MAPK and JAK-STAT pathways dysregulation in plasmablastic lymphoma

Joan Enric Ramis-Zaldivar,[1,2*] Blanca Gonzalez-Farre,[1,2*] Alina Nicolae,[3] Svetlana Pack,[3] Guillem Clot,[1,2] Ferran Nadeu,[1,2] Anja Mottok,[4] Heike Horn,[5,6,7] Joo Y. Song,[8] Kai Fu,[9] George Wright,[10] Randy D. Gascoyne,[4] Wing C. Chan,[8] David W. Scott,[4,11] Andrew L. Feldman,[12] Alexandra Valera,[1] Anna Enjuanes,[1,2] Rita M. Braziel,[13] Erlend B. Smeland,[14,15] Louis M. Staudt,[16] Andreas Rosenwald,[17] Lisa M. Rimsza,[18] German Ott,[5,6,7] Elaine S. Jaffe,[3] Itziar Salaverria[1,2,#] and Elias Campo[1,2,19,#] for the Leukemia and Lymphoma Molecular Profiling Project (LLMPP)

[1]Hematopathology Unit, Hospital Clínic of Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain; [2]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain; [3]Hematopathology Section, Laboratory of Pathology, National Cancer Institute, Bethesda, MD, USA; [4]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada; [5]Department of Clinical Pathology, Robert-Bosch-Krankenhaus, Stuttgart, Germany; [6]Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany; [7]University of Tübingen, Tübingen, Germany; [8]Department of Pathology, City of Hope National Medical Center, Duarte, CA, USA; [9]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, USA; [10]Biometric Research Branch, Division of Cancer Diagnosis and Treatment, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; [11]Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; [12]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; [13]Department of Clinical Pathology, Oregon Health & Science University, Oregon, OR, USA; [14]Department of Immunology and Centre for Cancer Biomedicine, University of Oslo, Oslo, Norway; [15]Oslo University Hospital, Oslo, Norway; [16]Lymphoid Malignancies Branch, Center for Cancer Research, National Institutes of Health, Bethesda, MD, USA; [17]Institute of Pathology, University of Würzburg, Würzburg, Germany; [18]Department of Laboratory Medicine and Pathology, Mayo Clinic, Phoenix, AZ, USA and [19]University of Barcelona, Barcelona, Spain

**JER-Z and BG-F contributed equally as co-first atuhors.*

#IS and EC contributed equally as co-senior atuhors.

 doi:10.3324/haematol.2020.271957

Received: September 24, 2020.

Accepted: March 31, 2021.

Pre-published: May 6, 2021.

Correspondence: *ELIAS CAMPO* - ecampo@clinic.cat

Supplementary Material

Ramis-Zaldivar& Gonzalez-Farré

Supplementary Methods

Copy number (CN) analysis

The study of the CN alterations (CNA) and CNN-LOH was performed using OncoScan FFPE Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). Evaluation and visual inspection of the CN data was performed using Nexus Biodiscovery 9.0 (Nexus Biodiscovery, El Segundo, USA) using standard settings of two different algorithms, TusCan and FASST2. Human reference genome was GRCh37/hg19. Only CNA larger than 100kb and CNN-LOH larger that 5Mb were considered. Alterations targeting genes involved in lymphomagenesis were considered without meeting the aforementioned criteria. DNA gains and losses arising from B-cell antigen receptor gene rearrangements at 2p11.2 (IGK), 14q32.33 (IGH), and 22q11.22 (IGL) were excluded from the analysis.[1,2]

According to the literature, cases were considered to carry chromothripsis-like patterns when at least 7 switches between two or more CN states were observed on an individual chromosome.[1]

The proportion of tumor cells (or cancer cell fraction, CCF) carrying each CNA was estimated from the OncoScan data using the following formula in which *BAF* is the mean B-allel frequency of the CN locus, *minor* is the minor number of copies of the least frequent allele ,and *major* the major number of copies of the most frequent allele.[3] The information of the *major* and *minor* were obtained from ASCAT (R package v2.5.2).

Losses

$$CCF_{loss} = \frac{(-2 * BAF + 1)}{(1 - BAF)}$$

Gains and amplifications

$$CCF_{gain} = \frac{(2 * BAF - 1)}{\Big((major - 1) - \big((major - 1) * BAF\big) - ((minor - 1) * BAF)\Big)}$$

Copy number neutral loss of heterozygosity

$$CCF_{CNN-LOH} = 2 * BAF - 1$$

The predicted CCFs were finally corrected by the tumor purity of the respective samples obtained by ASCAT (R package v2.5.2). CCF were only obtained for those samples with predicted diploid genotype. CNA were considered as clonal if their CCF was ≥85%, while subclonal otherwise.[4]

For comparison, CN data of 35 Burkitt lymphoma, BL (SNP 250k sty Gene chip, ThermoFisher Scientific inc.)[5], 41 Plasma cell myeloma, PCM (Genome-Wide Human SNP array 6.0 and 750k Cytoscan, ThermoFisher inc),[6,7] 49 Activated B-cell like diffuse large B-cell lymphoma, ABC-DLBCL (Cytoscan HD, ThermoFisher Scientific inc)[8] were reanalyzed using the same software and criteria described above.

Library preparation and targeted sequencing approach

A total of 27 FFPE DNA samples from plasmablastic lymphoma (PBL) were processed using SureSelectXT (Agilent Technologies, Santa Clara, CA) and using a panel design described in **Supplementary Table 2**. A total of 100ng of genomic DNA was sheared using the Covaris S220 focused-ultra sonicator (Covaris, Woburn, MA) to a target peak size of 150–200 bp. Library preparation were 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 27 cases across regions was 374x (range 66-1688x) and over 94% of the targeted regions were covered by at least 50 reads.

