Supplemental Methods

Immunohistochemistry

Thin sections (<10 µm) of patients' formalin-fixed paraffin-embedded (FFPE) tissues were prepared for screening of protein expression of CD10 (clone 56C6, Dako), BCL6 (clone PG-B6p, Invitrogen), and MUM1 (clone MUM1p, Dako), using the Dako Autostainer Link 48 (Dako, Glostrup, Denmark), according to standard procedures. Protein expression by the tumor cells for these markers were estimated by hematopathologists and scored with a cutoff value of ≥30% for CD10, BCL6, and MUM1, according to Hans' algorithm.

Fluorescence in situ hybridization

For O-DLBCL, rearrangement of *MYC*, *BCL2*, and *BCL6* were assessed with Vysis Dual Color Break Apart Rearrangement Probes from Abbott using the Dako Histology FISH Accessory Kit, according to standard procedures. For NO-DLBCL-GCB cases, rearrangement-status of *MYC* was analyzed and if *MYC* was rearranged, *BCL2* and *BCL6* rearrangement were additionally analyzed. For the analysis, a minimum of 100 tumor cells were manually scored by expert pathologists and a split of the signals in ≥10% of the tumor cells was considered as rearranged as reported before.[1]

DNA isolation for sequencing

Genomic DNA and RNA of O-DLBCL, was isolated from fresh frozen (n=18) and from FFPE biopsies (n=85). NO-DLBCL-GCB cases were additionally isolated from fresh frozen (n=1) and FFPE biopsies (n=62). DNA from fresh frozen biopsies were isolated from 25µm cryosections with the QIAamp DNA Mini Kit (Qiagen) and RNA was isolated with TRIzol (Invitrogen) and RNeasy Mini Kit (Qiagen). FFPE biopsies were decalcified, followed by a fully automated DNA and RNA isolation from deparaffinized 10 µm sections with the Tissue Preparation System (TPS) robot (Siemens Healthcare Diagnostics), as described previously.[2] DNA concentrations were quantified with the Qubit™ dsDNA HS Assay Kit and RNA was quantified with the Qubit™ RNA BR Assay Kit with the Qubit™ Fluorometer (Invitrogen).

Gene-expression-profiling (GEP)

GEP was performed with a NanoString system and an extended custom-made probe set, covering 20 genes of the Lymph2Cx-assay for COO classification and additionally 219 genes related to DLBCL. The genes of this custom-made probe set were initially selected to validate five previously published gene expression signatures: Lymph2Cx classifier for COO calling[3,4], MYC activity GEP score[5], copy number variation (CNV) associated GEP signature[6], consensus clustering[7] and the immune-ratio signature[8]. At the recent EHA annual meeting 2020 (abstract #EP1303), the reproducibility of these published GEP classifier in DLBCL has been presented.

Targeted next-generation sequencing (tNGS) - LYMFv1 panel design

All cases were sequenced with the LYMFv1 panel. This targeted next-generation sequencing (tNGS) panel was compiled from a comprehensive review of ~300 articles published before December 2017, for frequency and clinical relevance (diagnostic, prognostic or therapeutic markers) of genetic mutations in diffuse large B-cell lymphomas (DLBCLs) and other B-cell lymphomas. The initial reviews were selected based on their focus on next-generation sequencing data regarding DLBCL and all genes with a prevalence higher than 5% were included in the design. The literature search was performed in the PubMed Database with the following search-parameters:

((((((((((((((((((CLL) OR "Leukemia, Lymphocytic, Chronic, B-Cell"[Mesh]) OR ((DLBCL) OR "Lymphoma, Large B-Cell, Diffuse"[Mesh])) OR ((FL) OR "Lymphoma, Follicular"[Mesh])) OR ((BL) OR "Burkitt Lymphoma"[Mesh])) OR ((MZL) OR "Lymphoma, B-Cell, Marginal Zone"[Mesh])) OR ((MM) OR "Multiple Myeloma"[Mesh])) OR ((MCL) OR "Lymphoma, Mantle-Cell"[Mesh])) OR (PCNSL)) OR (primary central nervous system lymphoma)) OR (PCDLBCL-LT)) OR (PC-DLBCL-LT)) OR (Primary cutaneous DLBCL, leg type)) OR ((MALT) OR "Lymphoma, B-Cell, Marginal Zone"[Mesh])) OR (NMZL)) OR (nodal marginal zone lymphoma)) OR (SMZL)) OR (Splenic marginal zone lymphoma)) OR (primary testicular lymphoma)) OR (PTL)) OR ((WM) OR "Waldenstrom Macroglobulinemia"[Mesh]) AND ((((((((((((sanger[All Fields] AND sequencing[All Fields]) OR sanger-sequencing[All Fields])))) OR PCR[All Fields]) OR sequencing[All Fields]) OR (("high-throughput nucleotide sequencing"[MeSH Terms] OR ("high-throughput"[All Fields] AND "nucleotide"[All Fields] AND "sequencing"[All Fields]) OR "high-throughput nucleotide sequencing"[All Fields] OR ("next"[All Fields] AND "generation"[All Fields] AND "sequencing"[All Fields]) OR "next generation sequencing"[All Fields] OR (“next-generation”[All Fields] AND “sequencing”[All Fields])))) OR ((ngs[All Fields] OR NGS[All Fields]))) OR (("whole genome sequencing"[MeSH Terms] OR ("whole"[All Fields] AND "genome"[All Fields] AND "sequencing"[All Fields]) OR "whole genome sequencing"[All Fields] OR (“whole-genome”[All Fields] AND “sequencing”[All Fields])))) OR ((wgs[All Fields] OR WGS[All Fields]))) OR (("whole exome sequencing"[MeSH Terms] OR ("whole"[All Fields] AND "exome"[All Fields] AND "sequencing"[All Fields]) OR "whole exome sequencing"[All Fields] OR (“whole-exome”[All Fields] AND “sequencing”[All Fields])))) OR ((wes[All Fields] OR WES[All Fields]))) OR (("mutation"[MeSH Terms] OR "mutation"[All Fields]))) OR Alteration[All Fields])

