Supplementary Data

1. **Supplementary Methods**
2. **Supplementary Figures**

   Supplementary Figure S1 – Time between diagnosis and number of mutations.

   Supplementary Figure S2 – Example of a clonal sample pair (patient #31).

   Supplementary Figure S3 – Example of a clonal sample pair (patient #31) as determined with Euroclonality ARResT/Interrogate.

   Supplementary Figure S4 – Circosplots of alterations for each patient.

   Supplementary Figure S5 – aSHM frequently differed between primary and relapse LBCL-IP samples.

   Supplementary Figure S6 – Unique and shared linker histone mutations in paired samples.

3. **Supplementary Tables (separate excel sheets)**

   Supplementary Table S1 – A. NGS sample characteristics.
   B. Clinical parameters.

   Supplementary Table S2 – Genes in capture panel for mutation analysis Roche Nimblegen SeqCapEZ IRN 1000002633.

   Supplementary Table S3 – Genomic regions in capture panel for translocation analysis.

   Supplementary Table S4 – Functional somatic mutations per patient.

   Supplementary Table S5 – Mutation frequencies per cohort.

   Supplementary Table S6 – Translocations per patient.

   Supplementary Table S7 – Common clonal IGH, IGK and IGL recombination.

Supplementary Methods

Patient cohort

The IP-LBCL patient cohort is enriched for patients first diagnosed many years ago (first diagnosis 1985-2014, median 2002; relapse 1990-2022, median 2008), due to selection of relapsed LBCL-IP patients and the nature of the disease that is characterized by late relapses. As a result, clinical data were only limitedly accessible and/or available from the various hospitals. In the Netherlands, primary CNS lymphoma and testicular lymphoma were largely treated per protocol from the early 1990s on. Before 2004, primary treatment for testicular lymphoma consisted of CHOP chemotherapy, intrathecal MTX and generally radiotherapy to the contralateral testis irrespective of stage (n=18). After 2004, Rituximab was added to this regimen (n=10). Primary CNS lymphomas were treated per protocol with MBVP with or without intrathecal MTX and generally with radiotherapy to both testis as prophylaxis. Treatment for relapsed disease was not protocolized for either testicular (n=24) or for CNS relapses (n=9). As a result of the patient selection (patients with relapse), treatment of the first disease episode resulted in complete remission followed by relapse with an interval of 0.5 to 18 years (median 4 years). Treatment outcome of the relapse presentation likely largely depended on the site of relapse (CNS versus testis).

Immunohistochemistry

Immunohistochemistry (IHC) for IgM and IgG was performed on 33 primary and relapse LBCL-IP cases using standard diagnostic protocols. In brief, 3 µm sections were cut from FFPE tissue. IHC was performed using the OptiView DAB IHC Detection Kit (Roche Diagnostics) on the Ventana Benchmark Ultra (Roche, Rotkreuz, Switzerland). Primary polyclonal antibodies against IgM and IgG (IgM, 1:20000, Dako A0425, IgG, 1:80000, Dako A0423) were applied for 32 minutes. IGM and IGG expression was visually assessed as positive or negative by an expert pathologist (D.d.J.).

Laboratory procedure sequencing

Genomic DNA was extracted from formalin fixed paraffin embedded (FFPE) tumor samples with the QIAamp DNA FFPE tissue kit (Qiagen, Hilden, Germany) and fragmented with a covaris ME220 (Covaris Inc, Woburn, MA, USA). Fragmented DNA (100 ng) was used for KAPA library prep with IDT adapters (KAPA Biosystems, Wilmington, MA; IDT, Coralville, IA). An aliquot was subjected for shallow whole genome sequencing single read 50 base pairs on

the Hiseq 4000 (Illumina, San Diego, CA) in equimolar pools of 24 samples of 10 nmol/l per library.

For mutation and translocation analysis a 2.5 Mb capture panel was designed containing exonic regions of 311 genes and 23 genomic regions(Roche NimbleGen, Madison, WI; IRN 1000002633)(Supplemental Table 5 and 6). For the capture, 125 ng library per sample was used and 8 samples were pooled, 3 pools were sequenced together paired-end 150 base pairs on a Hiseq 4000 sequencer (Illumina, San Diego, CA).

Copy number alteration analysis

CNA analysis was performed as previously described.[1] In short, DNA was aligned against the reference genome (hg19) using BWA aln (v.0.7.12) and duplicates were marked with picard tools markduplicates (v.2.15). For CNA analysis the following R packages were used in R (v 3.4.1), reads were divided in 100 kb bins with QDNAseq (v1.12)[2] and content of guanine–cytosine basepairs (gc-basepairs), mappability and germline correction was performed. Waves were removed with NoWaves (v0.6)[3] and segmentation was performed with DNAcopy (v1.50.1)[4], ACE [5](v0) was used to estimate the tumor cell percentage and absolute copy numbers. Calling was performed with CGHcall [6](v2.38) with the tumor cell percentage as correction factor to a minimum of a tumor cell percentage of 0.2 to minimize overcalling. To compare the primary and relapse cohort the number of regions was reduced with CGHregions (v1.34)[7] with a maximum information loss of 20% resulting in 190 regions. These regions were compared with CGHtest (v1.1)[8] using a Wilcoxon rank test with 10000 permutations and a false discovery rate test for multiple testing correction.

Regions for the oncoprint were determined with GISTIC (v2.0)[9]. Multiple regions on the same chromosomal arm were taken together. To determine if a CNA was different between the pairs the segments in the region were compared. If the number of segments was different between the pairs or when the beginning or end point of the segment was more than 5Mb (50bins) apart, a difference in CNA was called.

Focal regions smaller than 3 Mb in the 100kb analysis were determined by GISTIC[9] and analyzed on 15kb bins by CGHcall.[6] When the frequency of the focal gain or loss was >10% different in the entire cohort for the 15kb bins compared to the 100kb bins, it was added as a separate feature to the oncoprint and frequency table. One location (chr 7p12.2) met these criteria, but by visual inspection it was clear that this was an artefact that was only present in lower quality samples and therefore we excluded this.

Mutation Analysis

Mutation analysis was performed as previously described.[10] Paired-end data 150 bp data was demultiplexed with Bclfastq (illumina) and adapters were trimmed with seqpurge (v0.1-104)[11]. Thereafter the reads were aligned with BWA-mem (v.0.7.12)[12] and realignment was performed with ABRA (v2.19)[13] to improve the alignment. The aligned reads were querysorted with sambamba (v0.5.6)[14] so that also secondary alignments were included in the mark duplicate analysis of Picardtools (v2.4.1)(which is important for translocation analysis). Next, reads were coordinate sorted and only the captured exonic regions were kept for the mutation calling. Three samples were sequenced twice due to low coverage after one sequence run and data was merged during the mark duplicate step. Samples with a tumor cell percentage <10% or <20% in combination with a mean target coverage <100 were excluded for further analysis.

Mutation calling was performed in two different ways. First, all samples were analyzed in tumor-only mode by Lofreq (V2.1.3.1)[15] and Mutect2 (V4.1.7)[16]. For Lofreq indel quality scores were obtained by lofreq indelqual, after which calling was performed using the settings [--min-cov 10 --min-mq 20 --min-bq 20 --min-alt-bq 20 --max-depth 1000 --sig 0.05] and subsequently filtered with the settings [--af-min 0.05 --cov-min 15 --sb-alpha 0.05], and minimal 2 supporting reads on both the forward and reverse read. Furthermore, calls should not be present in a blacklist that was created using an in house set of 25 normal casus including 13 matched normals from this cohort, 6 reactive lymph nodes and 6 hyperplasia samples using the same pipeline. Variants present in at least 2 samples were added to the blacklist. In addition, Mutect2 was ran with the setting [--max-mnp-distance 0] and a panel of normal file based on the 25 normals. Subsequently, the read orientation and contamination was calculated. Mutation calls were filtered by gatk filtermutectcalls with the settings [--max-events-in-region 50 --min-allele-fraction 0.05 --min-reads-per-strand 3 --unique-alt-read-count 4]. The Mutect2 germline filter was not used, since multiple *MYD88* mutations were predicted as germline when they were in conjunction with a chromosomal gain on that locus. In addition, mutations were filtered when they had a ROQ score <20 or TLOD score <20 to reduce noise. Mutations detected by both tools were annotated by snpSift (v4.3)[17] and funcotator (v4.1.7)[18] using the databases of COSMIC (v84)[19], Gnomad (v2.0.2)[20], gencode(v19)[21], dbsnp (build 151)[22], clinvar (20180401)[23], and HMF panel of normals (v2.0). Mutations labeled as common in dbsnp or present >3 times in HMF panel of normals were excluded. Also intronic, silent, UTR or flanking mutations were removed for further analysis.

For 12 samples, a matched normal sample was available. These samples were analyzed with Mutect2 in tumor-normal mode.[16] Called mutations of the tumor only and tumor-normal pipeline were taken together, discrepancies of the pipelines were manually curated using IGV, germline

events, noisy events and events with only 2 reads evidence were excluded (105/1871 mutations). Non-shared mutations were manually scrutinized by visual inspection in IGV to confirm presence/absence of mutations. Comparisons of the number of mutations was performed with a Wilcoxon test of paired data. Frequencies of alterations were compared between primary and relapse samples using a Fisher-exact test on all alterations present in >10% of the samples, false discovery rate was calculated according to Benjamini & Hochberg method. A correlation between time of diagnosis and number of shared mutations was calculated by fitting linear regression models using the R stats package v(3.6.1).