FASTQ files were generated by MiSeq control software and quality control of the raw data was performed using the FastQC tool. Sequencing reads were subsequently aligned to the human reference genome (GRCh37/hg19) using the Burrows-Wheeler Aligner–MEM algorithm.[9] Variant calling was performed using two different variant callers, Somatic Variant Caller (Illumina inc.) and Mutect2 (Genome Analysis Toolkit-GATK version 4.0.3)[10] and annotated using the VariantStudio software v3.0 and ANNOVAR, respectively.[11] 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 non-interrogated variants (non-exonic) and known polymorphisms described in the Single Nucleotide Polymorphism Database (dbSNP138) or ExAC database (release 2015) with more than 0.1% frequency and synonymous variants. Finally, each variant was also visually inspected with the Integrative Genomics Viewer (IGV, Broad Institute, version 2.3) software to exclude artifacts. We ended up with 194 exonic mutations and 165 *MYC* intronic variants (**Supplementary Fig. 1**). *MYC* intronic mutations were identified but excluded for further analysis. The CCF carrying each specific mutation (CCF_{mut}) was calculated using the following formula: $CCF_{mut}=(((q–2)*CCF_{CNA}+2)*VAF_{mut})/p$, where q is the copy locus number for the sample, CCF_{CNA} the CCF of the copy number alteration (0 to 1), VAF_{mut} the VAF of the mutation, and p the tumor purity of the sample (0 to 1).

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 in which the 90% of the mutations classified as functional were demonstrated to be somatic.[8] Inclusion criteria were: 1) any variant 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[12] or SIFT predictor if a definitive score was not

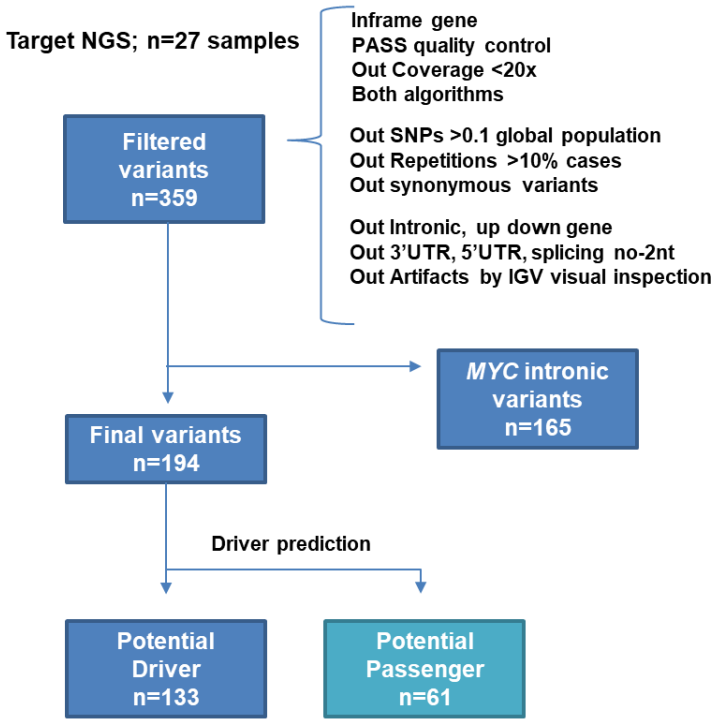
provided by Mutation Assessor.[13] Other predictors such as Polyphen-2 (Polymorphism Phenotyping-2)[14] and CADD (Combined Annotation Dependent Depletion)[15] were also used.

Nanostring PanCancer Pathways 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 12 samples (8 EBV-positive and 4 EBV-negative PBL or 5 STAT3-mutated and 7 STAT3-wt PBL) that passed QC, differential expression (DE) and pathway scoring analyses were performed using the nCounter Advanced Analysis (version 2.0.115)(NanoString Technologies inc.). Pathway scores were fit using the first principal component of each gene set's data. They are oriented such that increasing score corresponds to mostly increasing expression. For STAT3 mutated and non-mutated comparison EBV status was defined as a confounder variable. Data normalization was done using the geNorm algorithm, automatically performed by the software. Low count probes were removed using default threshold settings of the program pipeline. P-values associated with the fold change were derived using the Benjamin-Yekutieli FDR method. Genes with a fold change in absolute value greater than 1and FDR lower than 0.2 were considered to be differentially expressed. Gene ontology pathway enrichment (KEGG and BIOCARTA) analyses for the differentially expressed genes were performed using the DAVID web-tool.

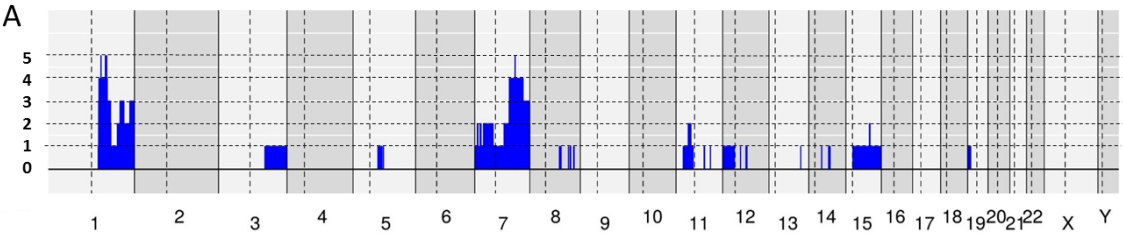
SUPPLEMENTARY FIGURES

Supplementary Figure S1. Next generation sequencing (NGS) analysis pipeline to identify potential driver mutations in 27 PBL.



Supplementary Figure S2. Recurrent amplifications in PBL. (A) Cumulative plot of amplified regions in PBL where y axis is the number of cases with the amplification. (B) Minimal regions of amplification with potencial target genes.

