Supplemental Material

Ramis-Zaldivar *et al*

Supplemental Methods

Library preparation SureSelectXT and Targeted sequencing approach

A total of 55 DNA samples, including 47 extracted from formalin fixed paraffin embedded (FFPE) tissue and 8 frozen materials, were processed using SureSelectXT (Agilent Technologies, Santa Clara, CA) a custom panel interrogating 96 genes, that was designed according to previous literature in DLBCL and other B-cell lymphomas in both adult and children population (**Supplemental Table 2**). Data from four cases have been previously published.[1] A total of 100ng of genomic DNA were sheared using the Covaris S220 focused-ultra sonicator (Covaris, Woburn, MA) to a target peak size of 150–200 bp. Library preparation was performed using SureSelectXT Custom Capture Library baits as described in SureSelectXT Target Enrichment System protocol (Agilent Technologies inc.). For amplification of the post capture libraries, 10 to 13 cycles were performed depending on the initial sample quality. The libraries were qualified using the Bioanalyzer HS (Agilent technologies inc.), quantified with the KAPA Library Quantification Kit (Kapa Biosystems, Wilmington, Massachusetts) and sequenced in a MiSeq instrument (Illumina, San Diego, CA) in a paired-end run of 150 bp. The average sequencing coverage of the 55 LBCL cases across regions was 447x (range 28-1439x) and over 84% of the targeted regions were covered by at least 100 reads.

FASTQ files were generated by MiSeq control software and quality control of the raw data was performed using the FastQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequencing reads were subsequently aligned to the human reference genome (GRCh37/hg19) using the Burrows-Wheeler Aligner–MEM algorithm.[2] Variant calling was performed using two different variant callers, Somatic Variant Caller (Illumina inc.) and Mutect2 (Genome Analysis Toolkit (GATK), version 4.0.3)[3] and variants were annotated using the VariantStudio software v3.0 and ANNOVAR respectively.[4] For Somatic Variant Caller (Illumina inc.), default settings were used to analyze sequencing results and to call the variants. Low quality or low coverage calls (total depth <20) were excluded. For Mutect2 variants, low quality variants were also excluded using FilterMutectCalls (GATK) with default thresholds. Only variants identified by both

algorithms were considered. We excluded variants affecting non-interrogated regions and known polymorphisms described in the GnomAD, 1000 Genomes and/or ExAC database (release 2015) with more than 0.1% frequency in normal population. In order to exclude artifacts, each variant was also inspected with the Integrative Genomics Viewer (IGV, Broad Institute, version 2.3) software. We ended up with 781 mutations including synonymous variants (**Supplemental Figure 1**).

Since DNA was extracted from FFPE or frozen tissue, FFPE DNA quality was evaluated using a qPCR in which fragments of 100bp and 200bp were amplified using primers previously described[5] and Fast SYBR Green Master Mix (Life Technologies inc.). ΔCt value was calculated for each sample comparing CT value for tumor DNA with the Ct value of gDNA control. The higher was the value, the more degraded was the FFPE DNA.

Driver prediction by mutation effect

Since no germline DNA was available and in order to select somatic variants, potential driver mutations were predicted according to previously published criteria[6] in which the 90% of the mutations classified as functional were demonstrated to be somatic. Inclusion criteria were: 1) variants described previously as somatic or functional on previous reports or COSMIC, 2) all truncating variants (nonsense, frameshift, splice donor or acceptor mutations; and 3) the remaining missense variants that were predicted to be functionally deleterious using Mutation Assessor[7] or SIFT predictor if a score was not provided by Mutation Assessor.[8] Other functional predictors as Polyphen-2 (Polymorphism Phenotyping-2)[9], CADD (Combined Annotation Dependent Depletion)[10] and CHASM-3.1[11] were also applied.

Mutational signatures

Mutational signatures contribution was interrogated for both all variants found in 47 primary LBCL and per frequently mutated genes using a fitting approach (MutationalPatterns R package).[12] Only mutational signatures described in adult DLBCL[13,14] corresponding to SBS1,

SBS2, SBS3, SBS4, SBS5, SBS8, SBS9, SBS12, SBS17a, SBS17b, SBS18, SBS19, SBS28 and SBS.C1 [15] were considered. Only single nucleotide variants not reported as polymorphisms, including synonymous and MYC intron 1, were considered for this analysis.

Signature bleeding between genes was corrected by iteratively removing the least contributing signature (if removal reduced the cosine similarity less than 0.01) with the exception of age-related signatures (SBS1 and SBS5) which were always included due to its known presence in all normal and tumor samples.

Aberrant somatic hypermutation (aSHM) hallmarks

Hallmarks of aSHM were studied in frequently mutated genes (>10 single nucleotide variants including both driver and non-driver predicted mutations, synonymous and MYC-intronic) and included location within AID-target region (defined as 2kb region after 150bp from a transcript start site), a higher transition over transversion ratio and location within AID-motifs (corresponding to sequences WA/TW/WRCY/RGYW/WGCW).[16]

Significance value was calculated comparing the observed proportion of variants with the expected by randomness using Test of Equal or Given Proportions. Therefore, proportion of variants in AID-target region was compared with the expected proportion of interrogated bases that occur in AID-target region, transition over transversion ratio compared with the expected 1/2 ratio and proportion of variants in AID-motifs compared with the expected proportion of interrogated bases within AID-target region that occur in AID-motif.

Verification using Ampliseq: Library preparation, amplification and quality control of unenriched libraries

A total of 37 samples including two relapsed samples were processed (17 DLBCL, 17 LBCL-*IRF4* and 3 HGBCL, NOS) using AmpliSeq (Illumina inc) with a custom panel interrogating recurrent regions of mutation in our series affecting 29 genes (**Supplemental Table 3**). In brief, each DNA probe was mixed with AmpliSeq HiFi Mix and one of the two primer pools, containing

201 or 198 amplicons, respectively. PCR was then transferred to a thermocycler (Applied Biosystems, Foster City, CA). PCR cycling conditions were as follows: Initial denaturation: 99°C for 2 min, cycling: 20 cycles of 99°C, 15 sec and 60°C, 4 min. After the end of the PCR reaction, primer end sequences were partially digested using FuPa reagent, followed by the ligation of Illumina dual sequencing indexes. The final library was purified using AMPure XP magnetic beads (Beckman Coulter, Krefeld, Germany) and quantified using qPCR and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA). Libraries were diluted to the same concentration (4nM) and an equimolar pool was prepared for each run (n=12 and n= 30, respectively). Each 4nM pool was diluted and load onto MiSeq sequencer, according to Illumina's protocol. The analysis pipeline applied for the identification of *BCL6* mutations in intron 1 was the same described in supplemental Figure 1.

It has been already described in the literature, that long indels or SNV in genes affected by aSHM are not appropriately amplified by PCR-based libraries as Ampliseq system due to an increase mismatch of the primers used for the amplification or to alterations of the amplicon size.[17] These technical aspects of the PCR based amplification result in a lower-than-expected coverage of the amplicon and increased incorrect assessment of variant allele frequencies. This primer mismatch is not only applicable to Ampliseq but also to conventional PCR amplification used Sanger Sequencing analysis. We have then excluded all non verified variants in genes known to be affected by aSHM (*IRF4, SOCS1, PIM1, SGK1 and BCL2*) or variants affected by long indels in the calculation of our verification rate.

Copy number analysis

DNAs were hybridized on Oncoscan FFPE, SNP6 or Cytoscan array (ThermoFisher Scientific inc.) depending of the tissue. Gains and losses and copy number neutral-loss of heterozygosity (CNN-LOH) regions were evaluated and visually inspected using Nexus Biodiscovery version 9.0 software (Biodiscovery, Hawthorne, CA). Human reference genome was GRCh37/hg19. The copy number alterations (CNAs) with minimum size of 100 kb and CNN-LOH larger than 5

Mb were considered informative. Published Cytoscan CN data from 116 adult DLBCL were used for age related analysis.[6]

Nanostring PanCancer Immune Profiling Panel

RCC files from the NanoString Digital Analyzer were imported into nSolver4.0 software (NanoString, Seattle, WA, USA) and checked for data quality using default QC settings. In the 21 samples (11 LBCL-*IRF4* and 10 DLBCL, NOS) that passed QC, differential expression (DE) and gene set analyses were performed using the nCounter Advanced Analysis (version 2.0.115)(NanoString Technologies inc.). Data normalization was done using the geNorm algorithm, automatically performed by the software. Genes with <20 counts in 100% of the cases were removed for differential expression analysis. *P*-values associated with the fold change were derived using the Benjamin-Yekutieli FDR method. Those genes with a log2FC>±1 and FDR< .05 were considered to be differentially expressed.

Supplemental Results

DNA and RNA quality analysis according to sample age

Time frame sample collection was from 1993 to 2019. In our series, we observed that the FFPE antiquity affected DNA quality according to qPCR values. The FFPE antiquity and the DNA quality according to qPCR values did not affect the number of variants or the C>T changes rate **(Supplemental Figure 2)**. The antiquity affected the coverage, nevertheless 83% of the FFPE samples had ≥90% of the regions covered at ≥50x. Regarding RNA when we correlated binding density quality score from nSolver software (NanoString inc.) with sample antiquity, more ancient samples tend to have less binding density.

Mutational signatures analysis

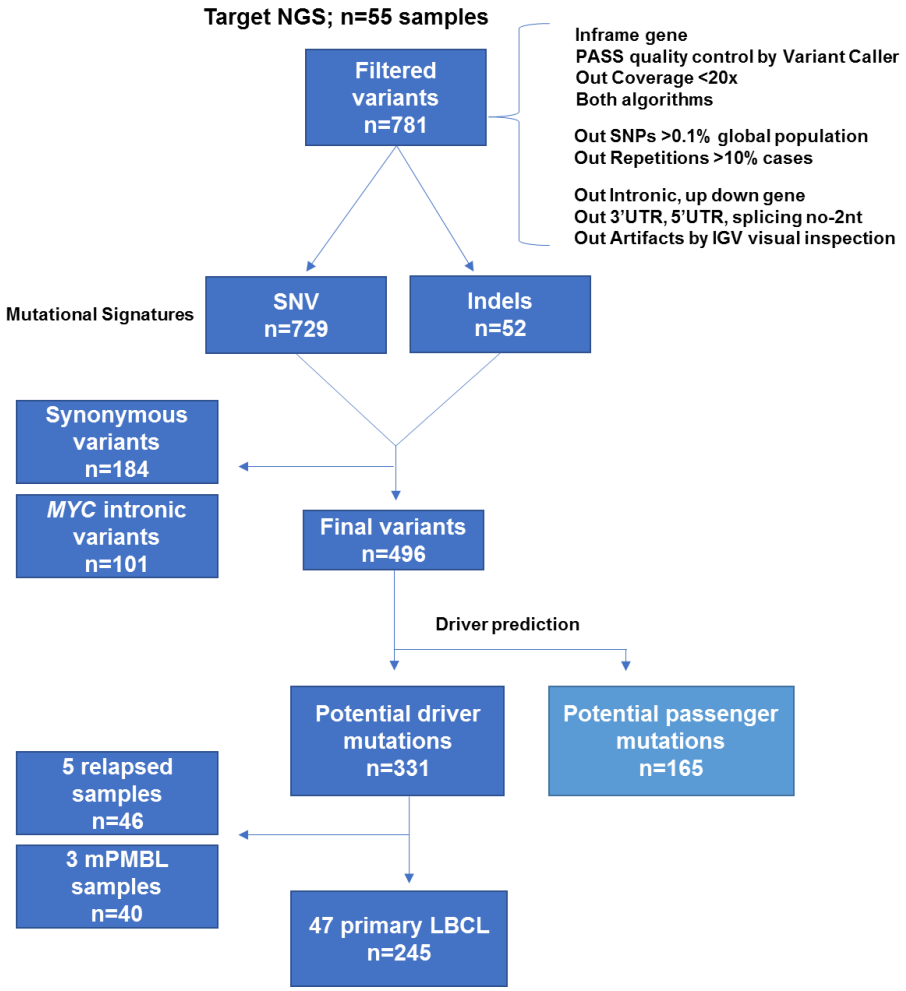
The relative contribution of mutational signatures previously described in DLBCL[13,14] was investigated for the global cohort of 47 primary LBCL (415 variants including driver, synonymous and *MYC* intron1). Three mutational signatures were identified, including the canonical activation-induced cytidine deaminase (cAID) (SBS.C1 in 43% of the variants) and two age related signatures (SBS1 and SBS5 in 10 and 47% respectively) with a cosine similarity of 0.92. When the relative contribution of these three mutational signatures was investigated for genes with at least 10 mutations, AID signature was detected in genes previously known[18] to be targets of aSHM such as *SOCS1*, *IRF4, PIM1, SGK1* and *MYC* (**Supplemental Figure 5, Supplemental Table 9**).

