Supplemental Data

1. **Supplemental Methods**
2. **Supplemental Figures**

supplemental Figure 1 – Kaplan Meier curve PFS and OS stage I

supplemental Figure 2 – Translocation breakpoints of BCL2 and BCL6

supplemental Figure 3 – Boxplot copy number load and number of nonsynonymous and splice site mutations per stage

supplemental Figure 4 – Analysis to determine number of clusters

supplemental Figure 5 – Barplot frequencies of the mutations and translocations per cluster stage I

supplemental Figure 6 – Boxplot copy number load per cluster

supplemental Figure 7 – Oncoprint sorted per cluster

supplemental Figure 8 – Consensus index

supplemental Figure 9 – Barplot frequencies of the mutations and translocations per cluster stage III/IV

supplemental Figure 10 – Hierarchic clustering plot stage III/IV

supplemental Figure 11– Barplot frequencies of the mutations and translocations per stage for the cases with microenvironment and NGS data complete

supplemental Figure 12 – Copy number landscape per stage for the cases with microenvironment and NGS data complete

supplemental Figure 13 – Hierarchic clustering plot per stage for the cases with microenvironment and NGS data complete

supplemental Figure 14 – Boxplots of microenvironment per cluster of follicular lymphoma

supplemental Figure 15 – Mutations in STAT6, HIST1H1C and HIST1H1E

3. **Supplemental Tables**

supplemental Table 1 – Inclusion criteria, treatment options and follow-up per trial

supplemental Table 2 – Antibodies used for IHC microenvironment and tumor cells

supplemental Table 3 – Genes in capture panel for mutation analysis

supplemental Table 4 – Regions in capture panel for translocation analysis

supplemental Table 5 – Case overview and data availability

supplemental Table 6 – Clinical characteristics of the cases fulfilling the clinical criteria compared to cases with complete NGS-data

supplemental Table 7 – Immunohistochemistry staining results tumor cells stage I

supplemental Table 8 – Microenvironment results per cohort

supplemental Table 9 – Translocation results

supplemental Table 10 – Copy number Alteration frequencies per cohort

supplemental Table 11 – Functional somatic mutations per patient

supplemental Table 12 – Mutation frequencies per cluster

supplemental Table 13 – Microenvironment frequencies per cluster

Supplemental Methods

Tissue Microarrays construction

On Haematoxylin and Eosin (H&E) stained slides from the original block represented tumor areas were marked. Two one mm cores were pounced out with semi-automatically and put into the tissue microarray (TMA). Reactive tonsils or kidneys were used as controls and orientation. A standard microtome technique was used for sectioning 3 µm sections into slides, which were marked with the appropriate identification.

Immunohistochemistry

For section adhesion the TMA slides were placed in racks at 60°C overnight. For de-waxing and de-hydration slides were incubate in xylene for 2 consecutive periods of 5 minutes (min). Subsequently, slides were transferred to Industrial Methylated Spirits (IMS) solution for 5 min and further 2 periods of 5 min in hydrogen peroxide in order to dehydrate tissue and reduce non-specific staining from the action of endogenous peroxidases on the chromogen. A final incubation of 5 min in IMS and rinsing in running tap water is required prior to antigen retrieval.

For the first incubation steps, 3000 ml of antigen unmasking solution was warmed up in a pressure cooker. When boiling, the racks with slides were immersed and left for 10 min at high heat (120-130°C) from the time a steady flow of steam escaped the outlet valve. After that the pressure cooker was cold down and opened. Once opened the slides were cooled under cold running tap water for 5 min and then quickly transferred to a wash buffer pot ensuring the slides don't dry.

A hydrophobic pen marked the edge of the array field on the slides and wash buffer was applied on the array field to keep it wet throughout the remaining procedure. The DAKO Autostainer System was used for timed dispensing of reagents into the slides run for 2-3 hours, according to the programmed software (Dako Autostainer Plus) for the number of slides, reagents and incubation times and rinse steps. The Super SensitiveTM Polymer-HRP IHC Detection System (Biogenex) was used for signal detection. After finishing all slides were replaced in racks and rinsed in tap water for 5 min. As a counterstain, the slides were placed

in haematoxylin solution for 5 min, rinsed for 2 min in running water and immersed quickly into acid alcohol solution for 5 times, after which were rinsed again for 2 min in running tap water. The slides were re-hydrated using IMS for 3 periods of 2 min and clarified by incubation in xylene baths. DPX xylene was used as mounting media, cover slips were applied without trapped air bubbles and left to dry.

Automated Image analysis using the Ariol SL-50 visual analysis software

Slides were scanned with the Olympus BX61 microscope on an automated platform (Prior). All cores were reviewed manually to exclude cores with less than 50% of tumor tissue, due to fibrotic or necrotic areas or technical artifacts. Training was done on representative areas. Positive stained cells or areas acquire a brown/black color characteristic of DAB. To allow contrast with the background the color hue, saturation and intensity were manipulated by selection individual pixels from positive events and not included negative or non-specific stained pixels. Training improved by limiting the size and shape of the areas considered positively. This procedure was also done for identifying the negative stained cells or areas.

Automated Image analysis using the Pannoramic Viewer System

The Pannoramic 250 Flash II scanner (3DHISTECH) was used for scanning the slides. Each core was observed on a computer screen using the Pannoramic Viewer computer interface for bookmarking the representative tumour areas and quantify the areas of interest. After this selection, the DensitoQuant module was used to quantify the number of DAB stained pixels. Only the top red and orange levels were used for identification of stained areas. For each antibody an optimal script was saved, after adjusting the brown tolerance and the score levels, and applied for analysis the selected areas. With the system we were able to calculate the % of positive cells or positive area of the total number of cells or area in the core.

DNA isolation and Library preparation

Genomic DNA was extracted from all LLBC tumor cores with a QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) and quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Carlsbad CA, USA).[1] The DNA was sheared by ultrasound with a Covaris ME220 (Covaris Inc, Woburn MA, USA), with settings adjusted to DNA from FFPE tissue, as previously described.[2] NGS-libraries were prepared with an input of 100ng sheared DNA using KAPA or KAPA Hyper Library Preparation (KAPA Biosystems, Wilmington MA, USA). In short, uniquely 8-basepairs (bp) indexed adapters (IDT, Coralville IA, USA) were ligated to the FFPE-extracted DNA, followed by purification using AMPure XP beads (Beckman Coulter, Brea CA, USA), which resulted in a fragment size in the range of 150 and 400bp.

Shallow whole genome sequencing (WGS) and copy number analysis (CNA)

For shallow WGS, 10nM of up to 24 barcoded samples NGS-libraries were equimolarly pooled and 12.5pM was loaded on one lane of a HiSeq Single End Flowcell (Illumina, San Diego CA, USA). Sequencing was performed on a HiSeq 4000 (Illumina, San Diego CA, USA) in a single-read 50-cycle run mode (SR50). Copy number analysis was performed as described previously.[3,4] Reads were aligned to the human reference genome build GRCh37/hg19 with BWA (v0.7.5) [5,6] and duplicates were marked with Picard (v2.15). Further analysis is performed in R (v3.4.1) using the Bioconductor package QDNAseq (v1.12.0)[2], if a sample had two bam files these were merged by QDNAseq, than the genome was divided into nonoverlapping bins of 100kb, followed by correction of GC content and mappability. Filtering of artefacts and germline variations was performed by a previously constructed blacklist containing regions with low mappability, common germ-line copy number variants and other regions with large deviations in genomes from the 1000 Genome project.[7] Wave correction was performed with NoWaves (v0.6).[8] Based on QDNAseq segmentation created by DNAcopy (v1.50.1)[9] ACE (v0) estimated the cellularity and absolute copy numbers.[10] CGHcall (v2.38.0)[11] used the cellularity as correction, with a minimum cellularity of 0.2, to call the CNAs. To reduce the number of data points CGHregions (v1.34) was used, with a maximal information loss of 1% allowed. Stage I and Stage III/IV follicular lymphoma (FL) were compared with CGHtest (v1.1)[12], which implements a two-sided Wilcoxon Rank-Sum

Test with 10,000 permutations including a false discovery rate (FDR) correction for multiple testing.

Targeted capture and deep sequencing for mutation and translocation analysis.

A custom targeted panel was designed using NimbleGen design software (Roche) to detect mutations and translocations of interest. All exons of 369 genes and 12 translocation targets were captured, including genic and intergenic regions (Roche ID 43712; supplemental Table 3 and 4). The panel was designed with the aim to cover most important driver genes of Follicular Lymphoma and Diffuse Large B-cell Lymphoma. Genes included in two commercially available lymphoma panels (Foundation One Hemo and HemoSeq 1.0) and genes annotated as driver genes in literature [13-18] The capture was performed according to NimbleGen EZ SeqCap library protocol (Roche Nimblegen, Madison WI, USA). 125ng was used from NGS-libraries to create equimolar pools with a total mass of 1µg DNA. Sequencing of the captured NGS-libraries was performed on the HiSeq 4000 (Illumina, San Diego CA, USA) in paired-end 150bp mode. This resulted in a mean target coverage of 246x. Paired-end 150bp reads were de-multiplexed by Bcl2fastq (Illumina) and Seqpurge (v0.1-104) trimmed the adapter sequences.[19] The reads were aligned to the human reference genome (GRCh37/hg19) with BWA mem (v0.7.12).[2] Mapped reads were realigned with ABRA (v2.19) and picardtools MarkDuplicates (v2.4.1) marked duplicate reads[20], to include secondary alignments in in the deduplication the setting ASSUME SORT ORDER=queryname was used (this is particularly important for translocation calling). Samples with a mean target coverage < 30 reads were excluded for further analysis. LoFreq (v2.1.3.1)[21] and Mutect2 in combination with filterMutect2 (v4.1.7.0)[22] were used for mutation calling using the following criteria: coverage depth >15x, minimal read and base quality >20, variant supporting reads >2 in each direction, variant allele frequency (VAF) >0.05 and the Mutect2 Phred-scaled qualities that alt allele are not due to read orientation artifact (ROQ) or the Log 10 likelihood ratio score of variant existing versus not existing (TLOD) must be >20 to reduce background noise. Further mutations present at least 2 times in the panel of normals (an in house AmsterdamUmc set consisting of 25 non-tumor samples (12 blood samples, 4 FFPE hyperplasia lymph node, 6 FFPE reactive lymph node and 3 FFPE epithelial tissues)). Mutations must be called by both callers to be included in further analysis.