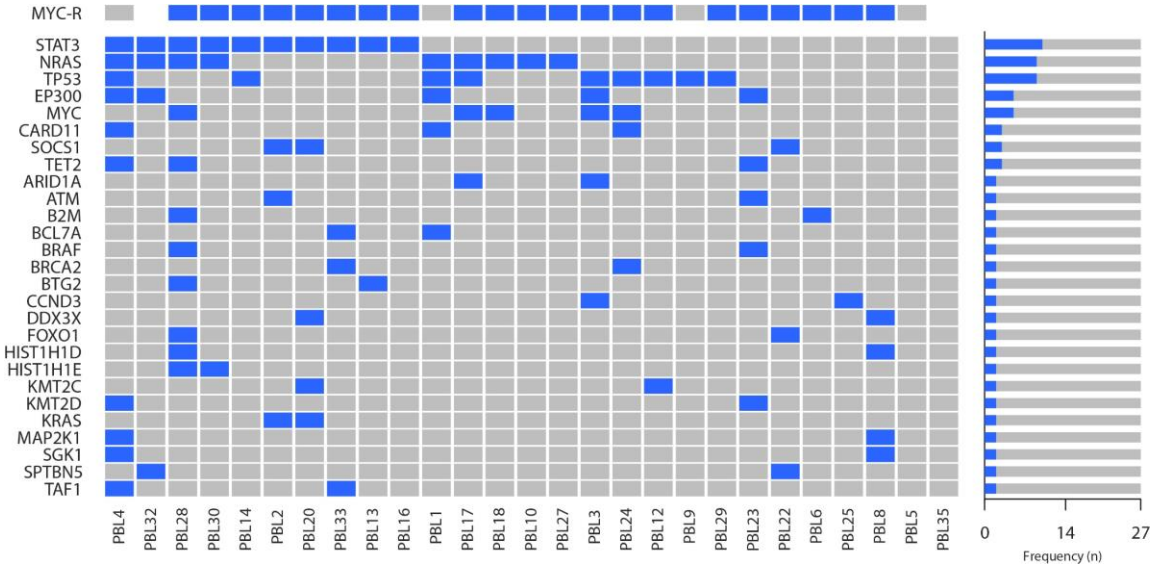
A



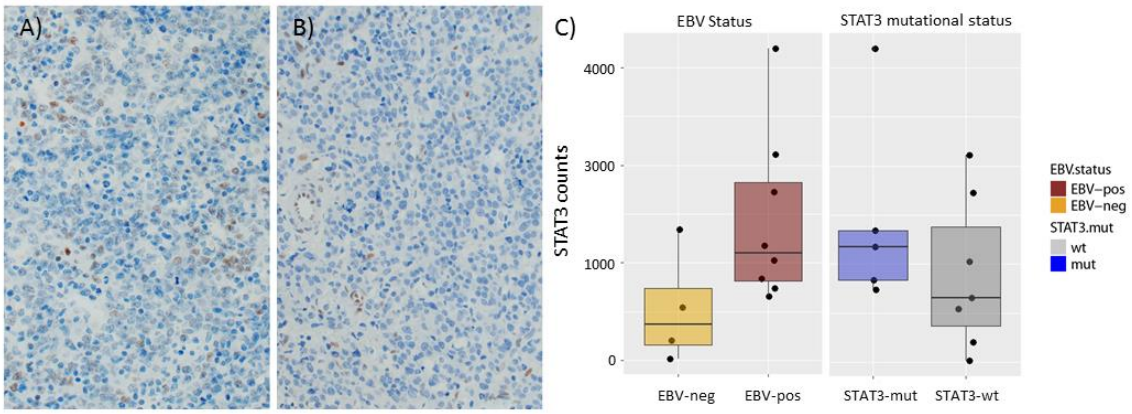
B

Minimal region (GRCh37/hg19)	Cases	cytoband	Target genes
chr1: 151759039-153010431	5	1q21,3	
chr1: 154294656-160349098	4	1q21.3-q23.2	*IL6R, ADAR, MEF2D, CKS1B*
chr1: 164529666-169452090	5	1q23.3-q24.2	*POU2F1*
chr1: 207162433-219230459	3	1q32.1-q41	*TRAF5, HHAT, IRF6, PTPN14*
chr1: 235728896-249212878	3	1q42.3-q44	
chr7: 7161234-9502381	2	7p22.1-p21.3	
chr7: 14993376-18769482	2	7p21.2-p21.1	*HDAC9*
chr7: 24788241-30118130	2	7p15.3-p14.3	
chr7: 31058272-53487804	2	7p14.3-p12.1	
chr7: 116417876-117613226	5	7q31.2-q31.31	*MET, WNT2*
chr11: 34942838-43540820	2	11p13-p11.2	*CD44, TRAF6*
chr15: 67685160-70140137	2	15q23	*MAP2K5*