RHOA mutational analysis by Sanger sequencing

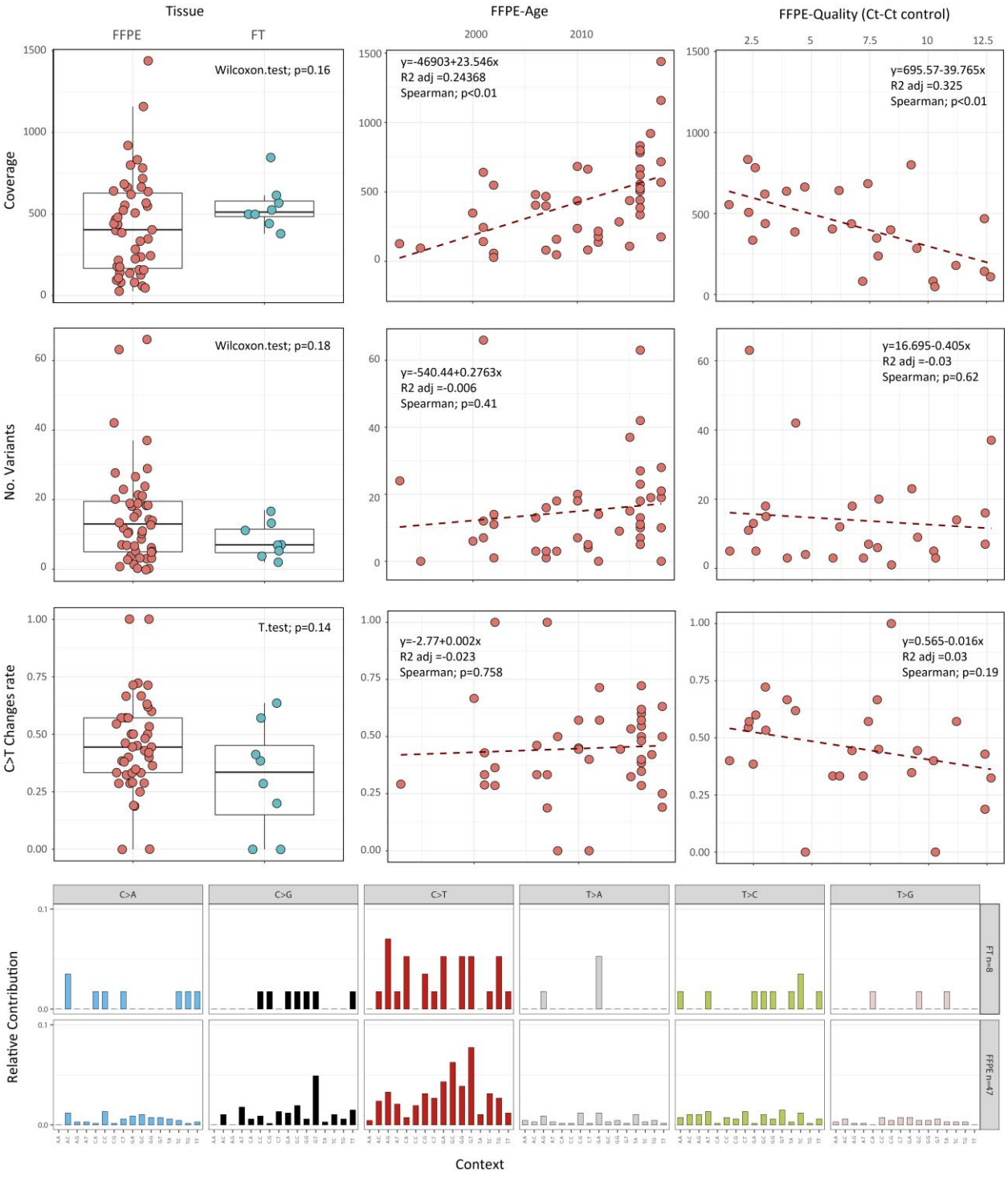
Due to the presence of *GNA13* mutations in 14% and 13% of DLBCL and HGBCL, NOS of our series respectively, we hypothesized that another gene of the same pathway as *RHOA* could be also mutated in our series. Mutations of *RHOA* have been rarely observed in DLBCL (1-3%).[13,19,20] but is highly recurrent in BL.[21] We then performed Sanger sequencing of *RHOA* exon 2 and 3 in 16 cases (9 DLBCL, 5 HGBCL, NOS and 2 DLBCL relapsed samples with DNA available). Only one DLBCL case presented a SNV in 5R>W position.

Supplemental Figures

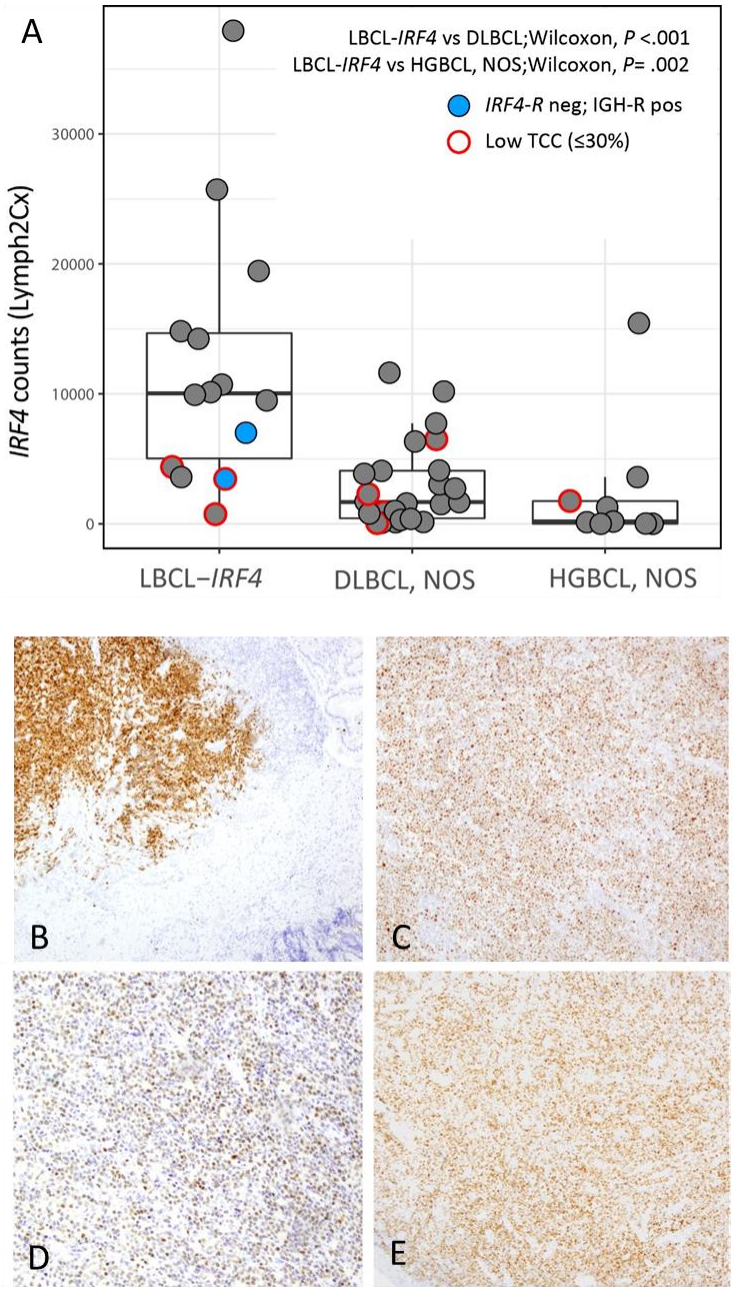
Supplemental Figure 1. Next generation sequencing (NGS) analysis pipeline followed to identify potential driver mutations in the 55 large B-cell lymphomas analyzed. Two different variant callers were used, Somatic Variant Caller (Illumina inc.) and Mutect2 (GATK version 4.0.3). Potential driver mutations were predicted according to previously published criteria.[6]



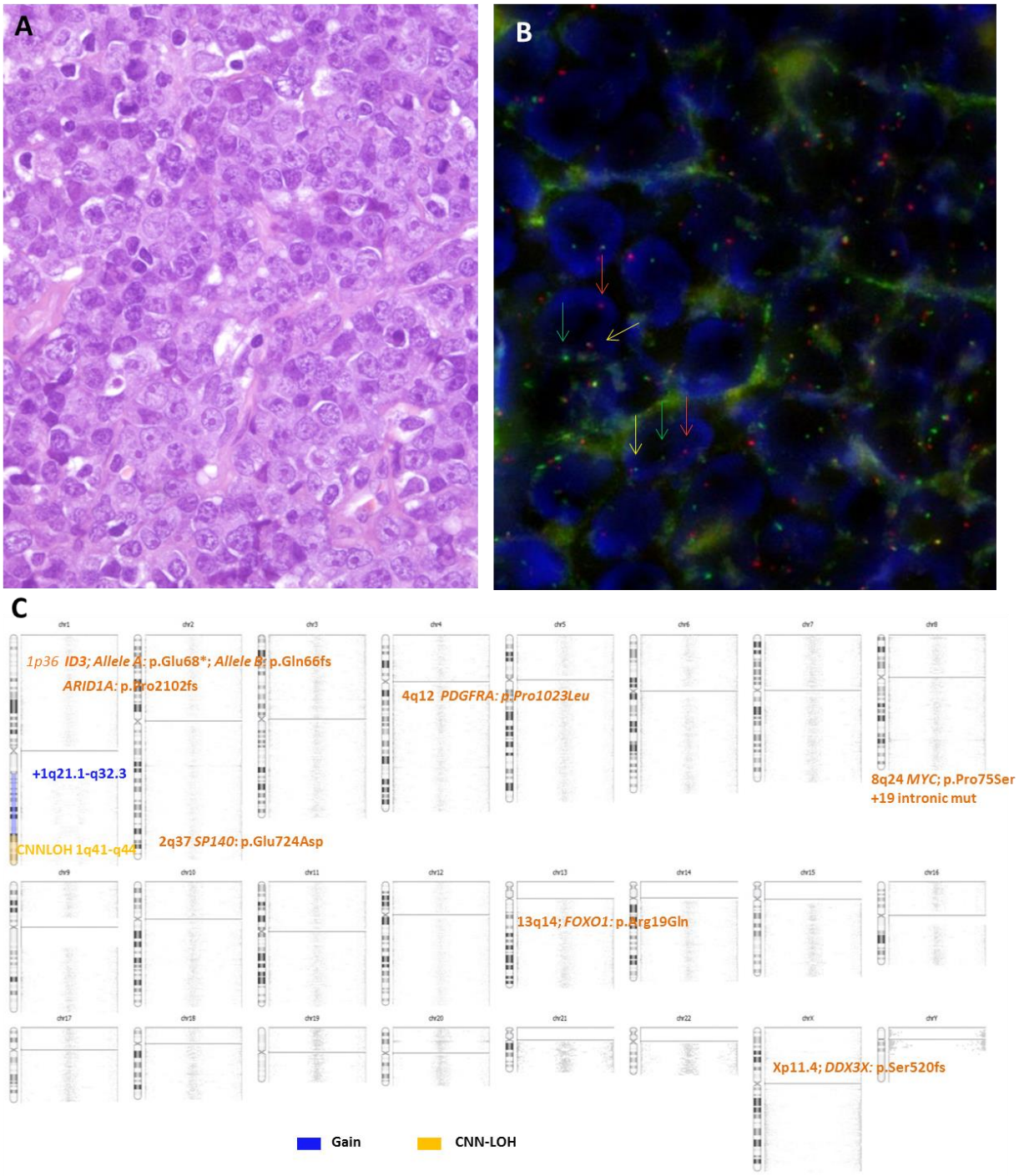
Supplemental Figure 2. Quality control of NGS data after filtering according to pipeline. **(A)** Coverage, **(B)** number of variants and **(C)** C<T proportion were assessed in terms of tissue type (FFPE and FT DNA samples) and FFPE DNA antiquity and quality. **(D)** Relative contribution of SNV changes of 47 FFPE and 8 frozen DNA samples. FFPE: formalin-fixed paraffin-embedded; FT: frozen tissue.