Following this search, ~260 original scientific reports with sequencing data of DLBCLs or other B-cell lymphomas were screened. Articles were selected based on the presence of sequencing data from B-cell lymphoma subtypes with more than 5 human cases. The data collected from these articles was used in designing a custom multiplex panel for FFPE DNA. Genes were selected based on their frequency and impact on diagnostics, prognostics and therapeutics for these malignancies. Supplemental table-8 lists the genes included in our diagnostic panel (LYMFv1) for tNGS and referenced publications. Genomic regions of interest were selected based on data from the articles, as well as biological databases as the Catalogue of Somatic Mutations in Cancer (COSMIC)[9]. The panel, called LYMFv1, was designed using the online Ion AmpliSeq Designer Tool (ThermoFisher Scientific), listing our regions of interest. The LYMFv1 panel covered 95.05% of our regions of interest from 52 different genes and consists of 1362 amplicons, divided over two primer pools.

Targeted next-generation sequencing procedure

Initially, libraries for sequencing were generated following manufacturer procedures. For PCR, 5 µL of 2.5 ng/µL sample DNA was mixed with 2 µL of 5x Ion AmpliSeq HiFi Mix (Ion AmpliSeq™ Library Kit 2.0; ThermoFisher Scientific 4480441), 2 µL of 5x LYMFv1 Ion AmpliSeq primer pool (ThermoFisher Scientific; IAD_143519_173), and 1 µL of Ion AmpliSeq™ ID panel (ThermoFisher Scientific; IAD_85912) for each pool of the LYMFv1 panel. Amplification was performed with a Bio-Rad C1000 Thermal Cycler with the following steps: 99 °C for 2 minutes, 22 cycles of 99 °C for 15 seconds and 60 °C for four minutes. After the PCR procedure both pools were combined and primers were digested by adding 2 µL of FuPa reagent (Ion AmpliSeq™ Library Kit 2.0) and heating in a Thermal Cycler at 50 °C for 10 minutes, 55 °C for 10 minutes, and 60 °C for 20 minutes. After the primer-digestion, samples were barcoded by adding 2 µL of the barcode primers (Ion AmpliSeq™ Library Kit 2.0), 4 µL of switch solution (Ion AmpliSeq™ Library Kit 2.0) and 2 µL of DNA ligase (Ion AmpliSeq™ Library Kit 2.0) and heated at 22 °C for 60 minutes and 72 °C for 10 minutes.

After barcoding, samples were purified in a TECAN robot with low TE solution and AMPure XP beads (Beckman Coulter; A63882), according to the library purification protocol. Purification was followed by qPCR using 5 µL of Ion Library TaqMan qPCR Master Mix (2x) and 0.5 µL of Ion Library TaqMan Quantitation Assay (20x; Ion Library TaqMan™ Quantitation Kit; ThermoFisher Scientific; 4468802) with 4.5 µL of 100 times diluted sample purified library. Together with a calibration curve of 10-, 100-, and 1000-times diluted *E. coli* (68 pM), qPCR was performed on a Bio-Rad S1000 thermal Cycler with the following settings: 50 °C for 2 minutes, 95 °C for 20 seconds, followed by 32 cycles of 95 °C for 3 seconds and 60 °C for 20 seconds. Based on qPCR data, samples were normalized and pooled and subsequently another qPCR was performed with similar conditions. A final concentration of 60 pM pooled libraries was used for sequencing preparation with the Ion Chef™ instrument (ThermoFisher Scientific; 4484177) with the Ion 540™ Chip Kit (ThermoFisher Scientific; A27765). Sequencing was performed with the Ion S5™ Sequencing platform (ThermoFisher Scientific; A27212), according to the manufacturer's protocol (ThermoFisher Scientific).

After sequencing with the LYMFv1 panel, the sequencing data was aligned to the human reference genome (GRCh37/hg19) with the TMAP 5.0.7 software (default parameters, https://github.com/iontorrent/TS).[10] Variants were called by the Torrent Variant Caller (Thermo Fisher Scientific) and added to the Geneticist Assistant NGS Interpretive Workbench (SoftGenetics) for functional annotation by a clinical molecular biologist. Variants with a population frequency of above 1% in the 1000 Genomes Project or variants in three DNA mixtures of 'healthy' individuals (n=288; sequenced during validation) were assigned as a single nucleotide polymorphism (SNP). These SNPs, together with sequencing-artifacts induced by homopolymeric regions and variants with a high strand bias (>90%), were excluded from further analysis. All remaining variants with a variant allele frequency (VAF) of ≥10% and read depth of ≥100 reads, were categorized based upon their pathogenicity: Class 1 - benign, class 2 - likely benign, class 3 - unknown significance, class 4 - likely pathogenic, and class 5 - pathogenic. Classification was performed with data from dbSNP[11], ClinVar[12], COSMIC[9], and available literature.

