Nat. Genet. **doi:10.1038/ng.892**

Analysis of the coding genome of diffuse large B-cell lymphoma

Laura Pasqualucci, Vladimir Trifonov, Giulia Fabbri, Jing Ma, Davide Rossi, Annalisa Chiarenza, Victoria A Wells, Adina Grunn, Monica Messina, Oliver Elliot, Joseph Chan, Govind Bhagat, Amy Chadburn, Gianluca Gaidano, Charles G Mullighan, Raul Rabadan & Riccardo Dalla-Favera

In the version of this supplementary file originally posted online, Supplementary Figures 2, 4 and 6 and Supplementary Tables 12, 14, 18 and 19 contained errors. The errors have been corrected in this file as of 19 August 2011.

Supplementary Information for

Analysis of the Coding Genome of Diffuse Large B-Cell Lymphoma

Laura Pasqualucci[1,2,3*], Vladimir Trifonov[4], Giulia Fabbri[1], Jing Ma[5], Davide Rossi[6], Annalisa Chiarenza[1], Victoria A. Wells[1], Adina Grunn[1], Monica Messina[1], Oliver Elliot[4], Joseph Chan[4], Govind Bhagat[2,3], Amy Chadburn[7], Gianluca Gaidano[6], Charles G. Mullighan[5], Raul Rabadan[4], and Riccardo Dalla-Favera[1,2,3,8*]

[1]*Institute for Cancer Genetics,* [2]*Herbert Irving Comprehensive Cancer Center,* [3]*Department of Pathology & Cell Biology,* [4]*Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA.*

[5]*Department of Pathology, St Jude Children's Research Hospital, Memphis, TN 38105, USA.*

[6]*Division of Hematology, Department of Clinical and Experimental Medicine and IRCAD, Amedeo Avogadro University of Eastern Piedmont, Novara 28100, Italy.*

[7]*Department of Pathology, Northwestern University. Feinberg School of Medicine, Chicago, IL 60611, USA.*

[8]*Department of Genetics & Development, Columbia University, New York, NY 10032, USA.*

[*]To whom correspondence should be addressed. E-mail: lp171@columbia.edu; rd10@columbia.edu

Nature Genetics: doi:10.1038/ng.892

SUPPLEMENTARY NOTE

Cell lines. The DLBCL cell lines used in the study have been described previously[1] and were all maintained in Iscove's Modified Dulbecco Medium (IMDM) supplemented with 10% FBS, 100 μg/mL penicillin, 100 μg/mL streptomycin and 2 mM L-glutamine, except for OCI-Ly10 and OCI-Ly4, which were cultured in IMDM with 20% heparinized human plasma and 55 μM β-mercaptoethanol. Supplementary Table 19 reports the results of the mutation analysis (and FISH analysis, where indicated) performed on commonly affected genes in 21 cell lines.

Primary cases (Screening and Extension Panels). The DLBCL Screening Panel comprised 48 cases (n=22 GCB-DLBCL, 24 ABC-DLBCL and 2 unclassified), which were analyzed for all 56 candidate genes shown in Figure 4, and was extended to 57 additional cases for selected genes found mutated at high frequency (total number, up to 105). The detailed phenotypic characterization of this large dataset has been described previously[2]. For the mutation analysis and SNP array analysis, high molecular weight genomic DNA was extracted from tumor and matched normal tissue (available for 8 cases) according to standard procedures. High quality copy number data were available for 73 of the 105 cases.

Identification of focal minimal common regions (MCR) of aberration. To facilitate the identification of copy number alterations that might have pathogenetic relevance, we first scanned through the curated segmented data across the affected samples to determine all consensus regions that are targeted by overlapping events of similar nature (gains and

losses, separately) in two or more samples. To determine MCRs of aberration, we then curated the list of consensus regions by visual inspection using the Integrative Genomics Viewer (IGV) software (http://www.broadinstitute.org/igv) to retain only the core consensus regions. In particular, we focused on *focal* MCRs of gain or loss, which were defined according to the following criteria: i) the consensus interval across all affected tumors is smaller that 1Mb and/or encompasses 1-3 genes; and ii) in at least one of the affected tumors, the segment of abnormal gene dosage encompasses 1-3 genes. These criteria exclude MCRs that contain ≤ 3 genes but derive from the overlap between large segments, as well as regions devoid of coding genes, which were not considered in the present study.