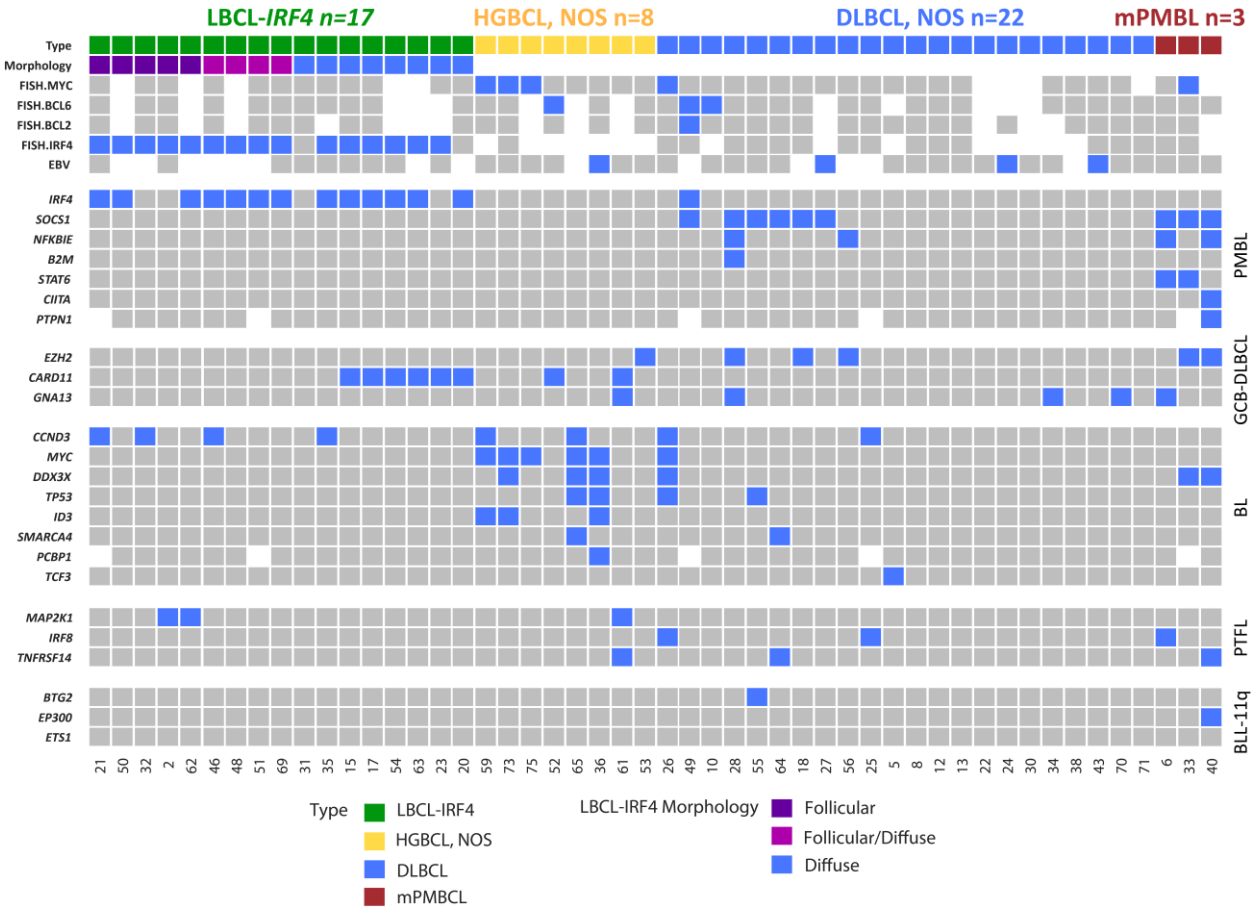
Supplemental Figure 3. *IRF4* expression at RNA and protein level between morphological groups. **(A)** Boxplot representing *IRF4* gene expression (number of counts of IRF4_NM_002460.1 on Lymph2Cx assay) between LBCL-*IRF4* (n=14), DLBCL, NOS (n=25) and HGBCL, NOS (n=9). *IRF4*-R: *IRF4* rearrangement; IGH-R: IGH rearrangement; neg: negative; pos: positive; TCC: tumor cell content. Immunostaining of MUM1/IRF4 from **(B-C)** D20 and D31 which are *IRF4*-negative IGH-positive LBCL-*IRF4* cases, **(D-E)** cases D17 and D15 which are *IRF4*-positive LBCL-*IRF4*.



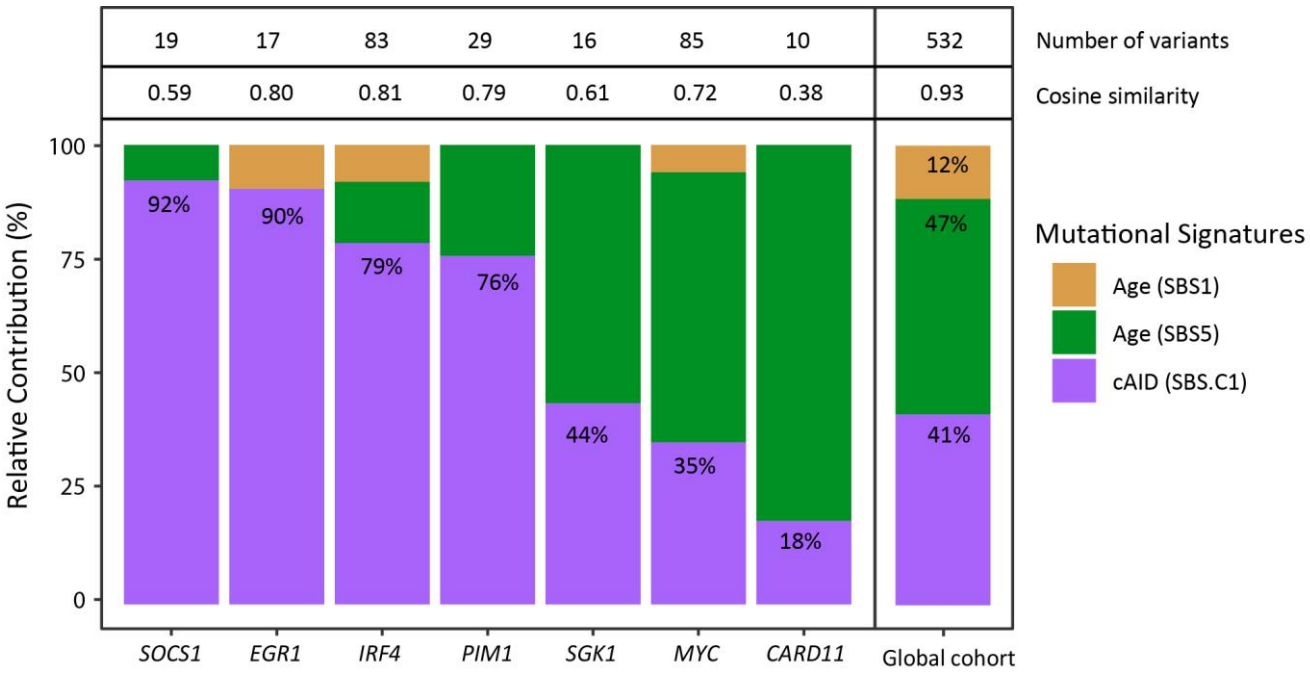
Supplemental Figure 4. Morphological, immunophenotypical and genetic features of a high grade B-cell lymphoma, NOS with *MYC* rearrangement. (Case D73) **(A, H&E)** showing cytological features intermediate between DLBCL and BL and **(B)** FISH with *MYC* break-apart depicting a signal constellation of one colocalization (yellow arrow) and one split signal (red and green arrows), and **(C)** ideogram of the copy number, copy number neutral-loss of heterozygosity (CNN-LOH) and mutational features of this case.