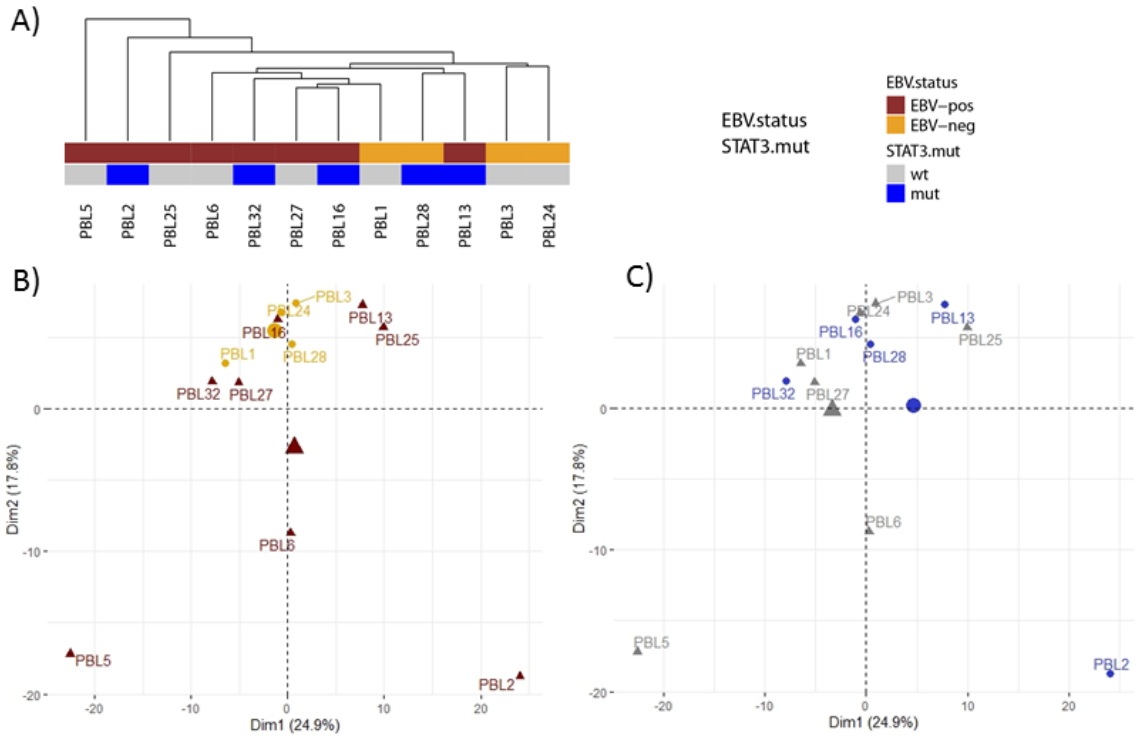
Supplementary Figure S3. Mutational landscape of 27 PBL. The heatmap shows the case specific pattern of driver mutations found by NGS. Each column represents a case and each row represents a gene. The right bar graph illustrates the mutation frequency of each gene. Only driver mutations are represented. *MYC* translocation status has been added at the top of the oncoprint.



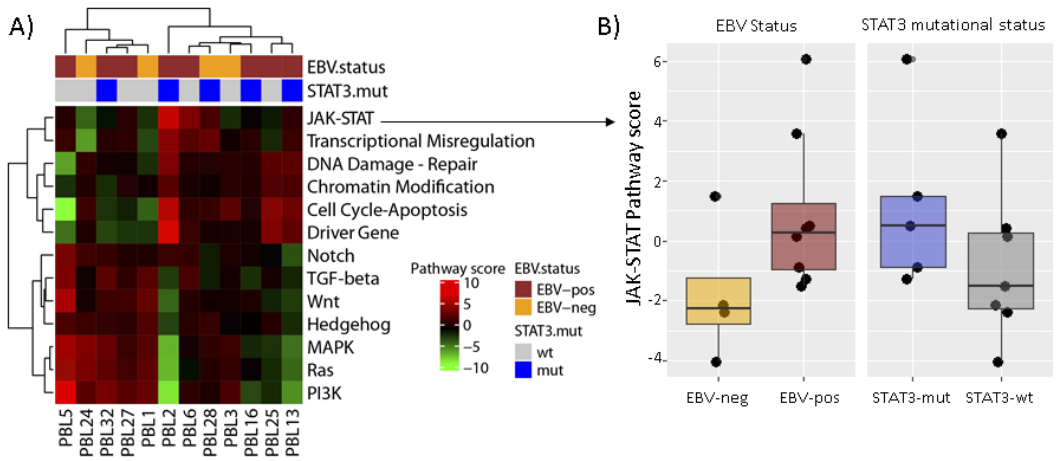
Supplementary Figure S4. STAT3 activation and expression depending on mutational status. A) pSTAT3 stain in case #2 with positive neoplastic cells. B) pSTAT3 stain in case #1 with positive internal control in endothelial cells and negativity in tumor cells. C) Boxplot representing STAT3 gene expression (number of counts of STAT3 on Nanostring PanCancer Pathways assay) between EBV-positive PBL (n=8) and EBV-negative PBL (n=4) and STAT3-mutated (n=5) and STAT3-wt (n=7) PBL cases. No differences were observed between STAT3 mutational status groups (2457 vs 1897 counts; Wilcoxon-test, P=0.27) whereas higher STAT3 expression was observed in EBV-positive cases compared to EBV-negative cases (2433 vs 1525 counts; Wilcoxon test, p=0.07).