Genomic Identification of Significant Targets in Cancer (GISTIC) analysis. In order to identify relevant regions of copy number alteration, we applied the GISTIC method described by Beroukhim R. *et al*. (2007)[3] to the curated segmentation results. GISTIC identifies significantly amplified or deleted regions of the genome across a set of samples. Each aberration is assigned a G-score that considers the amplitude of the aberration as well as the frequency of its occurrence across samples. GISTIC then assesses the statistical significance of each aberration by comparing the observed G-scores to the results that would be expected by chance, using a permutation test that is based on the overall pattern of aberrations seen across the genome. The method accounts for multiple-hypothesis testing using the false-discovery rate (FDR) framework to assign a *q* value to each aberration. In this study, we used the default *q*-value cutoff of 0.25 (shown by the green vertical line in Supplementary Figure 2). For each significant region, a peak region was

Nature Genetics: doi:10.1038/ng.892

identified, which is the part of the aberrant region with greatest amplitude and frequency of alteration and with minimal q-value. In addition, a wide peak was determined using a leave-one-out algorithm to allow for errors in the boundaries in a single sample (the wide peak boundaries are more robust for identifying the most likely gene targets in the region). Significant regions of loss and gain are listed in Supplementary Tables 7 and 8, respectively.

Algorithm for detection of Common Focal genomic alterations (ComFocal). As an additional tool to rank genes found affected by copy number alterations, we developed an algorithm that assigns a score to individual genes in the genome, based on the size of the aberration encompassing that gene, the amplitude of the aberration (copy number value), and the number of cases involved (V.T. and R.R., manuscript in preparation). Specifically, each aberrant region (in the curated segmentation data) is first assigned a score S_A that is equal to the inverse of its size L_A (i.e., the number of genes that are included in the region) weighted by the amplitude, defined as the deviation D_A from normal (diploid) copy number. The individual scores of each aberration encompassing a particular gene across different samples are then aggregated into a score S_G for that given gene:

$$S_G = \sum_A S_A = \sum_A D_A/L_A$$

This algorithm was applied to the set of curated segmentation data obtained after circular binary segmentation, as described in the SNP array analysis methods section.

Functional annotation clustering of DLBCL target genes. In order to determine whether genes affected by genetic lesions in DLBCL were enriched in specific functional

categories, the genes resulting from the WES analysis (n=93) and those included in focal MCRs of aberration encompassing 1-3 genes (n=480) were merged in a unique gene list, and functional annotation clustering analysis was performed using the DAVID software (http://david.abcc.ncifcrf.gov)[4,5] based on Gene Ontology terms. This tool groups terms with similar annotations in clusters, and subsequently ranks them by Enrichment Score; clusters were then selected by retaining only the ones including at least one significant ($p<0.05$) term. The DAVID software was also used to run a Functional Annotation chart analysis based on KEGG, BIOCARTA, PANTHER, REACTOME and BBID pathways. Moreover, the overlap between the gene list and the canonical pathways of the Molecular Signatures Database (MSigDB, Broad Institute) was calculated by the compute overlap tool of the GSEA software (http://www.broadinstitute.org/gsea/msigdb/index.jsp)[6].

Statistical analysis. Statistical significance in the frequencies of nucleotide and dinucleotide targeting was assessed by the Poisson distribution [dinucleotide counts that are less or equal than expected (underrepresented) or more or equal than expected (overrepresented)]. After correcting for multiple hypotheses (Sidak), only CpG appears significantly (over) represented ($p<0.01$, $p(Sidak)=0.038$). Similar values were obtained by the Fisher's exact test.