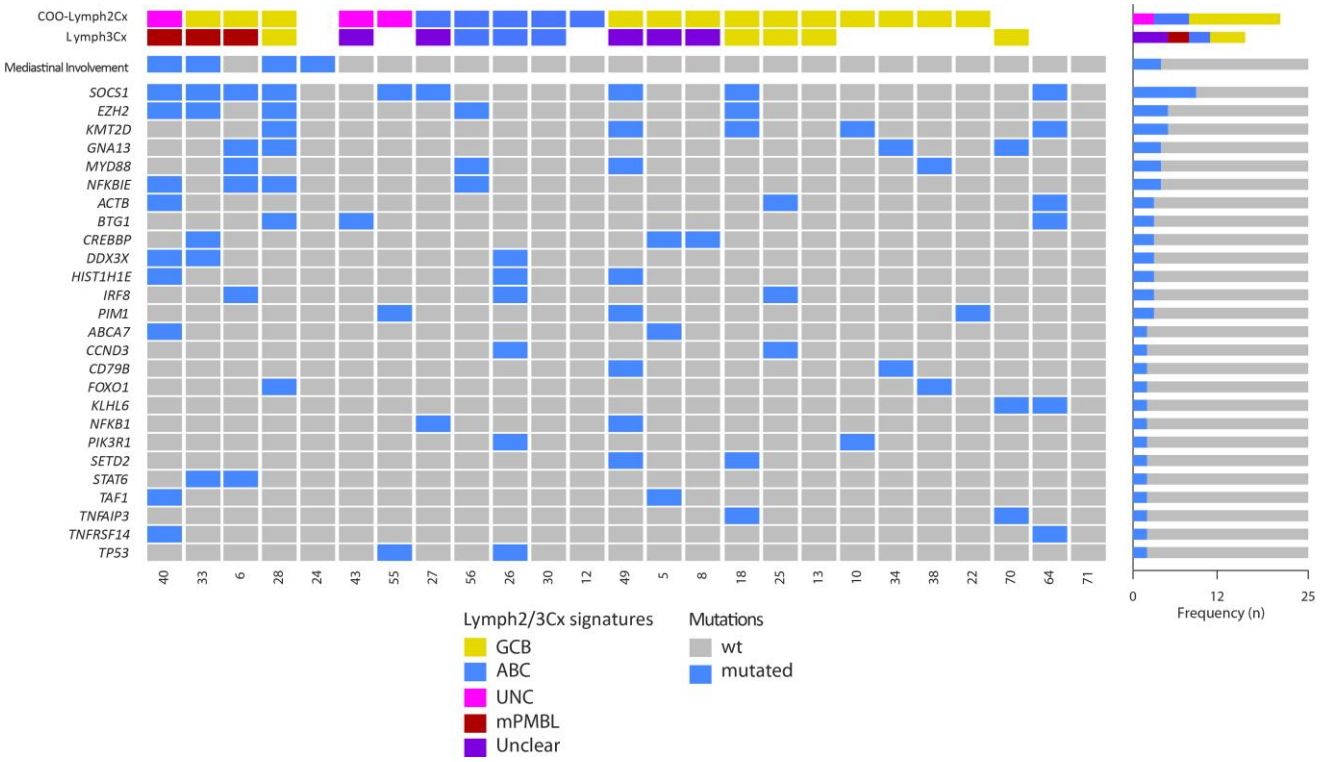
Supplemental Figure 5. Molecular features of current LBCL series according to recurrent mutational profiles on Primary mediastinal large B-cell lymphoma (PMBL), Burkitt Lymphoma (BL), pediatric type follicular lymphoma (PTFL) and Burkitt like lymphoma with 11q aberration (BLL-11q).



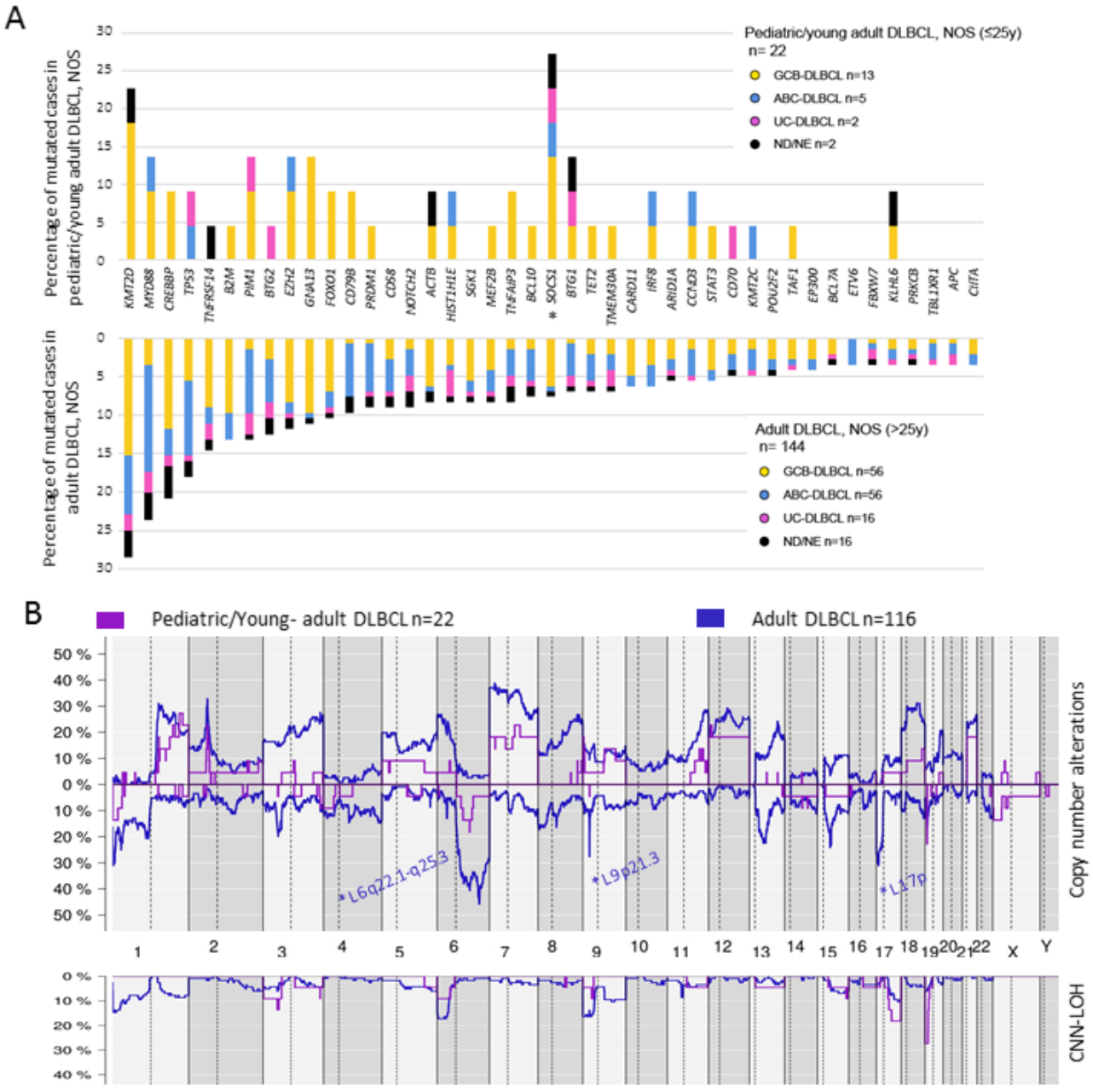
Supplemental Figure 6. Relative contribution of previously described signatures in DLBCL, NOS[13,14] for both all the variants found in 47 primary LBCL and per frequently mutated genes (at least 10 mutations). Single nucleotide variants including both driver and non-driver predicted mutations, synonymous and *MYC*-intronic mutations were considered for the analysis.



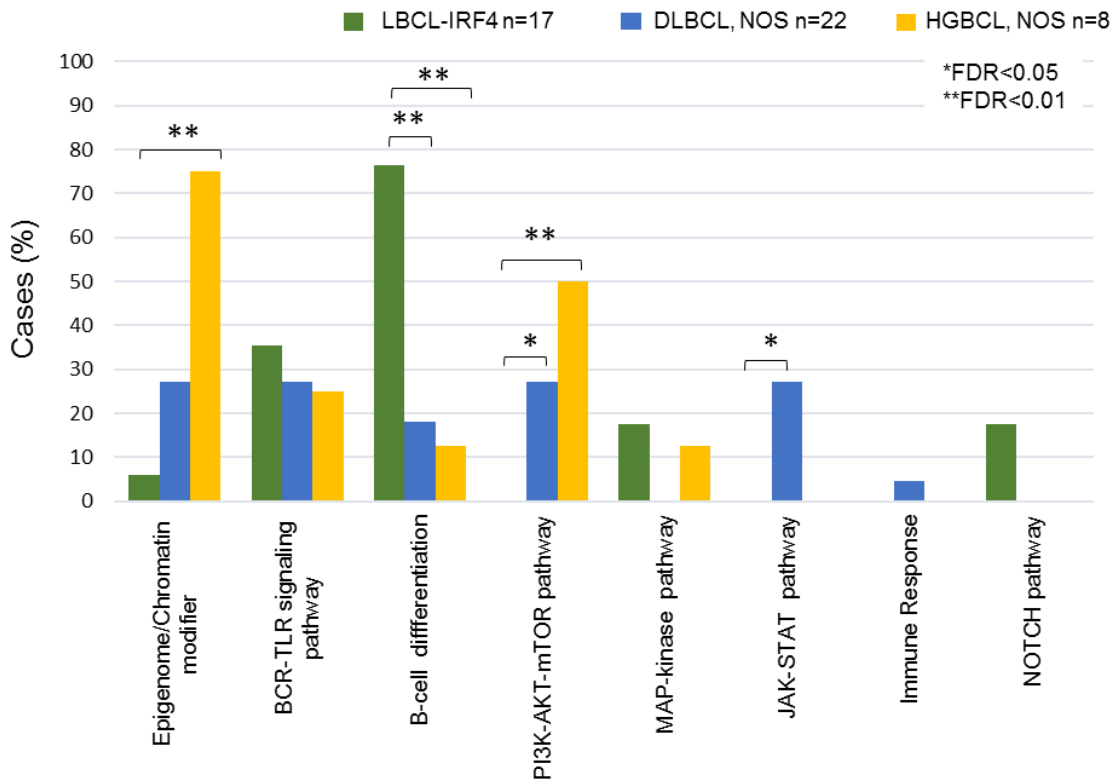
Supplemental Figure 7. Mutational profile of 25 pediatric/young-adult DLBCL, NOS according to cell of origin (Lymph2Cx and Lymph3Cx) determination. No significant differences in terms of mutation frequencies were observed between ABC and GCB (Fisher-test; NS).



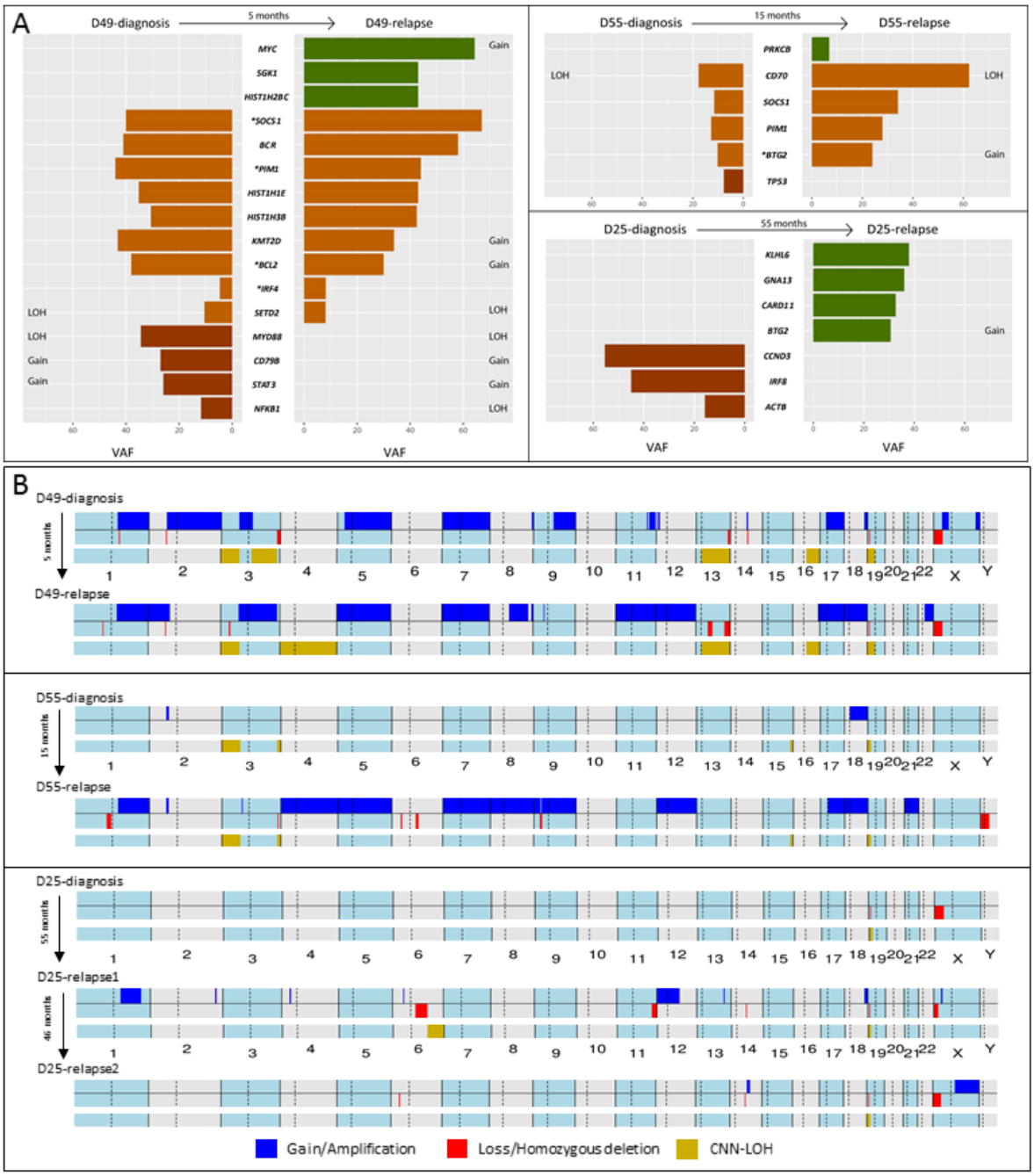
Supplemental Figure 8. Comparison of pediatric/young-adult versus adult DLBCL, NOS. (A) Percentage of mutated cases in pediatric/young-adult (upper panel) and adult DLBCL, NOS (lower panel) of the most frequently mutated genes interrogated in both series (at least 5 cases). Asterisk indicates differentially mutated genes between age groups (*P* <.05). **(B)** Comparative plot of copy number and copy number neutral-loss of heterozygosity (CNN-LOH) between 22 pediatric/young-adult DLBCL, NOS and 116 adult DLBCL, NOS. Significant different regions are indicated in the plot and the color denotes the enriched group (FDR<.1). GCB: germinal center B-cell; ABC: activated B-cell; UC: unclassified/intermediate; ND/NE: not done/not evaluable.