Effect prediction of called variants and functional annotation was performed with Funcotator (v4.1.7.0)[23] and SnpSift (v.4.3)[24] using the database of COSMIC (v84)[25], gnomAD

(v2.0.2)[26], gencode(v19)[27], dbsnp (build 151)[28], clinvar (20180401)[29], and the HMF panel of normals (v2.0).

Single nucleotide variants (SNVs) and small indels were labeled somatic if they were not common in dbsnp and not present > 3 times in the HMF panel of normals. Mutations marked by funcotator as intronic, silent, UTR or flanking mutations were removed for the analysis. All downstream analyses were performed in the custom script of programming language R (version 3.6.1). The Oncoprint is created using the ComplexHeatmap package (2.7.1.1016).[30] Somatic hypermutations (SHM) was called when a known target gene (*BCL2, BCL6, MYC* and *PIM1*) contained two or more mutations.

Complete-linkage hierarchical clustering was performed with the function 'hclust' of the 'stats' package. Distances were defined as $1\text{-cor}_{spearman}$ for both the genes and the patient samples, implemented by the 'cor' function, also from the 'stats' package. Dunn-index was calculated with clValid (version 0.7.1) for determining the ratio of mean intra-cluster distances to inter-cluster distances for 2-10 clusters. Lower scores indicate a better separation of the clusters. Stability of the clustering was tested by the method described in Monti et al. (2013), performing the clustering 1000 times on 61/84 randomly selected samples, and evaluating the consensus index.[31] All analysis was performed in R (version 3.5.1).

For translocation detection, four bioinformatic tools were combined including BreaKmer, GRIDDS, Wham and novoBreak[32-35] as previously described in detail.[36] Translocations detected by at least two tools were visual confirmed using the Integrative Genome Viewer (IGV).[37]

Data availability

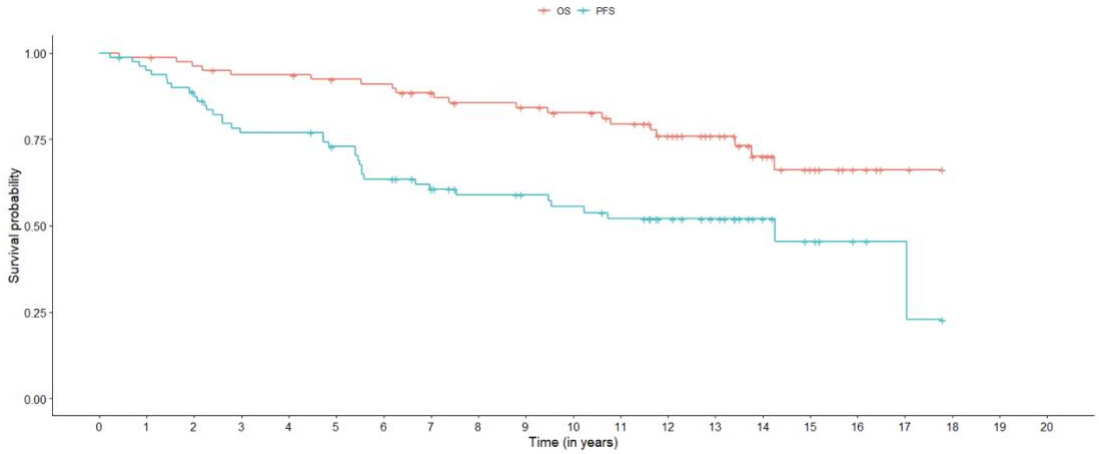
All sequence data has been uploaded to the European Genome-phenome Archive (EGA; accession number EGAS00001005755

Acknowledgement:

We thank the Hartwig Medical Foundation (Amsterdam, The Netherlands) for generating, analyzing and providing access to reference whole genome sequencing data of the Netherlands population.

1. van Essen HF, Ylstra B. High-resolution copy number profiling by array CGH using DNA isolated from formalin-fixed, paraffin-embedded tissues. *Methods Mol Biol*. 2012;838:329-341.
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 Res*. 2014;24(12):2022-2032.
3. Stevens WBC, Mendeville M, Redd R, et al. Prognostic relevance of CD163 and CD8 combined with EZH2 and gain of chromosome 18 in follicular lymphoma: a study by the Lunenburg Lymphoma Biomarker Consortium. *Haematologica*. 2017;102(8):1413-1423.
4. 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*. 2020;136(25):2927-2932.
5. Flicek P, Ahmed I, Amode MR, et al. Ensembl 2013. *Nucleic Acids Res*. 2013;41(Database issue):D48-55.
6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
7. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65.
8. van de Wiel MA, Brosens R, Eilers PH, et al. Smoothing waves in array CGH tumor profiles. *Bioinformatics*. 2009;25(9):1099-1104.
9. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. 2007;23(6):657-663.
10. 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*. 2019;35(16):2847-2849.
11. van de Wiel MA, Kim KI, Vosse SJ, van Wieringen WN, Wilting SM, Ylstra B. CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics*. 2007;23(7):892-894.
12. van de Wiel MA, Smeets SJ, Brakenhoff RH, Ylstra B. CGHMultiArray: exact P-values for multi-array comparative genomic hybridization data. *Bioinformatics*. 2005;21(14):3193-3194.
13. Zhang J, Grubor V, Love CL, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A*. 2013;110(4):1398-1403.
14. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011;476(7360):298-303.
15. Lohr JG, Stojanov P, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A*. 2012;109(10):3879-3884.
16. Pasqualucci L, Trifonov V, Fabbri G, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet*. 2011;43(9):830-837.
17. Pastore A, Jurinovic V, Kridel R, et al. Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: a retrospective analysis of a prospective clinical trial and validation in a population-based registry. *Lancet Oncol*. 2015;16(9):1111-1122.
18. Pasqualucci L, Khiabanian H, Fangazio M, et al. Genetics of follicular lymphoma transformation. *Cell Rep*. 2014;6(1):130-140.
19. Sturm M, Schroeder C, Bauer P. SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics*. 2016;17:208.

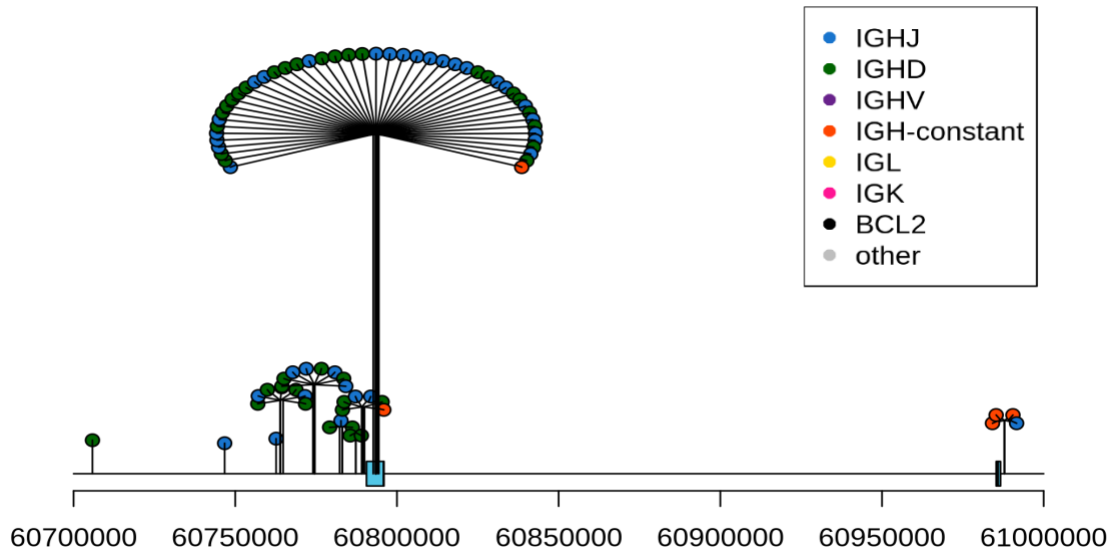
20. Mose LE, Wilkerson MD, Hayes DN, Perou CM, Parker JS. ABRA: improved coding indel detection via assembly-based realignment. *Bioinformatics*. 2014;30(19):2813-2815.
21. Wilm A, Aw PP, Bertrand D, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res*. 2012;40(22):11189-11201.
22. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213-219.
23. 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.
24. 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.
25. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. 2019;47(D1):D941-D947.
26. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443.
27. Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019;47(D1):D766-D773.
28. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-311.
29. 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.
30. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847-2849.
31. Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning*. 2003;52(July):91-118.
32. Abo RP, Ducar M, Garcia EP, et al. BreaKmer: detection of structural variation in targeted massively parallel sequencing data using kmers. *Nucleic Acids Res*. 2015;43(3):e19.
33. Cameron DL, Schroder J, Penington JS, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res*. 2017;27(12):2050-2060.
34. Kronenberg ZN, Osborne EJ, Cone KR, et al. Wham: Identifying Structural Variants of Biological Consequence. *PLoS Comput Biol*. 2015;11(12):e1004572.
35. Chong Z, Ruan J, Gao M, et al. novoBreak: local assembly for breakpoint detection in cancer genomes. *Nat Methods*. 2017;14(1):65-67.
36. Allahyar A, Pieterse M, Swennenhuis J, et al. Robust detection of translocations in lymphoma FFPE samples using targeted locus capture-based sequencing. *Nat Commun*. 2021;12(1):3361.
37. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178-192.