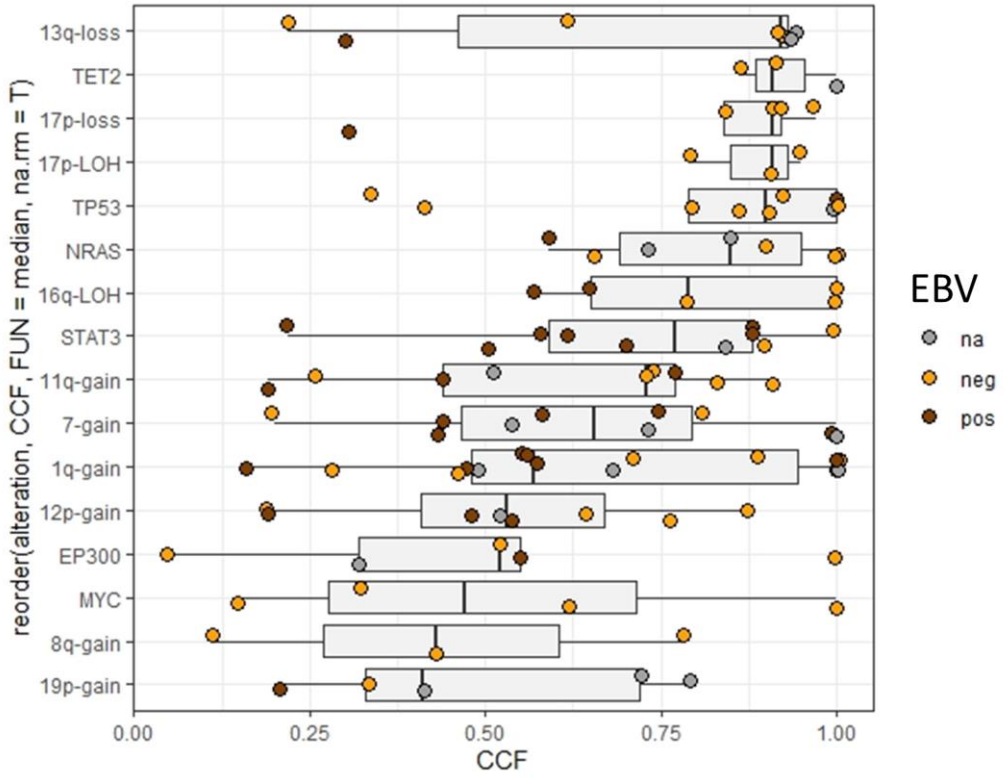
Supplementary Figure S5. Unsupervised analysis of PBL. A) Dendrogram shows the unsupervised hierarchical clustering of PBL samples based on the expression of 503 RNAs. B) Principal component analysis in terms of EBV status and C) STAT3 mutational status.



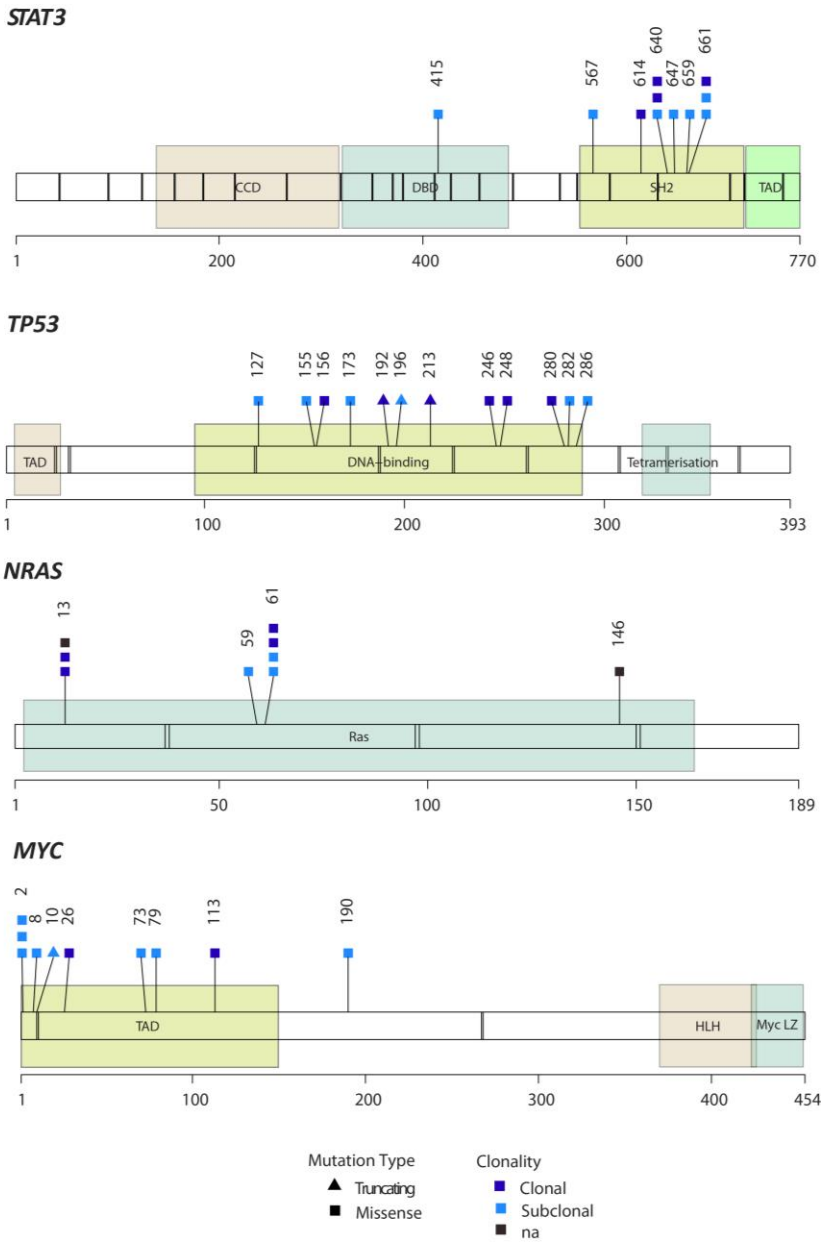
Supplementary Figure S6. Pathway enrichment analysis. A) Heatmap of the unsupervised hierarchical clustering based on the pathways scores calculated from nCounter Advanced Analysis (version 2.0.115). B) Pathway score comparison between EBV-positive and EBV-negative PBL and between STAT3 mutated and STAT3 non-mutated PBL. No significant differences were found in relation to the *STAT3* mutational status (Wilcoxon test; P=0.15), although a tendency to higher expression was observed in EBV-positive cases (Wilcoxon test; P=0.1).