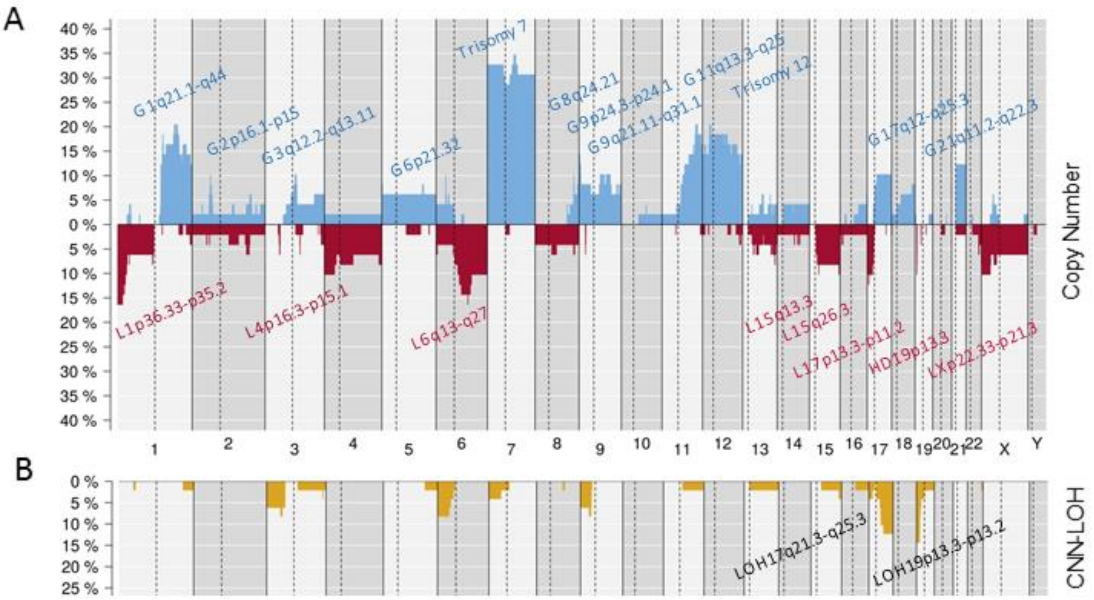
Supplemental Figure 9. Recurrent mutated pathways[6] in 47 LBCL primary tumors. Bar-graph shows the total number of mutated cases for each pathway. Each color bar indicates morphological subtypes. Asterisks represents significant mutated pathway in a morphological subtype.



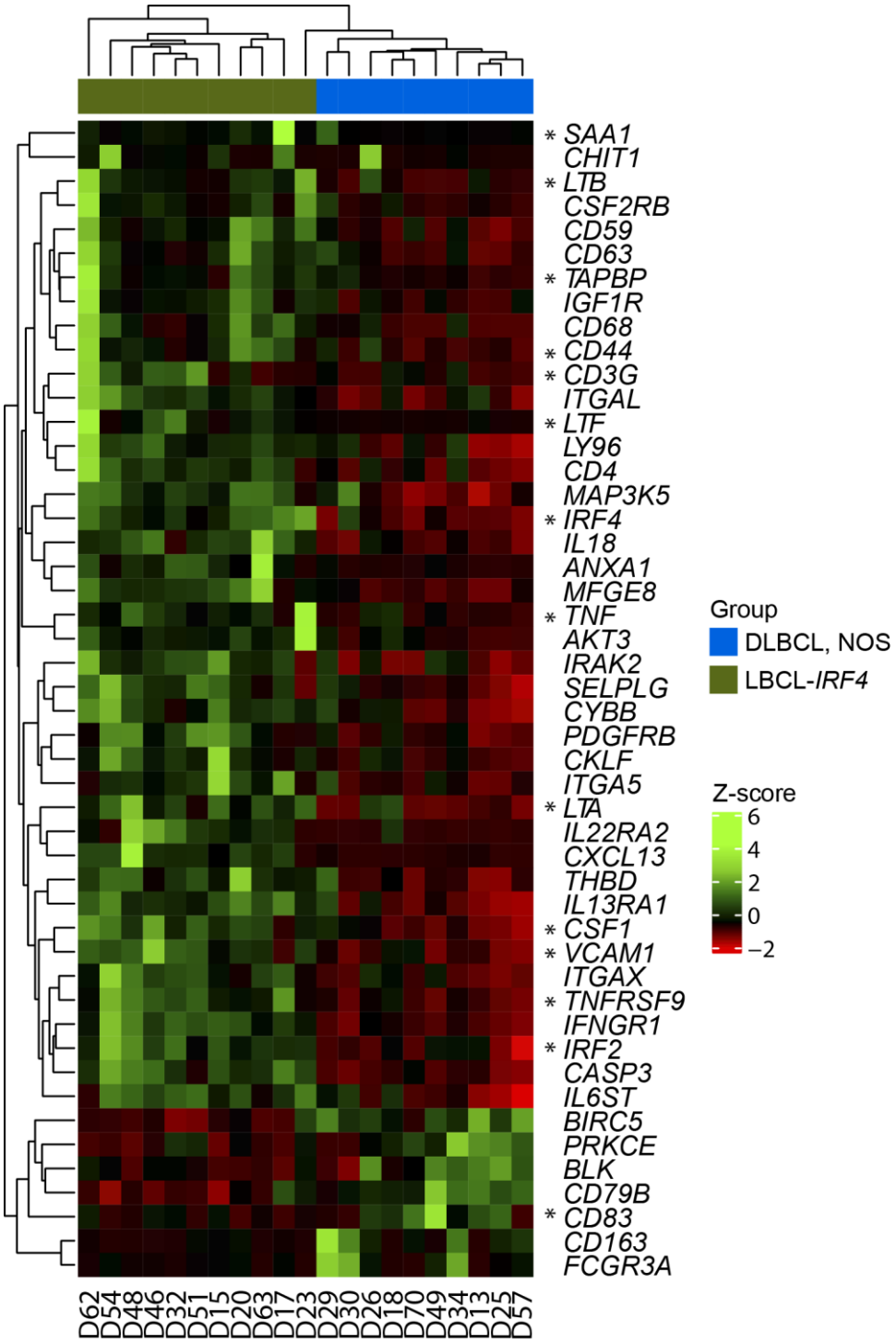
Supplemental Figure 10. (A) Detailed representation of mutational landscape by variant allele frequency (VAF) **(B)** and CN profiles of three paired cases (primary tumor-relapsed samples). In A, multiple mutations on a single gene are represented as the mean of mutation VAFs. CN information for each locus is indicated behind each gene. Asterisk denotes that some variants of specific gene are not shared between primary tumor and relapse sample. Acquired mutations are depicted in green, mutations only observed at diagnosis in red and shared mutations in orange. Note that VAF has not been corrected by tumor cell content since it was not available for all samples.



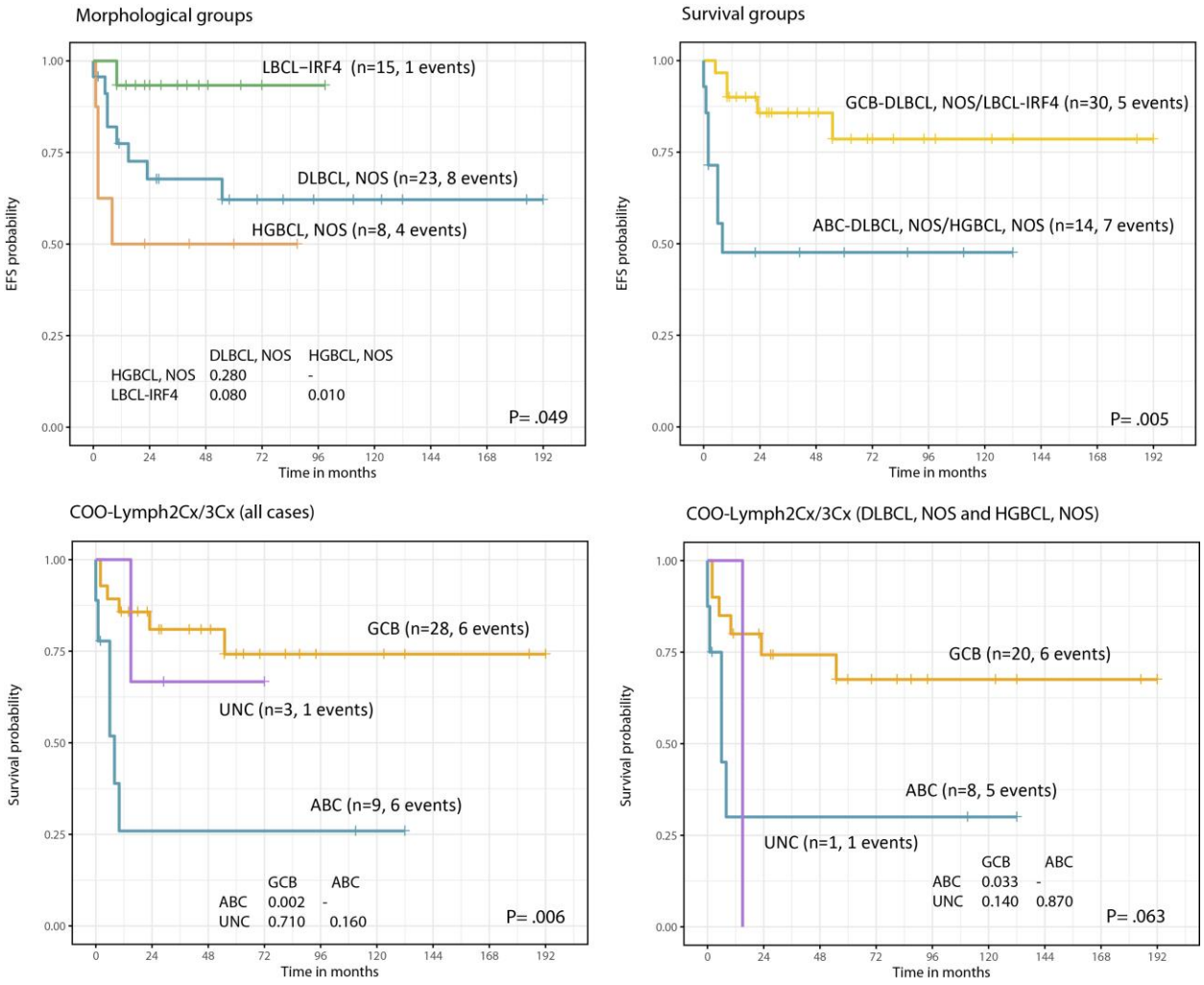
Supplemental Figure 11. Copy number (CN) analysis. **(A)** Global copy number and **(B)** copy number neutral loss of heterozygosity (CNN-LOH) profile of 49 pediatric/young-adult LBCL primary tumors excluding the four cases predicted as molecular PMBL. X-axis indicates chromosomes from 1 to Y and p to q. The vertical axis indicates frequency of the genomic aberration among the analyzed cases. Gains are depicted in blue, losses are depicted in red and CNN-LOH are depicted in yellow. Recurrent CN and CNN-LOH regions (>10% of cases) are indicated.