All variants of unknown pathogenicity (n=861) were discriminated in not-pathogenic or pathogenic based on the following prediction-score protocol: Variants with a CADD-phred score >25 were classified as likely-pathogenic (n=251), variants with a CADD-phred score between 10 and 25 were only classified as likely-pathogenic (n=256) if two or more prediction scores identified the variant as deleterious (SIFT, Polyphen2_HDIV, LRT, and MutationTaster). Remaining variants were either designated as not-pathogenic (n=338) or variants of unknown significance (n=16). Frameshift and splice-site mutations with unknown significance (n=46) were subsequently analyzed using additional information from the franklin.genoox.com website, assigning 40 likely-pathogenic, 5 unknown, and 1 not-pathogenic variants. The variants of unknown significance were not included in the oncoprint plot.

Targeted next-generation sequencing validation of the LYMFv1 panel

Since June 2018, the LYMFv1 panel was successively implemented in the routine diagnostics of B-cell lymphomas at the pathology department of the LUMC (ISO 15189 certified). Before implementation, the LYMFv1 tNGS panel and the used bioinformatic analytical pipeline were validated and optimized with various procedures. *First,* three mixtures of DNA from 96 blood samples of 'healthy' volunteers (n=288) were sequenced with the LYMFv1 panel to identify potential non-pathogenic variants, such as SNPs occurring frequently in the residing population of the immediate area. All identified potential non-pathogenic variants were excluded from further analysis. *Second,* as whole-exome-sequencing (WES) data was available from eight DLBCL cell lines and seven primary follicular lymphomas (FL), these samples were consecutively sequenced with the LYMFv1 panel. Variants identified by both sequencing methods were mutually compared to validate the variant identification of the LYMFv1 panel. Of these 15 samples, our tNGS/LYMFv1 bioinformatic pipeline identified 214 unique variants, and 173 unique variants with identical target-regions were determined by a WES bioinformatic pipeline. In total 72% of variants identified by tNGS/LYMFv1 were concordant with WES-analysis and vice versa 92% of WES-identified variants overlapped with the tNGS/LYMFv1-determined variants. The mean read count of tNGS (>1000 reads) was much higher than WES (~122 reads) and as such low occurring polyclonal variants were missed with WES but identified with tNGS. Furthermore, the discrepancy of 8% between variants identified by WES and tNGS were mostly found in *PIM1* and *BCL2*. Both genes are susceptible to somatic hypermutation that induces multiple variants and as such creates possible issues for an amplicon-based (tNGS) sequencing methods. These variants were mainly located on the primer binding region (PBR) of a LYMFv1 amplicon. Variants on PBR can influence the binding of the primer and impeding appropriate tNGS sequencing. As such, 8% of the missing WES-identified variants could be explained, resulting in a total concordance of 100%. *Third,* nine DLBCL cases were initially sequenced with another so-called 'cancer hotspot panel'[10], and subsequently analyzed with the LYMFv1-panel. This 'cancer hotspot panel' is used by the Pathology department for sequencing of solid tumors and 12 genes of this panel overlap with the LYMFv1 panel. The cancer hotspot panel identified 67 unique variants of which 100% were also identified using the LYMFv1 panel.

In conclusion, based on these results, we concluded that the validity of the LYMFv1 panel was appropriate.

Supplemental Results

Radiological assessments at diagnosis

A PET(-CT) scan was performed in most patients (n=125, 75%) and combined with MRI or skeletal scintigraphy in 25 and 4 cases, respectively. The CT scan was done in 36 cases (22%), and in 12 and 6 cases respectively combined with MRI or skeletal scintigraphy. MRI only was performed on 3 patients (2%). A conventional X-ray as only staging procedure was executed in two relatively older cases (1%) diagnosed in 2004 and 2007 in a palliative setting.

High-grade B-cell lymphoma and double expressor by IHC

The majority of (N)O-DLBCL cases (n=118; 81%) were successfully analyzed for *MYC*-FISH, demonstrating 24 positive (16%), 94 negative (64%) and 28 failed cases (20%). IHC for MYC, of the 28 cases with unknown (or failed) *MYC*-FISH status, demonstrated a negative (n=2) or a positive expression (n=7) or unknown status (n=19; no adequate tissue left). Five of the seven cases with MYC-IHC positivity also showed positive expression for BCL2 and can be classified as dual expressors (DE), including two PB-DLBCL, two disseminated-DLBCL, and one NO-DLBCL-GCB. In O-DLBCL, 42 cases (51%) were evaluated for both IHC MYC and BCL2. Besides a low frequency of *MYC*-rearrangements (n=8; 10%) in O-DLBCL, the overall contribution of DE in this cohort is relatively small (~20%, 18 DE cases of which 14 are *MYC*-FISH negative), as would be expected given the superior prognosis. Additionally, this study is overrepresented with early stage DLBCL cases, this corroborates with a study of Barraclough *et al.*, demonstrating that DE status did not have an impact on survival in DLBCL with Ann Arbor stage I/II and treated with R-CHOP-like therapy.[13]

Survival of young PB-DLBCL patients with tibia localization or localizations in the lower extremities

In contrast to a previous report,[22] young PB-DLBCL patients (≤40 years) were not significantly more likely to have a localization in the lower than upper limbs of the appendicular skeleton compared to older PB-DLBCL patients (>40 years; 92% vs 67%; P=0.193). Accordingly, no difference in survival was observed between patients with lower or upper extremities in each age group (≤40: P=0.120, >40: P=1).