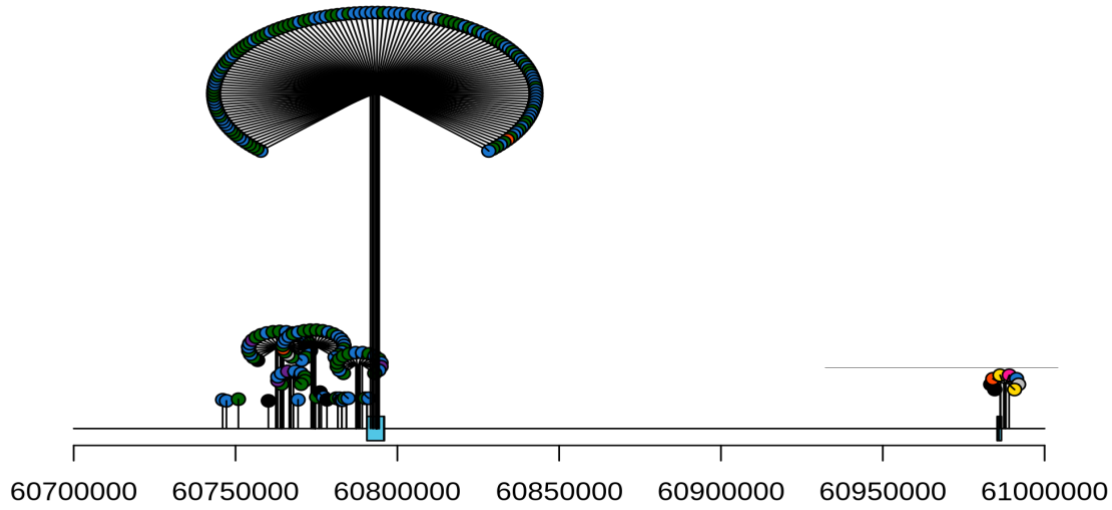
241

242 **Supplemental Figure 1 – Kaplan Meier curve PFS (blue line) and OS (red line) stage I**

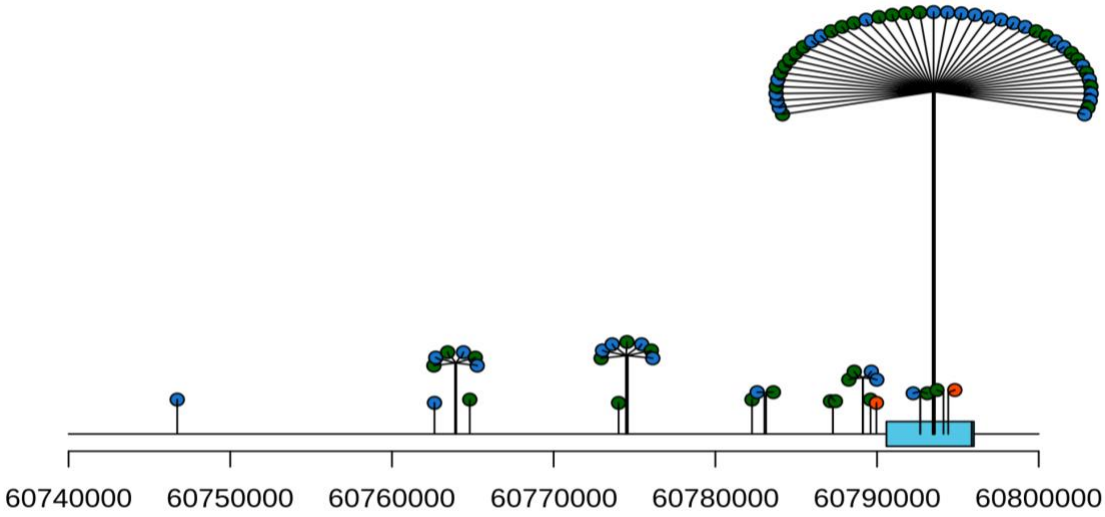
A Stage I – BCL2 region



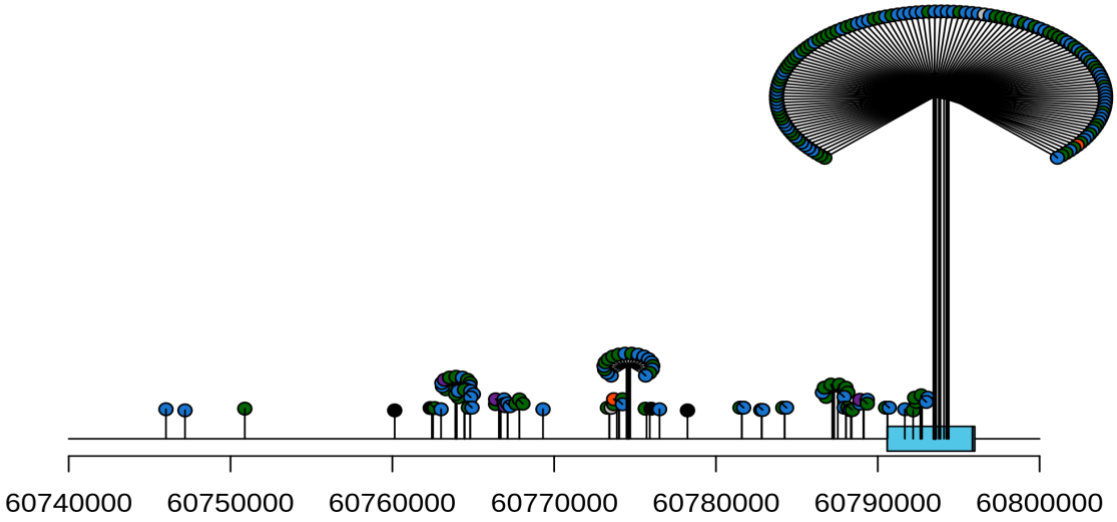
B Stage III/IV – BCL2 region



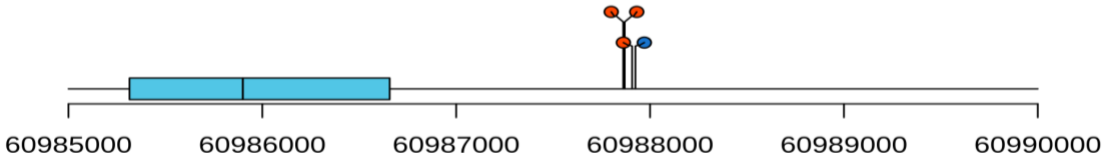
C Stage I - zoomed BCL2 exon 3 including MBR and downstream region



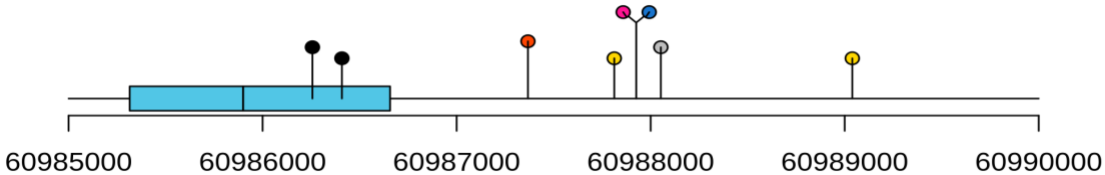
D Stage III/IV - zoomed BCL2 exon 3 including MBR and downstream region



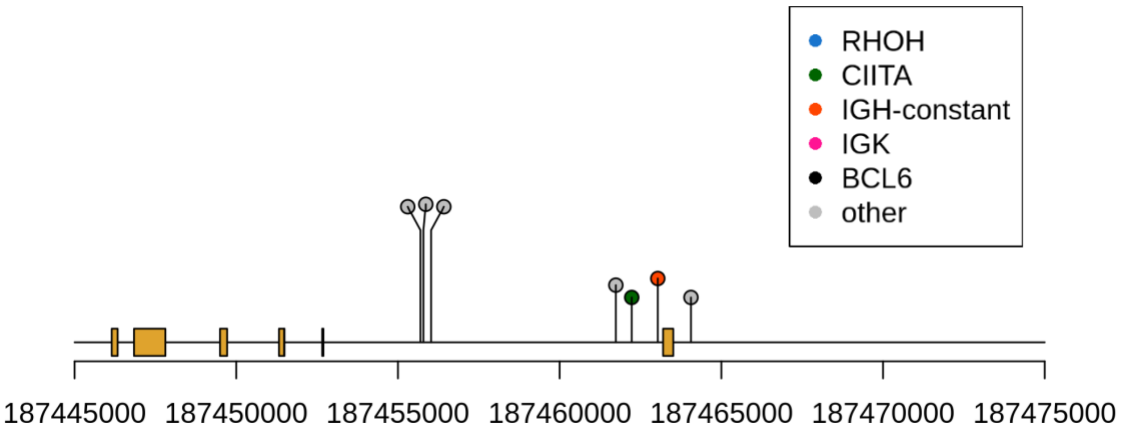
E Stage I - zoomed BCL2 exon 1 and 2 and upstream region



