ ID NO: 34)
35	PHF23	NM_024297.2	Housekeeping	-2.97	-4. 14	CTGTCTGTGTCCCGACACATAATCTCTGTCTCTTGGACCT GCCACCATCACTTTCTG GGTCAGGATTGG AATTGGG ATGG AATGGGACAGTTGTCTATAA (SEQ ID NO: 35)
36	PIM2	NM_006875.2	ABC Gene	0	7 1.8	GCCATCCAGCACTGCCATTCCCGTGGAGTTGTCCATCGTG ACATCAAGGATGAGAACATCCTGATAGACCTACGCCGTG GCTGTGCCAAACTCATTGATT (SEQ ID NO: 36)
37	PRDX2	NM_005809.4	DLBCL gene	-1.28	0	GCATGGGGAAGTTTGTCCCGCTGGCTGGAAGCCTGGCAG TGACACGATTAAGCCCAACGTGGATGACAGCAAGGAATA TTTCTCCAAACACAATTAGGCT (SEQ ID NO: 37)
38	PRKCB	NM_212535. 1	DLBCL gene	- 1.83	0	GCATTTGGAGTCCTGCTGTATGAAATGTTGGCTGGGCAGG CACCCTTTGAAGGGGAGGATGAAGATGAACTCTTCCAAT CCATCATGGAACACAACGTAG (SEQ ID NO: 38)
39	PRR6	NM_1 8 1716.2	PMBCL gene	1.33	0	TTCATTGTTCCAGCTTCTCGCTTCAAGCTCCTGAAGGGAG CTGAGCACATAACGACTTACACGTTCAATACTCACAAAG CCCAGCATACCTTCTGTAAGA (SEQ ID NO: 39)

	A	B	C	D	E	
	Gene Symbol	Accession	Gene Type	PMBCL/ DLBCL Coefficient	ABC/GCB Coefficient	Gene Target Sequence
40	PTGIR	NM_000960.3	PMBCL gene	2.06	0	CTGACATTTCAAGCTGACCCTGTGATCTCTGCCCTGTCTT CGGGCGACAGGAGCCAGAAAATCAGGGACATGGCTGAT GGCTGCGGATGCTGGAACCTTG (SEQ ID NO: 40)
41	QSOX1	NM_002826.4	PMBCL gene	2.85	0	TAGGGCAGCTCAGTCCCTGGCCTCTTAGCACCACATTCCT GTTTTTCAGCTTATTTGAAGTCCTGCCTCATTCTCACTGGA GCCTCAGTCTCTCCTGCTT (SEQ ID NO: 41)
42	R3HDM1	NM_015361.2	Housekeeping	-2.97	-4.14	CCTGTGTTCCCAAGAGAATTACATTATTGACAAAAGACTC CAAGACGAGGATGCCAGTAGTACCCAGCAGAGGCGCCAG ATATTTAGAGTTAATAAAGAT (SEQ ID NO: 42)
43	RAB7L1	NM_001135664.1	ABC Gene	0	70.45	CATTTGAATTGTCTCCTGACTACTGTCCAGTAAGGAGGCC CATTGTCACTTAGAAAAGACACCTGGAACCCATGTGCATT TCTGCATCTCCTGGATTAGC (SEQ ID NO: 43)
44	RCL1	NM_005772.3	PMBCL gene	1.32	0	TGGTGAATCATTTGAACTGAAGATTGTGCGACGGGGAAT GCCTCCCGGAGGAGGAGGCGAAGTGGTTTTCTCATGTCCT GTGAGGAAGGTCTTGAAGCCC (SEQ ID NO: 44)
45	RHOF	NM_019034.2	PMBCL gene	2.48	0	CTGCGGCAAGACCTCGCTGCTCATGGTGTACAGCCAGGG CTCCTTCCCCGAGCACTACGCCCCATCGGTGTTCGAGAAG TACACGGCCAGCGTGACCGTT (SEQ ID NO: 45)
46	S1PR2	NM_004230.2	GCB gene	0	-78.74	TCCCGCCAGGTGGCCTCGGCCTTCATCGTCATCCTCTGTT GCGCCATTGTGGTGGAAAACCTTCTGGTGCTCATTGCGGT GGCCCGAAACAGCAAGTTCC (SEQ ID NO: 46)
47	SERPINA9	NM_00104251 8.1	GCB gene	0	-61.81	CCACTAAATCCTAGGTGGGAAATGGCCTGTTAACTGATG GCACATTGCTAATGCACAAGAAATAACAAACCACATCCC TCTTTCTGTTCTGAGGGTGCAT (SEQ ID NO: 47)
48	SLAMF1	NM_003037.2	PMBCL gene	1.18	0	GTGTCTCTTGATCCATCCGAAGCAGGCCCTCCACGTTATC TAGGAGATCGCTACAAGTTTTATCTGGAGAATCTCACCCT GGGGATACGGGAAAGCAGGA (SEQ ID NO: 48)
49	SNX11	NM_013323.2	PMBCL gene	2.79	0	TCATTTGTATGTAGGACCAGGAGTATCTCCTCAGGTGACC AGTTTTGGGGACCCGTATGTGGCAAATTCTAAGCTGCCAT ATTGAACATCATCCCACTGG (SEQ ID NO: 49)

	A	B	C	D	E	
	Gene Symbol	Accession	Gene Type	PMBCL/ DLBCL Coefficient	ABC/GCB Coefficient	Gene Target Sequence
50	TFPI2	NMJ)06528.2	PMBCL gene	1.06	0	TTTAATCCAAGATACAGAACCTGTGATGCTTTCACCTATA CTGGCTGTGGAGGGAATGACAATAACTTTGTTAGCAGGG AGGATTGCAAACGTGCATGTG (SEQ ID NO: 50)
5 1	TMOD1	NM_003275.2	PMBCL gene	1.15	0	AGATGCTCAAGGAGAACAAGGTGTTGAAGACACTGAATG TGGAATCCAACTTCATTTCTGGAGCTGGGATTCTGCGCCT GGTAGAAGCCCTCCCATACAA (SEQ ID NO: 51)
52	TNFRSF1 3B	NM_01 2452.2	ABC Gene	0	66.49	TGCAAAACCATTTGCAACCATCAGAGCCAGCGCACCTGT GCAGCCTTCTGCAGGTCACTCAGCTGCCGCAAGGAGCAA GGCAAGTTCTATGACCATCTCC (SEQ ID NO: 52)
53	TRAF1	NM_005658.3	PMBCL gene	1.98	0	CGAGTGATGGGTCTAGGCCCTGAAACTGATGTCCTAGCA ATAACCTCTTGATCCCTACTCACCGAGTGTTGAGCCCAAG GGGGGATTTGTAGAACAAGCC (SEQ ID NO: 53)
54	TRIM56	NM_030961.1	Housekeeping	-2.97	-4. 14	GTGGAGGCCGAGGACATTTTCCTGAAGGGCAGGGGTTGG CAACTTTTCAACATGGAGTGCCAAACTGCTAACCCGTCTT CTAGTGTGTGAGAATAGGGAC (SEQ ID NO: 54)
55	UBXN4	NM_0 14607.3	Housekeeping	-2.97	-4. 14	CATCGCGACGGCCAAAAGGAGCGGCGCGGTCTTCGTGGT GTTCGTGGCAGGTGATGATGAACAGTCTACACAGATGGC TGCAAGTTGGGAAGATGATAAA (SEQ ID NO: 55)
56	VRK3	NM_0 16440.3	Housekeeping	-2.97	-4. 14	ACAGACAAGAGTGGGCGACAGTGGAAGCTGAAGTCCTTC CAGACCAGGGACAACCAGGGCATTCTCTATGAAGCTGCA CCCACCTCCACCCTCACCTGTG (SEQ ID NO: 56)
57	WAC	NM_100486.2	Housekeeping	-2.97	-4. 14	CCTCTGGACTGAACCCCACATCTGCACCTCCAACATCTGC TTCAGCGGTCCCTGTTTCTCCTGTTCCACAGTCGCCAATA CCTCCCTTACTTCAGGACCC (SEQ ID NO: 57)
58	WDR55	NM_0 17706.4	Housekeeping	-2.97	-4. 14	CTACCTCTTCAATTGGAATGGCTTTGGGGCCACAAGTGAC CGCTTTGCCCTGAGAGCTGAATCTATCGACTGCATGGTTC CAGTCACCGAGAGTCTGCTG (SEQ ID NO: 58)

[0031] The above set of genes comprise the Lymph3Cx assay, which includes probe sets for 30 PMBCL/DLBCL predictive genes that were identified by being strongly differentially expressed between the PMBCL and DLBCL in U133+ data, strongly differentially expressed between the PMBCL and DLBCL in Lymph5Cx data, and differentially expressed between the DLBCL and PMBCL/cHL cell lines; 15 probes sets that had been included on the Lymph2Cx assay as being predictive of the ABC and GCB subtypes within DLBCL; probe sets for 13 housekeeping genes (including the 5 housekeeping genes from the Lymph2Cx assay) that showed strong expression and low variability in both U133+ and Lymph5Cx data and so were suitable for normalization. The data are discussed in the Example below.