Sensitivity analysis

The inclusion of high-grade B-cell lymphoma (HGBCL) with *MYC*- and *BCL2*- and/or *BCL6*-rearrangement cases could significantly contribute to a specific molecular profile and inferior survival and thus potentially bias the identified results. A total of ten HGBCL cases were excluded in a sensitivity analysis. Concerning molecular profiles, the four most frequently mutated genes in PB-DLBCL (*B2M*, *EZH2*, *IRF8*, and *TNFRS14*) remained significantly different compared to NO-DLBCL-GCB (P=0.025, P=0.009, P=0.047, and P<0.001, respectively). For survival outcomes, after excluding 4 HGBCL cases, PB-DLBCL retained their superior OS compared to NO-DLBCL-GCB with Ann Arbor I/II (P=0.034). Likewise, no significant difference in OS was found between O-DLBCL and NO-DLBCL-GCB with Ann Arbor stage III/IV when HGBCL cases (n=6) were excluded (P=0.17). The survival data are consistent with two independent studies, demonstrating a superior OS for limited Ann Arbor Stage I(/II) of DLBCL with a double or triple hit phenotype as compared to advanced stage (II/)III/IV.[23,24] Considering this sensitivity analysis and the two mentioned studies, we concluded that the inclusion of a relatively small number of HGBCL did not significantly bias our results.

Multivariate analysis

The heterogeneity in patient characteristics and therapies might have biased our survival outcomes. To investigate this potential confounding in our results, we performed a Cox-regression analysis focusing on the identified results between PB-DLBCL and NO-DLBCL-GCB subgroups. The combination of PB-DLBCL and NO-DLBCL-GCB subgroups resulted in a relatively small number of events (n=18), limiting the number of confounders in the model to achieve a reliable model fit. The most likely confounding factors gender, age (younger or older than 60 years), and treatment type (R-CHOP/R-CHOP-like, other non-R-CHOP-like chemotherapy, or palliative treatment) were separately fitted in to the Cox model together with disease subtype (PB-DLBCL and NO-DLBCL-GCB) (Supplemental Table-8). Comparing Cox-models with and without confounders shows a drop in HR from 4.1 to 3.7, indicating that the effect of age and treatment type is roughly independent of disease subtype (PB-DLBCL vs NO-DLBCL-GCB). This justifies the comparison of smaller subgroups without correction for confounders. Fitting patients with either R-CHOP or other chemotherapy into the model with disease subtype and adjuvant radiotherapy showed no significant influence of adjuvant radiotherapy if corrected for disease subtype. The role of adjuvant radiotherapy for PB-DLBCL remains controversial as a small retrospective cohort study showed prolonged survival outcomes for patients treated with a combined modality (R-CHOP + radiotherapy),[25] whereas two independent studies recently demonstrated that adjuvant radiotherapy did not improve survival outcomes for limited-stage (Ann Arbor I/II) DLBCL patients in general.[26,27]

References

1. Vermaat JS, Somers SF, de Wreede LC, et al. MYD88 mutations identify a molecular subgroup of diffuse large B-cell lymphoma with an unfavourable prognosis. *Haematologica.* 2019.
2. van Eijk R, Stevens L, Morreau H, van Wezel T. Assessment of a fully automated high-throughput DNA extraction method from formalin-fixed, paraffin-embedded tissue for KRAS, and BRAF somatic mutation analysis. *Exp Mol Pathol.* 2013;94(1):121-125.
3. Scott DW, Mottok A, Ennishi D, et al. Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin Determined by Digital Gene Expression in Formalin-Fixed Paraffin-Embedded Tissue Biopsies. *J Clin Oncol.* 2015;33(26):2848-2856.
4. Scott DW, Wright GW, Williams PM, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood.* 2014;123(8):1214-1217.
5. Carey CD, Gusenleitner D, Chapuy B, et al. Molecular classification of MYC-driven B-cell lymphomas by targeted gene expression profiling of fixed biopsy specimens. *J Mol Diagn.* 2015;17(1):19-30.
6. Chan FC, Telenius A, Healy S, et al. An RCOR1 loss-associated gene expression signature identifies a prognostically significant DLBCL subgroup. *Blood.* 2015;125(6):959-966.
7. Monti S, Savage KJ, Kutok JL, et al. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood.* 2005;105(5):1851-1861.
8. Keane C, Vari F, Hertzberg M, et al. Ratios of T-cell immune effectors and checkpoint molecules as prognostic biomarkers in diffuse large B-cell lymphoma: a population-based study. *Lancet Haematol.* 2015;2(10):e445-455.
9. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 2019;47(D1):D941-D947.
10. Cohen D, Hondelink LM, Solleveld-Westerink N, et al. Optimizing Mutation and Fusion Detection in NSCLC by Sequential DNA and RNA Sequencing. *J Thorac Oncol.* 2020;15(6):1000-1014.
11. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* 1999;9(8):677-679.
12. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062-D1067.
13. Barraclough A, Alzahrani M, Ettrup MS, et al. COO and MYC/BCL2 status do not predict outcome among patients with stage I/II DLBCL: a retrospective multicenter study. *Blood Adv.* 2019;3(13):2013-2021.
14. Jawad MU, Schneiderbauer MM, Min ES, Cheung MC, Koniaris LG, Scully SP. Primary lymphoma of bone in adult patients. *Cancer.* 2010;116(4):871-879.
15. Pellegrini C, Gandolfi L, Quirini F, et al. Primary bone lymphoma: evaluation of chemoimmunotherapy as front-line treatment in 21 patients. *Clinical lymphoma, myeloma & leukemia.* 2011;11(4):321-325.
16. Wu H, Bui MM, Leston DG, et al. Clinical characteristics and prognostic factors of bone lymphomas: focus on the clinical significance of multifocal bone involvement by primary bone large B-cell lymphomas. *BMC cancer.* 2014;14:900.
17. Bruno Ventre M, Ferreri AJ, Gospodarowicz M, et al. Clinical features, management, and prognosis of an international series of 161 patients with limited-stage diffuse large B-cell lymphoma of the bone (the IELSG-14 study). *The oncologist.* 2014;19(3):291-298.
18. Tao R, Allen PK, Rodriguez A, et al. Benefit of consolidative radiation therapy for primary bone diffuse large B-cell lymphoma. *International journal of radiation oncology, biology, physics.* 2015;92(1):122-129.
19. Messina C, Christie D, Zucca E, Gospodarowicz M, Ferreri AJ. Primary and secondary bone lymphomas. *Cancer treatment reviews.* 2015;41(3):235-246.