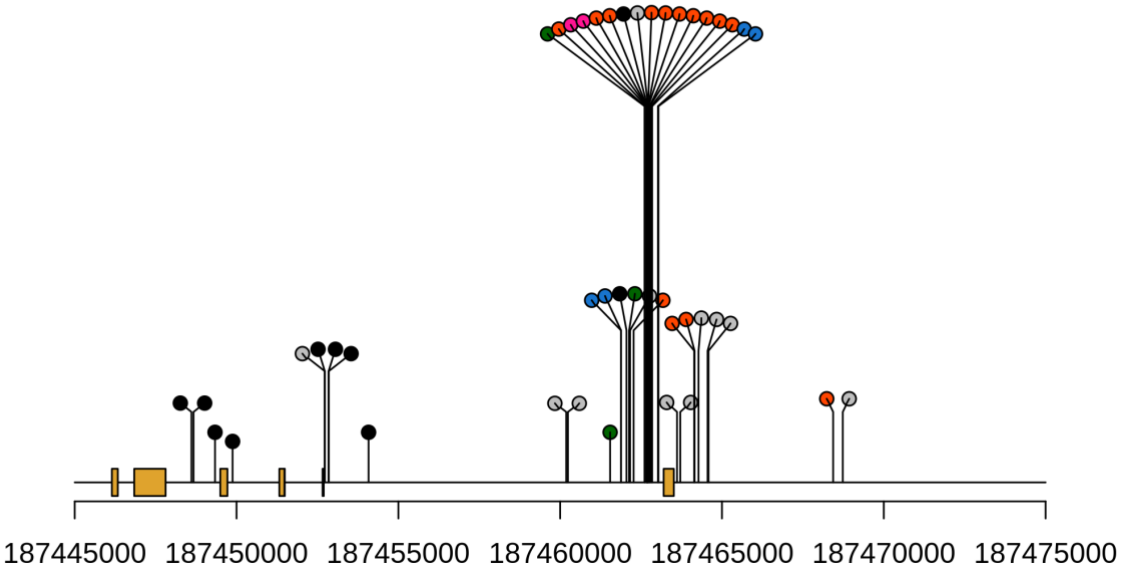
F Stage III/IV - zoomed BCL2 exon 1 and 2 and upstream region



G Stage I – BCL6 region, exon 6 to 1 and upstream

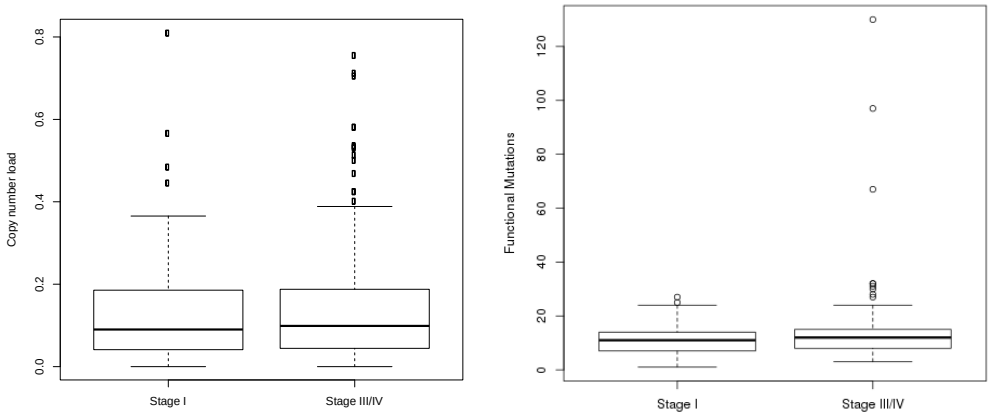


H Stage III/IV – BCL6 region, exon 6 to 1 and upstream



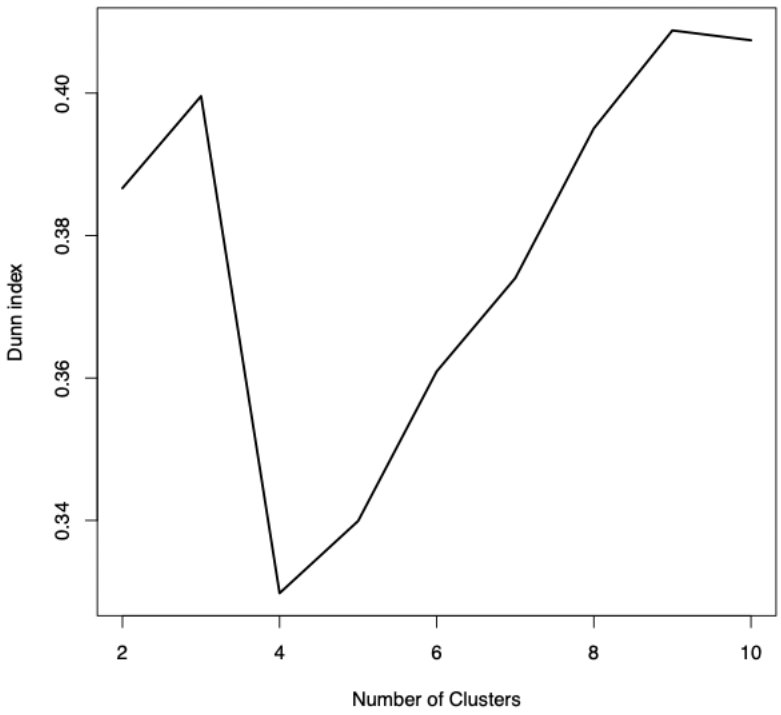
Supplemental Figure 2 - Translocation breakpoints of BCL2 and BCL6
Breakpoints are depicted on chromosomal region around BCL2 and BCL6, x-axis represent the position on the chromosome (hg19). Each pin is an unique breakpoint, of most samples 2 breakpoints are depicted, which represent both sides of the translocation breakpoint. The colours represent the partner gene of the translocation indicated in the legends of **A** and **G**. Stage I and Stage III and IV are depicted in separate figures indicated above. **A-B** entire BCL2 gene with surrounding region of chromosome 18, BCL2 exons are indicated in blue. **C-D** Zoomed region of exon 3 including the Major breakpoint region (MBR) of BCL2 indicated in blue and the downstream region. **E-F** Zoomed region of exon 1 and 2 of BCL2 indicated in blue and the upstream region. **G-H** exon 6 to 1 of BCL6 indicated in yellow with the upstream region on chromosome 3.

A B



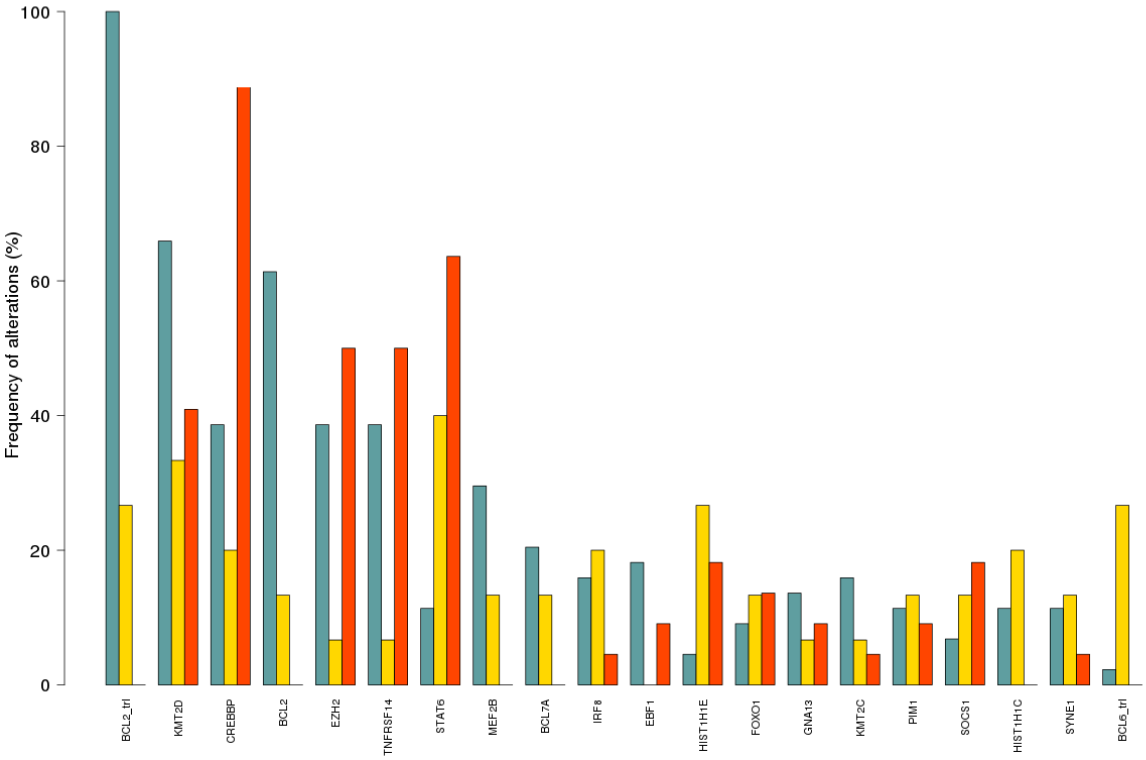
Supplemental Figure 3 – Boxplot copy number load and number of nonsynonymous and splice site mutations per stage

A: The copy number load per stage depicted as boxplot stage I (n=82) median 0.09 (mean 0.13) stage III/IV (n=139) median 0.10 (mean 0.15) (p=0.52) and **B:** The number of nonsynonymous and splice site mutations per stage depicted as boxplots. Stage I (n=82) median =11 (mean 11.1) stage III/IV (n=139) median=12 (mean 14.1) (p=0.28)