Nature Genetics: doi:10.1038/ng.892

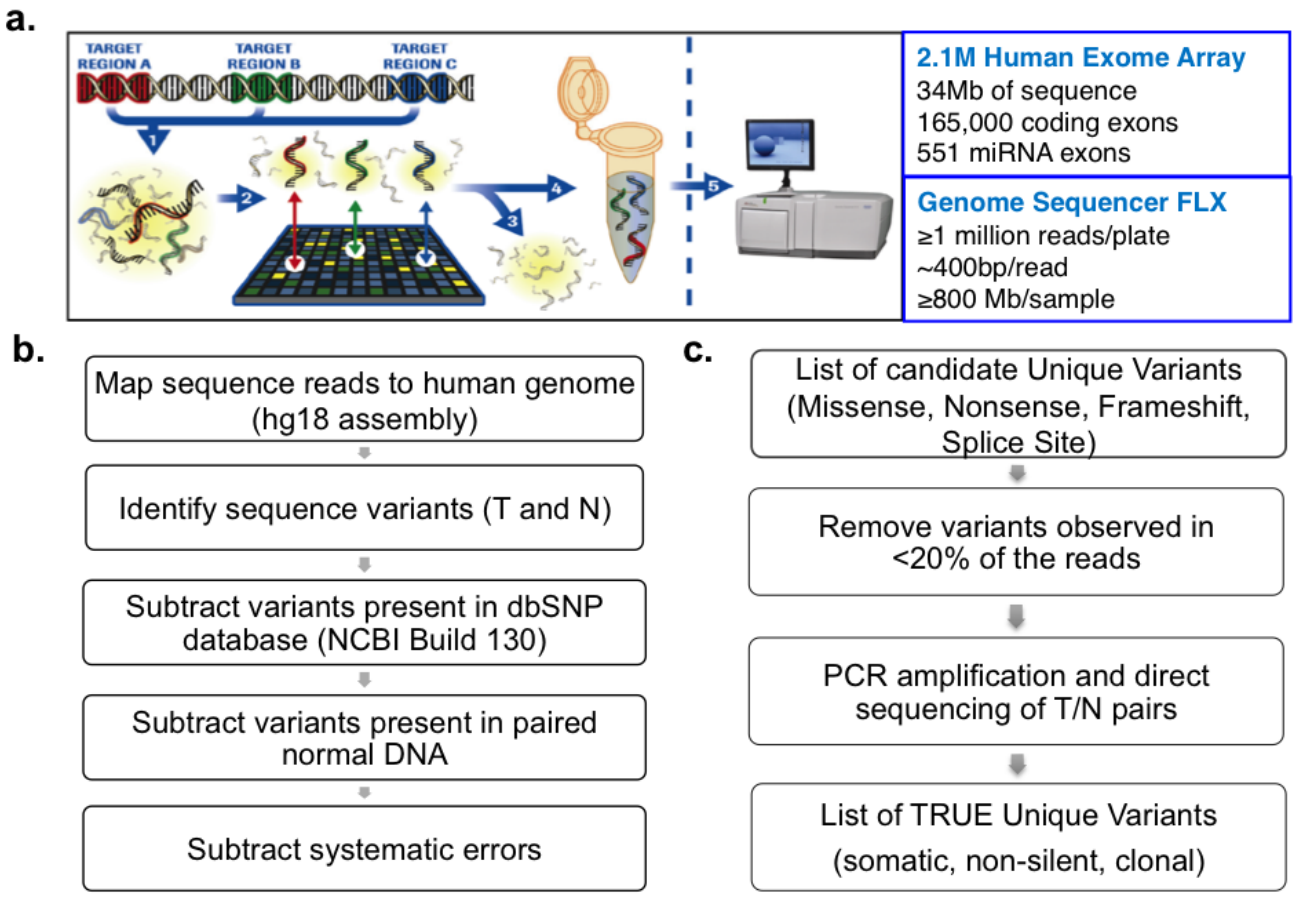
Validation of candidate somatic mutations identified by WES

Overall, 386 variants were predicted by the WES approach to be specifically associated with the tumor sample in the six patients analyzed. Of these, 96 (corresponding to 93 genes) were confirmed to be somatic in origin by Sanger sequencing of the corresponding paired tumor and normal genomic DNA, while 106 were also found in the matched normal DNA, thus representing non previously annotated germline polymorphisms that escaped detection by the high throughput sequencing analysis, possibly due to the relatively low mean depth of coverage (Supplementary Table 4). The remaining 184 variants were absent in both tumor and normal genomic DNA, when tested by Sanger sequencing (data not shown).

Supplementary References

1. Pasqualucci, L. et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189-95 (2011).
2. Compagno, M. et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**, 717-21 (2009).
3. Beroukhim, R. et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* **104**, 20007-12 (2007).
4. Huang da, W. et al. Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* **Chapter 13**, Unit 13 11 (2009).
5. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
6. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).