20. Li X, Xu-Monette ZY, Yi S, et al. Primary Bone Lymphoma Exhibits a Favorable Prognosis and Distinct Gene Expression Signatures Resembling Diffuse Large B-Cell Lymphoma Derived From Centrocytes in the Germinal Center. *The American journal of surgical pathology.* 2017;41(10):1309-1321.
21. Cleven AHG, Hogendoorn PCW. Hematopoietic Tumors Primarily Presenting in Bone. *Surg Pathol Clin.* 2017;10(3):675-691.
22. Subik MK, Herr M, Hutchison RE, et al. A highly curable lymphoma occurs preferentially in the proximal tibia of young patients. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2014;27(11):1430-1437.
23. Oki Y, Noorani M, Lin P, et al. Double hit lymphoma: the MD Anderson Cancer Center clinical experience. *British journal of haematology.* 2014;166(6):891-901.
24. Petrich AM, Gandhi M, Jovanovic B, et al. Impact of induction regimen and stem cell transplantation on outcomes in double-hit lymphoma: a multicenter retrospective analysis. *Blood.* 2014;124(15):2354-2361.
25. Pilorge S, Harel S, Ribrag V, et al. Primary bone diffuse large B-cell lymphoma: a retrospective evaluation on 76 cases from French institutional and LYSA studies. *Leukemia & lymphoma.* 2016;57(12):2820-2826.
26. Poeschel V, Held G, Ziepert M, et al. Four versus six cycles of CHOP chemotherapy in combination with six applications of rituximab in patients with aggressive B-cell lymphoma with favourable prognosis (FLYER): a randomised, phase 3, non-inferiority trial. *Lancet.* 2019;394(10216):2271-2281.
27. Lamy T, Damaj G, Soubeyran P, et al. R-CHOP 14 with or without radiotherapy in nonbulky limited-stage diffuse large B-cell lymphoma. *Blood.* 2018;131(2):174-181.
28. Dybkaer K, Bogsted M, Falgreen S, et al. Diffuse large B-cell lymphoma classification system that associates normal B-cell subset phenotypes with prognosis. *J Clin Oncol.* 2015;33(12):1379-1388.
29. Michaelsen TY, Richter J, Brondum RF, et al. A B-cell-associated gene signature classification of diffuse large B-cell lymphoma by NanoString technology. *Blood Adv.* 2018;2(13):1542-1546.
30. Tripodo C, Zanardi F, Iannelli F, et al. A Spatially Resolved Dark- versus Light-Zone Microenvironment Signature Subdivides Germinal Center-Related Aggressive B Cell Lymphomas. *iScience.* 2020;23(10):101562.

Supplemental Tables

Supplemental Table-1 – Individual clinical characteristics of all patients: age at diagnosis, Ann Arbor stage, IPI-score, extranodal (non-)osseous localization, first line therapy, imaging technique, biopsy type. Furthermore, important information regarding sequencing data: tumor cell percentage, average amplicon read count, and transition/transversion ratio of all variants and variants above 0.1 allele frequency per sample of all patients with O-DLBCL and NO-DLBCL-GCB sub-entities

Supplemental Table-2 - LYMFv1 tNGS panel

The LYMFv1-panel amplicons used for sequencing of lymphoma samples. The exact chromosomal location and corresponding gene of all 1362 amplicons are listed.

Supplemental Table-3 - Cell-of-origins of O-DLBCL and NO-DLBCL-GCB

Cell-of-origins of O-DLBCL and NO-DLBCL, as classified by the Hans' criteria with immunohistochemistry or the Lymph2Cx-assay with NanoString.

Supplemental Table-4 - NanoString raw data

Raw data as generated by the NanoString of the O-DLBCL and NO-DLBCL-GCB cases.

Supplemental Table-5 - tNGS identified variants

Variants identified in O-DLBCL and NO-DLBCL, which passed the sequencing quality criteria, and filtered for rare variants (found in <1% of the population databases) and variants occurring in a local healthy population. Pathogenicity is classified into five classes ranging from non-pathogenic to pathogenic. Furthermore, class 3 – unknown pathogenicity and other unknown pathogenic variants were classified into benign or pathogenic variants by prediction scores (CADD, SIFT, Polyphen, LRT, and MutationTaster).

Supplemental Table-6 - Univariate survival analysis

Univariate survival analysis of patient characteristics and molecular aberrations from all O-DLBCL and NO-DLBCL-GCB sub-entities.

Supplemental Table-7 - O-DLBCL and literature-based molecular clusters

Allocation of O-DLBCL sub-entities and NO-DLBCL-GCB, based on mutational profiles associated to literature-based molecular clusters.