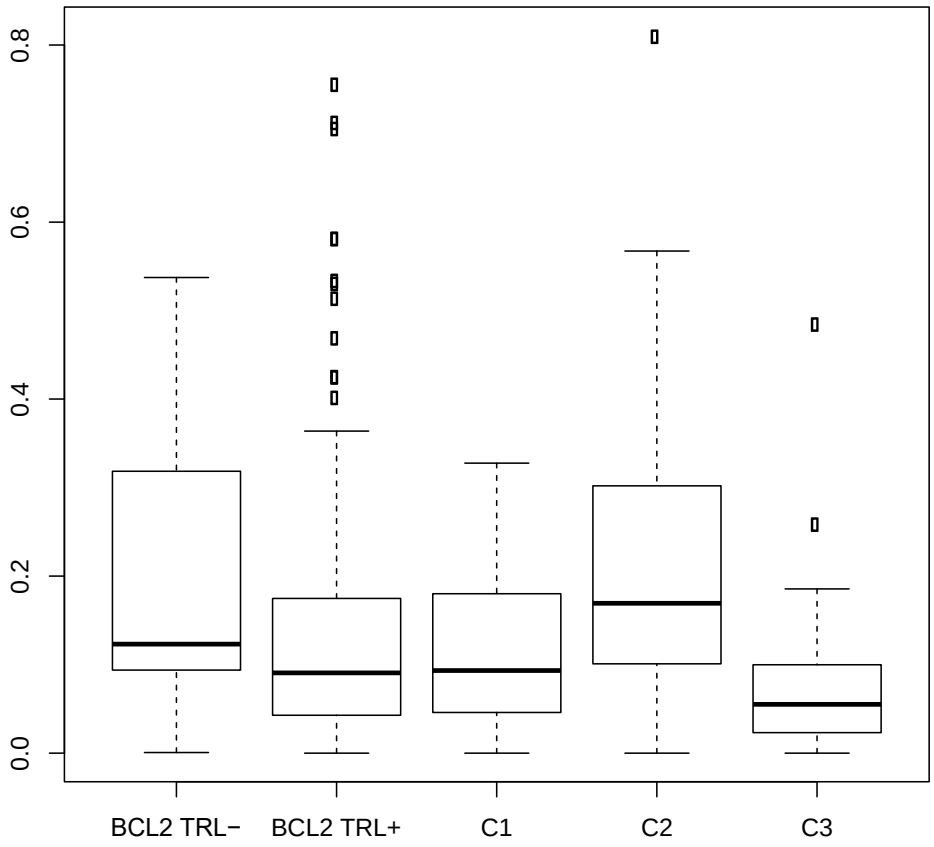
Supplemental Figure 4 – Analysis to determine number of clusters.

Dunn index (y-axis) versus number of clusters (x-axis) for stage I



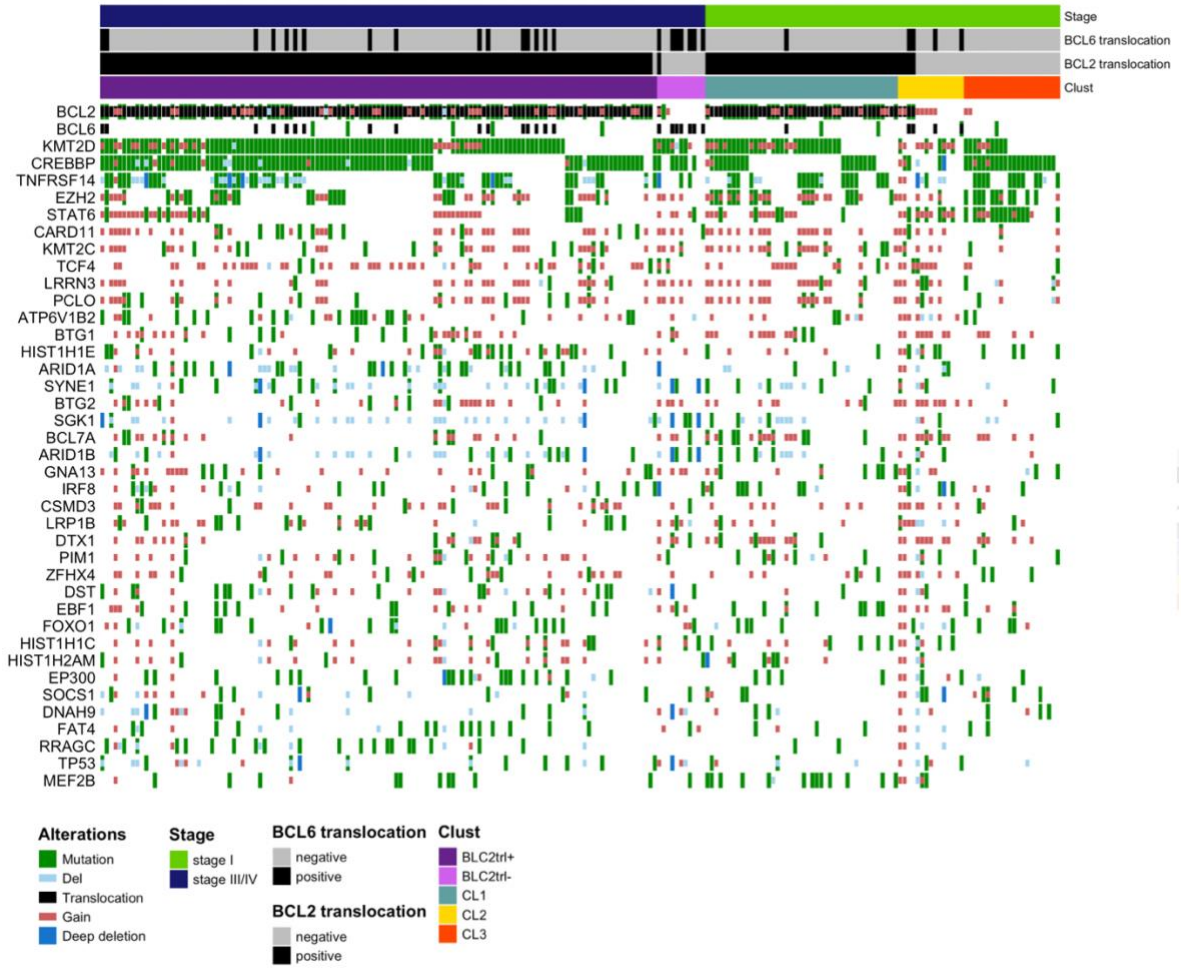
Supplemental Figure 5 – Barplot frequencies of the mutations and translocations per cluster stage I

frequency of BLC2 and BLC6 translocations and top 20 mutated genes depicted, stage I CL1 (green, n=44), CL2 (yellow, n=15) and CL3 (orange, n=22).



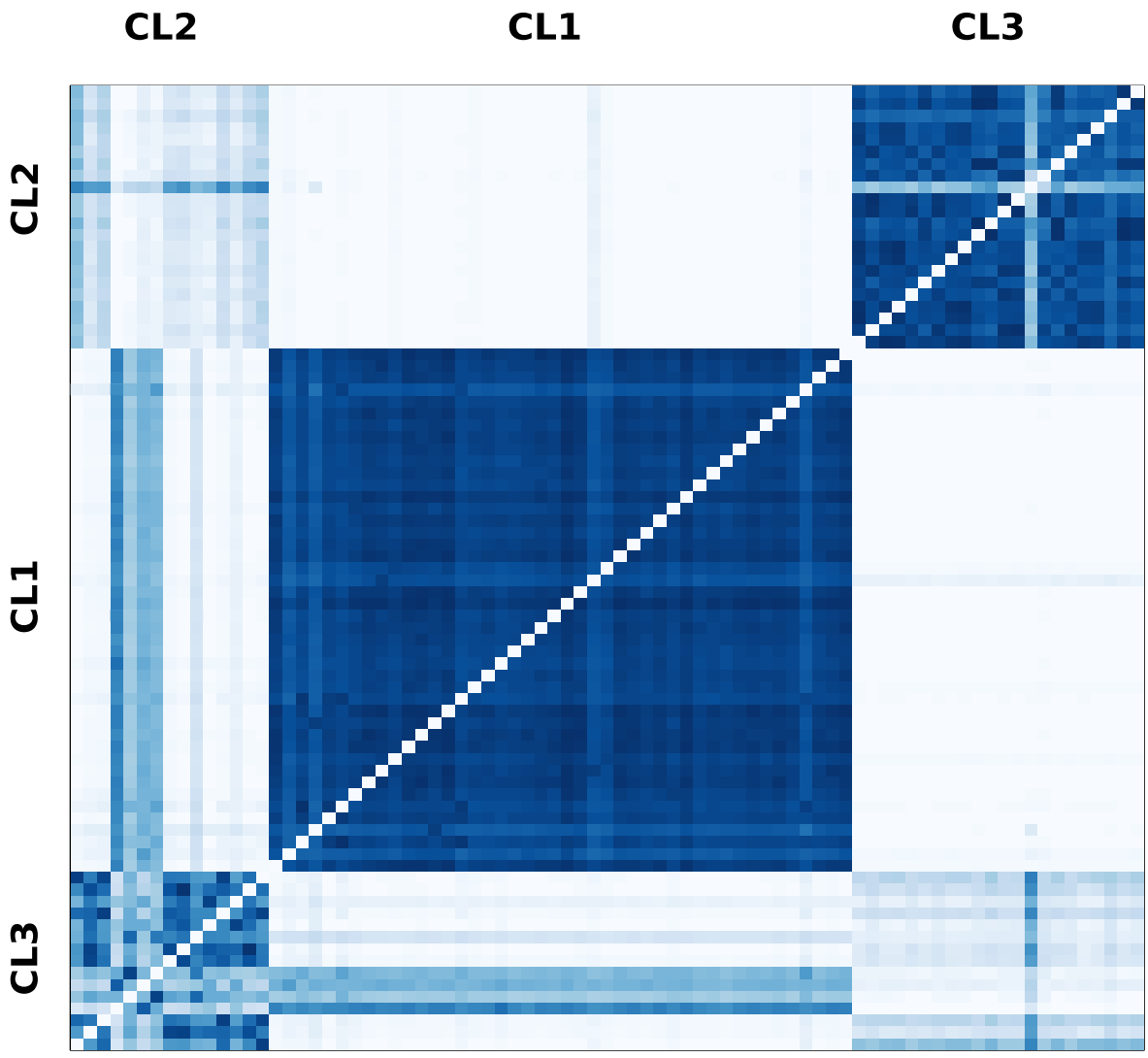
Supplemental Figure 6 – Boxplot copy number load per cluster