[0032] In an embodiment, an equation used to determine a tumor predictor score is:

$$S = \sum_{fe=0}^{58} a_{i^{\times}\ i},$$

wherein a_i is the model coefficient value for gene i as listed in Table 1, column D for determining whether the sample is PMBCL or DLBCL and as listed in Table 1, column E for determining whether the sample is ABC DLBCL or GCB DLBCL, and x_i is the log_2 transformed expression signal value for gene i.

[0033] In an embodiment, the coefficients used to generate a tumor predictor score may be refined, and tumor predictor score cut-points used to produce model scores may be refined. For example, using methods as described herein with the same genes as those in Table 1, the coefficients for each gene may be determined to be different than as listed in Table 1 based on, e.g., the use of different types of sample (e.g., fresh) or use of different microarrays that provide different signal values. In an embodiment, the above methods may be incorporated into other methods, for example a Bayesian method as described in International Patent Application Publication No. WO 2015/069790, which is incorporated herein by reference.

[0034] In an embodiment, the present invention also provides a method for classifying a lymphoma according to the following tumor predictor scores (S value cut-points): (i) DLBCL when S is less than -57.95, (ii) PMBCL when S is greater than -23.57, (iii) uncertain DLBCL/PMBCL when S is between -57.95 and -23.57, (iv) GCB DLBCL when S is less than 798.5, (v) ABC DLBCL when S is greater than 1324.5, or (vi) uncertain ABC/GCB DLBCL when S is between 798.5 and 1324.5. Such an embodiment uses the equation

$$S = \sum_{k=0}^{58} a_i x_i$$

as defined above.

[0035] In an embodiment, an equation used to determine the probability score P(PMBCL) of a sample is

$$P(PMBCL) = \frac{\langle P(.SpMBCL/DLBCL\rangle_{PPMBCL'} \circ_{PMBC} L)}{\varphi(S; V\text{-}PMBCL' \sigma_{PMBCl.}) + \psi(\beta : \mu_{DLBCL'} \sigma_{DLBCL})}$$

wherein $S_{PMBCL/DLBCL}$ is the tumor predictor score; $\mu_{PM}\beta_{CL>}\mu_{DLBCL}$ and σ_{PMBCL}, σ_{DLBCL} represent the mean and standard deviations of the PMBL and DLBCL subtypes as indicated in Table 2, herein.

[0036] In an embodiment, an equation used to determine the probability score P(ABC) of a sample is

$$P(ABC) = \frac{\varphi(S_{ABC\,I\,GCB}; \mu_{ABC'} \sigma_{ABc)}}{\varphi(S; \mu_{ABC}, \sigma_{ABC}) + \varphi(S; \mu_{GCB}, \sigma_{GCB})}$$

wherein $S_{ABC\ /\ GCB}$ is the tumor predictor score; μ_{ABC}, μ_{GCB} and $\sigma_{ABC>\circ GCB}$ represent the mean and standard deviations of the ABC and GCB subtypes as indicated in Table 2; and wherein $\phi(\chi; \mu, \sigma)$ is the standard normal density calculated by

$$\varphi(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

[0037] In an embodiment, after the P(PMBCL) and P(ABC) have been calculated for a particular sample, the sample is classified according Table 3, provided herein.

[0038] In an embodiment, the present invention entails the development of a set of nucleic acid probes that are able to measure the abundance of particular mRNA species using the NanoString Technologies® platform for the purpose of gene expression profiling of PMBCL in order to distinguish this lymphoma subtype from other subtypes of aggressive B

cell lymphoma. In this embodiment, RNA is extracted from, e.g., FFPE, samples using standard commercial kits and then hybridized and detected. The resultant digital RNA counts reflect the relative abundance of mRNAs transcribed from different genes. These expression levels are then combined in statistical algorithms to create a tumor predictor score that provides a probability that a tumor is PMBCL, ABC DLBCL, or GCB DLBCL.

[0039] In an embodiment, the present invention provides a composition consisting of probes to the target sequences described herein. In another embodiment, the present invention also provides a kit comprising the probes, for example, a kit comprising components suitable for performing NanoString Technologies® nCounter® digital gene expression assays.

[0040] The following includes certain aspects of the invention.

[0041] 1. A method for classifying the lymphoma type of a sample, the method comprising:

(a) providing a formalin-fixed and paraffin-embedded (FFPE) lymphoma sample from the subject;

(b) isolating RNA from the sample;

(c) obtaining gene expression data from the RNA,

wherein the gene expression data comprises signal values that represent expression levels for each gene of Table 1; and

(d) determining a tumor predictor score from the gene expression data, wherein the tumor predictor score is calculated by

$$S = \sum_{k=0}^{58} a_i x_i,$$

wherein a_i is the model coefficient value for gene i as listed in Table 1, column D for determining whether the sample is PMBCL or DLBCL and as listed in Table 1, column E for determining whether the sample is ABC DLBCL or GCB DLBCL,

and x_i is the log_2 transformed expression signal value for gene i; and

(e) when the coefficient values in column D of Table 1 are used, classifying the lymphoma as:

(i) DLBCL when S is less than -57.95,

(ii) PMBCL when S is greater than -23.57,

(iii) uncertain DLBCL/PMBCL when S is between -57.95 and -23.57,

(e') when the coefficient values in column E of Table 1 are used, classifying the lymphoma as:

(iv) GCB DLBCL when S is less than 798.5,

(v) ABC DLBCL when S is greater than 1324.5, or

(vi) uncertain ABC/GCB DLBCL when S is between 798.5 and 1324.5.

[0042] 2. The method of aspect 1, wherein the method further comprises determining the probability that the sample is PMBCL or ABC DLBCL, wherein the probability is determined by

(g) determining the probability that the sample is PMBCL by calculating the probability score of

$$P\text{^}PMBCL \quad) = \frac{\varphi(_{S\,p\,M\,BCL/D\,L_B\,CL}; \mu_{p\,M\,B\,CL}, \tau_{p\,M\,B\,CL})}{\psi(S; \mu_{PMBCL}, \sigma_{PMBc\ L}) + \varphi(s\ ; \mu_{DLBCL}, \sigma_{DLBCL})},$$

wherein $S_{PMBCL/DLBCL}$ is the tumor predictor score; $\mu_{PMB\alpha\ L}, \mu_{DLBCL}$ and $\sigma_{PMBCL}, \sigma_{DLBCL}$ represent the mean and standard deviations of the PMBCL and DLBCL subtypes as indicated in Table 2;

(h) determining the probability that the samples is ABC DLBCL by calculating the probability score of

$$P(ABC) = \frac{\varphi(S_{ABCl\,GCB}; \mu_{ABC}, \sigma_{A\,B_c})}{\varphi(S; \mu_{ABC}, \sigma_{ABc}) + \varphi(S; \mu_{G\,CB}, \sigma_{GCB})},$$

wherein $S_{ABC\ /\ GCB}$ is the tumor predictor score; $\mu_{A\,BC}$, V-GCB and $\sigma_{ABC}, \sigma_{GCB}$ represent the mean and standard deviations of the ABC and GCB subtypes as indicated in Table 2; and

wherein $\phi(\chi; \mu, \sigma)$ is the standard normal density calculated by

$$\varphi(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

[0043] 3. The method of aspect 2, wherein a score of P(PMBCL) $\geq$ 0.9 indicates that the sample is PMBCL, regardless of the P(ABC) score.

[0044] 4. The method of aspect 2, wherein a score of P(PMBCL) $\leq$ 0.1 and a score of P(ABC) <0.1 indicates that the tumor is GCB DLBCL.

[0045] 5. The method of aspect 2, wherein a score of P(PMBCL) $\leq$ 0.1 and a score of P(ABC) $\geq$ 0.9 indicates that the tumor is ABC DLBCL.

[0046] 6. The method of aspect 2, wherein a score of P(PMBCL) $\leq$ 0.1 and a score of P(ABC) greater than 0.1 and less than 0.9 indicates that the tumor is an unclassified DLBCL.

[0047] 7. The method of any one of aspects 1-6, wherein the RNA gene expression data is obtained using a NanoString Technologies® nCounter® assay.