Supplemental Table-8 - Cox-regression models of PB-DLBCL and NO-DLBCL-GCB

The most likely confounding factors, gender, age (younger or older than 60 years), and treatment type (R-CHOP/R-CHOP-like, non-R-CHOP-like chemotherapy, or palliative treatment), were separately fitted in the Cox model together with disease subtype (PB-DLBCL vs NO-DLBCL-GCB). The P-value of PB-DLBCL vs NO-DLBCL-GCB show that the effect of disease subtype remains, during correction for gender, age, or therapy.

Supplemental Table-9 - Articles used for design of the tNGS panel

Frequencies found of B-cell lymphomas in ~300 articles used to design the LYMFv1 (diagnostic) panel and the BLYMF200 (Research) panel for targeted NGS. All genes are listed with the percentage range in which it occurs in each B-cell lymphoma according to the articles screened. Additionally, the articles from which this data was gathered are indicated with an x in their corresponding column and the row of the corresponding gene.

Supplemental Figures

Supplemental Figure-1 - Principal component analysis of gene expression from NanoString with the Lymph2Cx extended probeset, based on six individual measurements.

Each dot represents a sample and are categorized based on their presence on one of the six measurements (A, B, C) or if the sample belongs to an O-DLBCL subgroup or NO-DLBCL-GCB group (D). The PCA was used to identify outliers and 'batch effect'. Outliers were identified by an 'abnormal' distance from the main group of samples and consequently removed from further analysis. Additionally, differences in gene expression might be induced using cartridges of NanoString and imbalanced randomized loading of the samples on these cartridges. As such, PCA can be overlaid with information regarding the cartridge allocation of each sample, to identify potential different clusters based on cartridges.

A) Includes all measured samples and one sample had a largely different GEP compared to all other samples and as such, #24 - PB-DLBCL has been removed as outlier.

B) Includes all samples except for the previous outlier, demonstrating another outlier (#85 - NO-DLBCL-GCB) which was additionally removed from further analysis.

C) Excluding both outliers demonstrated no significant batch effect based on the six different measurements with PCA, subsequently all cases were included for further analysis.

D) Categorization of three O-DLBCL subgroups and NO-DLBCL-GCB with PCA showed that based on GEP NO-DLBCL-GCB appears to differ from O-DLBCL, and the O-DLBCL subgroups seems to conglomerate together.

E) PCA with division of fresh frozen and FFPE tissue types.

Abbreviations: DLBCL, diffuse large B-cell lymphoma; GCB, germinal center B-cell like; NO-DLBCL-GCB, non-osseous DLBCL of the GCB phenotype; O-DLBCL, DLBCL with osseous involvement; PB-DLBCL, primary bone DLBCL; PCA, principal component analysis.

Supplemental Figure-2 - Heatmap of four clusters identified by unsupervised hierarchical clustering of the extended NanoString dataset (234 genes).

This clustering analysis provided four different clusters, that partially overlaps with WHO-defined O-DLBCL subgroups and NO-DLBCL-GCB. A significant difference was found between cluster A, allocating eight PB-DLBCL and one NO-DLBCL-GCB, and cluster B, with three PB-DLBCL and 12 NO-DLBCL-GCB (P=0.001). Cluster C contained several O-DLBCLs, but no NO-DLBCL-GCBs and cluster D was an agglomeration of several O-DLBCL subtypes and NO-DLBCL-GCBs. Disseminated-DLBCL was observed across all four clusters, indicating its heterogeneity and wide variety in disease origins of individual cases.

A) Heatmap of all endogenous genes included in the NanoString analysis, which separates four clusters based on gene-expression data.

B) Barplot of the (N)O-DLBCL composition of each cluster. The barplots show a separation between PB-DLBCL and NO-DBLCL-GCB based on gene-expression.

Abbreviations: DLBCL, diffuse large B-cell lymphoma; GCB, germinal center B-cell like; NO-DLBCL-GCB, non-osseous DLBCL of the GCB phenotype; O-DLBCL, DLBCL with osseous involvement; PB-DLBCL, primary bone DLBCL; WHO, world health organization.

Supplemental Figure-3 - Differential gene expression of centroblast and centrocyte genes.