Supplemental Figure 12. Heatmap of differentially expressed genes in LBCL-*IRF4* (n=11) versus DLBCL (n=10) based on the nCounter PanCancer Immune Profiling Panel (NanoString inc.). Asterisk indicates NF-κB target gene according to http://www.bu.edu/nf-kb/gene-resources/target-genes/.



Supplemental Figure 13. Clinical and molecular parameters associated to worse EFS in the 46 current series with available follow up. COO: Cell of origin; UNC: unclassified/intermediate.



Supplemental Tables

Supplemental Table 1. Details of all antibodies used source and conditions of use.

Antibody	Clone	Source	Antigen retrieval/visualization	Dilution
CD20	L26	DAKO, (Copenhagen, Denmark)	EDTA 1 mM pH 9/ ENVISION FLEX (DAKO)	RTU
CD79a	JCB 117	DAKO	EDTA 1 mM pH 9/ ENVISION FLEX (DAKO)	RTU
CD3	Polyclonal	DAKO	EDTA 1 mM pH 9/ ENVISION FLEX (DAKO)	RTU
CD5	4C7	DAKO	EDTA 1 mM pH 9/ ENVISION FLEX (DAKO)	RTU
CD10	56C6	DAKO	EDTA 1 mM pH 9/ ENVISION FLEX (DAKO)	RTU
BCL6	PG-B6p	DAKO	EDTA 1 mM pH 9/ ENVISION FLEX (DAKO)	RTU
BCL2	124	DAKO	EDTA 1 mM pH 9/ ENVISION FLEX (DAKO)	RTU
Ki67	Mib-1	DAKO	Citrate 10 mM pH 6/ ENVISION FLEX (DAKO	RTU
MUM1	MRQ-43	Ventana, Roche (Oro Walley, AR, USA)	CC1 solution / ultraView Universal DAB Detection Kit. Automated immunostainer (Benchmark XT; Ventana)	RTU
MYC	Y69	Ventana, Roche	CC1 solution / ultraView Universal DAB Detection Kit. Automated immunostainer (Benchmark XT; Ventana)	RTU
CD21	EP3093	Ventana, Roche	CC1 solution / ultraView Universal DAB Detection Kit. Automated immunostainer (Benchmark XT; Ventana)	RTU

RTU, ready to use.

According to previous reports, BCL2,[14] BCL6 and MUM1[15] were considered positive when ≥70%, ≥30% or ≥60% of the cells were positive. MYC was considered positive when more than 40% of positive tumor cells were observed, following the criteria Johnson et al.[22]

Supplemental Table 2. Ninety-six genes analyzed using SureSelectXT Target NGS panel including references for inclusion in the mutational analysis and mean coverage by gene and amplicon.

Provided in excel format

Supplemental Table 3. Ampliseq Target NGS design for sequencing 29 selected genes.

Provided in excel format

Supplemental Table 4. Primers used for variant verification.

SOCS1

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
SOCS1-M1-F	CACCCCCGGACGCTATG	299	D28, D64
SOCS1-M1-R	AGGGGCCCCCAGTAGAAT		
SOCS1-N2-F	CTGAAAGTGCACGCGGATG	167	D55R
SOCS1-N2-R	CCTGCGGATTCTACTGGGG		
SOCS1-A5-F	GAGGAGGAGGAAGAGGAGGA	114	D49
SOCS1-A5-R	GGCTGGCCCCTTCTGTAG		
SOCS1-A4-F	AGGGGCCCCCAGTAGAAT	176	D49
SOCS1-A4-R	TCCTCCTCTTCCTCCTCCTC		
SOCS1-A2-F	CTGCCATCCAGGTGAAAGC	93	D33
SOCS1-A2-R	GAACTGCTTTTTCGCCCTTA		
SOCS1-A1-F	AGGGGAAGGAGCTCAGGTAG	201	D33
SOCS1-A1-R	AGAGCTTCGACTGCCTCTTC		

IRF4

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
IRF4-N1-F	CAGCTCTTCTCCCCGCAG	213	D15, D32, D54
IRF4-N1-R	TTCTCCTCGTTCTCCCACAC		
IRF4-N2-F	CGGAGAGTTCGGCATGAG	200	D17, D54, D50
IRF4-N2-R	GGCCGGAGACCTTGAAGAG		
IRF4-N3-F	GTGTGGGAGAACGAGGAGAA	236	D15, D54
IRF4-N3-R	GACGCCACCTGATGCCTC		
IRF4-A1-F	CGCAGTGCAGAGCAGAGC	200	D21
IRF4-A1-R	TTCTCCTCGTTCTCCCACAC		

BCR

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
BCR-1-F	GCTTCCTGAAGGACAACCTG	93	D49
BCR-1-R	CGACGTAGATGCTCTGGTAGG		
BCR-2-F	CACCACCTACCGCATGTTC	106	D49
BCR-2-R	GTGCCGCTTATGGCACTG		

CCND3

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
CCND3-1-F	GCCCCTCCTCTGCTTAGTG	89	D25
CCND3-1-R	AGCCAGACCAGCACTCCTAC		
CCND3-2-F	TCTGTAGGAGTGCTGGTCTGG	116	D21
CCND3-2-R	GAAGCTGCACTCAGGGAGAG		

KMT2D

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
KMT2D-F	GGCCATTGACTCAGGGGTAG	90	D49
KMT2D-R	GACCCAGCCGTTTCTTCAG		

IRF8

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
IRF8-F	GCAGAAGAGGCTGGGAAGAG	92	D25
IRF8-R	TCTGGAAACATCCGGAAGAC		

EZH2

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
EZH2-F	GAATACAGGTTATCAGTGCCTTACC	91	D33
EZH2-R	AGGCTGGGGGATTTTTATCA		

CREBBP

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
CREBBP-F	TGAGGCTGCTGGAACTGG	160	D33
CREBBP-R	TGGCGAGTATGAATCCACAG		

MYD88

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
MYD88-F	CCAGCGACATCCAGTTTGT	161	D49
MYD88-R	ACCCCGTGGCCTTCTAGC		

DDX3X

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
DDX3X-F	AAACACTGTCATCTACCAATGTCTG	154	D33
DDX3X-R	CTTCATGGCCCTCAAAGC		

CD79B

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
CD79B-2-F	CTGGCACACACCCATGACT	158	D49
CD79B-2-R	GCATGGAAGAGTCCCAGAAC		
CD79B-3-F	GTTCTGGGACTCTTCCATGC	163	D33
CD79B-3-R	GCAGAGCCCACGTTTCATAG		
CD79B-4-F	GCTATGAAACGTGGGCTCTG	150	D49
CD79B-4-R	GGACTAAGCCCAGGGAGTCT		

RHOA

Primer name	Seq (5'-3')	Amplicon size (pb)	Case
RHOA-2-F	GCTTTCCATCCACCTCGATA	200	
RHOA-2-R	GACTTCTTGTGCATTGCAGGT		
RHOA-3-F	AGCTACACAGGCAGTGACAAA	209	
RHOA-3-R	GGGGGATTAACCTTGCACTC		

Supplemental Table 5. Clinicopathological features of 31 Diffuse large B-cell lymphoma, NOS.

Case	Age, gender	Biopsy site	Immunophenotype					*In situ* hybridization			Stage*	COO Nanostring (Lymph2Cx)	Treatment	Outcome, Follow-up
			CD10	BCL6	MUM1	BCL2	EBER	*MYC*	*BCL2*	*BCL6*				
D3R	20, M	Supraclavicular LN	+	+	+	-	-					GCB	CT-A	CR, 197m**
D5	20, F	LN	-	+	-	+	-	N	N			GCB		
D6	25, F	Axillary LN	-	+	-		-	N	N	N	II	GCB[#]	CT-A	CR, 164m
D8	25, M	Inguinal LN	-	+	-		-	N	N	N	II	GCB	CT-A	CR, 185m
D9	22, F	LN	-	+	+	+	-	N	N	N	II	ABC	CT-A	DwD, 20m**
D10	21, M	Cervical LN	-	-	-	+	-	N	N	R	II	GCB	CT-A	CR, 132m
D12	17, M	LN	-	+	+	+	-	N	N	N	I	ABC	CT-A	CR, 111m
D13	4, F	Supraclavicular LN	+	+	+	+	-	N	N	N	I	GCB	CT-P	CR, 70m
D14R	11, M	Abdominal LN	+	+	-	-	-					GCB	CT-P	DwD, 80m**
D18	16, M	Inguinal LN	+	+	+	-		N	N	N	I-A	GCB	CT-A	CR, 94m
D22	24, M	Liver	-	+	-		-				IV	GCB	CT-A	CR, 55m
D24	2, M	Lung	-	+	-	+	+		N				CT-P	CR, 58m
D25	5, F	Cervical LN	+	+	-	-		N	N	N	I	GCB	CT-P	CR, 145m**
D26	12, M	Spleen	+	+	+	+	-	R	N	N		ABC	CT-P	DwD, 6m**
D27	11, M	Cervical LN		-	-	-	+	N			IV-B	ABC	CT-P	CR, 132m

D28	14, M	Cervical LN	+	+	-	-	-	N	N	N	IV-A	GCB	CT-P	CR, 81m
D29	1, M	Small bowel	-	+	-	-	+	N	N	N	III	GCB	CT-P	CR, 192m
D30	1, F	LN	-	-	-	+	+				IV	ABC		DwD, 0m**
D33	9, F	Mediastinal mass	-	+	+	+	-	R	N	N	IV	GCB[#]	CT-P	CR, 23m**
D34	11, M	Cervical LN	+	+	-	-	-	N		N		GCB	CT-P	CR, 27m
D38	12, F	Tonsil and submandibular LN	+	+	-	+		N	N	N	II	GCB	CT-P	CR, 28m
D40	24, M	Supraclavicular LN	-	+	+	+	-			N	III	UNC[#]	CT-A	DwD, 11m**
D42	22, M	Mediastinal LN	+	+	-	+		N	N	N	IV	GCB[#]	CT-A	CR, 184m**
D43	22, F	Cervical LN	-	-	+		+	N	N	N	II	UNC		
D49	21, M	Tonsil	+	+	+	-		N	R	R		GCB	CT-A	CR, 34m**
D55	25, M	Axillary LN	+	+	-		-	N	N	N	IV	UNC	CT-A	DwD, 31m**
D56	1, M	Cervical LN	-	-	+	+		N	N	N	IV	ABC	CT-P	CR, 2m
D57	5, M	Abdominal mass	-	+		+	-	R	N	N	III		CT-P	CR, 123m
D64	16, M	Inguinal LN	-	+	-	+	-	N	N	N				
D70	9, M	Axillary LN	+	+	+	+	-	N	N	N				
D71	10, M	Appendix	+	+	+	+	-		N	N	III	GCB	CT-P	CR, 8m

Abbreviations: R: relapse; M: male; F: Female; LN: Lymph node; R: rearrangement; N: Normal; COO: Cell of origin; GCB: Germinal center B-cell; ABC: Activated B-cell; UNC: Intermediate/unclassified; CT-P: Chemotherapy with pediatric schema protocol; CT-A: Chemotherapy with adult schema protocol (R-CHOP/ESHAP); CR: Complete response; DwD: Dead with disease; m: months.