[0048] It shall be noted that the preceding are merely examples of embodiments. Other exemplary embodiments are apparent from the entirety of the description herein. It will also be understood by one of ordinary skill in the art that each of these embodiments may be used in various combinations with the other embodiments provided herein.

[0049] The following example further illustrates the invention but, of course, should not be construed as in any way limiting its scope.

EXAMPLE

[0050] This example demonstrates the ability of the newly developed and validated assay to distinguish a PMBCL sample from a DLBCL sample, and to distinguish between ABC DLBCL and GCB DLBCL samples based on gene expression signatures, in accordance with embodiments of the invention.

[0051] All cases included in this study were retrieved from the tissue archives of participating centers. Conventional and immunohistochemically stained slides were reviewed by at least 4 members of the Lymphoma/Leukemia Molecular Profiling Project (LLMPP) expert hematopathology panel. Information on age, sample location and clinical presentation were made available whenever possible. If 3 pathologists independently agreed, a definite diagnosis was established - all other cases were subjected to panel discussion. Pathological diagnosis is referred to as the "gold standard" for DLBCL/PMBCL classification herein. This study was conducted with approval from Institutional Review Boards according to the Declaration of Helsinki.

[0052] Details on study design, gene selection and model building are provided in Figure 1 and below. Between the training and validation cohorts, in total 108 PMBCL and 127 DLBCL cases were studied. Following deparaffinization, RNA was extracted from up to five 10 µm formalin-fixed, paraffin-embedded (FFPE) tissue sections (tumor content ≥ 60% of tissue area) using the Qiagen DNA/RNA FFPE Kit (Hilden, Germany) according to the manufacturer's instructions. RNA was quantified using a spectrophotometer (Nanodrop, ThermoFisher, Germany). Gene expression analysis was performed on 200 ng of RNA using a custom codeset on the NanoString platform (NanoString Technologies, Seattle, WA, USA) at the "high sensitivity" setting on the Prep Station and 555 fields of view on second generation nCounter analyzers.

Feature/gene selection

[0053] To select genes with high discriminative power to distinguish between PMBCL and DLBCL, a large set of existing gene expression data, previously generated using different platforms (Lenz et al., N. Engl. J. Med., 359: 2313-2323 (2008) and Scott et al., Blood, 124: 3016-3019 (2014), both incorporated herein by reference), was interrogated.

[0054] In total 248 DLBCLs were used, all with gene expression profiling performed on fresh-frozen tissue using the Affymetrix U133 plus 2.0 microarray platform (Thermo Fisher Scientific, Waltham, MA, USA). A subset of 79 of the cases had matching gene expression analysis performed using NanoString technology in conjunction with a customized codeset containing 814 genes (termed Lymph5Cx) on formalin-fixed, paraffin-embedded (FFPE) material (Scott et al., Blood, 124: 3016-3019 (2014)).

[0055] The dataset for PMBCL consisted of 36 gene expression profiles (GEP) generated with the Affymetrix platform and 42 targeted gene expression data derived from the Lymph5Cx assay, with 13 cases overlapping between the two platforms.

[0056] Affymetrix samples were normalized with the Affymetrix MAS5.0 algorithm and log2-transformed. NanoString gene expression counts for a given sample were normalized by dividing the counts for each gene by the geometric mean of counts for the housekeeping genes for that sample and then multiplying them by 512. The normalized counts were then log2-transformed. Lymph3Cx samples for which the geometric mean of the housekeeping genes was less than 60 were excluded from analysis.

[0057] In order to identify genes which were strongly predictive of subtype (i.e. DLBCL or PMBCL), a "Z-score" was calculated for each gene represented on both the Affymetrix and Lymph5Cx platform:

$$Z_i = \frac{\mu_{Pi} - \mu_{Di}}{\sqrt{\frac{n_P\sigma_{Pi}^2 + n_D\sigma_{Di}^2}{n_P + n_D}}}$$

where $\{n_P, n_D\}$ represent the number of PMBCL and DLBCL specimens available on a particular platform and $\{\mu_{Pi}, \mu_{Di}, \sigma_{Pi}^2, \sigma_{Di}^2\}$ represent their respective means and variances of the normalized signal values of gene i. The genes considered as significant were those for which both the Affymetrix and the Lymph5Cx Z-score were either greater than 1.25 or less than -1.0. In this way, the larger Affymetrix data set was used to identify potential genes, while at the same time ensuring that similar results are likely when analyzing FFPE samples. A less stringent negative cut-point than positive cut-point was used in order to avoid a large imbalance between the number of predictor genes that were characteristic of PMBCL (i.e., higher in PMBCL than DLBCL) rather than characteristic of DLBCL (i.e., higher in DLBCL than PMBCL). These cut-offs resulted in 9 DLBCL characteristic genes and 35 PMBCL characteristic genes.

[0058] To guard against a potential bias of selecting predominantly microenvironment/stroma-related genes, also included was GEP (generated using Affymetrix arrays) of cell lines (42 DLBCL-derived in addition to 3 PMBCL and 3 classical Hodgkin lymphoma (cHL) cell lines). A t-test between the DLBCL and the combined set of PMBCL and cHL lines was performed and any gene which did not show a significant p-value (one-sided $p<0.05$) in the same direction as it was observed in the patient data was excluded. In this way, the differential expression observed was due to differences in the tumor cells rather than due to infiltrating bystander cells. Fourteen genes were excluded in this step, resulting in a final list of 6 DLBCL characteristic genes and 24 PMBCL characteristic genes (Figure 2).

Model building and score formulation

[0059] The core training set of cases to build the model and define score thresholds to distinguish PMBCL from DLBCL consisted of 68 cases (48 DLBCL (19 ABC, 19 GCB, 10 unclassified DLBCL) and 20 PMBCL, according to consensus review). Data were

normalized as described above. Samples with low normalizer values were excluded from further analyses. Based on the Lymph3Cx data set, a naive linear discriminant analysis score was calculated as follows:

$$S_j = \sum_{i=1}^{58} a_i x_{ij}$$

where x_{ij} is the Lymph3Cx signal (log_2-transformed nCounter expression) for a predictive gene i on sample j, and a_i is a scaling factor.

[0060] The scaling factor coefficients were calculated as follows for the DLBCL/PMBCL model:

$$a_i \sim \frac{\mu_{Pi} \sim \mu_{Di}}{(20\ \sigma_{Pi}^2 + 48\sigma\text{^})/68}$$

$\{\mu_{Pi}, \mu_{Di}, \sigma_{Pi}^2, \sigma_{Di}^2\}$ represent the observed mean and variance of the $\log_2$-transformed normalized counts for gene i within the PMBCL or DLBCL specimens on the Lymph3Cx platform, and the numbers 20 and 48 are derived from the number of histopathological reviewed DLBCL and PMBCL samples available on the Lymph3Cx platform. Coefficients for housekeeping genes were set to the constant value that would make the total sum of coefficients (both housekeeping and predictive genes) equal to 0. In this way, the model is automatically normalized so that a uniform fold increase or decrease of expression across all genes will have no effect on the model score.

[0061] The scaling factor coefficients for the predictive ABC/GCB genes in the Lymph3Cx ABC/GCB predictor were set equal to those in the Lymph2Cx predictor. Coefficients for housekeeping genes were set to that constant value that would make the total sum of coefficients (both housekeeping and predictive genes) equal to 0.

[0062] The values of the a_i coefficients are provided in Table 1 for the DLBCL/PMBCL and ABC/GCB models.

[0063] All 38 DLBCL samples for which Lymph3Cx data was available were additionally analyzed by the Lymph2Cx array. For these samples the Lymph3Cx model score was compared to the Lymph2Cx model score and were found to be nearly identical except for shift by a constant value of 1109. Therefore, the ABC and GCB variance values

were used, as in the Lymph2Cx predictor, but the ABC and GCB means were shifted by 1109 points in the Lymph3Cx predictor, as indicated in Table 2.

[0064] In order to estimate the distribution of the DLBCL vs PMBCL score within the DLBCL and PMBCL subsets score, it was important to draw from as many available cases as possible. To this end, predictor scores were generated for all 313 samples (248 DLBCL and 65 PMBCL) available.

[0065] Of the DLBCL samples, 86 were of the ABC DLBCL type, 121 were of the GCB DLBCL type and 41 were unclassified DLBCL. Of the samples, 284 samples (36 PMBCL, and all DLBCL samples) were analyzed with an Affymetrix U133+ array, 121 of the samples were analyzed with the 814 gene Lymph5Cx array (42 PMBCL, 28 ABC, 38 GCB, 13 unclassified DLBCL). Of the samples, 92 samples (13 PMBCL, 28 ABC, 38 GCB, 13 unclassified DLBCL) were analyzed with both the Affymetrix U133+ and Lymph5Cx arrays.