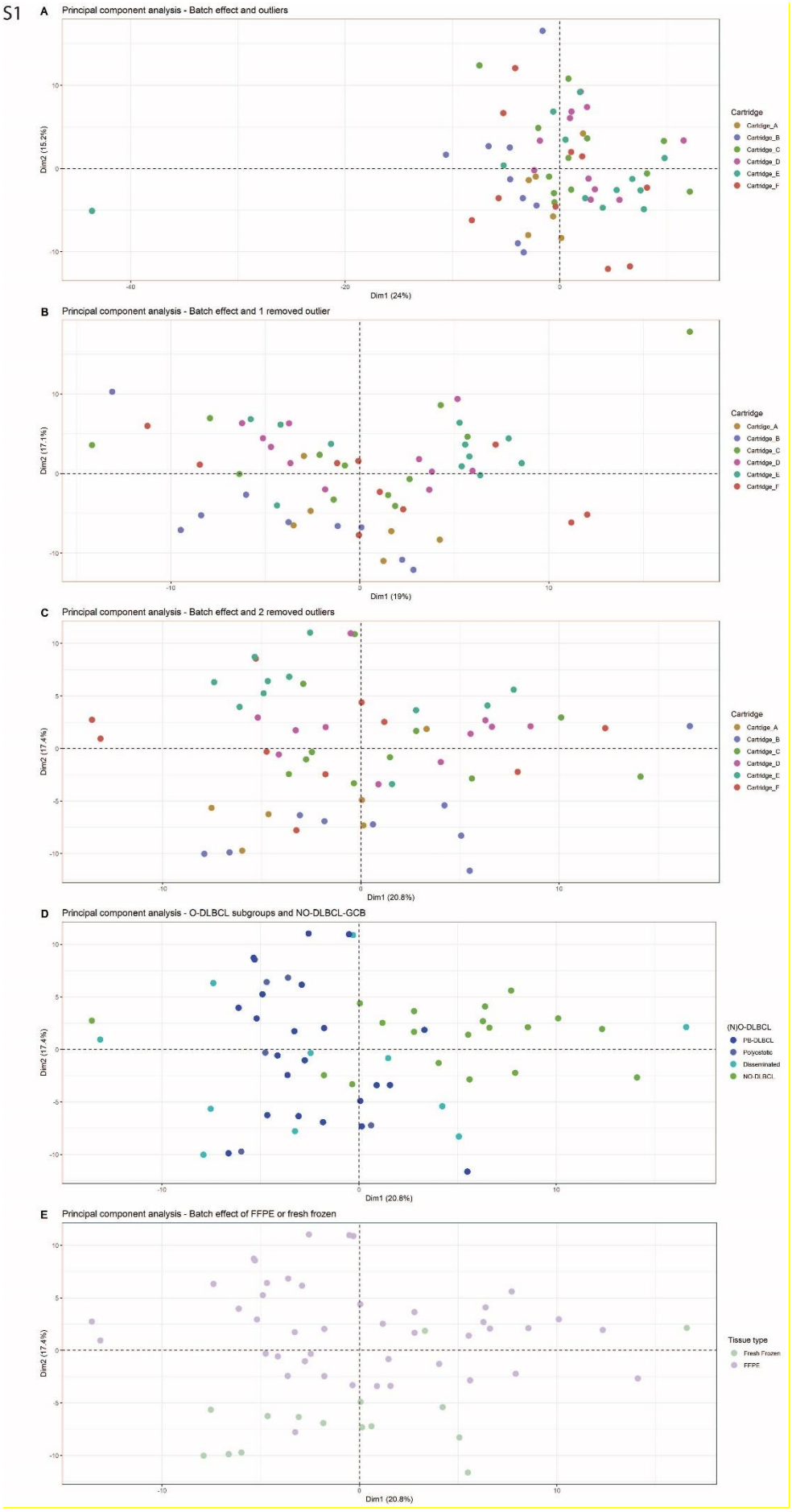
A) Boxplots of 13 genes with differential gene expression between centroblast and centrocyte B cells, that overlapped with the DNA-microarray panel from Dybkaer, *et al*. (J. Clin. Oncol. 2015)[28] and NanoString panel from Michaelsen, *et al*. (Blood Adv. 2018)[29]. Anova statistical testing was performed to identify significant differences in gene expression between the three clusters identified by GEP.

B) The genes with their corresponding expression levels (from the BAGS(2CLINIC)-assays) in centroblast and centrocyte are listed in this table.

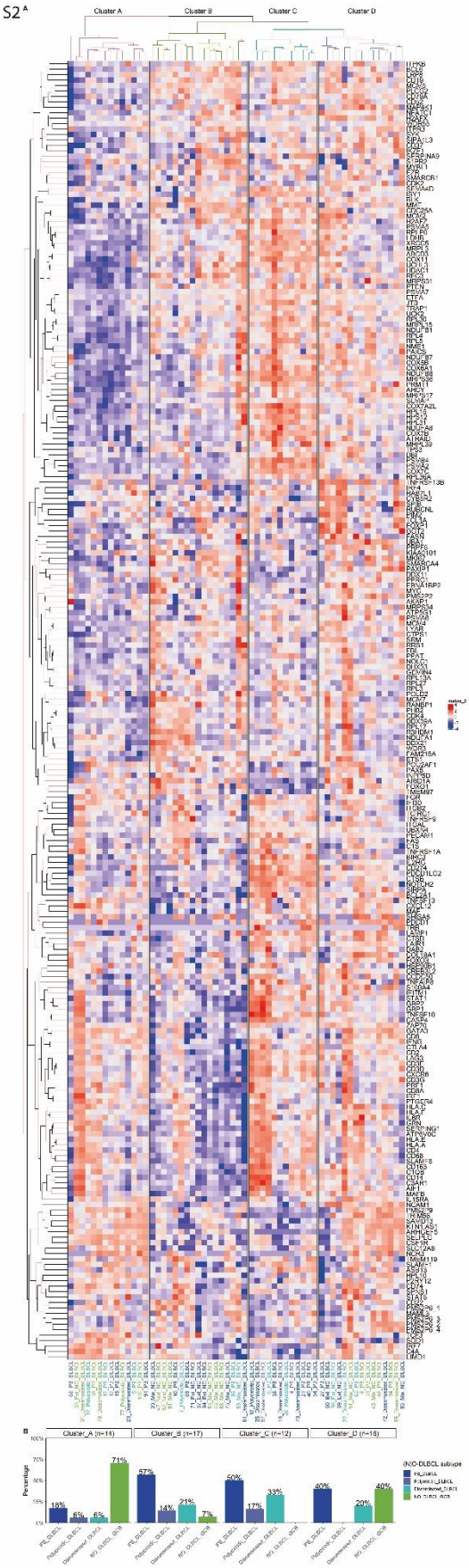
C) Boxplot of 11 genes with differential gene expression from a recently published article by Tripodo, *et al*. (IScience 2020)[30] in which they identified 53 differentially expressed genes between dark and light zone of the germinal center.

D) The genes with corresponding expression levels from DZ/LZ spatial signature for classification of centrocyte and centroblast cells are listed in the table.