*Stage was established according St. Jude/International pediatric NHL staging system (IPNHLSS) or Ann Arbor staging system for pediatric and adult patients respectively.

**Patients who had a relapse/progression

[#]Predicted as mPMBL by Lymph3Cx and are also described in Supplementary Table 7.

Supplemental Table 6. Pathological and clinical features of 12 High grade B-cell lymphoma, NOS.

Case	Age, gender	Biopsy site	Morphology	Immunophenotype						*In situ* hybridization				Stage*	COO Nanostring (Lymph2Cx)	Treatment	Outcome, Follow-up
				CD10	BCL6	MUM1	BCL2	TdT	MYC	EBER	*MYC*	*BCL2*	*BCL6*				
D19	7, M	Retroperitoneum	Blastoid	+		-	+	-	+		R	N				CT-P	CR, 41m
D36	14, M	Abdominal tumor	DLBCL/BL	+	+	-	+			+	N	N	N	III	GCB	CT-P	DwD, 2m
D37	3, M	Abdominal LN	DLBCL/BL	+	+	-	+	-		-	N	N	N	III	GCB	CT-P	CR, 22m
D52	23, F	Breast	DLBCL/BL	-	+	+	+	-		-	N	N	R	IV-A	ABC	CT-P	CR, 191m**
D53	13, F	Stomach	DLBCL/BL	+	+	-	-	-		-	N	N	N	IV	GCB	CT-P	CR, 87m
D58	4, M	LN	DLBCL/BL	-			-			+				IV	ABC	CT-P	DwD, 1m
D59	5, F	Kidney	DLBCL/BL	+	+		+	-	+	-	R	N	N				D, 1m[#]
D61	12, M	Intestine	Blastoid	+	+		-	-	-	-	N						
D65	4, M	Retroperitoneum	Blastoid	+				-			N			III	GCB	CT-P	DwD, 7m**
D72	19, M	Palate	DLBCL/BL	+	+	+	-		+	-	N	N	N		GCB		
D73	12, M	Tonsil	DLBCL/BL	+			-		-	-	R	N	N		GCB		
D75	6, F	Intestine	Blastoid	+	+	-	-	-		-	R			II	GCB	CP-P	CR, 60m

Abbreviations: M: male; F: Female; LN: Lymph node; R: rearrangement; N: Normal; COO: Cell of origin; GCB: Germinal center B-cell; ABC: Activated B-cell; CT-P: Chemotherapy with pediatric schema protocol; CR: Complete response; D: died: DwD: Dead with disease; m: months.
*Stage was established according St. Jude/International pediatric NHL staging system (IPNHLSS) or Ann Arbor staging system for pediatric and adult patients respectively.
** Patients who had a relapse/progression and need rescue treatment.
[#] This patient died before treatment due to cardiac arrest during port-a cath insertion, so this event is excluded of survival analysis.

Supplemental Table 7. Clinical, morphological and genetic data of four cases predicted as PMBL and 5 cases predicted as uncertain between mPMBL and DLBCL by Lymph3Cx assay.

Case	Age, Gender	Diagnosis	Biopsy Site	Other sites	PMBL morphological features	Hans	EBV	FISH *MYC*	FISH *CIITA*	9p24 *JAK2*	2p16 *REL*	Lymph3Cx	Lymph2Cx
D6	25, F	DLBCL	Axillary LN		Not typical	GC	-	-		Tri 9	WT	PMBL	GCB
D33	9, F	DLBCL	Mediastinal mass	BM, pelvis and ovary	Not typical	Non-GC	-	+		WT	Ampli	PMBL	GCB
D40	24, M	DLBCL	Supraclavicular LN	Multiple LN in thoracic region and mediastinum, bone and spleen	Typical	Non-GC	-		-	Tri 9	WT	PMBL	UNC
D42	22, M	DLBCL	Mediastinal LN	BM, lung, liver, suprarenal and renal	Not typical	GC		-	-	WT	WT	PMBL	GCB
D5	20, F	DLBCL	LN		Not typical	GC	-	-		WT	WT	Uncertain PMBL/GCB	GCB
D8	25, M	DLBCL	Inguinal LN		Not typical	GC	-	-				Uncertain PMBL/GCB	GCB
D27	11, M	DLBCL	Cervical LN	BM and spleen	Not typical	Non-GC	+	-		Ampli	WT	Uncertain PMBL/UNC	ABC
D43	22, F	DLBCL	Cervical LN		Not typical	Non-GC	+	-		Gain	WT	Uncertain PMBL/GCB	UNC
D49	21, M	DLBCL	Tonsil		Not typical	GC		-		WT	Ampli	Uncertain PMBL/GCB	GCB

Abbreviations: F: Female; M: Male; LN: Lymph node; BM: bone marrow; GC: germinal center; Tri: trisomy; Ampli: amplification; GCB: Germinal Center B-cell derived; ABC: Activated B-cell derived; PMBL: Primary mediastinal large B-cell lymphoma; UNC: Unclassified; WT: wild type.
Copy number alterations obtained from Oncoscan/SNP-array analysis.

Supplemental Table 8. List of somatic mutations in LBCL including prediction of amino acid changes that affect protein function (MA, SIFT, Polyphen2, CADD and CHASM).

Provided in excel format.

Supplemental Table 9. Aberrant somatic hypermutation (aSHM) hallmarks in frequently mutated genes[#]

Gene name	Total SNV	Mutated cases	Mutations/ case	Variants within AID target region	AID-target region bias (*P*-value)	Transition over transversions bias (*P*-value)	AID-motif bias (*P*-value)
MYC	85	11	7.73	75	0.88 (<0.001)	0.74 (0.111)	0.59 (0.002)
IRF4	83	16	5.19	80	0.96 (<0.001)	1.10 (<0.001)	0.55 (0.172)
PIM1	29	12	2.42	29	1.00 (0.001)	1.64 (0.002)	0.66 (0.758)
SOCS1	19	7	2.71	19	1.00 (1.000)	0.58 (0.935)	0.42 (0.470)
EGR1	17	6	2.83	16	0.94 (0.091)	2.20 (0.006)	0.69 (1.000)
SGK1	16	6	2.67	16	1.00 (<0.001)	2.20 (0.006)	0.81 (0.044)
CARD11	10	10	1.00	0	-	-	-

[#] Ten single nucleotide variants including driver and passengers predicted, synonymous and *MYC*-intronic.
AID-target region is defined as 2Kb region after 150bp from a transcript start site)
AID-motifs correspond to sequences WA/TW/WRCY/RGYW/WGCW)[16]
Significance value was calculated using Test of Equal or Given Proportions.

Supplemental Table 10. Intron1 *BCL6* mutational analysis results using Ampliseq Target NGS. Potential IRF4-binding sites within the 5′ flanking sequences and the first intron of the *BCL6* gene were previously defined by Saito et al.[23]

Provided in excel format

Supplemental Table 11. Global copy number and copy number neutral loss of heterozigosity alterations in 59 LBCL cases

Provided in excel format

Supplemental Table 12. Differentially expressed genes in LBCL-*IRF4* (n=11) vs DLBCL (n=10) based on the Nanostring PanCancer Immune Profiling Panel. *P*-value from False Discovery Rate test for multiple testing). Asterisk indicates NF-κB target gene according to http://www.bu.edu/nf-kb/gene-resources/target-genes/.

Provided in excel format

Supplemental Table 13. Clinicopathological features of the different age groups.

Characteristics	0-18 y	19-25 y	All LBCL	>25 y ¶	Significance (*P*-value) Age (0-18y) vs (19-25y)	Pediatric/young-adult vs adult (>25 y)
Num. Patients	45	14	59	144		
Type						
DLBCL, NOS	18/45 (40%)	9/14 (64%)	27/59 (46%)		0.134	
HGBCL, NOS	10/45 (22%)	2/14 (14%)	12/59 (20%)		0.712	
LBCL-*IRF4*	17/45 (38%)	3/14 (21%)	20/59 (34%)		0.342	
Ratio M:F	28:17	9:5	37:22	74:70	1.000	0.163
Primary extranodal involvement	15/45 (33%)	4/14 (29%)	19/59 (32%)		1.000	
Head and neck	21/40 (53%)	6/12 (50%)	27/52 (52%)		1.000	
Stage III/IV	14/29 (48%)	4/10 (40%)	18/39 (46%)	78/143 (55%)	0.726	0.371
LDH high	8/25 (32%)	4/7 (57%)	12/32 (38%)	65/142 (46%)	0.379	0.436
In situ Hybridization						
EBV (EBER)	6/32 (19%)	1/12 (8%)	7/44 (16%)		0.653	
FISH MYC	6/34 (18%)	0/11 (0%)	6/45 (13%)	11/119 (9%)	0.311	0.566
FISH BCL2	0/29 (0%)	1/9 (11%)	1/38 (3%)	25/125 (20%)	0.237	**0.010**
FISH BCL6	0/29 (0%)	3/8 (38%)	3/37 (8%)	24/117 (21%)	**0.007**	0.134
FISH IRF4	14/31 (45%)	3/9 (33%)	17/40 (43%)		0.707	
COO-Hans						
GC	31/44 (70%)	8/14 (57%)	39/58 (67%)		0.514	
Non-GC	13/44 (30%)	6/14 (43%)	19/58 (33%)		0.514	
COO-Nanostring #						
GCB	27/36 (75%)	8/13 (62%)	35/49 (71%)	43/105 (41%)	0.476	**<0.001**
ABC	6/36 (17%)	3/13 (23%)	9/49 (18%)	49/105 (47%)	0.683	**0.009**
UC	3/36 (8%)	2/13 (15%)	5/49 (10%)	13/105 (12%)	0.598	1.000
Genetics						
No. Mutations	5.1 (0-16)	5.8 (1-20)	5.2 (0-20)		0.969	
No. CNA	4.5 (0-24)	12.8 (1-34)	6.2 (0-34)	20 (1-108)	**0.046**	**<0.001**
Chromothripsis	1/39 (3%)	3/10 (30%)	4/49 (8%)	28/116 (24%)	**0.023**	0.077
CR rate of first line treatment	29/35 (83%)	9/11 (82%)	38/46 (83%)	96/142 (68%)	1.000	0.061
Rituximab	7/35 (20%)	9/11 (82%)	16/46 (35%)	116/143 (81%)	**<0.001**	**<0.001**
No. of deads	6/35 (17%)	2/11 (18%)	8/46 (17%)	82/144 (57%)	1.000	**<0.001**
Relapse/Progress	7/35 (20%)	6/11 (55%)	13/46 (28%)	70/144 (49%)	0.051	**0.017**
Median Follow-up	38.5 months	40 months	40 months	84 months	0.234	0.581
5-year OS	86%	79%	83%	63%	0.910	**0.007**
5-year EFS	76%	46%	68%	65%	**0.044**	0.490

[#]Cell of origin by Lymph2Cx or Lymph3Cx.
Abbreviations; y: years; M: male; F: female; GC: germinal center; CR: Complete response; OS: Overall survival; EFS: Event free survival.
[¶] Clinical data from patients older than 25 years old was recruited and updated from Karube et al.[5]

Supplemental Table 14. Clinical and morphological features of the different lymphoma entitites.