The % of CNA per cluster depicted as boxplots. Stage III/IV BLC2trl+ (n=128), stage III/IV BCL2trl- (n=11) and stage I: CL1 (n=44), CL2 (n=15) and CL3 (n=22)



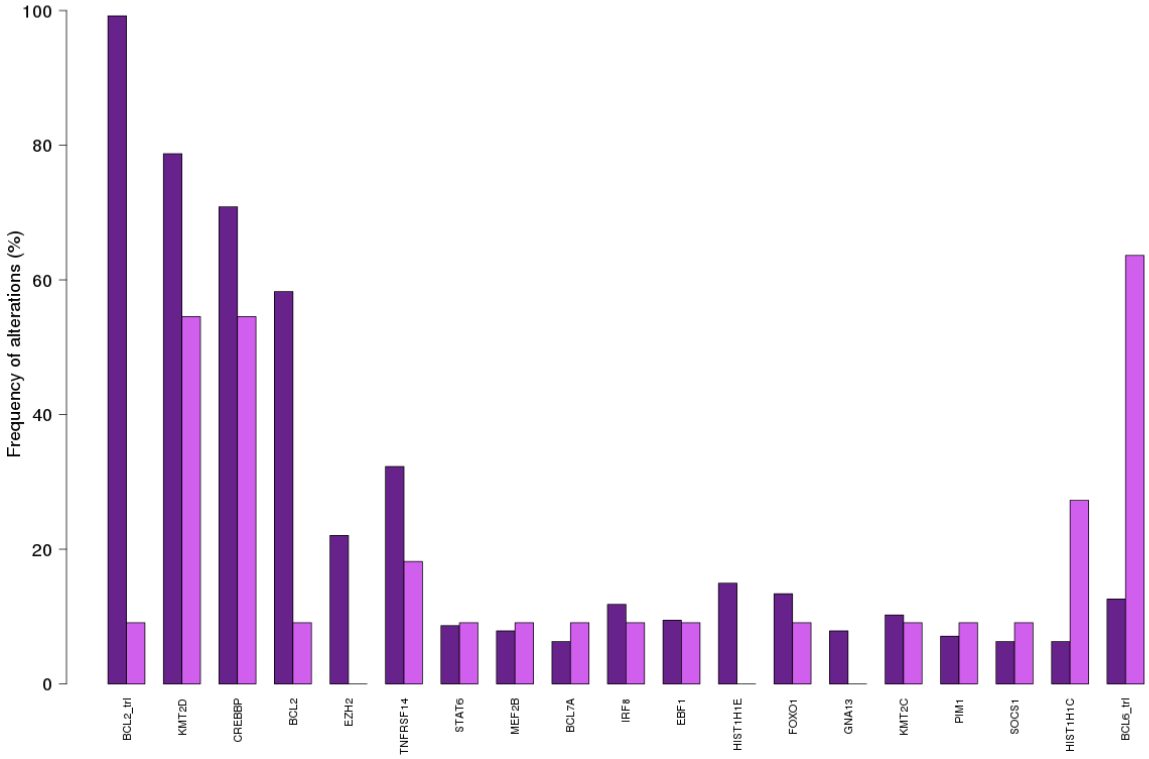
Supplemental Figure 7 – Oncoprint sorted per cluster

Distribution of mutations of the top 40 mutated genes stage I (n=81) and stage III/IV (n=139). Each column represents an individual case, stage III/IV BLC2trl+ (dark purple, n=128), stage III/IV BCL2trl- (light purple, n=11), CL1 (green, n=44), CL2(yellow, n=15) and CL3(orange, n=22). Each row represents a genes ordered based of frequency of mutations appearing within the whole cohort. Alterations are color-coded as indicated in the figure.



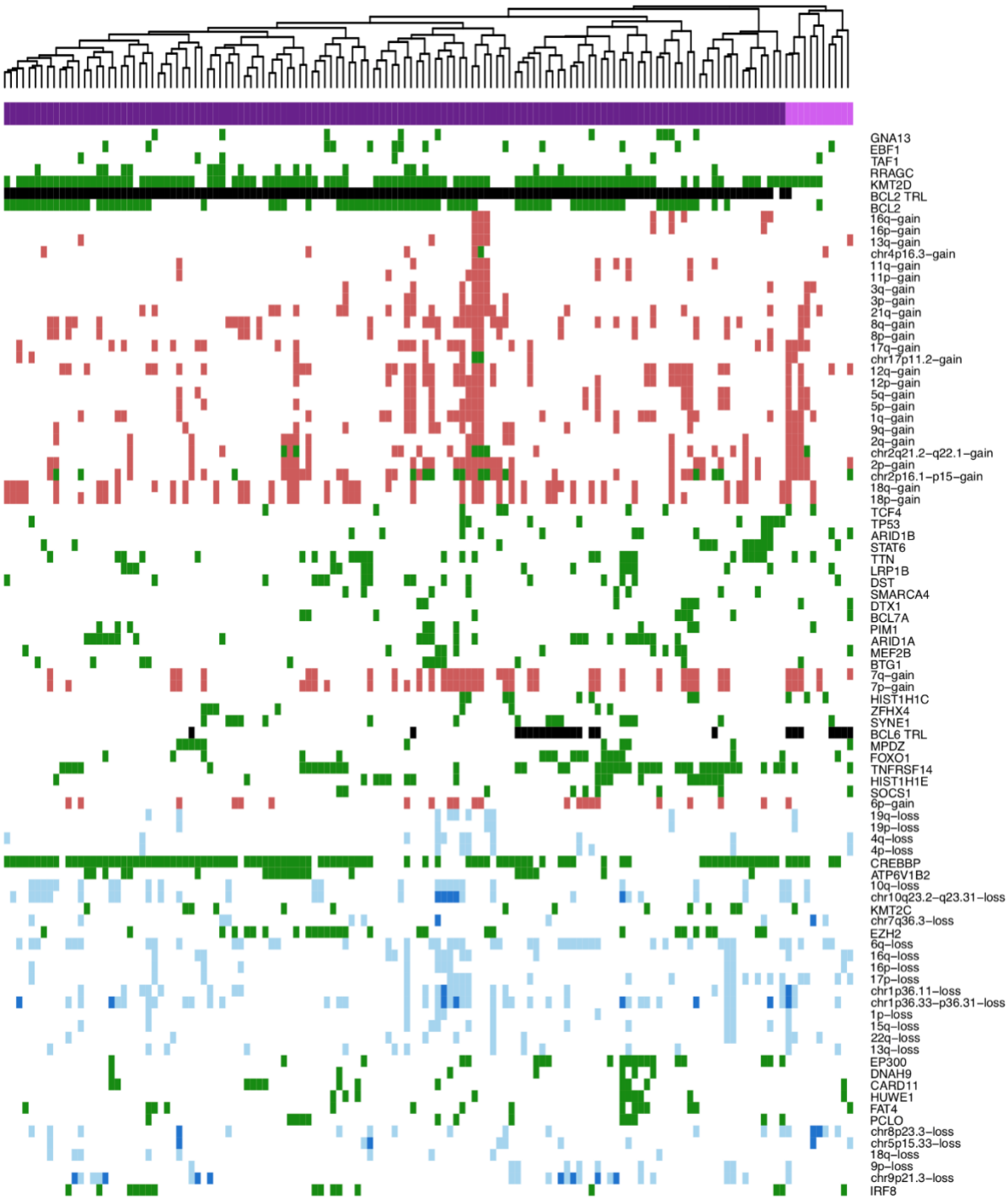
Supplemental Figure 8 – Consensus index

Stability analysis of the stage I clustering. Colours indicate probability of co-clustering of two samples, from 0 to 1 (consensus index). A clear separation between the clusters is found.



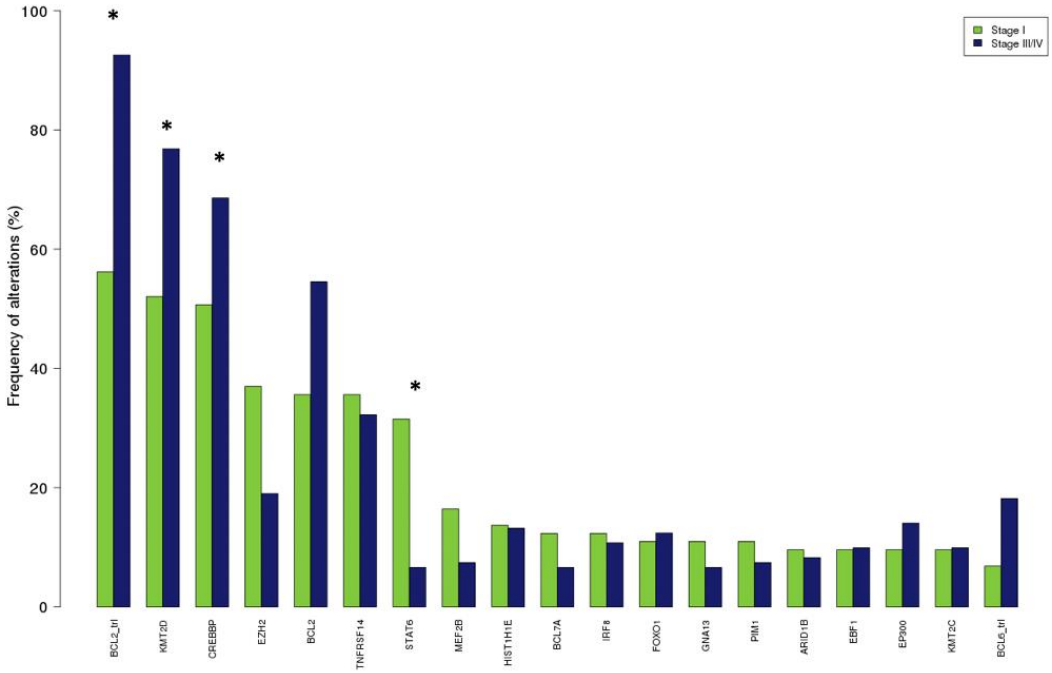
Supplemental Figure 9 – Barplot frequencies of the mutations and translocations per cluster stage III/IV

frequency of BLC2 and BLC6 translocations and top 20 mutated genes depicted, Stage III/IV BLC2trl+ (dark purple, n=128), stage III/IV BCL2trl- (light purple, n=11).