An oncoprint was created with Complexheatmap (2.7.1.1016)[24] and circosplots were made with circlize (0.4.13)[25]. For somatic hyper mutation (SHM) comparisons, intronic and silent mutations were included as well. SHM comparisons were performed for the known target genes IGLL5, PIM1, BTG1, BTG2 and OSBPL1.[26-28] Analyses were performed with R (v3.6.1).

Translocation analysis

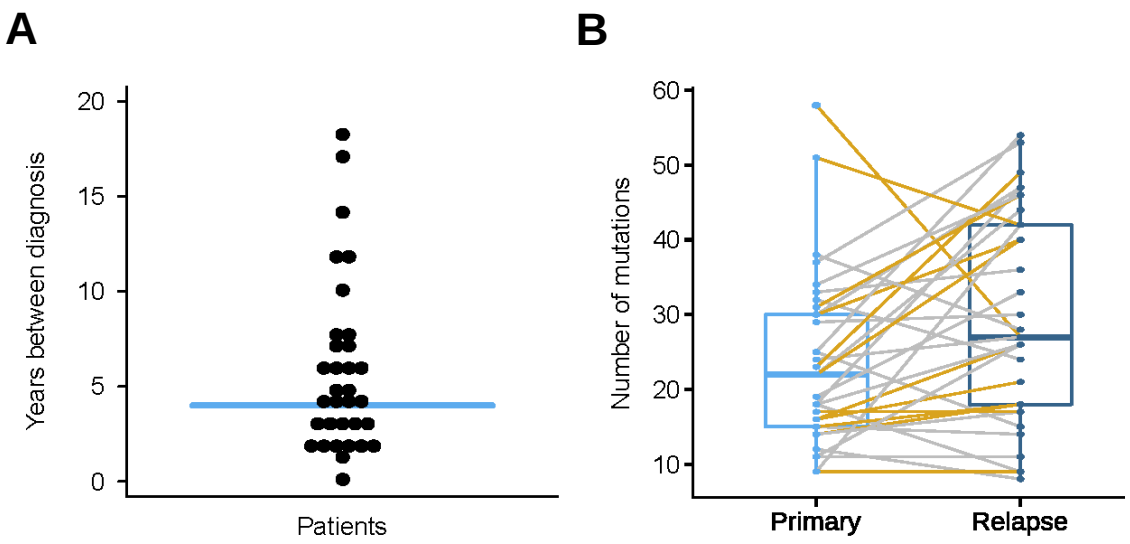
Translocation analysis was performed as previously described.[29] In short, 4 structural variant (SV) tools were combined using Snakemake: Breakmer (v0.0.4)[30], NovoBreak (v1.1.3)[31], GRIDSS (v1.4.2)[32] and Wham (v1.7.0)[33]. All tools were executed using default settings except Wham which was done with a mapping quality of 10 and base quality of 5. For BreaKmer, a target file containing all translocation regions of the capture panel was used, with the regions divided in non-overlapping 5kb regions. Subsequently, SVs were filtered based on the following criteria: at least one of the SV breakpoints should be within 300 bp of the capture area and SVs should meet the thresholds set for each tool (BreaKmer: at least 4 split reads and 3 discordant reads, Wham: at least 8 supporting reads, GRIDSS quality score above 450 and for NovoBreak an coverage of 4 high mapping quality translocation reads). Duplicate SVs detected with the same tool were removed. Then, SVs should be detected by at least 2 of the 4 tools and breakpoints within 10bp of each other were considered to be the same SV. In addition, a blacklist was created using the 25 normal samples described above. SVs detected less than 10bp apart were added to the blacklist and areas less than 50 bp apart were merged to one larger area. The latter was done because these were often highly repetitive regions or partly homologous regions, giving rise to many false positives. SVs with a breakpoint within the blacklisted region were removed from the output. The remaining translocations were manually inspected in IGV after which only true translocations were kept.

Immunoglobulin clonality analysis

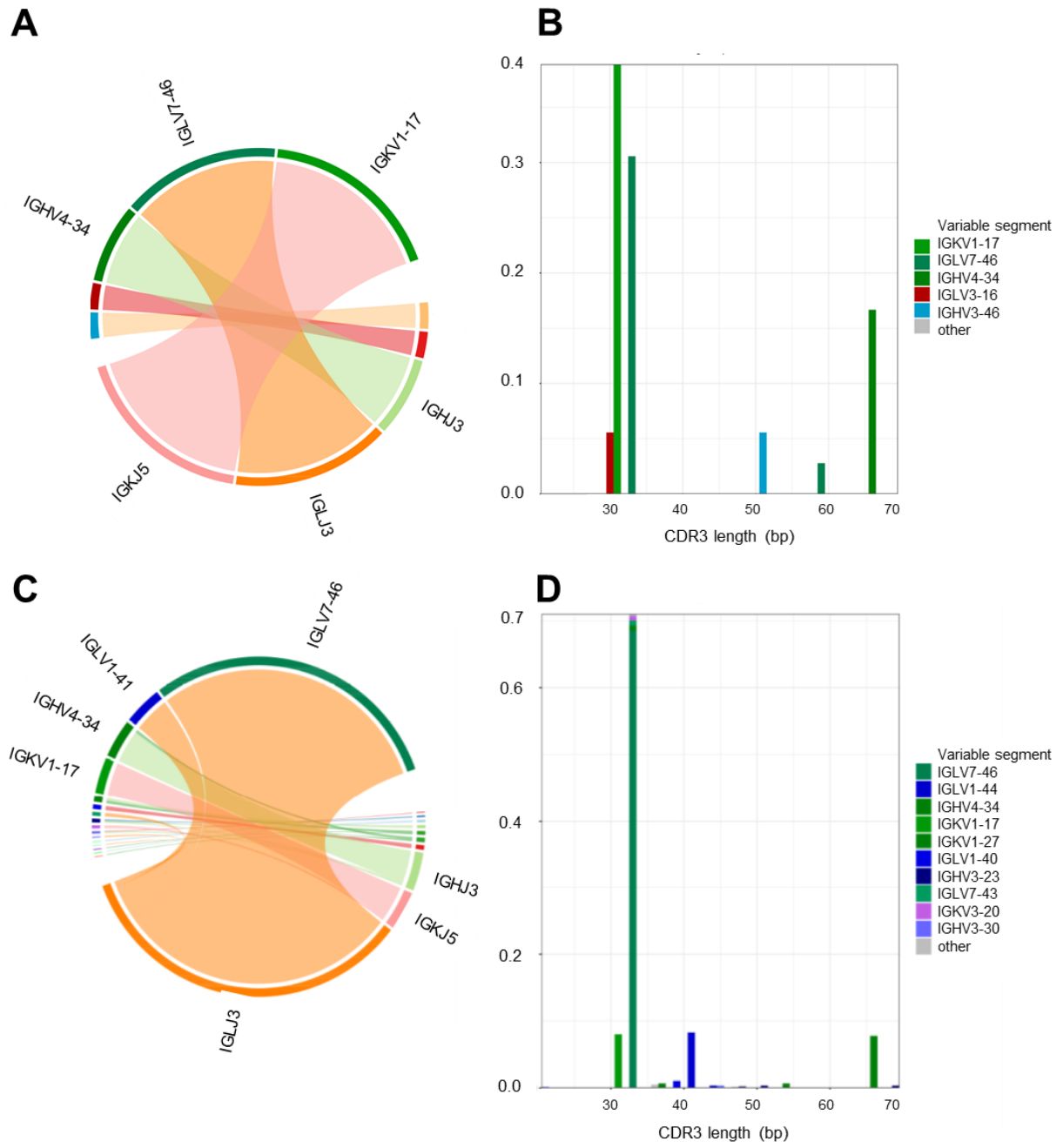
For the clonality assessment the paired-end 150bp aligned bam files were used. To improve the accuracy and speed of the analysis only reads aligned to the coding parts of IG genes were used that were partly soft-clipped as determined with samtools and samclip (v2.0). The V(D)J recombination was determined using Abstar (v0.3.5)[34]. The JSON output of Abstar was converted using R (v3.5.1) to csv format mimicking the IMGT output. All reads with exactly the same CDR3 were assigned to the V-gene with the longest fragment that map to the V-genes, because this fragment has the highest mapping reliability. The data was visualized with VDJtools (v1.2.1)[35]. Samples with the same CDR3-length and similar CDR3 sequence were considered clonal. Sample pairs for which no clonal V(D)J recombination was detected were manually inspected using IGV. If the same breakpoint in the IG region was found in both samples it was considered clonal as well.