[0066] For each sample on each platform, a linear model score was generated based on the coefficients defined above (column D of Table 1) but using the normalized gene signal values of the particular platform. To account for platform differences, the Lymph5Cx scores were linearly normalized such that the mean and variance of the normalized Lymph5Cx model score, over the set of 68 cases for which also Lymph3Cx data was available, matched the mean and variance of the Lymph3Cx model score over the same set of cases. Similarly, the Affymetrix model score was normalized to match the normalized Lymph5Cx model score over all cases for which both Lymph5Cx and Affymetrix data was available. As a final score, the Lymph3Cx score was used for all cases for which it was available, the normalized Lymph5Cx score was used for all cases for which Lymph5Cx but not Lymph3Cx was available, and the normalized Affymetrix model score was used for cases in which Lymph5Cx/3Cx was unavailable.

[0067] In more detail, for a given sample j analyzed with the Lymph3Cx platform, a Lymph3Cx model score was calculated as

$$T_j = \sum_{i=1}^{58} a_i x_{ij}$$

where x_{ij} is the log transformed Lymph3Cx for gene i on sample j. Similarly, a Lymph5Cx based score for sample j was calculated as:

$$F_j = \sum_{i=i}^{58} a_i y_{ij}$$

where y_{ij} is the log transformed Lymph5Cx counts for gene i on sample j. The model was then linearly adjusted so that it matches the mean and variance of the Lymph3Cx model as follows

$$F'j = \left(F_j + \underset{k\in B}{\text{Mean}}(T_{i} - F_k)\right) \sqrt{\frac{\underset{k\in A}{\text{Var}}(T_{i})}{\underset{k\in A}{\text{Var}}(F_{i})}}$$

where the mean and variance are taken over A, the set of samples for which both Lymph3Cx and Lymph5Cx data exists. The technique was then repeated for samples for which only U133+ data was available. The score was calculated as:

$$U_j = \sum_{i=1}^{58} a_i z_{ij}$$

where z_{ij} is the log transformed MAS5 U133+ measure of gene expression for gene i on sample j.

$$U'_j = \left(U_j + \underset{k\in B}{\text{Mean}}(F'_{i} - U_k)\right) \sqrt{\frac{\underset{k\in B}{\text{Var}}(F_k')}{\underset{i\in B}{\text{Var}}(i/_{i})}}$$

where the mean and variance are taken over B, the set of samples for which both Lymph5Cx and U133+ data exists. For each of the 313 samples a final model score representative was defined:

$$S_j = \begin{cases} T_j & \text{if sample j has Lymph3Cx data available} \\ F_j & \text{if sample j has Lymph5Cx data available but not Lymph3Cx} \\ U_j & \text{if sample j has only U133+ data available} \end{cases}$$

[0068] To account for a certain possibility of erroneous classification during the pathology review process, the final score formulation was adjusted. In modeling the relationship between model score and diagnosis, the following normal mixture model was used:

$$P(S_j \mid D_j = \text{DLBCL}) = (1 - \varepsilon_D)\varphi(S_j; \mu_D, \sigma_D^2) + \varepsilon_D \varphi(S_j; \beta\rho, \sigma_P^2)$$

$$P(S_j \mid D_j = \text{PMBCL}) = \varepsilon_P \varphi(S_j; \hat{\mu}_D, \hat{\sigma}_D^2) + (1 - \varepsilon_P)\varphi(S_j; \hat{\mu}_P, d_P^2)$$

where S_j is the model score for sample j; D_j is the pathological diagnosis for sample j; ϕ is the normal density function calculated by

$$\varphi(S_j; \hat{\mu}, \hat{\sigma}^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{(S_j - \hat{\mu})^2}{2\hat{\sigma}^2}\right)$$

and $\varepsilon\rho$ and ε_D are the probability of misclassification of a case as DLBCL or PMBCL; and $\hat{\mu}_D$, $\hat{\sigma}_P^2$, β_P, $\hat{\sigma}_P^2$ are the means and variances of the model scores for cases for which the diagnosis of DLBCL or PMBCL is correct (Table 2).

[0069] The values ε_P, ε_D, $\hat{\mu}_D$, $\hat{\sigma}_D^2$, $\hat{\mu}_P$, $\hat{\sigma}_P^2$ were estimated via maximum likelihood over the 313 normalized model scores, resulting in a very low rate of pathology misclassification with $\varepsilon\rho$ =0.033 and ε_D =0.0076. The principal of maximum likelihood is used to estimate the unknown parameters by choosing those values which maximized the following expression representing the likelihood of obtaining the observed data.

$$\prod_{\substack{j \text{ Path} \\ \text{reviewed as} \\ \text{PMBCL}}} P(S_j \mid D_j = \text{DLBCL}) + \prod_{\substack{j \text{ Path} \\ \text{reviewed as} \\ \text{DLBCL}}} P(S_j \mid D_j = \text{DLBCL})$$

Table 2: Model means and standard deviations

Model	Subtype	Mean	Standard deviation
PMBCL/DLBCL	PMBCL	-85.60	24.66
PMBCL/DLBCL	DLBCL	14.42	31.94
ABC/GCB	ABC	2107.24	468.93
ABC/GCB	GCB	-216.78	595.51

[0070] By estimating distributions for model scores of true DLBCL and PMBCL cases, a Bayesian algorithm was used to calculate the probability of a sample being PMBCL:

$$P\ (PMBCL) = \frac{\varphi\left(S_j;\ \hat{\mu}_P,\ \hat{\sigma}_P^2\right)}{\varphi(5_j;\ \hat{\mu}_P, \hat{\sigma}_P^2) + \phi\left(S_j;\ f_D, \hat{\sigma}_D^2\right)}$$

where $\varphi(5,-;\ \hat{\mu}, \hat{\sigma}^2)$ is the standard normal density.

[0071] Since it was unknown what the composition of the target population of this assay would be, the following was assumed: a flat prior with equal a priori likelihood of a sample being of the DLBCL or PMBCL subtype.

[0072] Similarly, the probability that a sample was ABC was calculated according to:

$$P\ (ABC) = \frac{\varphi(S_j; \hat{\mu}_{ABC}, \hat{\sigma}_{ABC})}{\varphi\{Sj; \hat{\mu}_{ABC}, {}^{\circ}ABC) + {}^{\varphi}(Sj; ficcB.\ \hat{\sigma}_{GCB}^2)}$$

where $\hat{\mu}_{ABC}$, $\hat{\sigma}_{ABC}$ and $\hat{\mu}_{GCB}$, $\hat{\sigma}_{GCB}$ represent the mean and standard deviations of the ABC and GCB subtypes as indicated in Table 2.

[0073] After the P(PMBCL) and P(ABC) have been calculated for a particular sample, the sample is classified according Table 3.

Table 3. Prediction result look-up table

		PMBCL/DLBCL Model result		
		P(PMBCL) ≤ 0.1	0.1 < P(PMBCL) < 0.9	P(PMBCL) ≥ 0.9
ABC/GCB Model Result	P(ABC) ≤ 0.1	GCB	Unclear PMBCL/GCB	PMBCL
	0.1 < P(ABC) < 0.9	Unclassified DLBCL	Unclear PMBCL/Unclassified DLBCL	PMBCL
	P(ABC) ≥ 0.9	ABC	Unclear PMBCL/ABC	PMBCL

[0074] As a final molecular diagnosis, those specimens were designated to represent PMBCL for which P(PMBCL) ≥ 0.9; as DLBCL for which P(PMBCL) ≤ 0.1; and as "uncertain DLBCL/PMBCL", those cases for which 0.1 < P(PMBCL) < 0.9.

[0075] Also, if P(PMBCL) ≥ 0.9 then the tumor sample will be predicted to be PMBCL regardless of the ABC/GCB predictor result. If P(PMBCL) ≤ 0.1 and P(ABC) ≤ 0.1 then that tumor is indicated as GCB DLBCL. If P(PMBCL) ≤ 0.1 and P(ABC) ≥ 0.9 then that tumor is indicated as ABC DLBCL. If (PMBCL) ≤ 0.1 and P(ABC) is greater than 0.1 and less than 0.9, then that tumor is indicated as an unclassified DLBCL.

[0076] If the PMBCL probability falls between 0.1 and 0.9, the tumor prediction results may be unclear. For example, if O.K P(PMBCL) < 0.9 and P(ABC) ≤ 0.1, then the sample is indicated as unclear PMBCL/GCB. If O.K P(PMBCL) < 0.9 and 0.1 < P(ABC) < 0.9, then the sample is indicated as unclear PMBCL/Unclassified DLBCL. If O.K P(PMBCL) < 0.9 and P(ABC) ≥ 0.9, then the sample is indicated as unclear PMBCL/GCB.