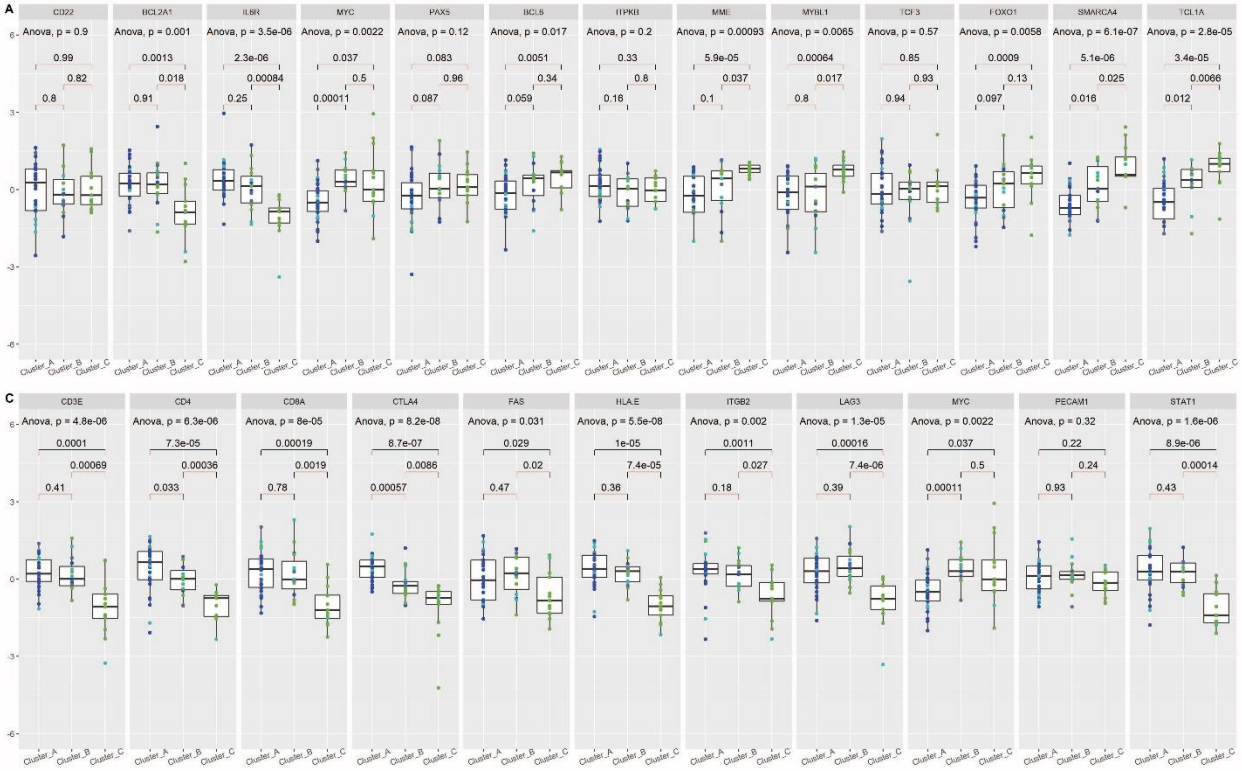
S1


A
Principal component analysis - Batch effect and outliers
Dim2 (15.2%)
Dim1 (24%)
Cartridge
Cartridge_A
Cartridge_B
Cartridge_C
Cartridge_D
Cartridge_E
Cartridge_F
B
Principal component analysis - Batch effect and 1 removed outlier
Dim2 (17.1%)
Dim1 (19%)
Cartridge
Cartridge_A
Cartridge_B
Cartridge_C
Cartridge_D
Cartridge_E
Cartridge_F
C
Principal component analysis - Batch effect and 2 removed outliers
Dim2 (17.4%)
Dim1 (20.8%)
Cartridge
Cartridge_A
Cartridge_B
Cartridge_C
Cartridge_D
Cartridge_E
Cartridge_F
D
Principal component analysis - O-DLBCL subgroups and NO-DLBCL-GCB
Dim2 (17.4%)
Dim1 (20.8%)
(N)O-DLBCL
PB-DLBCL
Polycystic
Disseminated
NO-DLBCL
E
Principal component analysis - Batch effect of FFPE or fresh frozen
Dim2 (17.4%)
Dim1 (20.8%)
Tissue type
Fresh Frozen
FFPE

S2 A


Cluster A
Cluster B
Cluster C
Cluster D
B
Cluster_A (n=14)
Cluster_B (n=17)
Cluster_C (n=12)
Cluster_D (n=15)
Percentage
(N)O-DLBCL subtype
PE_DLBCL
Plasmablastic_DLBCL
Denovo_DLBCL
MYC DLBCL_BCL6
18%
6%
6%
71%
57%
14%
21%
7%
50%
17%
33%
40%
20%
40%

S3



B

Genes	Centroblast	Centrocyte
CD22	Low	High
BCL2A1	Low	Average
IL6R	Low	Average
MYC	Low	Average
PAX5	Average	High
BCL6	High	Average
ITPKB	High	Average
MME	High	Average
MYBL1	High	Average
TCF3	High	Average
FOXO1	High	Low
SMARCA4	High	Low
TCL1A	High	Low

D

Genes	Centroblast	Centrocyte
CD3E	Low	High
CD4	Low	High
CD8A	Low	High
CTLA4	Low	High
FAS	Low	High
HLA.E	Low	High
ITGB2	Low	High
LAG3	Low	High
MYC	Low	High
PECAM1	Low	High
STAT1	Low	High