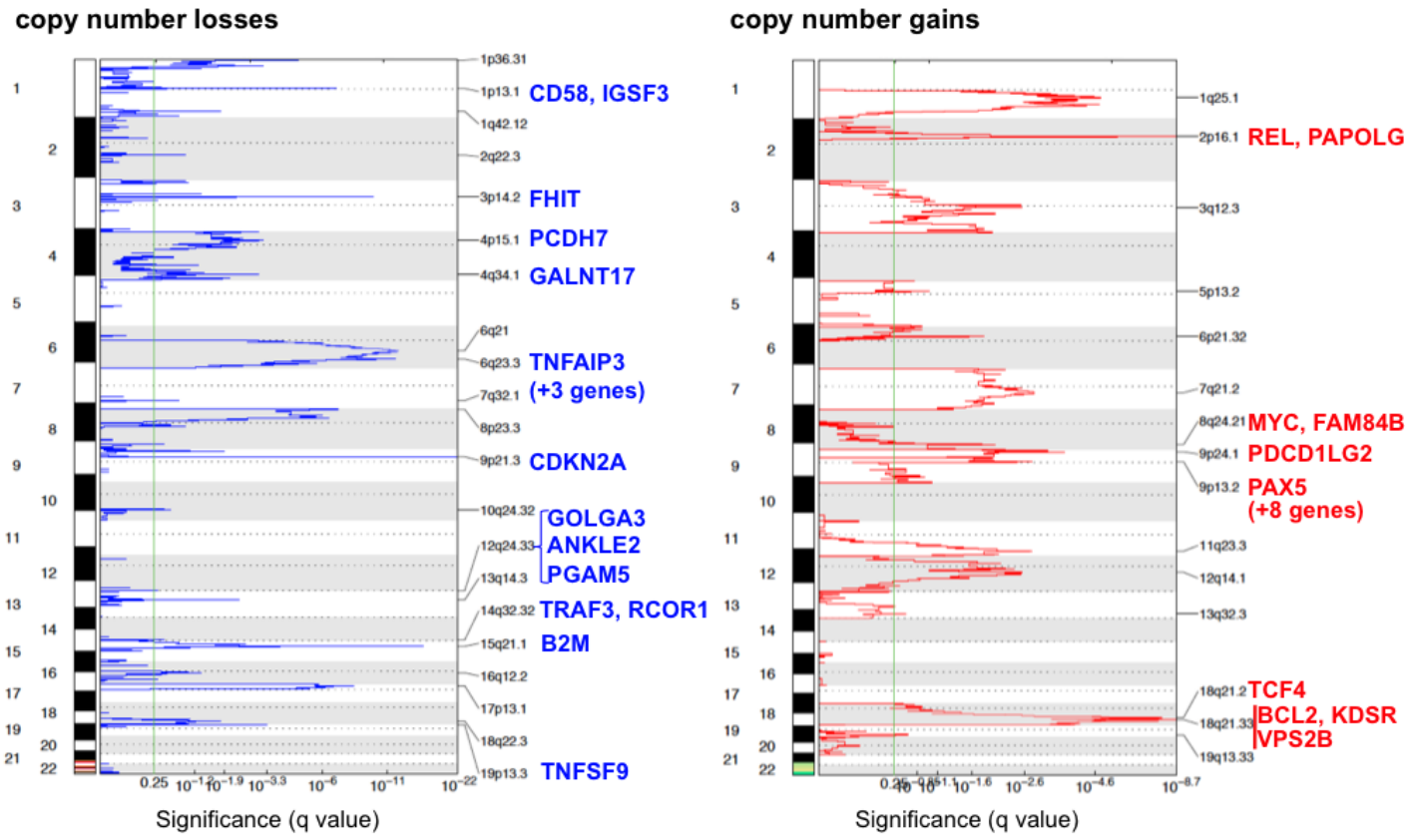
Supplementary Figure 1



Supplementary Figure 1. Whole Exome Sequencing approach. a, Schematics of the sequence capture protocol used for enrichment of non-repetitive coding regions and ultra-high-throughput sequencing of 6 DLBCL patients (matched tumor-normal DNAs). The technical features of the NimbleGen 2.1M capture array and the expected performance of the 454 GS FLX sequencer are indicated. **b,** Computational pipeline adopted for the identification of tumor-specific, non-silent sequence variants after mapping to the human genome assembly hg18 (NCBI Build 36). **c,** Final validation procedure of candidate sequence variants.