[0077] Eighteen PMBCL cases of the core training cohort were assigned to be PMBCL by the gene expression-based assay, whereas the remaining two cases were classified into the uncertain group. None of the PMBCL cases were misclassified as DLBCL. Eight DLBCL cases fell in the uncertain category and one GCB DLBCL case was classified as PMBCL by the Lymph3Cx assay. Notably, the score of this case (-15.54) was relatively close to the cut-off (-23.57). As the new assay contains the probes for DLBCL COO assignment as described previously (Scott et al., Blood, 123: 1214-1217 (2014), incorporated herein by reference), it was assessed whether the classification matches the COO as derived from gene expression profiling using snap-frozen tissue biopsies and the Affymetrix U133 plus 2.0 microarrays or

the Lymph2Cx assay, respectively. A linear model score was generated using the same probe sets and coefficients as were used in the Lymph2Cx predictor (Scott et al., Blood, 123: 1214-1217 (2014); the Lymph2Cx assay is a 2-way classifier that is focused on distinguishing between two subtypes of aggressive B-cell Non-Hodgkins lymphoma: ABC-DLBCL and GCB-DLBCL; the Lymph2Cx assay does not distinguish between PMBCL and DLBCL). Of note, there was no misclassification of cases with regards to switching between the ABC and GCB classes. Nine cases changed between the unclassified category and ABC or GCB, respectively.

Validation cases and re-review of misclassified cases

[0078] For the independent validation cohort, 118 cases diagnosed as PMBCL and collected over a period of 35 years at the Department of Pathology at the BC Cancer Agency (BCCA), Vancouver, Canada, were selected for pathology review. Five cases were re-classified (4 cases were diagnosed as DLBCL and 1 case as classical Hodgkin lymphoma). In 7 cases PMBCL was considered to be among the differential diagnoses and correlation with clinical data and presentation was recommended. In 1 case the material available for review was not sufficient to render a diagnosis, and in the remaining cases (105) the diagnosis of PMBCL was confirmed. Of these, 88 cases had sufficient material for further analysis.

[0079] For intra-laboratory comparison, RNA from 12 of the 18 biopsies selected were run a second time at BCCA, with each run performed using a different RNA aliquot and different NanoString cartridges and 6 cases were replicated at Mayo Clinic.

Results

[0080] To develop a classification assay, applicable to FFPE tissue specimens, that aims at a robust discrimination between PMBCL and DLBCL as well as the DLBCL subtypes (GCB, ABC, and unclassified, respectively), gene expression features were selected from previously published datasets (Rosenwald et al., J. Exp. Med., 198: 851-861 (2003); Lenz et al., N. Eng. J. Med., 359: 2313-2323 (2008); and Alizadeh et al., Nature, 403: 503-51 1 (2000); each incorporated herein by reference). The selection process, aiming at identification of genes with the highest discriminative power and good concordance between the different analytical platforms used for GEP, yielded 58 genes for subsequent assay development (Table 1). Of those, 30 genes were employed to distinguish PMBCL from DLBCL, with 24 being overexpressed in PMBCL and 6 genes showing higher expression levels in DLBCL.

This approach of "balanced" gene selection was chosen to make the model less vulnerable to normalization artifacts. Additionally, 15 genes from the Lymph2Cx assay (Scott et al., Blood, 123: 1214-1217 (2014)) were included, and the remaining 13 genes were chosen as housekeeping genes, including all 5 from the Lymph2Cx assay. To train a linear regression model and establish model thresholds to distinguish PMBCL from DLBCL, a customized NanoString codeset including these 58 genes was then applied to a training cohort of 68 cases, of which 20 were diagnosed as PMBCL by consensus review and 56 were classified as DLBCL. The performance of the Lymph3Cx assay in the training cohort is shown above. The gene expression-based model, including coefficients and thresholds was "locked" and subsequently applied to the independent validation cohort comprising 167 FFPE tissue biopsies (88 PMBCL and 79 DLBCL by consensus review). None of these specimens were part of the training cohort, nor had been previously used to train the Lymph2Cx assay.

[0081] The assay yielded gene expression data of sufficient quality in 160/167 cases (95.8%), leaving 88 PMBCL and 70 DLBCL cases for final analysis (2 additional DLBCL cases were excluded because of a mismatch to previously analyzed frozen biopsies). Among the pathologically-defined PMBCL, 75 cases (85%) were classified as such based on Lymph3Cx. Ten percent (9 cases) were assigned to the "uncertain" category and ~5% (4 cases) showed a molecular signature of DLBCL. Of note, scores for most of the 'misclassified' PMBCLs (3/4) were close to the cut-off (Figure 3). Among the pathologically-defined DLBCL cases, 58 (83%) were classified as DLBCL by the assay, 14% (10 cases) were "uncertain" and two cases (3%) were predicted to be PMBCL.

[0082] A pathological re-review was performed on the six misclassified cases from the validation cohort. For one DLBCL case the panel would have changed the diagnosis to PMBCL (as assigned by the assay) based on additional clinical information (young patient, mediastinal involvement), which was not available at the time of initial review. The other misclassified DLBCL case, presenting as an intra-parotid lymph node enlargement with no other manifestations, was still considered to represent DLBCL. Of the 4 mis-assigned PMBCL cases, two were still felt to fit the morphological spectrum of PMBCL with typical clinical presentation, one case was deemed unclassifiable based on the available information and material for review and one case was felt to rather represent DLBCL based on the clinical information of a generalized lymphadenopathy. Of note, none of these 4 cases harbored rearrangements or copy number alteration of *CUTA* or the PD1 receptor ligands PDL1 and PDL2. Similar to the results obtained in the training cohort, no misclassified cases were seen

with regards to DLBCL cell-of-origin (COO) subclassification. Eleven cases changed between the unclassified category and ABC or GCB, respectively.

[0083] For 66 cases with available COO predictions and model scores from the Lymph2Cx assay (Scott et al., Blood 123: 1214-1217 (2014) and Scott et al., J. Clin. Oncol., 33: 2848-2856 (2015), both incorporated herein by reference) the results were compared to Lymph3Cx and revealed a high correlation coefficient between both assays (Spearman r = 0.9937). No classification changes were observed, demonstrating the robustness of DLBCL COO-assignment across these assays (Figure 4).

[0084] Experiments were next performed to determine intra-laboratory reproducibility and inter-laboratory concordance of the Lymph3Cx assay. Eighteen biopsies were selected on the basis that their model scores were equally distributed across the population and thus are representative of the validation cohort. For inter-laboratory comparison, separate tissue scrolls or unstained sections were distributed to an independent laboratory (Mayo Clinic, Scottsdale) where RNA was extracted and run on the Lymph3Cx assay. The concordance was excellent with Spearman $r^2 = 0.996$ (Figure 5), demonstrating the robustness of the assay.

[0085] These results demonstrate that the newly developed and validated Lymph3Cx assay robustly distinguishes between PMBCL and DLBCL tumors based on gene expression signatures and shows high concordance with the pathological classification of an expert hematopathologist panel.

[0086] All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein.

[0087] The use of the terms "a" and "an" and "the" and "at least one" and similar referents in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The use of the term "at least one" followed by a list of one or more items (for example, "at least one of A and B") is to be construed to mean one item selected from the listed items (A or B) or any combination of two or more of the listed items (A and B), unless otherwise indicated herein or clearly contradicted by context. The terms "comprising," "having," "including," and "containing" are to be construed as open-ended terms (i.e., meaning "including, but not limited to,") unless

otherwise noted. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

[0088] Preferred embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Variations of those preferred embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventors expect skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

CLAIM(S):

1. A method for classifying the lymphoma type of a sample, the method comprising:

(a) providing a formalin-fixed and paraffin-embedded (FFPE) lymphoma sample from the subject;

(b) isolating RNA from the sample;

(c) obtaining gene expression data from the RNA,

wherein the gene expression data comprises signal values that represent expression levels for each gene of Table 1; and

(d) determining a tumor predictor score from the gene expression data, wherein the tumor predictor score is calculated by

$$S = \sum_{k=0}^{58} a_i x_i,$$

wherein a_i is the model coefficient value for gene i as listed in Table 1, column D for determining whether the sample is PMBCL or DLBCL and as listed in Table 1, column E for determining whether the sample is ABC DLBCL or GCB DLBCL,

and x_i is the $\log_2$ transformed expression signal value for gene i; and

(e) when the coefficient values in column D of Table 1 are used, classifying the lymphoma as:

(i) DLBCL when S is less than -57.95,

(ii) PMBCL when S is greater than -23.57,

(iii) uncertain DLBCL/PMBCL when S is between -57.95 and -23.57,

(e') when the coefficient values in column E of Table 1 are used, classifying the lymphoma as:

(iv) GCB DLBCL when S is less than 798.5,

(v) ABC DLBCL when S is greater than 1324.5, or

(vi) uncertain ABC/GCB DLBCL when S is between 798.5 and 1324.5.