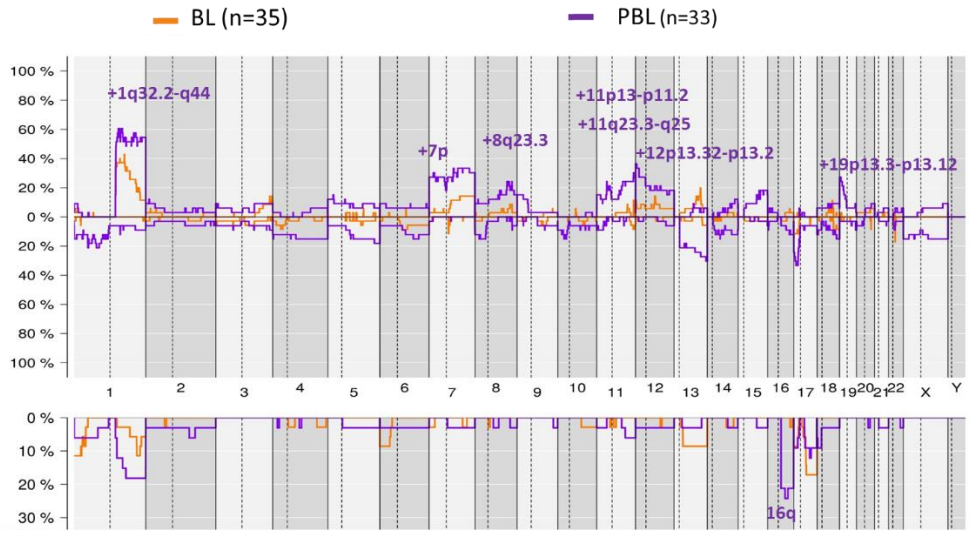
Supplementary Figure S7. Clonality analysis of recurrent PBL alterations. Boxplots represent the CCF of a given alteration in a case. There were included recurrent CNA (gain/loss, >25% of the cases) and mutated genes (>10% of the cases) with a minimum of 3 cases. The highest CCF is represented if a case have more than two mutations affecting the same gene.



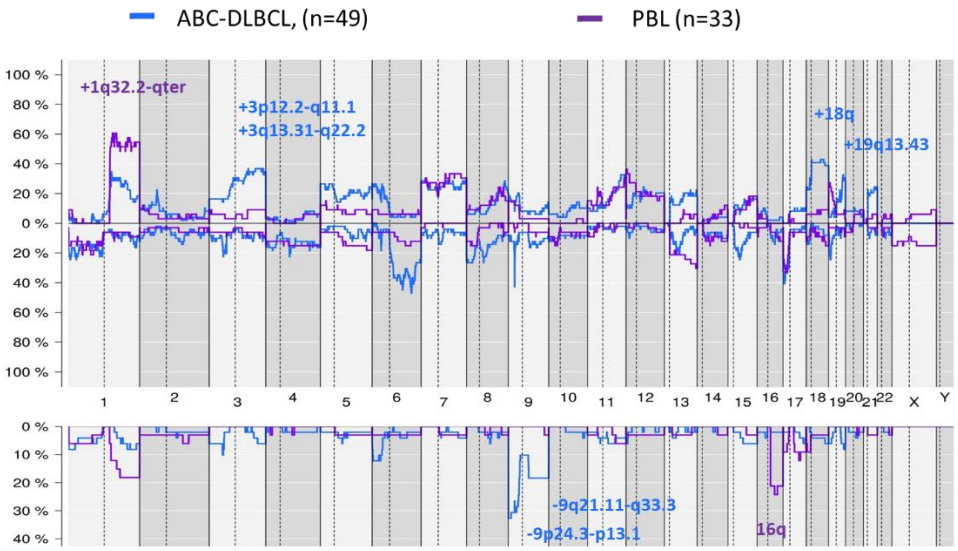
Supplementary Figure S8. Clonality of recurrent driver mutations in PBL. A diagram of the relative positions of driver mutations is shown for *STAT3*, *TP53*, *NRAS* and *MYC* genes. X-axes indicate aminoacid position. Square shape indicates missense mutations whereas triangle shape truncating mutations. Colors indicate clonality status of the mutation. CCD: coiled coil domain; DBD: DNA-binding domain; TAD: transcription activation domain; HLH: helix-loop-helix and LZ: leucine zipper.



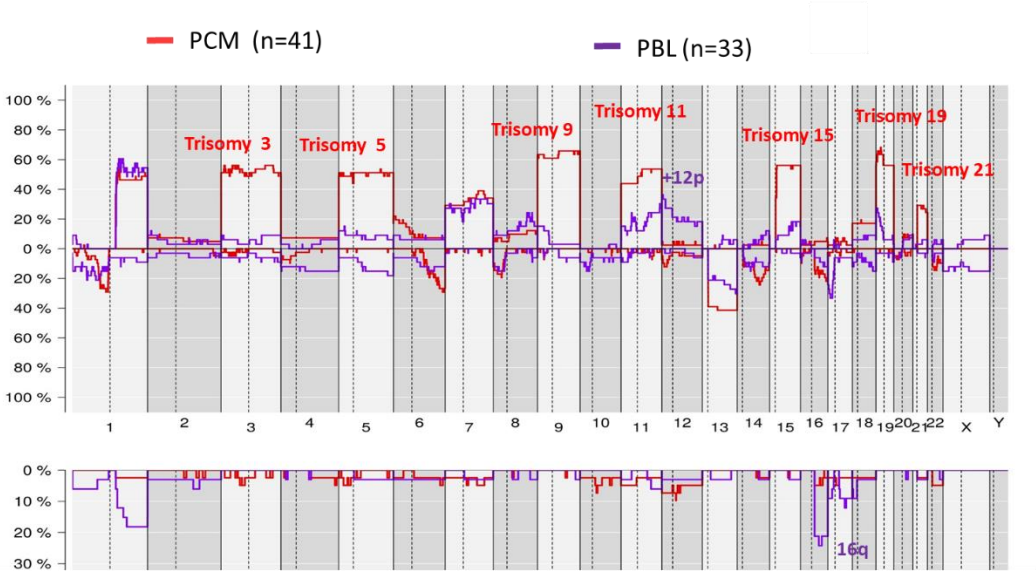
Supplementary Figure S9. Comparison of PBL versus BL. Comparative plot of copy number (CN) and copy number neutral-loss of heterozygosity (CNN-LOH) between 33 PBL and 35 BL (250K SNP array, Affymetrix).[5] Significant different regions are indicated in the plot and the color denotes the enriched group (Fisher's test; FDR<0.05; min 5 cases). Genomic complexity of PBL was 12.2 vs. 5.97 alterations/case (Wilcoxon test; *P*<0.01).