To confirm validity of this method and to study 3 cases that were inconclusive with our method, 12 samples were analyzed using the novel EuroClonality assay for clonality detection of immunoglobulins. The assay was performed using two-step PCR clonality protocol for Illumina sequencing, which is an attuned protocol for FFPE- derived DNA based on the previously described two-step PCR protocol and the IG-clonality protocol for Ion Torrent sequencing.[36-38] The data was evaluated by using ARResT/Interrogate and the evaluation criteria described in van den Brand et al. 2021.[39]

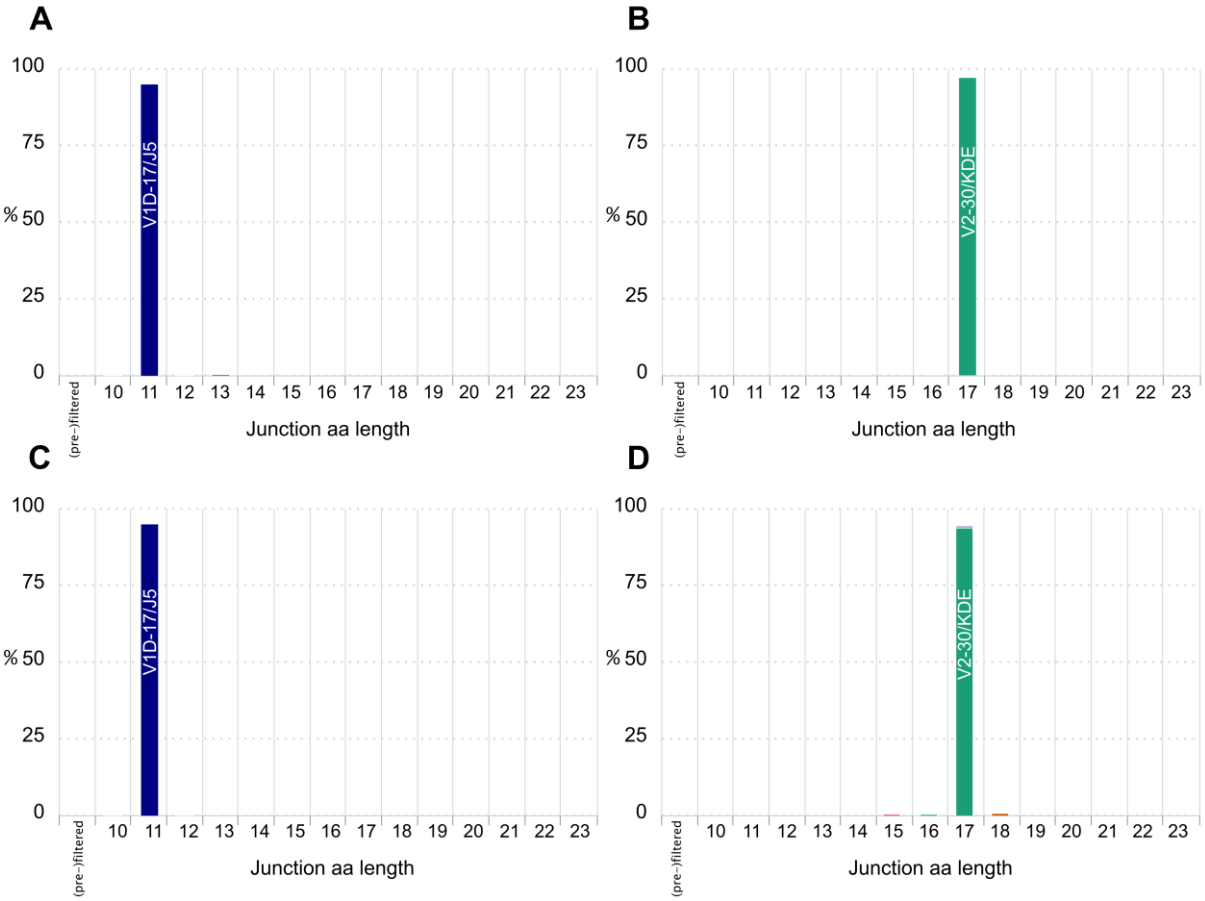
Supplementary Figures



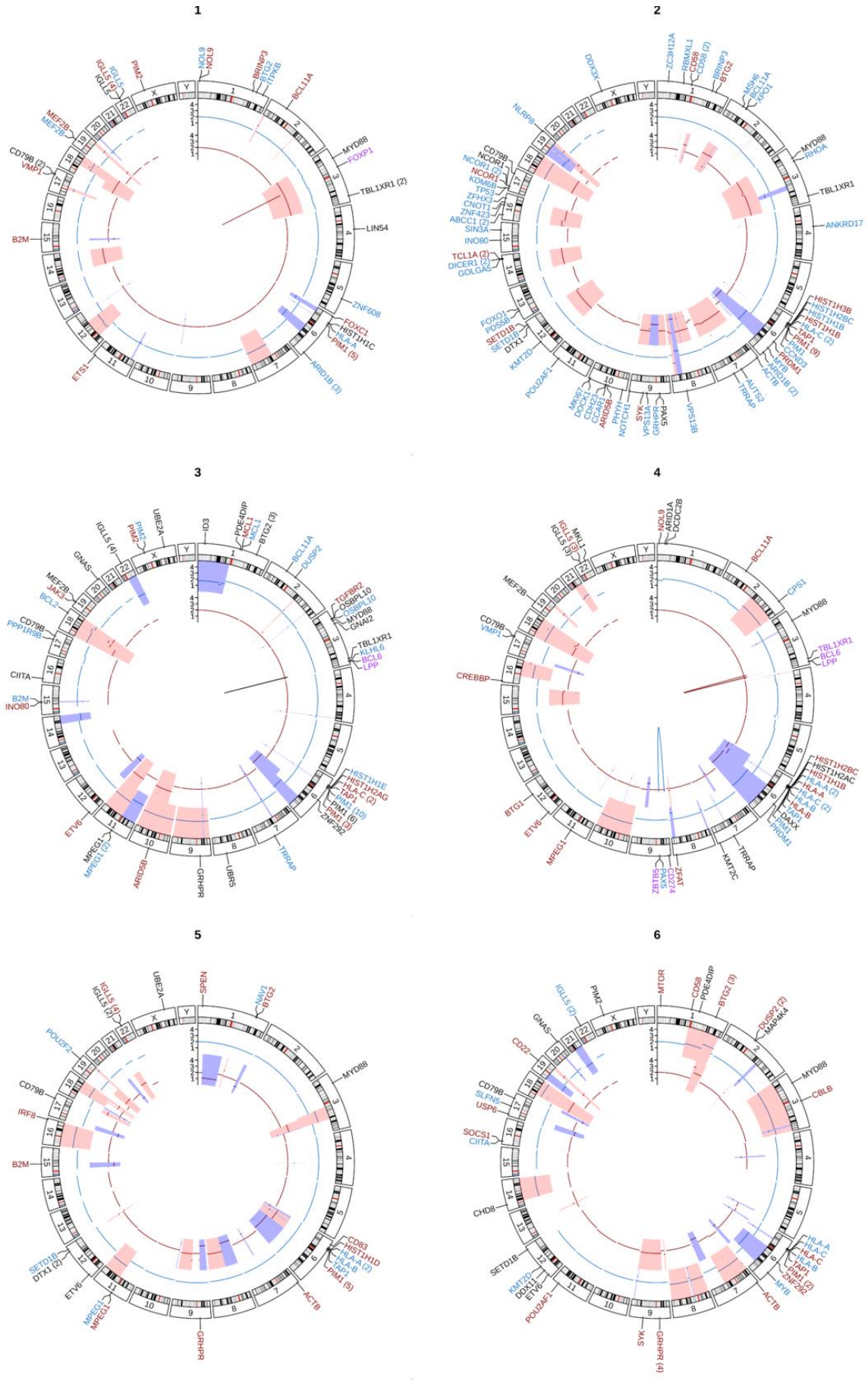
Supplementary Figure S1. Time between diagnosis and number of mutations. (A) The time between diagnosis depicted as beeswarm plot. Each dot is 1 patient, the median time between diagnosis is 4 years, (range 0-18 years). (B) The number of non-synonymous mutations in the primary samples and the relapse samples are depicted in boxplots. Sample pairs are connected by a line. Yellow samples indicate the presence of a matched normal and samples in grey are without matched normal. The fast majority of the relapse samples have a significantly higher number of mutations as determined with a Wilcoxon rank sum test in paired sample mode (23.5 vs 30 mutations on average, *P* value = 0.009).



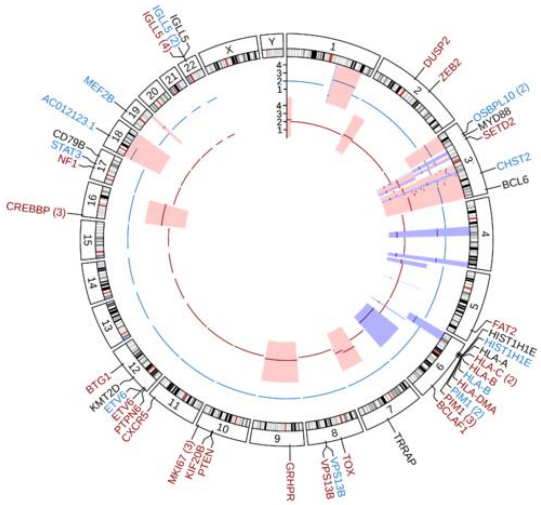
Supplementary Figure S2. Example of a clonal sample pair (patient #31). (A) Circular representation of immunoglobulin VJ recombinations determined with Abstar [34] and visualized with VDJtools [35] of the primary sample of patient #31, with in the lower part the J-segment and in the upper part the V-segment. (B) Barplots of the CDR3 lengths of the V(D)J recombination of the primary sample of patient #31. (C) Circular representation of immunoglobulin VJ recombinations of the relapse sample of patient #31. Details are as in A. Same segment combinations were seen in primary (A) and relapse (C) samples. (D) Barplots of the CDR3 lengths of the V(D)J recombination of the primary sample of patient #31. The major combinations in the primary and relapse sample are of the same length with the same V-J combination, indicating a clonal relationship.