2. The method of claim 1, wherein the method further comprises determining the probability that the sample is PMBCL or ABC DLBCL, wherein the probability is determined by

(h) determining the probability that the sample is PMBCL by calculating the probability score of

$$P(PMBCV) = \frac{\varphi(S_{PMBCL/DLBCL};\mu_{PMBCL},\sigma_{PMBCL})}{\varphi(S;\mu_{PMBCL},\sigma_{PMBcL})+\varphi(S;\mu_{DLBCL},\sigma_{DLBCL})},$$

wherein $S_{PMBCL/DLBCL}$ is the tumor predictor score; $\mu_{PMB\alpha} > \mu_{DLBCL}$ and $\sigma_{PMBCL}, \sigma_{DLBCL}$ represent the mean and standard deviations of the PMBL and DLBCL subtypes as indicated in Table 2;

(h) determining the probability that the samples is ABC DLBCL by calculating the probability score of

$$P(ABC) = \frac{\varphi(S_{ABCIGCB};\mu_{ABC},\sigma_{ABC})}{\varphi(S;\mu_{ABC},\sigma_{ABc})+\varphi(S;\mu_{GCB},\sigma_{GCB})},$$

wherein $S_{ABC/GCB}$ is the tumor predictor score; μ_{ABC}, μ_{GCB} and $\sigma_{ABC}, \sigma_{GCB}$ represent the mean and standard deviations of the ABC and GCB subtypes as indicated in Table 2; and

wherein $\phi(\chi;\mu,\sigma)$ is the standard normal density calculated by

$$\varphi(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

3. The method of claim 2, wherein a score of P(PMBCL) $\geq$ 0.9 indicates that the sample is PMBCL, regardless of the P(ABC) score.

4. The method of claim 2, wherein a score of P(PMBCL) $\leq$ 0.1 and a score of P(ABC) $\leq$ 0.1 indicates that the tumor is GCB DLBCL.

5. The method of claim 2, wherein a score of P(PMBCL) $\leq$ 0.1 and a score of P(ABC) $\geq$ 0.9 indicates that the tumor is ABC DLBCL.

6. The method of claim 2, wherein a score of P(PMBCL) ≤ 0.1 and a score of P(ABC) greater than 0.1 and less than 0.9 indicates that the tumor is an unclassified DLBCL.

7. The method of any one of claims 1-6, wherein the RNA gene expression data is obtained using a NanoString Technologies® nCounter® assay.

Figure 1

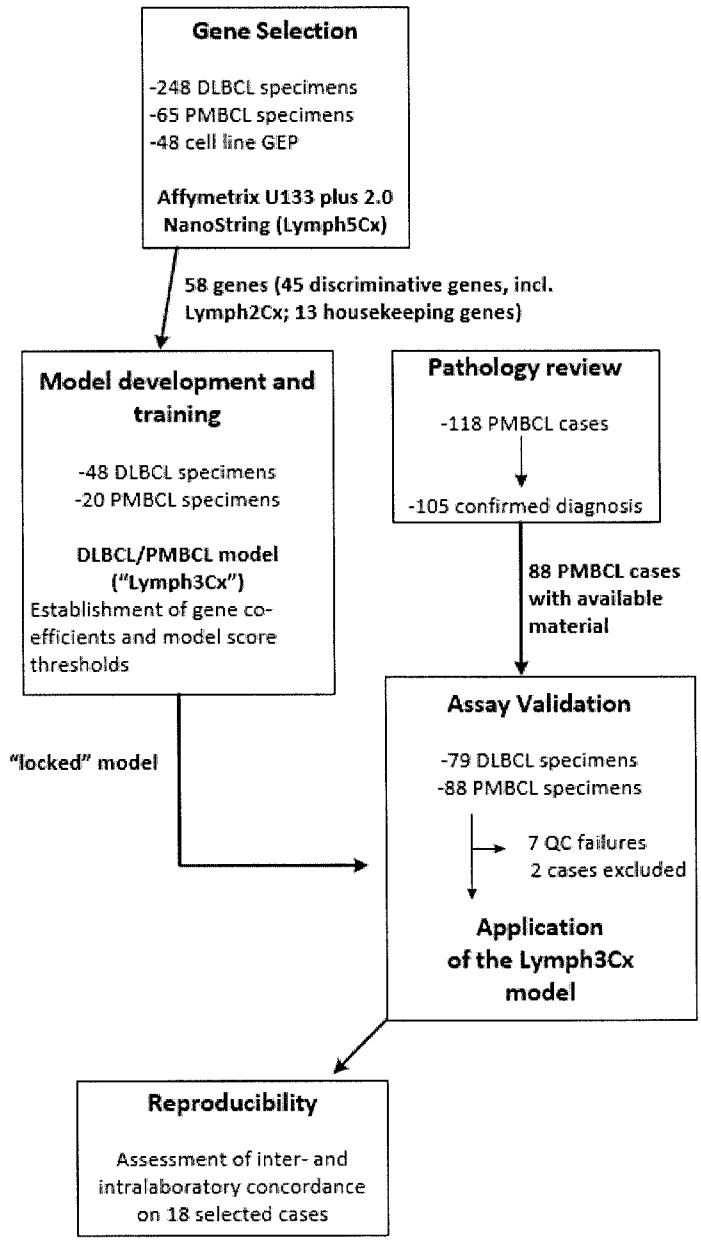

Gene Selection
-248 DLBCL specimens
-65 PMBCL specimens
-48 cell line GEP
Affymetrix U133 plus 2.0
NanoString (Lymph5Cx)
58 genes (45 discriminative genes, incl. Lymph2Cx; 13 housekeeping genes)
Model development and training
-48 DLBCL specimens
-20 PMBCL specimens
DLBCL/PMBCL model ("Lymph3Cx")
Establishment of gene co-efficients and model score thresholds
Pathology review
-118 PMBCL cases
-105 confirmed diagnosis
88 PMBCL cases with available material
"locked" model
Assay Validation
-79 DLBCL specimens
-88 PMBCL specimens
7 QC failures
2 cases excluded
Application of the Lymph3Cx model
Reproducibility
Assessment of inter- and intralaboratory concordance on 18 selected cases

Figure 2

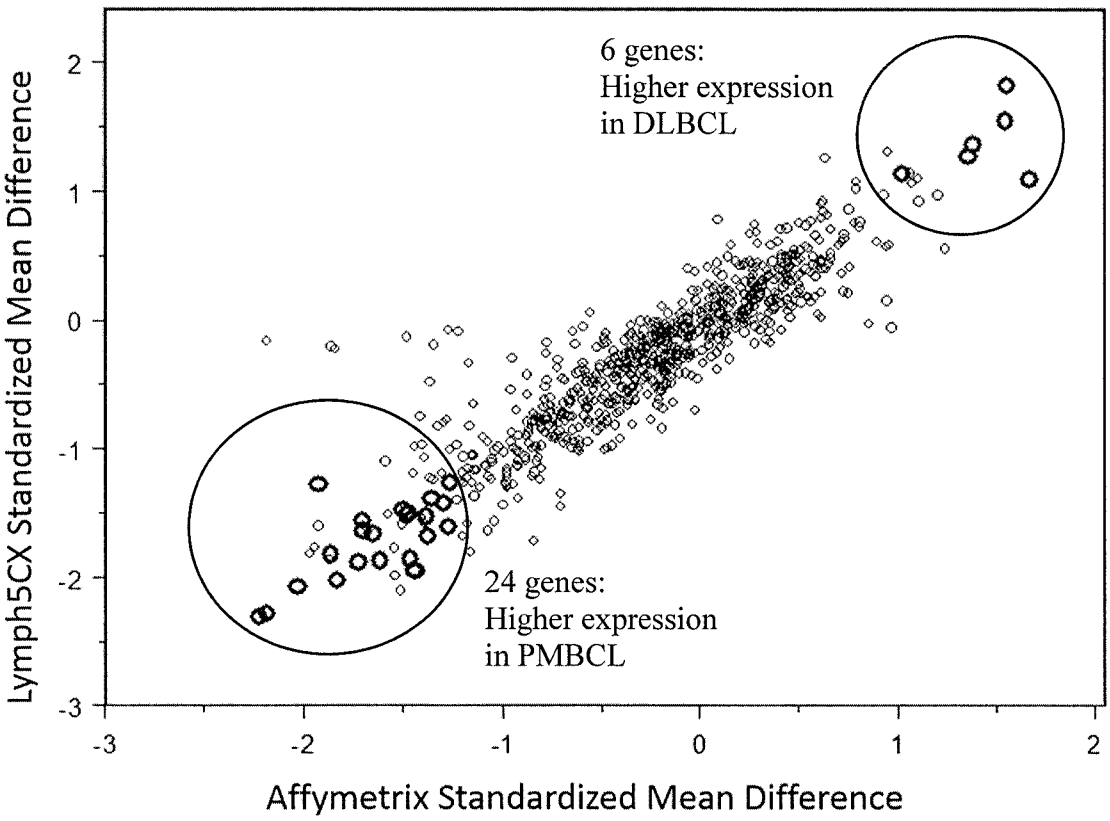

6 genes:
Higher expression
in DLBCL
24 genes:
Higher expression
in PMBCL
Lymph5CX Standardized Mean Difference
Affymetrix Standardized Mean Difference
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2