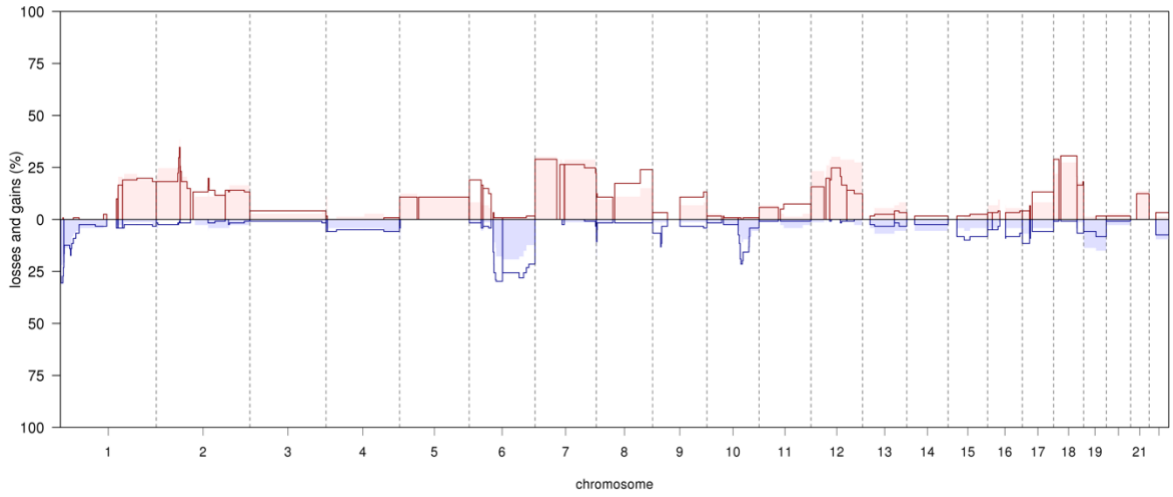
Supplemental Figure 10 – Hierarchic clustering plot stage III/IV

Features of stage III/IV (n=139) included in unsupervised hierarchical clustering are somatic mutations present in more than 5% of the cases, BCL2 and BCL6 translocations, and focal and chromosomal arm level aberrations present in more than 5% of the samples with Spearman correlation. Each column represents one patient, stage III/IV BCL2trl+ (dark purple, n=128) and stage III/IV BCL2trl- (light purple, n=11). Mutations (green), translocations (turquoise) and copy number aberrations (gains=red, losses=light blue and multiple losses=dark blue) are ordered in rows.



Supplemental Figure 11 – Barplot frequencies of the mutations and translocation per stage for the cases with microenvironment and NGS data complete

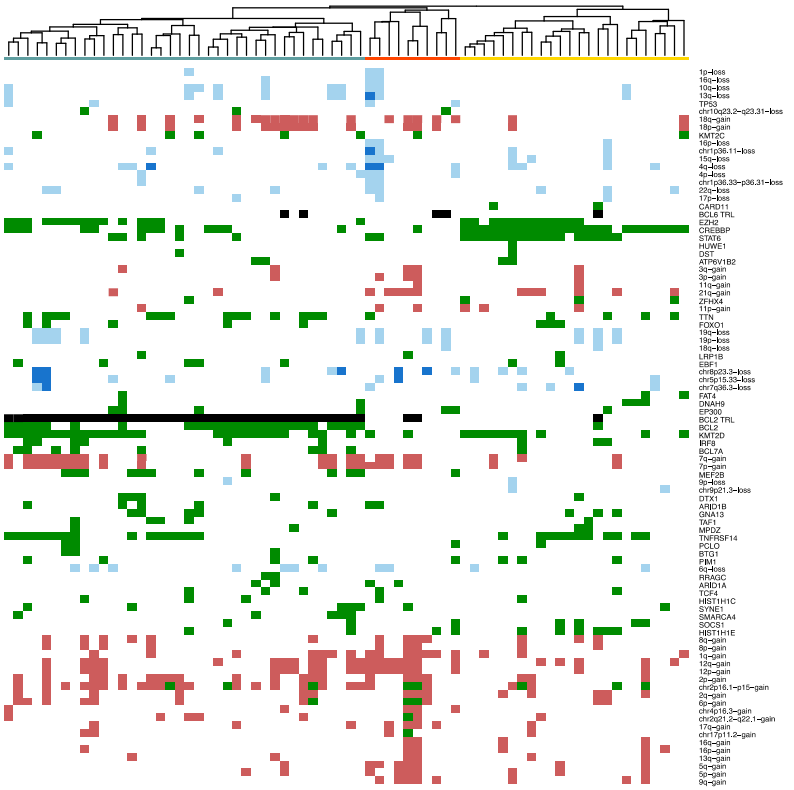
frequency of *BLC2* and *BLC6* translocations and top 20 mutated genes according to stage I in green (n=73) and stage III/IV in blue (n=120), significant differences are indicated by *q<0.05, (Fisher-exact test and false discovery rated using Benjamini & Hochberg method)



Supplemental Figure 12 – Copy number landscape per stage for the cases with microenvironment and NGS data complete

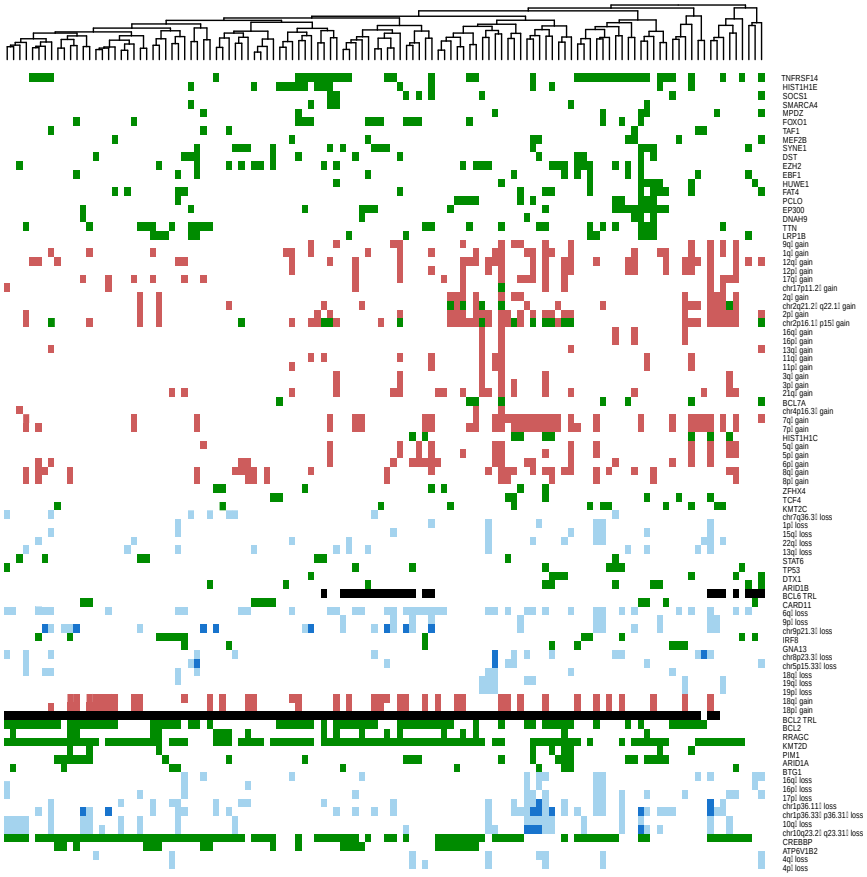
comparison plots for CNAs between stage I as filled areas (n=73) and stage III/IV as lines (n=120) are percentages of the number of cases with gains (positive value red) and losses (negative value blue), sorted for chromosome position (x-axis)

348 **A**



349

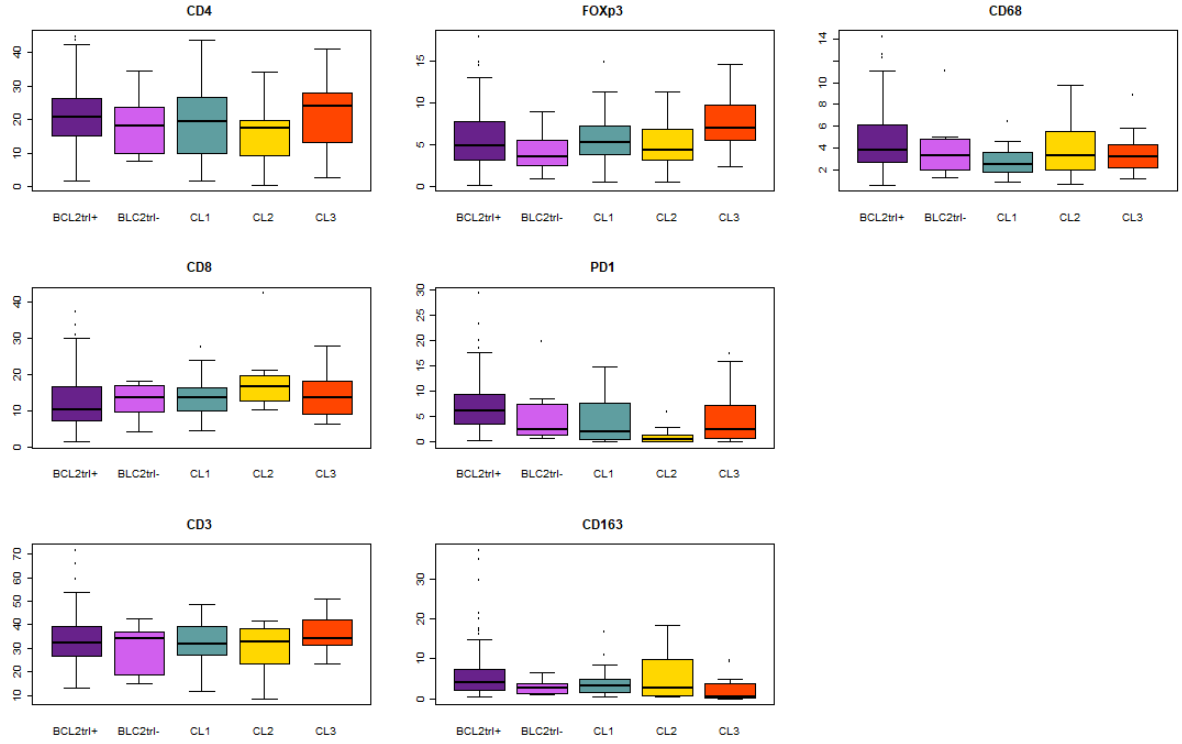
350 **B**



351

Supplemental Figure 13 – Hierarchical clustering plot for the cases with microenvironment and NGS data complete