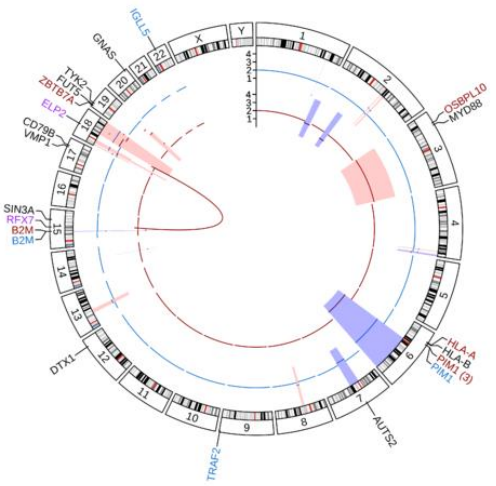
Supplementary Figure S3. Example of a clonal sample pair (patient #31) as determined with Euroclonality ARResT/Interrogate. Barplots of the amino acid junction length determined with Euroclonality ARResT/Interrogate of IGK-VJ in the primary (A) and relapse (C) sample of patient #31. Barplots of the amino acid junction length determined with Euroclonality ARResT/Interrogate of IGK-DE in the primary (B) and relapse (D) sample of patient #31.


1
NOL9
NOL9
BRINP3
BTG2
TFAP4
BCL11A
MYD88
FOXP1
TBL1XR1 (2)
LIN54
ZNF608
FOXO1
HIST1H1C
HLA-A
PIM1 (5)
ARID1B (3)
ETS1
B2M
CD79B (2)
VMP1
MEF2B
MEF2B
IGLL5
IGLL5
PIM2
2
DDX3X
ZC3H12A
RBM3XL1
CD58
CD58 (2)
BRINP3
BTG2
MSH6
BCL11A
RPO1
MYD88
RHOA
TBL1XR1
ANKRD17
HIST1H3B
HIST1H2BC
HIST1H1B
HLA-C (2)
PIM1 (9)
CCND3
PRDM1
ARID1B (2)
ACTB
AUTS2
TRRAP
VPS13B
SYK
GRHPR
VPS13A
PHYH
NOTCH1
ARID5B
CCAR2
DOCK11
MKI67
POU2AF1
KMT2D
DTX1
SETD1B
PDS5B
FOXO1
TCL1A (2)
DICER1 (2)
GOLGA5
SIN3A
INO80
ABCC1 (2)
ZNF423
CNOT1
ZFHX3
TP53
KDM6B
NCOR1
NCOR1 (2)
CD79B
NLRP8
3
UBE2A
ID3
PDE4DIP
MCL1
BTG2 (3)
BCL11A
DUSP2
TGFBR2
OSBPL10
OSBPL10
MYD88
GNAI2
TBL1XR1
KLHL6
BCL6
LPP
HIST1H1E
HIST1H2AG
HLA-C (2)
PIM1 (10)
PIM1 (9)
ZNF292
TRRAP
UBR5
GRHPR
ARID5B
MPEG1 (3)
MPEG1
ETV6
CIITA
B2M
INO80
CD79B
PPP1R9B
BCL2
MEF2B
JAK3
GNAS
IGLL5 (4)
PIM2
PIM2
4
NOL9
ARID1A
DDX23
IGLL5 (3)
IGLL5 (3)
MK1
BCL11A
CPS1
MYD88
TBL1XR1
BCL6
LPP
HIST1H2BC
HIST1H2AC
HLA-A (2)
HIST1H1B
HLA-A
HLA-C (2)
HLA-B
TAP1
DAXX
PRDM1
KMT2C
TRRAP
ZFAT
CD274
PDCD1LG2
RBM15
MPEG1
ETV6
BTG1
CREBBP
MEF2B
CD79B
VMP1
5
UBE2A
SPEN
NAV1
BTG2
IGLL5 (4)
IGLL5 (2)
MYD88
POU2F2
CD79B
IRF8
B2M
CD83
HIST1H1D
HLA-A (2)
HLA-B
TAP1
PIM1 (5)
ACTB
GRHPR
MPEG1
MPEG1
ETV6
SETD1B
DTX1 (2)
6
MTOR
CD58
PDE4DIP
BTG2 (3)
DUSP2 (2)
MAP4K4
IGLL5 (2)
PIM2
GNAS
CD22
MYD88
CBLB
CD79B
SLFN5
USP6
SOCS1
CIITA
HLA-A
HLA-C
HLA-B
TAP1
PIM1 (2)
ZNF292
MYB
ACTB
GRHPR (4)
SYK
POU2AF1
ETV6
DDX11
KMT2D
SETD1B
CHD8

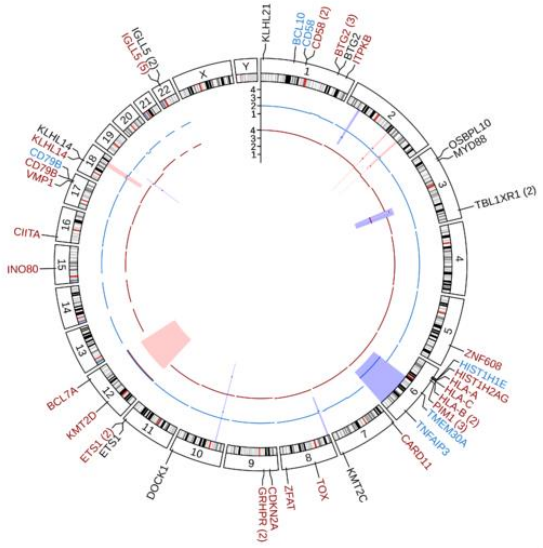
7


IGLL5
IGLL5 (4)
DUSP2
ZEB2
OSBPL10 (2)
MYD88
SETD2
CHST2
BCL6
FAT2
HIST1H1E
HLA-A
HLA-C (2)
HLA-B
HLA-DMA
PIM1 (2)
PIM1 (3)
BCLAF1
TRRAP
TOX
VPS13B
VPS13B
GRHPR
MKI67 (3)
KIF5B
PTEN
BTG1
KMT2D
ETV6
ETV6
PTPN6
CXCR3
CREBBP (3)
CD79B
STAT3
NF1
AC012123.1
MEF2B

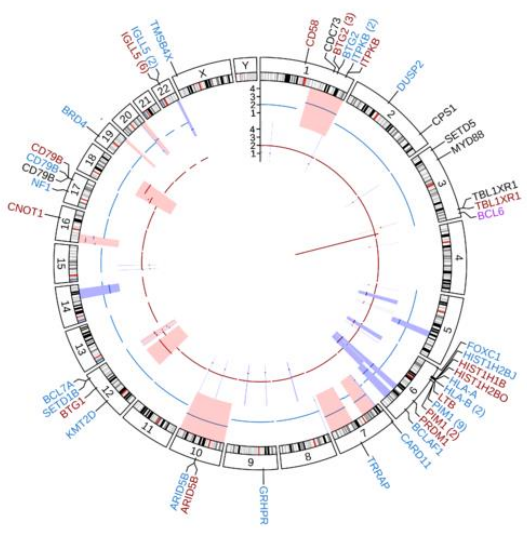
8


IGLL5
GNAS
TYK2
FLT3
ZBTB7A
ELP2
CD79B
VMP1
SIN3A
RFX7
B2M
B2M
DTX1
TRAF2
AUTS2
OSBPL10
MYD88
HLA-A
HLA-B
PIM1 (3)
PIM1

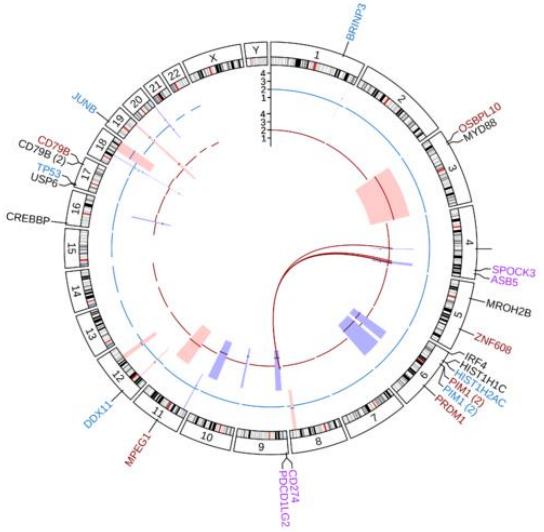
9


KLHL21
BCL10
CD58
CD58 (2)
BTG2 (3)
BTG2
ITPKB
IGLL5 (2)
IGLL5 (5)
OSBPL10
MYD88
TBL1XR1 (2)
ZNF608
HIST1H1E
HIST1H2AG
HLA-A
HLA-B (2)
PIM1 (3)
TNFAIP3
TMEM30A
CARD11
KMT2C
TOX
ZFAT
CDKN2A
GRHPR (2)
DOCK1
ETS1 (2)
ETS1
BCL7A
KMT2D
CIITA
INO80
KLHL14
KLHL14
CD79B
CD79B
VMP1

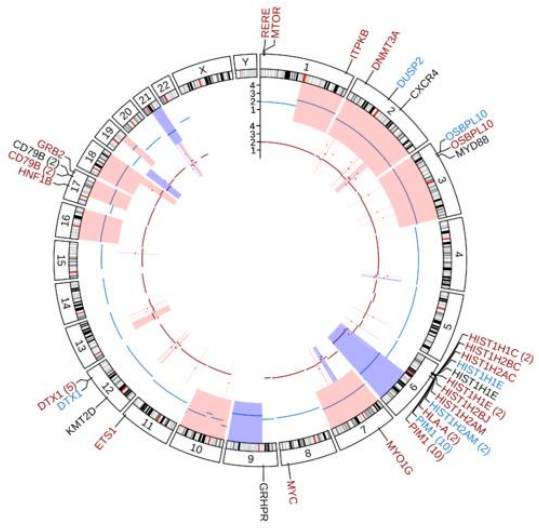
10


TMSB4X
IGLL5 (2)
IGLL5 (6)
CD58
CDC73
BTG2 (3)
BTG2
ITPKB (2)
ITPKB
DUSP2
GPS1
SET D5
MYD88
TBL1XR1
TBL1XR1
BCL6
FOXC1
HIST1H2BJ
HIST1H1B
HIST1H2BO
HLA-A
HLA-B (2)
LTB
PIM1 (9)
PIM1 (2)
PRDM1
BCLAF1
CARD11
TRRAP
GRHPR
ARID5B
ARID5B
BCL7A
SETD1B
BTG1
KMT2D
CNOT1
BRD4
CD79B
CD79B
CD79B
NF1