Figure 3

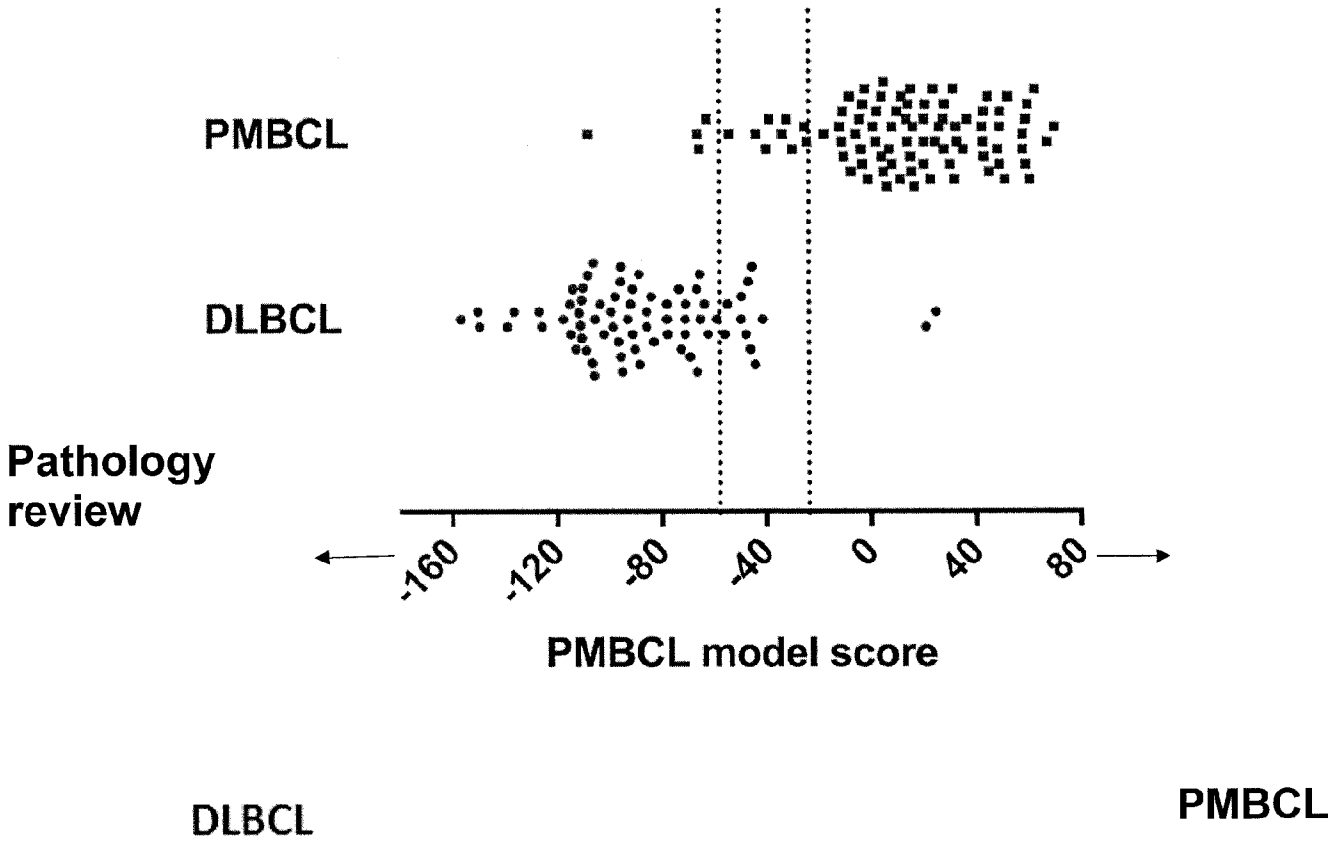

PMBCL
DLBCL
Pathology review
-160
-120
-80
-40
0
40
80
PMBCL model score
DLBCL
PMBCL