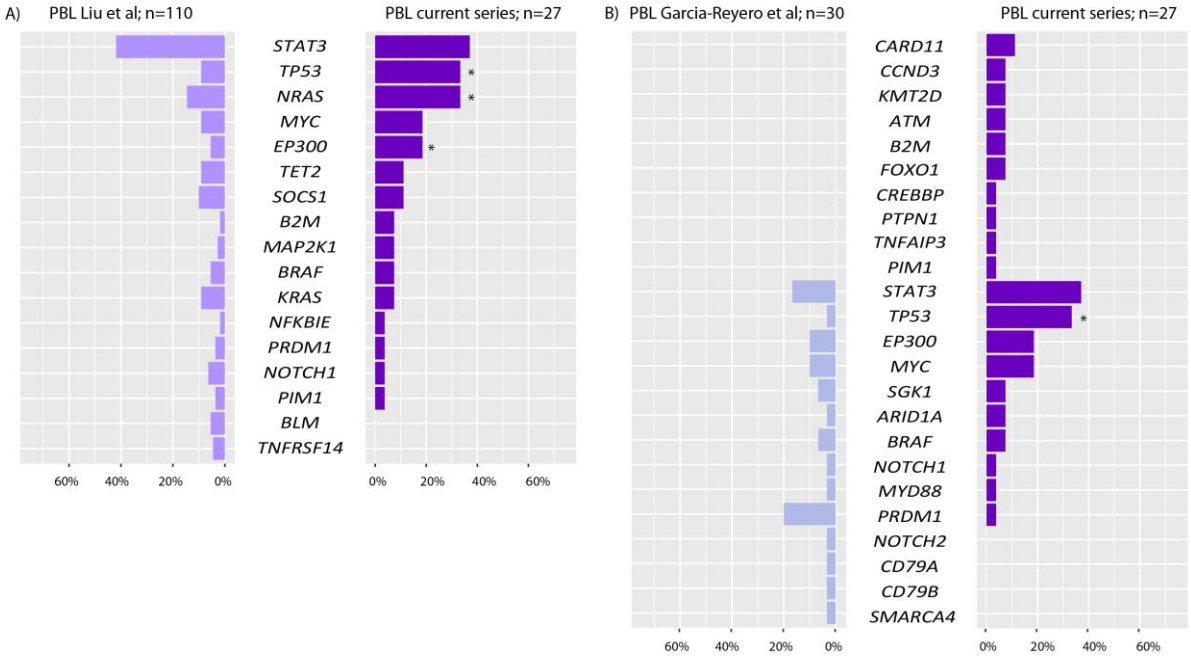
Supplementary Figure S10. Comparison of PBL versus ABC-DLBCL. Comparative plot of CN and CNN-LOH between 33 PBL and 49 ABC-DLBCL (Cytoscan HD array, Affymetrix).[8] Significant different regions are indicated in the plot and the color denotes the enriched group (Fisher's test; FDR<0.05; min 5 cases). Genomic complexity of PBL was 12.2 vs. 22.1 alterations/case (wilcoxon test; *P*<0.01).



Supplementary Figure S11. Comparison of PBL versus PCM. (A) Comparative plot of CN and CNN-LOH between 33 PBL and 41 PCM (750K cytoscan and SNP-array 6.0, Affymetrix).[6,7] Significant different regions are indicated in the plot and the color denotes the enriched group (Fisher's test; FDR<0.05; min 5 cases). Genomic complexity of PBL was 12.2 vs. 13 alterations/case (wilcoxon test; *P*<0.36).



Supplementary Figure S12. Comparison of mutational findings between current series and previous NGS analyses on PBL. (A) Liu et al PBL comparison of 17 commonly interrogated and mutated genes by both studies.[16] (B) Garcia-Reyero et al PBL comparison of 24 commonly interrogated and mutated genes by both studies.[17] Bars indicate the percentage of mutated cases. Asterisk indicates differentially mutated genes between groups (Fisher's test; P<0.05).



Supplementary Tables

Supplementary Table S1. Details of all antibodies used, source and conditions of use

Antibody	Clone	Source	Antigen retrieval	Dilution
CD20	L26	Ventana, Roche (Oro Walley, AR, USA)	CC1 36’	RTU
CD79a	SP18	Ventana, Roche	CC1 20’	RTU
PAX5	SP34	Ventana, Roche	CC1 52’	RTU
CD138	B-A38	Ventana, Roche	CC1 64’	RTU
CD56	123C3	Ventana, Roche	CC1 36’	RTU
HHV-8	13B10	Ventana, Roche	CC1 64’	RTU
MUM1/*IRF4*	MRQ-43	Ventana, Roche	CC1 64’	RTU
BCL-2	124	Ventana, Roche	CC1 64’	RTU
BCL-6*	GI19IE/A8	Ventana, Roche	CC1 48’	RTU
HHV8	13B10	Ventana, Roche	CC1 64’	RTU
p-STAT3	D3A7	DAKO	PT-LINK	1/100

RTU, ready to use.
According to previous reports, BCL2,[18] BCL6 and MUM1 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.[18] All cases were stained using an Automated immunostainer (Benchmark XT; Ventana) with ultraView or optiview* Universal DAB Detection Kit.