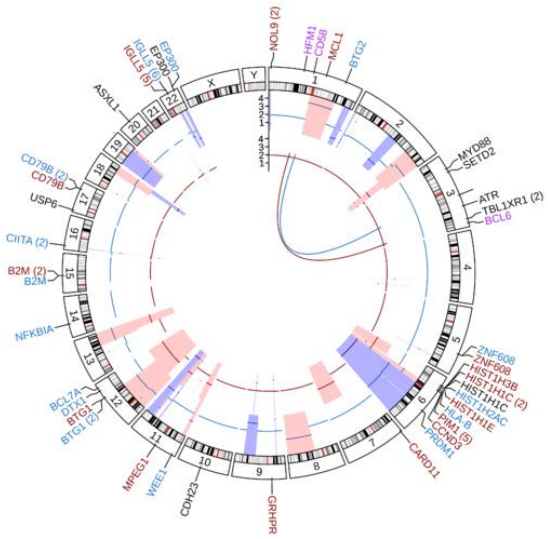
11


BRINP3
OSBPL10
MYD88
SPOCK3
ASB5
MROH2B
ZNF608
IRF4
HIST1H1C
HIST1H2AC
PIM1 (2)
PRDM1
CD274
PDCD1LG2
MPEG1
DDX11
JUNB
CD79B
CD79B (2)
TP53
USP6
CREBBP

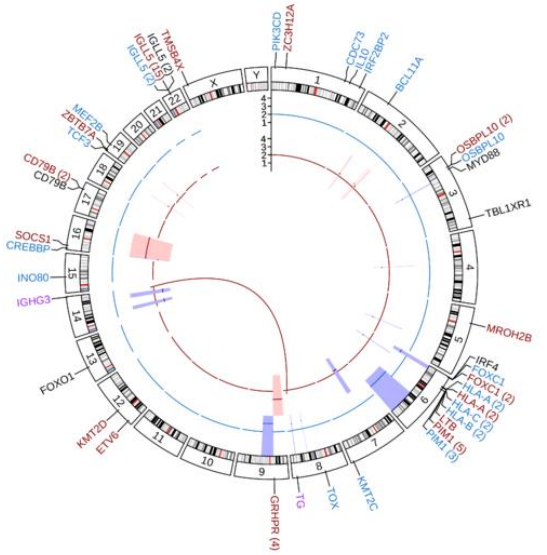
12


RERE
MTOR
ITPKB
DNMT3A
DUSP2
CXCR4
OSBPL10
OSBPL10
MYD88
HIST1H1C (2)
HIST1H2BC
HIST1H2AC
HIST1H1E
HIST1H1E
HIST1H2BJ (2)
HIST1H2AM
HIST1H2AM (2)
HLA-A (2)
PIM1 (10)
PIM1
MYO1G
MYC
GRHPR
ETS1
KMT2D
DTX1 (5)
DTX1
GRB2
CD79B
CD79B (2)
HNF1B