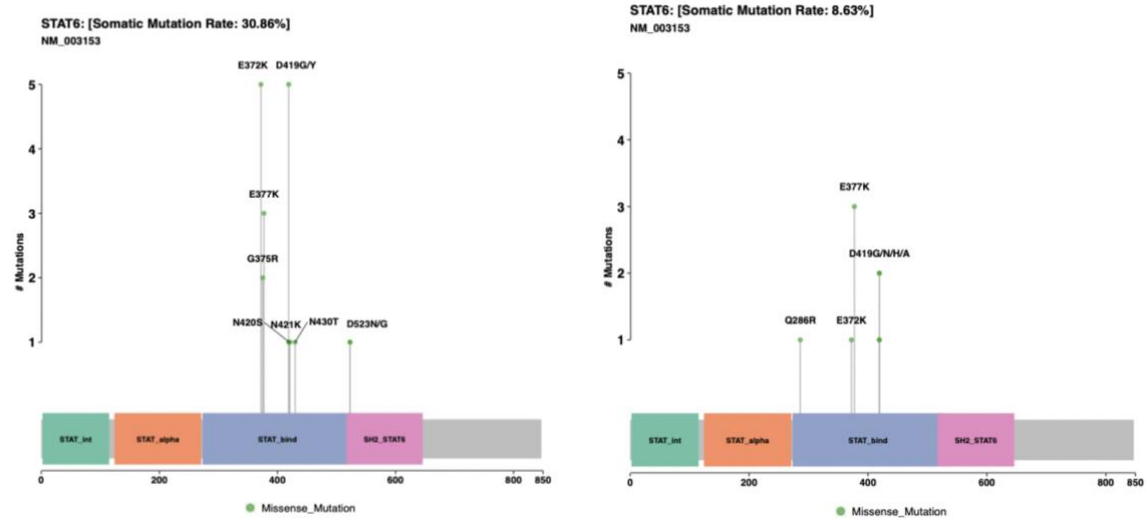
A: Features of stage I (n=72) **B:** Features of stage III/IV (n=120) included in unsupervised hierarchical clustering are somatic mutations present in more than 5% of the cases, BCL2 and BCL6 translocations, and focal and chromosomal arm level aberrations present in more than 5% of the samples with Spearman correlation. Each column represents one patient, stage I; cluster 1 (CL1) (green, n=38), cluster 2 (CL2) (yellow, n=10) and cluster 3 (CL3) (orange, n=24). stage III/IV; BCL2trl+ (dark purple, n=111) and stage III/IV BCL2trl- (light purple, n=9). Mutations (green), translocations (turquoise) and copy number aberrations (gains=red, losses=light blue and multiple losses=dark blue) are ordered in rows



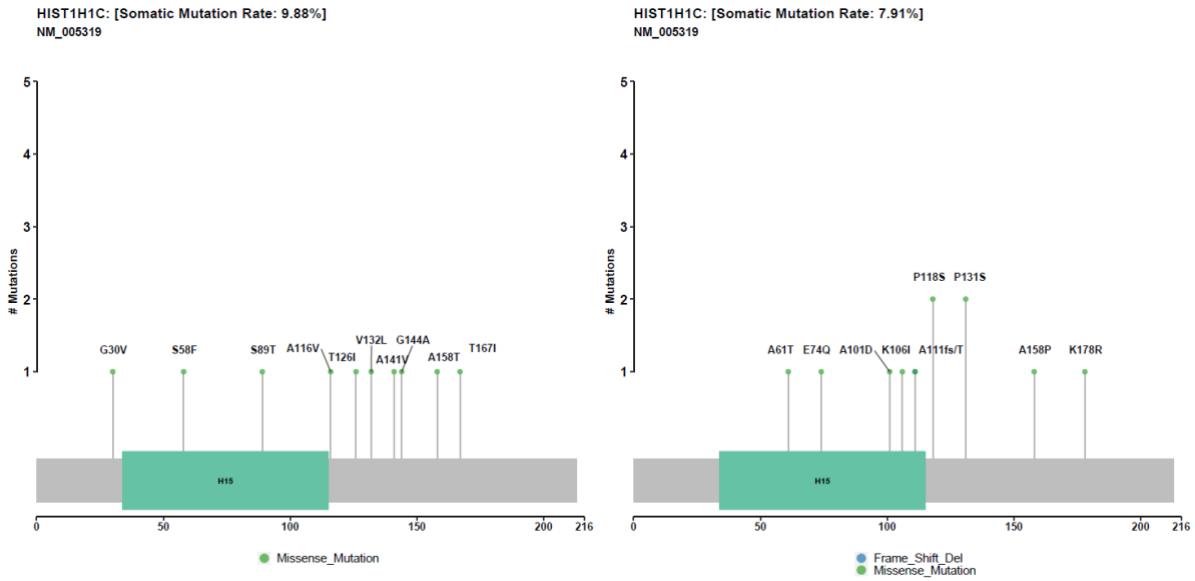
Supplemental Figure 14 - boxplots of microenvironment per cluster of follicular lymphoma For CD4, CD8, CD3, FOXP3 and PD1 the percentage of positive nucleated cells of all nucleated cells are depicted as boxplots and for CD163 and CD68 the percentage of positive area of the total cell area computer assisted scored are plotted in the boxplots with 25th and 75th percentile. Stage III/IV BLC2trl+ (dark purple, n=107), stage III/IV BCL2trl- (light purple, n=7), CL1 (green, n=37), CL2 (yellow, n=11) and CL3 (orange, n=21).

A **B**

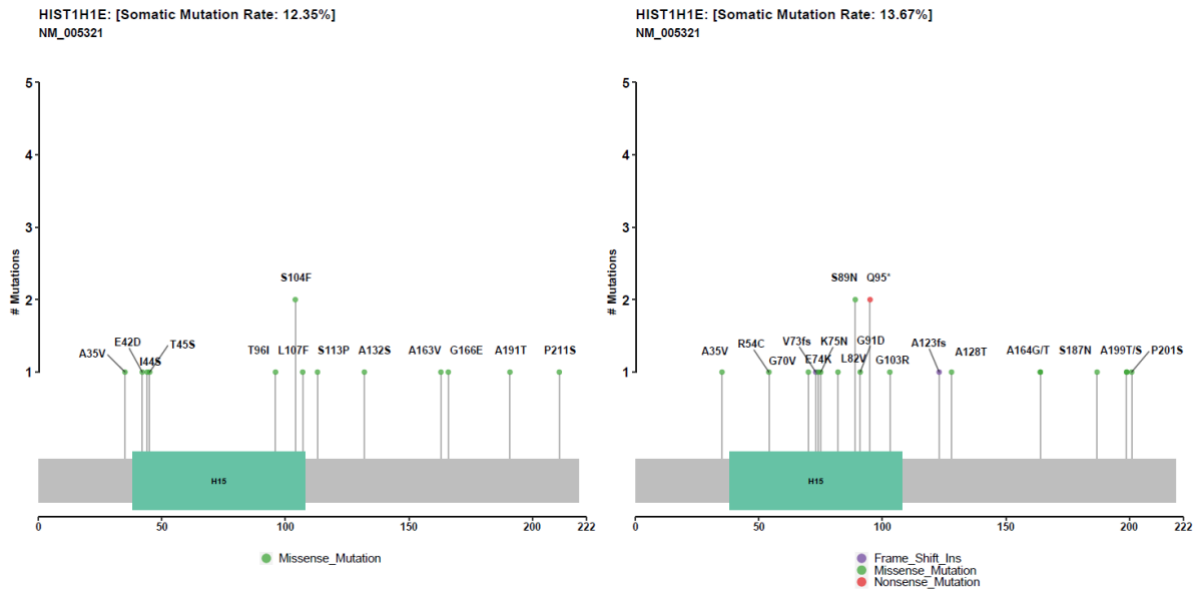
Stage I Stage III/IV



C **D**



E **F**



Supplemental Figure 15 – Mutations in STAT6, HIST1H1C and HIST1H1E
A/B. STAT6 per stage, hotspot mutations are E262K/A, E267K/A and D309G/N/H/V/Y/A Missense mutations are depicted in green, frame shift mutations in purple and nonsense mutations in red. Mutations are visualized by Mutation Mapper from cBioPortal (https://www.cbioportal.org/mutation_mapper)
C/D.HIST1H1C mutations per stage, dispersed pattern suggesting loss of function mutations.
E/F HIST1H1E mutations per stage, dispersed pattern suggesting loss of function mutations