Supplementary Table S2. Ninety-four genes sequenced using SureSelectXT Target NGS panel including references for inclusion in the mutational analysis and mean coverage by gene and amplicon.

Provided in excel format

Supplementary Table S3. Gene expression profile by NanoString PanCancer Pathways Panel. Information of the 12 cases analyzed. NanoString PanCancer Pathways Panel design and probe details. Differentially expressed genes in *STAT3*-mutated PBL (n=5) vs STAT3-non-mutated PBL (n=7) and EBV-negative PBL (n=4) vs EBV-positive PBL (n=8) PBL. Gene ontology and pathway enrichment of the differential expressed genes.

Provided in excel format

Supplementary Table S4. Copy number, copy number neutral loss of heterozygosity and copy number clonality information of 33 PBL cases (Human reference genome GRCh37/hg19)

Provided in excel format

Supplementary Table S5. List of 194 somatic mutations in 27 PBL including driver prediction of amino acid changes that affect protein function (MA, SIFT, Polyphen2, CADD).[8]

Provided in excel format.

Supplementary Table S6. p-STAT3 IHC of PBL and PCM according to STAT3 mutational and EBV status.

Aria	EBV status	STAT3 MUTATION	STAT3 IHC
PBL2	pos	S614R (44)	+
PBL4	neg	Y640F (34)	NV
PBL30		D641Y (42)	scattered
PBL6	pos	WT	scattered
PBL1	neg	WT	-
PBL5	pos	WT	-
PBL8	neg	WT	+
PCM1			-
PCM2			-
PCM3			-
PCM4			-

NV: not valorable; WT: wild type

Supplementary Table S7. 184 MYC mutations (including 14 exonic, 5 synonymous and 165 intronic mutations).

Provided in excel format.

Supplementary Table S8. Novel agents and targets in lymphoma treatment.

Pathway	Drug name	Mechanism	Entity	Clinical trials
MAPK			PCM	(Phase II) NCT01085214
	Selumetinib	MEK1/2 inhibitor		
	Sorafenib	Multikinase inhibitor	PCM	(Phase I) NCT00474929
	Vemurafenib	BRAF inhibitor	PCM	(Phase II) NCT01524978
	Cobimetinib	BRAF inhibitor	PCM	(Phase I) NCT03312530
JAK-STAT	Ruxolitinib	JAK2 inhibitor	PCM	(Phase I) NCT00639002
	SC99	JAK2/STAT3 inhibitor		Preclinical phase
	Siltuximab	anti-IL-6 Ab	PCM,B-NHL, solid tumor	Multiple (phase I/II)[19]
	Tocilizumab	anti-IL-6R Ab	CLL, solid tumor	Multiple (phase I/II)[19]
	OPB-51602	STAT3 SH2 binder	PCM, NHL, AML, ALL, CML	Multiple (phase I)[19]
	AZD9150	STAT3 antisense oligo-nt	DLBCL	(Phase I) NCT03527147
TP53	PRIMA-1/PRIMA-1MET	antimutant p53 agents	CLL, solid tumor	Multiple (phase I/II)[20]
MYC	JQ1	BET inhibitor	PCM	Preclinical phase[21]
EBV	Arginine butyrate + ganciclovir	antiviral agents	EBV-positive lymphomas	(Phase 2) NCT00917826

Supplementary References

1. Edelmann J, Holzmann K, Miller F, et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. Blood 2012;120(24):4783–4794.

2. Rausch T, Jones DTW, Zapatka M, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell 2012;148(1–2):59–71.

3. Nadeu F, Clot G, Delgado J, et al. Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. Leukemia 2018;32(3):645–653.

4. Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. Nature 2015;526(7574):525–530.

5. Scholtysik R, Kreuz M, Klapper W, et al. Detection of genomic aberrations in molecularly defined Burkitt's lymphoma by array-based, high resolution, single nucleotide polymorphism analysis. Haematologica 2010;95(12):2047–2055.

6. Lopez-Corral L, Sarasquete ME, Bea S, et al. SNP-based mapping arrays reveal high genomic complexity in monoclonal gammopathies, from MGUS to myeloma status. Leukemia 2012;26(12):2521–2529.

7. Paiva B, Corchete LA, Vidriales M-B, et al. Phenotypic and genomic analysis of multiple myeloma minimal residual disease tumor cells: a new model to understand chemoresistance. Blood 2016;127(15):1896–1906.

8. 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.

9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–1760.

10. 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.

11. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164.

12. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011;39(17):e118.

13. 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.

14. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods* 2010;7(4):248–249.

15. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46(3):310–315.

16. Liu Z, Filip I, Gomez K, et al. Genomic characterization of HIV-associated plasmablastic lymphoma identifies pervasive mutations in the JAK-STAT pathway. Blood cancer Discov 2020;1(1):112–125.

17. Garcia-Reyero J, Martinez Magunacelaya N, Gonzalez de Villambrosia S, et al. Genetic lesions in MYC and STAT3 drive oncogenic transcription factor overexpression in plasmablastic lymphoma. Haematologica [Epub ahead of print].

18. 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.

19. Johnson DE, O'Keefe RA, Grandis JR. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. Nat Rev Clin Oncol 2018;15(4):234–248.

20. Bykov VJN, Eriksson SE, Bianchi J, Wiman KG. Targeting mutant p53 for efficient cancer therapy. Nat Rev Cancer 2018;18(2):89–102.

21. Delmore JE, Issa GC, Lemieux ME, et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell 2011;146(6):904–917.