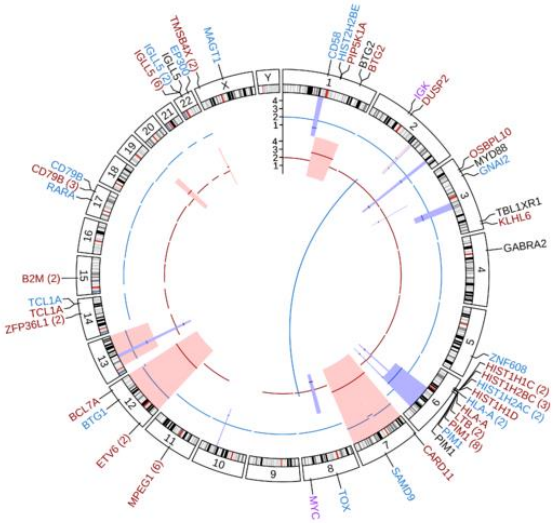
13



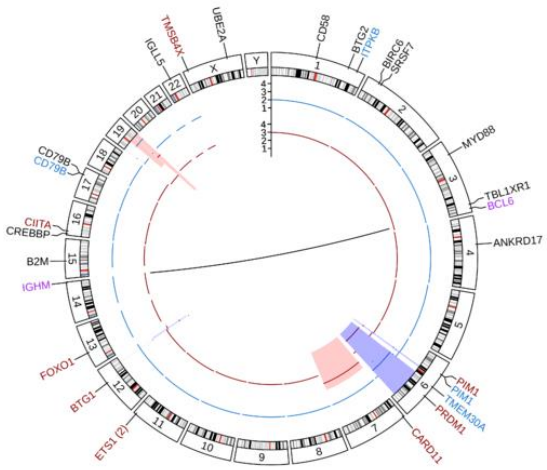
14



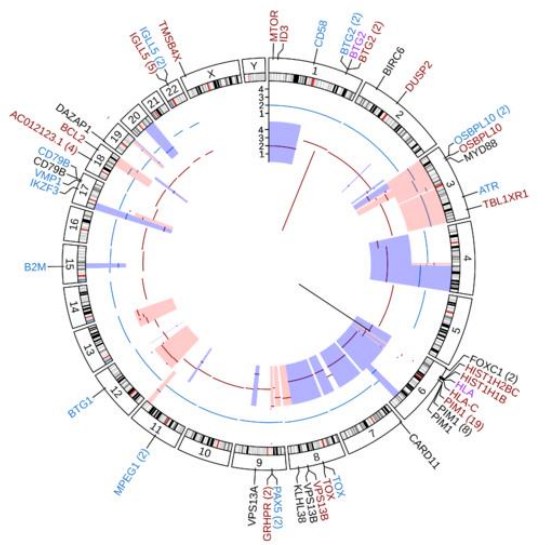
15



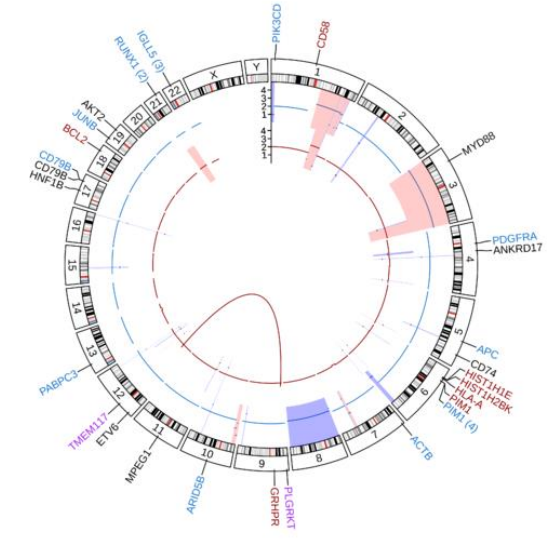
16



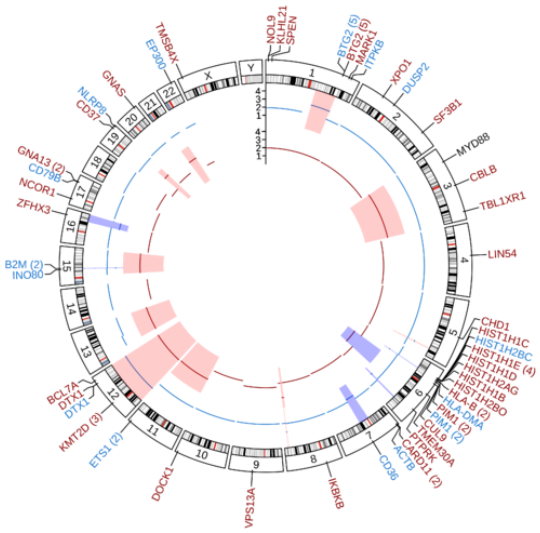
17



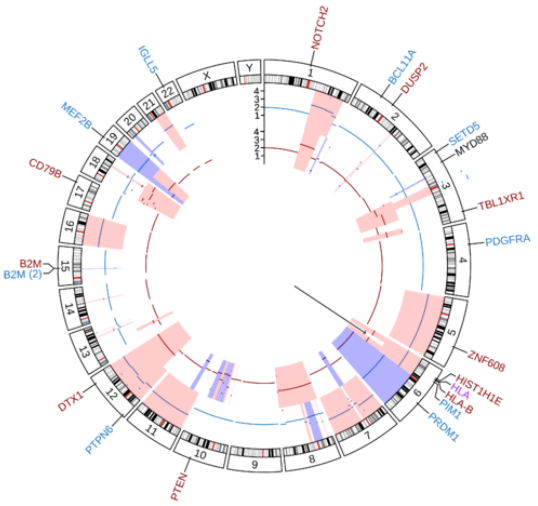
18



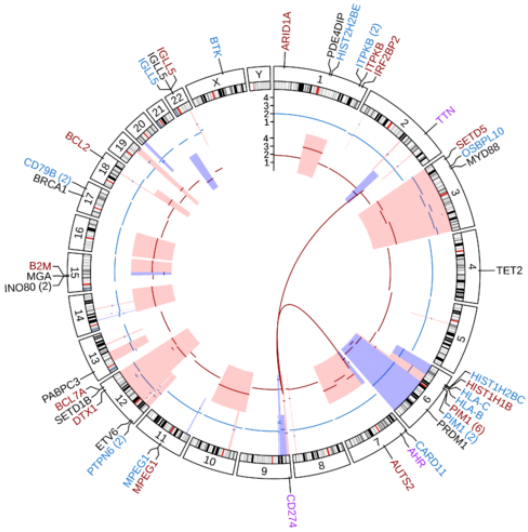
19



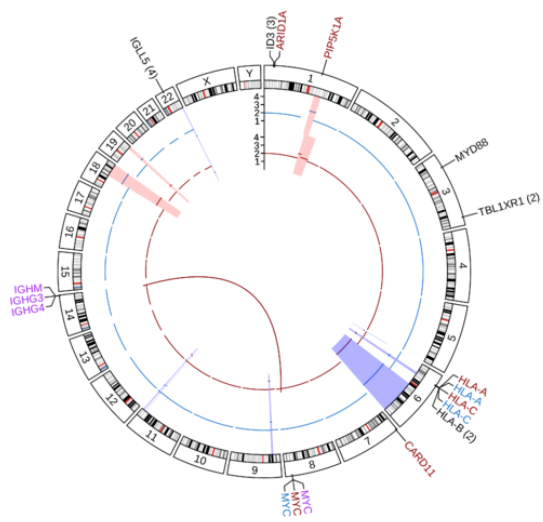
20



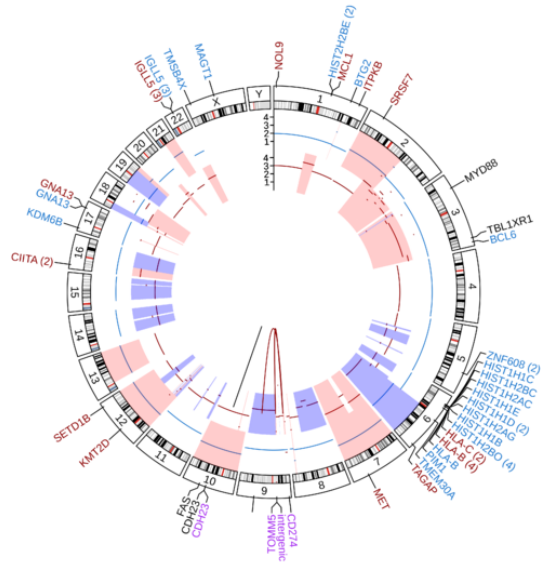
21



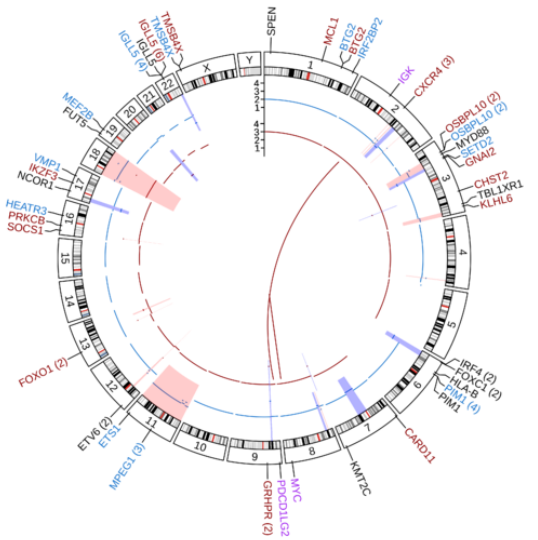
22



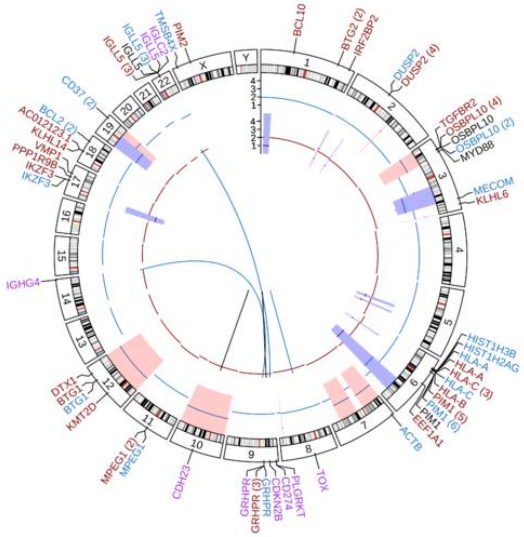
23



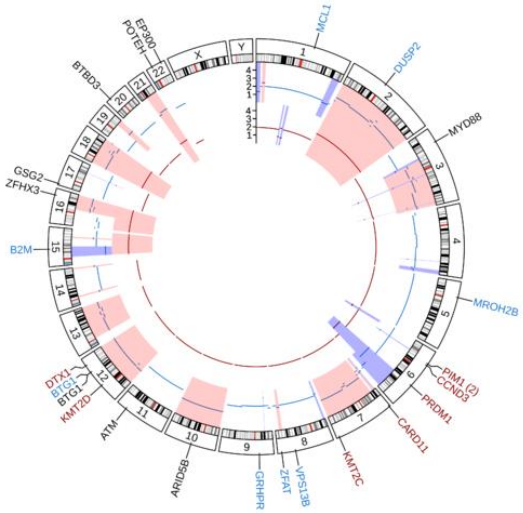
24



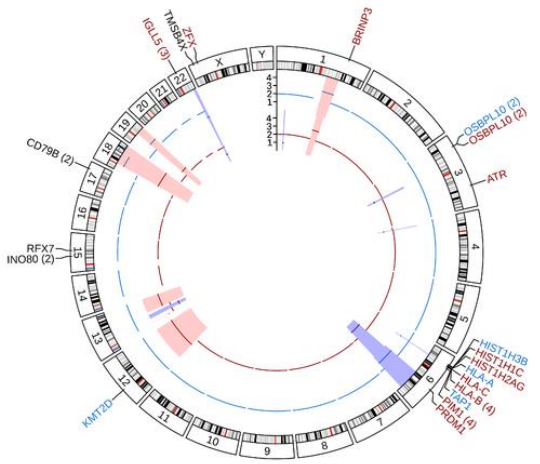
25



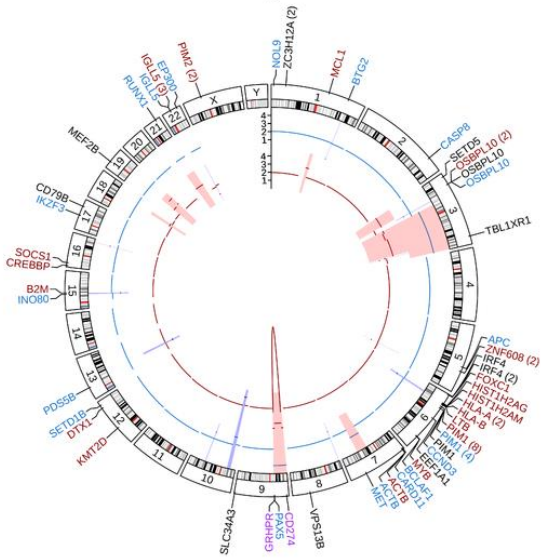
26



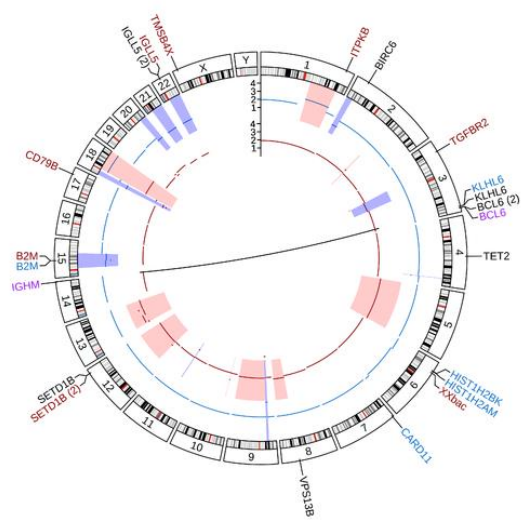
27



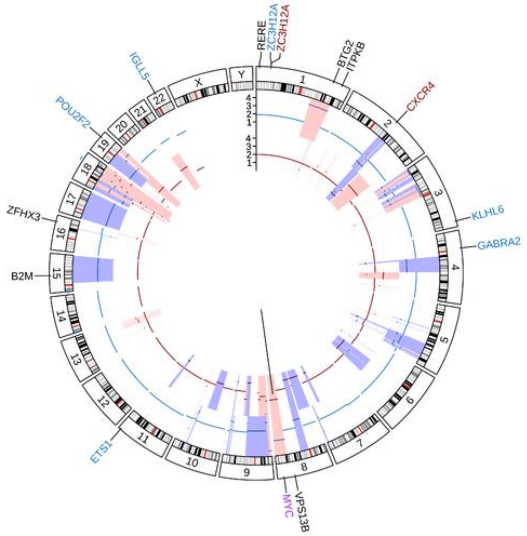
28

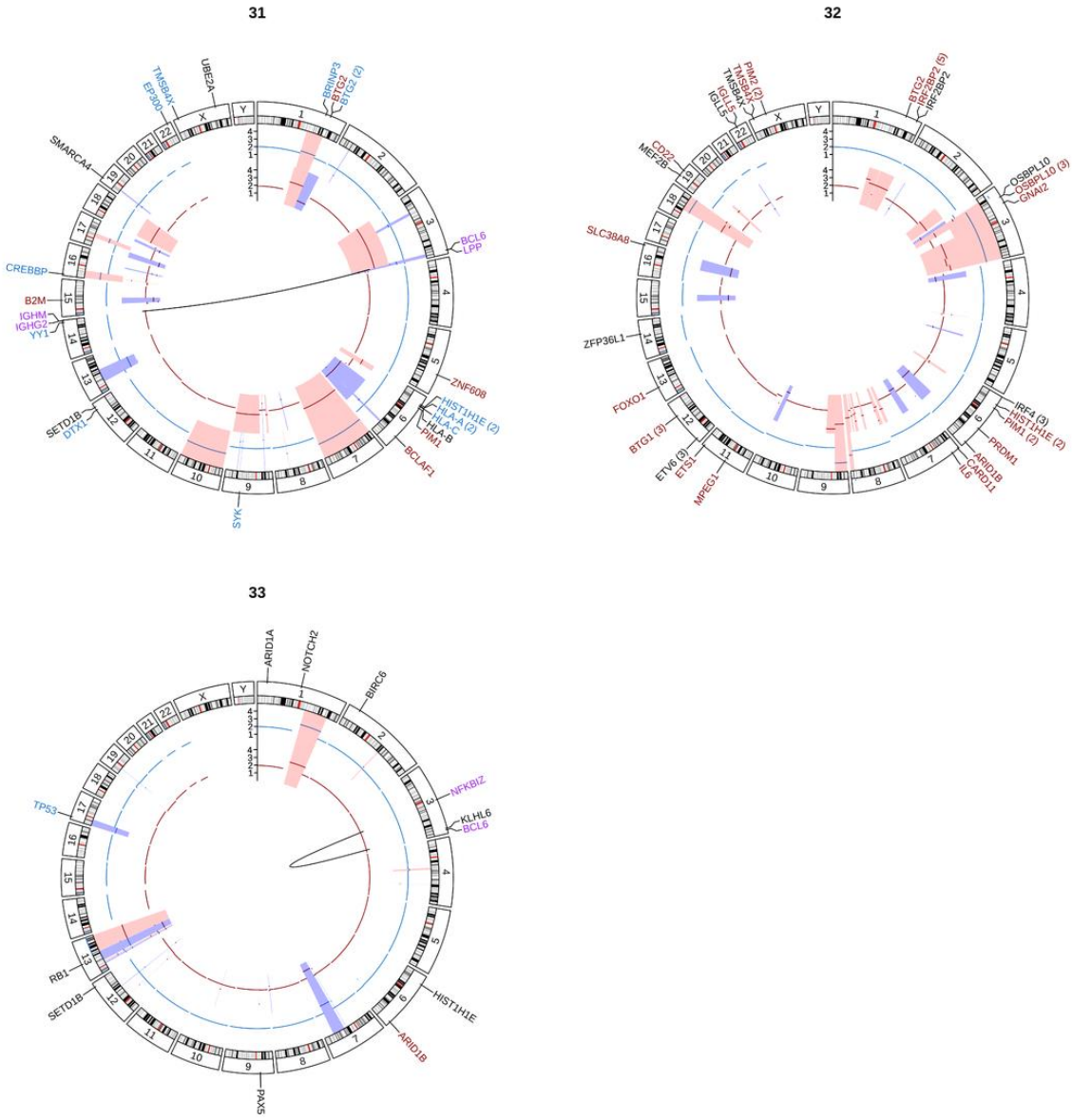


29

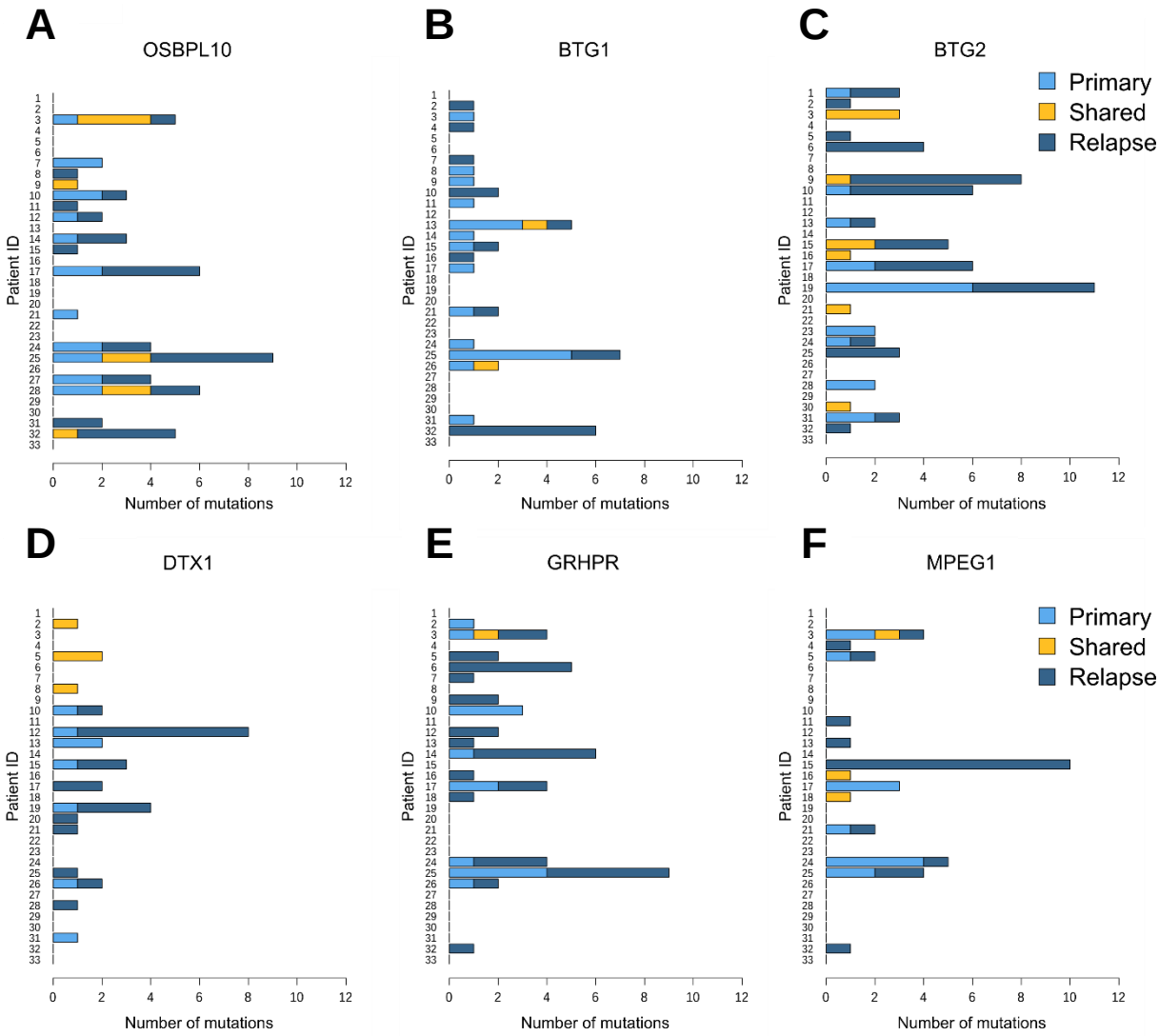


30

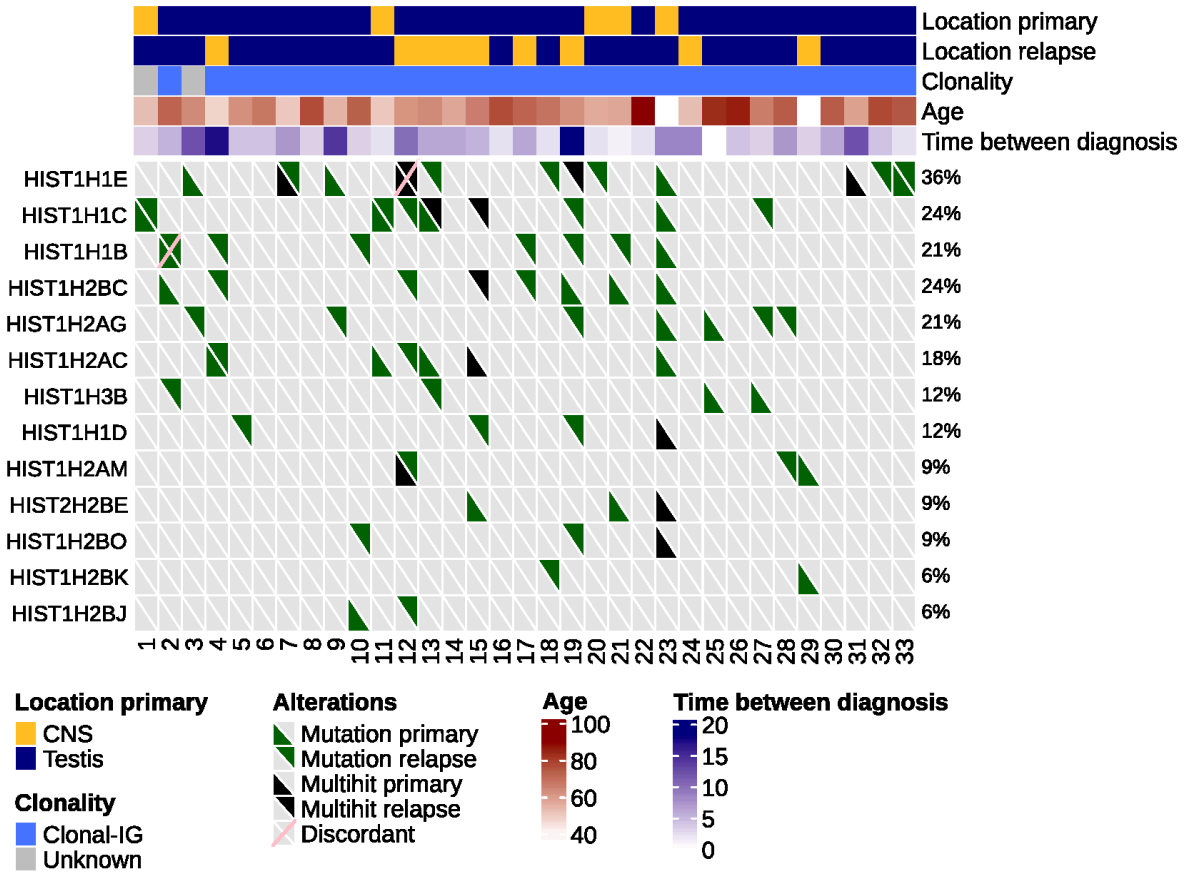




Supplementary Figure S4. Circosplots of alterations for each patient. Circosplot of all alterations in all primary and relapse samples (#1 - #33). Names of mutated genes are plotted around the ideogram of the chromosomes, in blue unique mutations of the primary sample, in red unique mutations of the relapse sample and in black the shared mutations between primary and relapse samples. If more than one mutation in the gene is present, the number of mutations is indicated in parentheses. Genes with translocations are indicated in purple, the connecting lines inside the ideogram indicate if the translocation is in the primary (blue), the relapse (red) sample or shared (black). Inside the circosplot the CNAs are depicted, in the outer circle the primary and in the inner circle the relapse CNAs. Copy number gains are depicted in red and losses in blue.



Supplementary Figure S5. aSHM frequently differed between primary and relapse LBCL-IP samples. Barplots of unique and shared aSHM. The mutations in the primary samples are indicated in light blue, the relapse samples in grey blue and the shared mutations in orange. (A) In OSBPL10 on average 14% of the SHM was shared, (B) in BTG1 on average 4% were shared mutations, (C) in BTG2 on average 23% were shared mutations, (D) in DTX1 on average 20% were shared mutations, (E) in GRHPR on average 1% were shared mutations and (F) in MPEG 1 on average 17% were shared mutations.



Supplementary Figure S6. Unique and shared linker histone mutations in paired samples. Oncoprint of linker histone alterations (rows) in sample pairs (columns). If present in the primary sample the bottom left corner is colored (according to the colors in the legend) and if present in the relapse sample the upper right corner is colored. If the alteration is discordant in the primary and the relapse sample a line is drawn across the box. In the upper panel, the location of the IP-LBCL, the clonality based on immunoglobulin analysis, age of the patient and time between diagnosis is indicated.

References

1. Los-de Vries GT, de Boer M, van Dijk E, et al. Chromosome 20 loss is characteristic of breast implant-associated anaplastic large cell lymphoma. *Blood*. Dec 17 2020;136(25):2927-2932. doi:10.1182/blood.2020005372