Figure 4

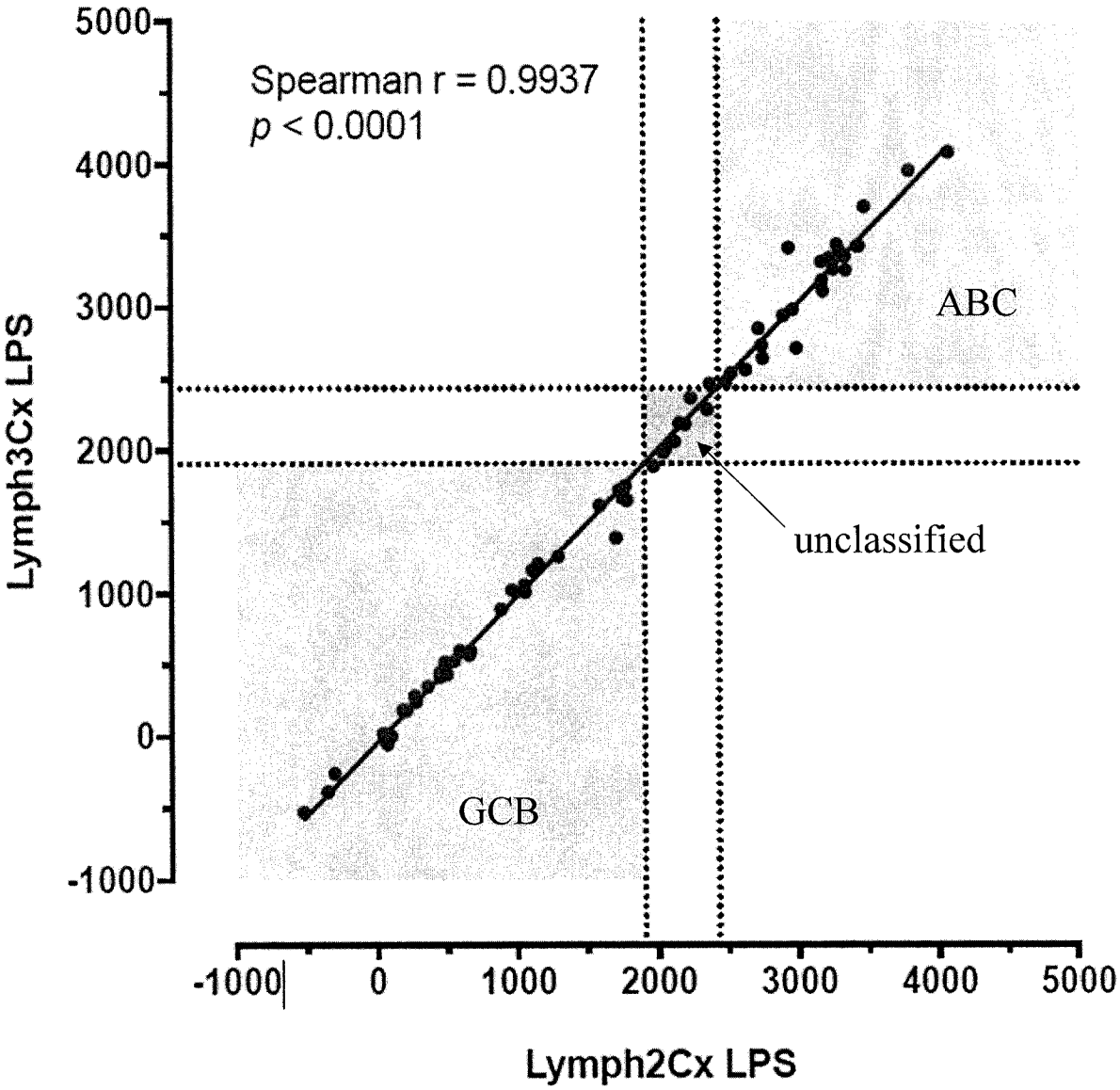
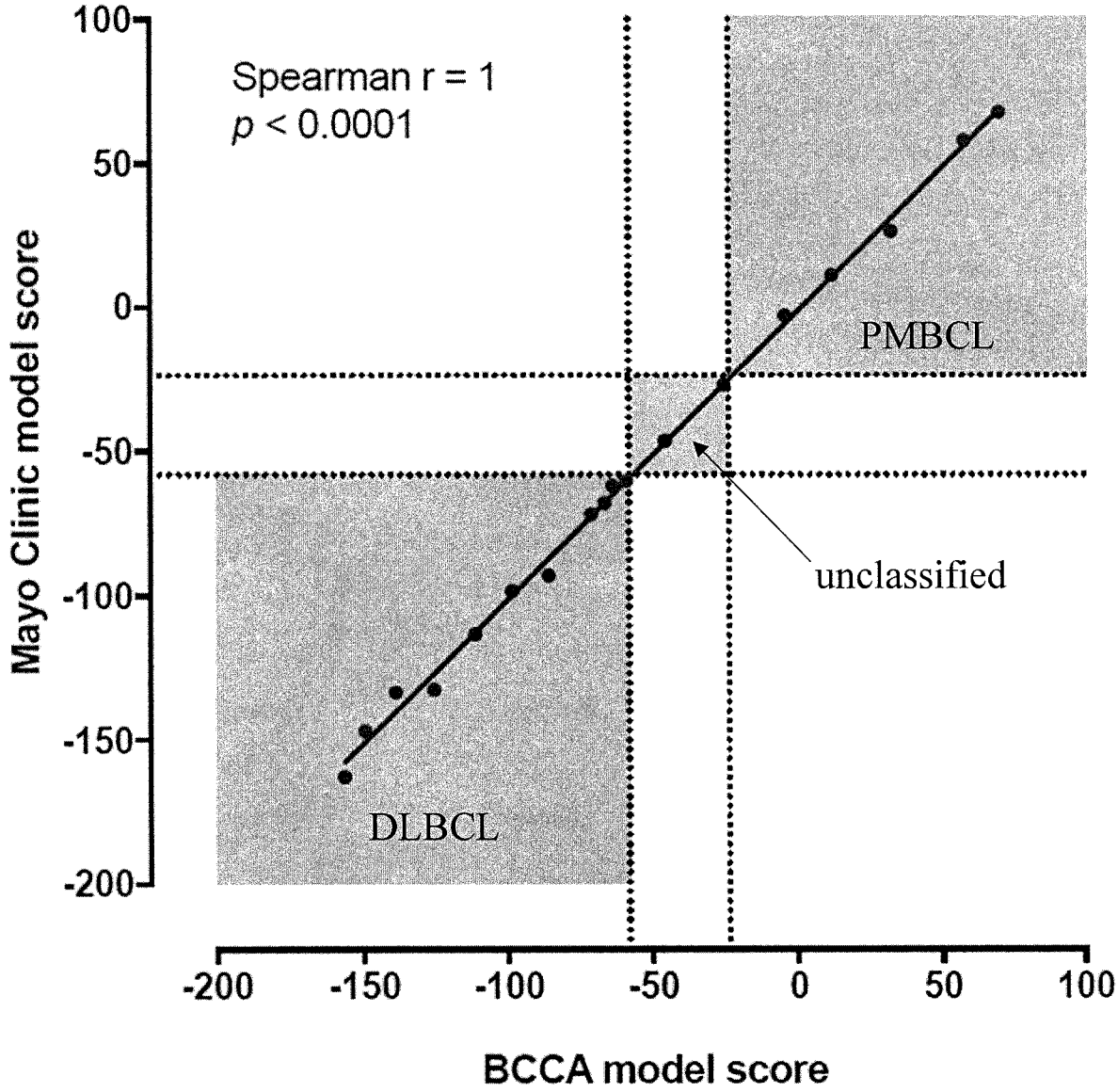

Spearman r = 0.9937
p < 0.0001
ABC
unclassified
GCB
Lymph3Cx LPS
Lymph2Cx LPS
5000
4000
3000
2000
1000
0
-1000
-1000
0
1000
2000
3000
4000
5000

Figure 5


Spearman r = 1
p < 0.0001
PMBCL
unclassified
DLBCL
Mayo Clinic model score
BCCA model score
100
50
0
-50
-100
-150
-200
-200
-150
-100
-50
0
50
100

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/036084

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6886
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 , WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2015/069790 Al (US HEALTH [US] ; BRITISH COLUMBIA CANCER AGENCY [CA] ; UNIV ARIZONA STAT) 14 May 2015 (2015-05-14) cited in the application abstract; claim 1; examples 1-2 paragraph [0054] ; table 2 -----	1-7
A	WO 2005/024043 A2 (NAT INST HEALTH [US] ; STAUDT LOUIS [US] ; WRIGHT GEORGE [US] ; TAN BRUCE) 17 March 2005 (2005-03-17) abstract; claims 39-95 pages 123-165 ; examples 13-17 ; tables 36-39 , 45 , 46, 56 ----- -/- ·	1-7

[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" documentwhich may throw doubts on priority claim(s) orwhich 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
3 September 2018	11/09/2018

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	Aguilera, Miguel

3

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.
A	A. MOTTOK ET AL: "MOLECULAR CLASSIFICATION OF PRIMARY MEDIASTINAL LARGE B CELL LYMPHOMA USING FORMALIN-FIXED, PARAFFIN-EMBEDDED TISSUE SPECIMENS - AN LLMPP PROJECT", HEMATOLOGICAL ONCOLOGY, vol. 35, 7 June 2017 (2017-06-07), pages 59-60, XP055495961, US ISSN: 0278-0232, D0I: 10.1002/hon.2437_46 abstract	1-7
A	D. W. SCOTT ET AL: "Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue", BLOOD, vol. 123, no. 8, 20 February 2014 (2014-02-20), pages 1214-1217, XP055495536, US ISSN: 0006-4971, D0I: 10.1182/blood-2013-11-536433 abstract; figure 1A	1-7
A	R0SENWALD A ET AL: "Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma", THE JOURNAL OF EXPERIMENTAL MEDICINE, ROCKEFELLER UNIVERSITY PRESS, US, vol. 198, no. 6, 15 September 2003 (2003-09-15), pages 851-862, XP002396795, ISSN: 0022-1007, D0I: 10.1084/JEM.20031074 abstract; figure 2A	1-7
A	LENZ G ET AL: "Stromal gene signatures in large-B-cell lymphomas", NEW ENGLAND JOURNAL OF MEDICINE, THE - NEJM, MASSACHUSETTS MEDICAL SOCIETY, US, vol. 359, no. 22, 27 November 2008 (2008-11-27), pages 2313-2323, XP002551776, ISSN: 1533-4406, D0I: 10.1056/NEJM0A0802885 abstract; figure 2	1-7

-/--

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	M. H. VELDMAN-JONES ET AL: "Reproducible, Quantitative, and Flexible Molecular Subtyping of Clinical DLBCL Samples Using the NanoString nCounter System", CLINICAL CANCER RESEARCH, vol. 21, no. 10, 9 October 2014 (2014-10-09), pages 2367-2378, XP055495781, US ISSN: 1078-0432, DOI: 10.1158/1078-0432.CCR-14-0357 abstract Supplemental Table 1 & Veldman-Jones ET AL: "Supplemental Table 1", Clinical Cancer Research 21(10): 2367-2378 (2014), 9 January 2014 (2014-01-09), XP055495789, Retrieved from the Internet: URL: http://clincancerres.aacrjournals.org/highwire/filestream/131853/field_highwire_adjunct_files/5/126978_2_supp_2633385_nfzhbv.pptx [retrieved on 2018-07-27] the whole document	1-7

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2018/036084

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015069790 A1	14-05-2015	AU 2014346788 A1	26-05-2016
		CA 2929826 A1	14-05-2015
		CN 106232831 A	14-12-2016
		EP 3066215 A1	14-09-2016
		JP 2017502686 A	26-01-2017
		KR 20160127713 A	04-11-2016
		US 2016283653 A1	29-09-2016
		WO 2015069790 A1	14-05-2015
WO 2005024043 A2	17-03-2005	AU 2004271192 A1	17-03-2005
		CA 2537254 A1	17-03-2005
		EP 1664762 A2	07-06-2006
		EP 2157524 A2	24-02-2010
		US 2005164231 A1	28-07-2005
		WO 2005024043 A2	17-03-2005