Characteristics	LBCL-*IRF4*	DLBCL, NOS	HGBCL, NOS	Significance (*P*-value) LBCL-IRF4 vs DLBCL	LBCL-IRF4 vs HGBCL, NOS	DLBCL vs HGBCL, NOS
Num. Patients	20	27	12			
Median Age	14 (5-22)	12 (1-25)	9.5 (3-23)	0.889	0.093	0.330
0-18y	17/20 (85%)	18/27 (67%)	10/12 (83%)	0.191	1.000	0.446
19-25y	3/20 (15%)	9/27 (33%)	2/12 (17%)	0.191	1.000	0.446
Ratio M:F	9:11	20:7	8:4	0.069	0.291	0.709
Primary extranodal involvement	4/20 (20%)	6/27 (22%)	9/12 (75%)	1.000	**0.004**	**0.004**
Head and neck	14/18 (78%)	11/23 (48%)	2/11 (18%)	0.063	**0.003**	0.140
Stage III/IV	3/14 (21%)	9/18 (50%)	6/7 (86%)	0.147	**0.016**	0.179
LDH high	2/12 (17%)	5/12 (42%)	5/8 (63%)	0.371	0.062	0.650
Immunohistochemistry						
CD10	11/20 (55%)	13/26 (50%)	10/12 (83%)	0.774	0.139	0.077
MUM1	20/20 (100%)	11/26 (42%)	2/7 (29%)	**<0.001**	**<0.001**	0.676
BCL6	20/20 (100%)	22/27 (81%)	8/8 (100%)	0.063	1.000	0.315
KI-67>75%	11/14 (79%)	15/19 (79%)	11/11 (100%)	1.000	0.230	0.268
BCL2	10/19 (53%)	14/23 (61%)	5/11 (45%)	0.756	1.000	0.475
CD20	19/19 (100%)	22/22 (100%)	9/9 (100%)	1.000	1.000	1.000
In situ Hybridization						
EBV (EBER)	0/12 (0%)	5/22 (23%)	2/10 (20%)	0.137	0.195	1.000
FISH MYC	0/13 (0%)	2/21 (10%)	4/11 (36%)	0.513	**0.031**	0.148
FISH BCL2	0/10 (0%)	1/21 (5%)	0/7 (0%)	1.000	1.000	1.000
FISH BCL6	0/11 (0%)	2/20 (10%)	1/6 (17%)	0.527	0.353	1.000
FISH IRF4	17/19 (89%)	0/17 (0%)	0/4 (0%)	**<0.001**	**0.002**	1.000
COO-Hans						
GC	11/20 (55%)	18/27 (67%)	10/11 (91%)	0.546	0.055	0.225
Non-GC	9/20 (45%)	9/27 (33%)	1/11 (9%)	0.546	0.055	0.225
COO-Nanostring[#]						
GCB	10/14 (71%)	17/25 (68%)	8/10 (80%)	1.000	1.000	0.686
ABC	1/14 (7%)	6/25 (24%)	2/10 (20%)	0.386	0.550	1.000
UC	3/14 (21%)	2/25 (8%)	0/10 (0%)	0.329	0.239	1.000
Genetics						
No. Mutations	5.2 (0-11)	4.7 (0-20)	6.6 (1-11)	0.247	0.278	0.143
No. CNA	6.2 (0-27)	5.8 (0-34)	7.1 (0-27)	0.518	0.697	0.939
Chromothripsis	2/20 (10%)	1/22 (5%)	1/7 (14%)	0.598	1.000	0.431
CR rate of first line treatment	14/15 (93%)	19/23 (83%)	5/8 (63%)	0.630	0.103	0.335
Rituximab	6/15 (40%)	10/23 (43%)	0/8 (0%)	1.000	0.058	**0.032**
No. of deads	0/15 (0%)	5/23 (22%)	3/8 (38%)	0.136	**0.032**	0.393
Relapse/Progress	1/15 (7%)	8/23 (35%)	4/8 (50%)	0.061	**0.033**	0.676
Median Follow-up	29 months	70 months	22 months	0.060	0.531	0.126
5-year EFS	93%	62%	50%	0.075	**0.010**	0.280

[#]Cell of origin by Lymph2Cx or Lymph3Cx.
Abbreviations; y: years; M: male; F: female; GC: germinal center; CR: Complete response; OS: Overall survival; EFS: Event free survival.

Supplemental Table 15. Clinical and morphological features of the different age groups of DLBCL

Characteristics	0-18 y	19-25 y	All pDLBCL, NOS	>25 y[¶]	Significance (*P*-value) Age (0-18y) vs (19-25y)	Pediatric/young-adult vs adult (>25 y)	Young-adult (19-25y) vs adult (>25 y)
Num. Patients	18	9	27	144			
Ratio M:F	14:4	6:3	20:07	74:70	0.653	**0.035**	0.721
Primary extranodal involvement	5/18 (28%)	1/9 (11%)	6/27 (22%)		0.628		
Head and neck	7/16 (44%)	4/7 (57%)	11/23 (48%)		0.667		
Stage III/IV	7/12 (58%)	2/6 (33%)	9/18 (50%)	78/143 (55%)	0.620	0.804	0.416
LDH high	3/8 (38%)	2/4 (50%)	5/12 (42%)	65/142 (46%)	1.000	1.000	1.000
In situ Hybridization							
EBV (EBER)	4/14 (29%)	1/8 (13%)	5/22 (23%)		0.613		
FISH MYC	2/14 (14%)	0/7 (0%)	2/21 (10%)	11/119 (9%)	0.533	1.000	1.000
FISH BCL2	0/14 (0%)	1/7 (14%)	1/21 (5%)	25/125 (20%)	0.333	0.125	0.595
FISH BCL6	0/14 (0%)	2/6 (33%)	2/20 (10%)	24/117 (21%)	0.079	0.364	1.000
FISH IRF4	0/12 (0%)	0/5 (0%)	0/17 (0%)		1.000		
COO-Hans							
GC	12/18 (67%)	6/9 (67%)	18/27 (67%)		1.000		
Non-GC	6/18 (33%)	3/9 (33%)	9/27 (33%)		1.000		
COO-Nanostring							
GCB	11/16 (69%)	6/9 (67%)	17/25 (68%)	43/105 (41%)	1.000	**0.009**	0.268
ABC	5/16 (31%)	1/9 (11%)	6/25 (24%)	49/105 (47%)	0.364	0.182	0.259
UC	0/16 (0%)	2/9 (22%)	2/25 (8%)	13/105 (12%)	0.120	1.000	0.241
Genetics							
No. Mutations	4.4 (0-16)	5.4 (1-20)	4.7 (0-20)		0.803		
No. CNA	4.9 (0-24)	8.2 (1-34)	5.8 (0-34)	20 (1-108)	0.682	**<0.001**	**0.001**
Chromothripsis	1/16 (6%)	0/6 (0%)	1/22 (5%)	28/116 (24%)	1.000	0.129	0.584
CR rate of first line treatment	13/16 (81%)	6/7 (86%)	19/23 (83%)	96/142 (68%)	1.000	0.221	0.665
Rituximab	4/16 (25%)	6/7 (86%)	10/23 (43%)	116/143 (81%)	**0.019**	**<0.001**	1.000
No. of deads	3/16 (19%)	2/7 (29%)	5/23 (22%)	82/144 (57%)	0.621	**0.003**	0.405
Relapse/Progress	4/16 (25%)	4/7 (57%)	8/23 (35%)	70/144 (49%)	0.182	0.264	1.000
Median Follow-up	75 months	55 months	70 months	84 months	0.504	0.287	0.138
5-year OS	87%	71%	81%	63%	0.740	**0.035**	0.190
5-year EFS	71%	43%	62%	65%	0.150	0.900	0.320

[#]Cell of origin by Lymph2Cx or Lymph3Cx.
Abbreviations; y: years; M: male; F: female; GC: germinal center; CR: Complete response; OS: Overall survival; EFS: Event free survival.
[¶]Clinical data from patients older than 25 years old was recruited and updated from Karube et al.[5]

Supplementary references

1. Gonzalez-Farre B, Ramis-Zaldivar JE, Salmeron-Villalobos J et al. Burkitt-like lymphoma with 11q aberration: A germinal center derived lymphoma genetically unrelated to Burkitt lymphoma. *Haematologica.* 2019 Sep;104(9):1822-1829.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
3. McKenna A, Hanna M, Banks E et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
4. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
5. van Dongen JJ, Langerak AW, Bruggemann M et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia.* 2003;17(12):2257-2317.
6. Karube K, Enjuanes A, Dlouhy I et al. Integrating genomic alterations in diffuse large B-cell lymphoma identifies new relevant pathways and potential therapeutic targets. *Leukemia.* 2018;32(3):675-684.
7. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 2011;39(17):e118.
8. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073-1081.
9. Adzhubei IA, Schmidt S, Peshkin L et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.
10. Kircher M, Witten DM, Jain P et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.
11. Carter H, Chen S, Isik L et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res.* 2009;69(16):6660-6667.
12. Blokzijl F, Janssen R, van BR, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* 2018;10(1):33.

13. Chapuy B, Stewart C, Dunford AJ et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med.* 2018;24(5):679-690.
14. Arthur SE, Jiang A, Grande BM et al. Genome-wide discovery of somatic regulatory variants in diffuse large B-cell lymphoma. *Nat Commun.* 2018;9(1):4001.
15. Alexandrov LB, Kim J, Haradhvala NJ et al. The Repertoire of Mutational Signatures in Human Cancer. *biorxiv.* 2018
16. Papaemmanuil E, Rapado I, Li Y et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet.* 2014;46(2):116-125.
17. Jennings LJ, Arcila ME, Corless C et al. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn.* 2017;19(3):341-365.
18. Khodabakhshi AH, Morin RD, Fejes AP et al. Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget.* 2012;3(11):1308-1319.
19. Reddy A, Zhang J, Davis NS et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell.* 2017;171(2):481-494.
20. Schmitz R, Wright GW, Huang DW et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N Engl J Med.* 2018;378(15):1396-1407.
21. Lopez C, Kleinheinz K, Aukema SM et al. Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. *Nat Commun.* 2019;10(1):1459.
22. Johnson NA, Slack GW, Savage KJ et al. Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J Clin Oncol.* 2012;30(28):3452-3459.
23. Saito M, Gao J, Basso K et al. A signaling pathway mediating downregulation of BCL6 in germinal center B cells is blocked by BCL6 gene alterations in B cell lymphoma. *Cancer Cell.* 2007;12(3):280-292.