2. Scheinin I, Sie D, Bengtsson H, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Research*. 2014;24(12):2022-2032. doi:10.1101/gr.175141.114
3. van de Wiel MA, Brosens R, Eilers PH, et al. Smoothing waves in array CGH tumor profiles. *Bioinformatics*. May 1 2009;25(9):1099-104. doi:10.1093/bioinformatics/btp132
4. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. Mar 15 2007;23(6):657-63. doi:10.1093/bioinformatics/btl646
5. Poell JB, Mendeville M, Sie D, Brink A, Brakenhoff RH, Ylstra B. ACE: absolute copy number estimation from low-coverage whole-genome sequencing data. *Bioinformatics*. Aug 15 2019;35(16):2847-2849. doi:10.1093/bioinformatics/bty1055
6. van de Wiel MA, Kim KI, Vosse SJ, van Wieringen WN, Wilting SM, Ylstra B. CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics*. Apr 1 2007;23(7):892-4. doi:10.1093/bioinformatics/btm030
7. van de Wiel MA, van Wieringen WN. CGHregions: Dimension Reduction for Array CGH Data with Minimal Information Loss. *Cancer Informatics*. 02/08 2007;3:55-63.
8. van de Wiel MA, Smeets SJ, Brakenhoff RH, Ylstra B. CGHMultiArray: exact P-values for multi-array comparative genomic hybridization data. *Bioinformatics*. Jul 15 2005;21(14):3193-4. doi:10.1093/bioinformatics/bti489
9. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12(4):R41. doi:10.1186/gb-2011-12-4-r41
10. Los-de Vries GT, Stevens WBC, van Dijk E, et al. Genomic and microenvironmental landscape of stage I follicular lymphoma, compared with stage III/IV. *Blood Adv*. Sep 27 2022;6(18):5482-5493. doi:10.1182/bloodadvances.2022008355
11. Sturm M, Schroeder C, Bauer P. SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC bioinformatics*. 2016;17:208-208. doi:10.1186/s12859-016-1069-7
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. Jul 15 2009;25(14):1754-60. doi:10.1093/bioinformatics/btp324
13. Mose LE, Wilkerson MD, Neil Hayes D, Perou CM, Parker JS. ABRA: Improved coding indel detection via assembly-based realignment. *Bioinformatics*. 2014;30(19):2813-2815. doi:10.1093/bioinformatics/btu376
14. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics*. 2015;31(12):2032-2034. doi:10.1093/bioinformatics/btv098
15. Wilm A, Aw PP, Bertrand D, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res*. Dec 2012;40(22):11189-201. doi:10.1093/nar/gks918
16. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. Mar 2013;31(3):213-9. doi:10.1038/nbt.2514
17. Cingolani P, Patel VM, Coon M, et al. Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in Genetics*. 2012;3(MAR):35-35. doi:10.3389/fgene.2012.00035
18. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. Sep 2010;20(9):1297-303. doi:10.1101/gr.107524.110

19. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. Jan 8 2019;47(D1):D941-D947. doi:10.1093/nar/gky1015
20. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. May 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7
21. Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. Jan 8 2019;47(D1):D766-D773. doi:10.1093/nar/gky955
22. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. Jan 1 2001;29(1):308-11. doi:10.1093/nar/29.1.308
23. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. Jan 4 2018;46(D1):D1062-D1067. doi:10.1093/nar/gkx1153
24. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. Sep 15 2016;32(18):2847-9. doi:10.1093/bioinformatics/btw313
25. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics*. Oct 2014;30(19):2811-2. doi:10.1093/bioinformatics/btu393
26. Vater I, Montesinos-Rongen M, Schlesner M, et al. The mutational pattern of primary lymphoma of the central nervous system determined by whole-exome sequencing. *Leukemia*. Mar 2015;29(3):677-85. doi:10.1038/leu.2014.264
27. Khodabakhshi AH, Morin RD, Fejes AP, et al. Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget*. 2012;3(11):1308-1319. doi:10.18632/oncotarget.653
28. Alkodsi A, Cervera A, Zhang K, et al. Distinct subtypes of diffuse large B-cell lymphoma defined by hypermutated genes. *Leukemia*. Nov 2019;33(11):2662-2672. doi:10.1038/s41375-019-0509-6
29. Allahyar A, Pieterse M, Swennenhuis J, et al. Robust detection of translocations in lymphoma FFPE samples using targeted locus capture-based sequencing. *Nat Commun*. Jun 7 2021;12(1):3361. doi:10.1038/s41467-021-23695-8
30. Abo RP, Ducar M, Garcia EP, et al. BreaKmer: detection of structural variation in targeted massively parallel sequencing data using kmers. *Nucleic Acids Research*. 2015;43(3):e19-e19. doi:10.1093/nar/gku1211
31. Chong Z, Ruan J, Gao M, et al. novoBreak: local assembly for breakpoint detection in cancer genomes. *Nature methods*. 2017;14(1):6-11. doi:10.1038/nmeth.4084
32. Cameron DL, Schroder J, Penington JS, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res*. Dec 2017;27(12):2050-2060. doi:10.1101/gr.222109.117
33. Kronenberg ZN, Osborne EJ, Cone KR, et al. Wham: Identifying Structural Variants of Biological Consequence. *PLOS Computational Biology*. 2015;11(12):e1004572-e1004572. doi:10.1371/journal.pcbi.1004572
34. Briney B, Burton DR. Massively scalable genetic analysis of antibody repertoires. *bioRxiv*. 2018;doi:10.1101/447813
35. Shugay M, Bagaev DV, Turchaninova MA, et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput Biol*. Nov 2015;11(11):e1004503. doi:10.1371/journal.pcbi.1004503
36. van Bladel DAG, van der Last-Kempkes JLM, Scheijen B, Groenen P, EuroClonality C. Next-Generation Sequencing-Based Clonality Detection of Immunoglobulin Gene Rearrangements in B-Cell Lymphoma. *Methods Mol Biol*. 2022;2453:7-42. doi:10.1007/978-1-0716-2115-8_2
37. Scheijen B, Meijers RWJ, Rijntjes J, et al. Next-generation sequencing of immunoglobulin gene rearrangements for clonality assessment: a technical feasibility study by EuroClonality-NGS. *Leukemia*. Sep 2019;33(9):2227-2240. doi:10.1038/s41375-019-0508-7
38. Bruggemann M, Kotrova M, Knecht H, et al. Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in

acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia*. Sep 2019;33(9):2241-2253. doi:10.1038/s41375-019-0496-7
39. van den Brand M, Rijntjes J, Mobs M, et al. Next-Generation Sequencing-Based Clonality Assessment of Ig Gene Rearrangements: A Multicenter Validation Study by EuroClonality-NGS. *J Mol Diagn*. Sep 2021;23(9):1105-1115. doi:10.1016/j.jmoldx.2021.06.005