Nature Genetics: doi:10.1038/ng.892

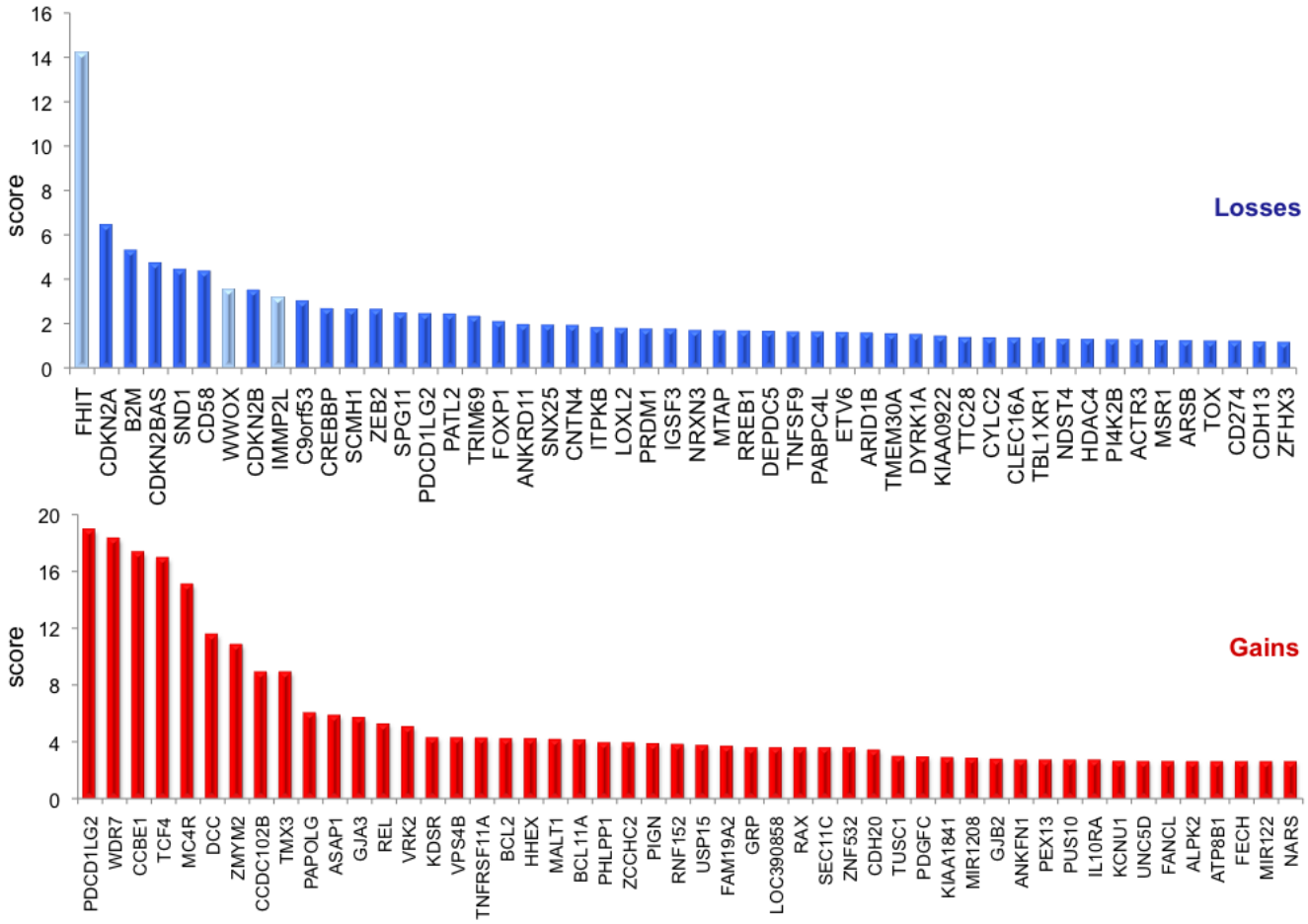
Supplementary Figure 2



Supplementary Figure 2. Significant regions of copy number alterations in the DLBCL genome. Statistical analysis of copy number gains (red) and copy number losses (blue) in 79 DLBCL cases according to GISTIC analysis. False discovery rates for each aberration (q values, X axis) are plotted at each genomic position across the 22 chromosomes (Y axis). The green line represents the q value cutoff for significance (0.25). Centromere positions in each chromosome are indicated by dotted lines, and the location of the significant peak regions (21 deletions and 15 gains) is indicated to the right of each panel; gene names are provided only for focal aberrations encompassing ≤3 genes, unless indicated.

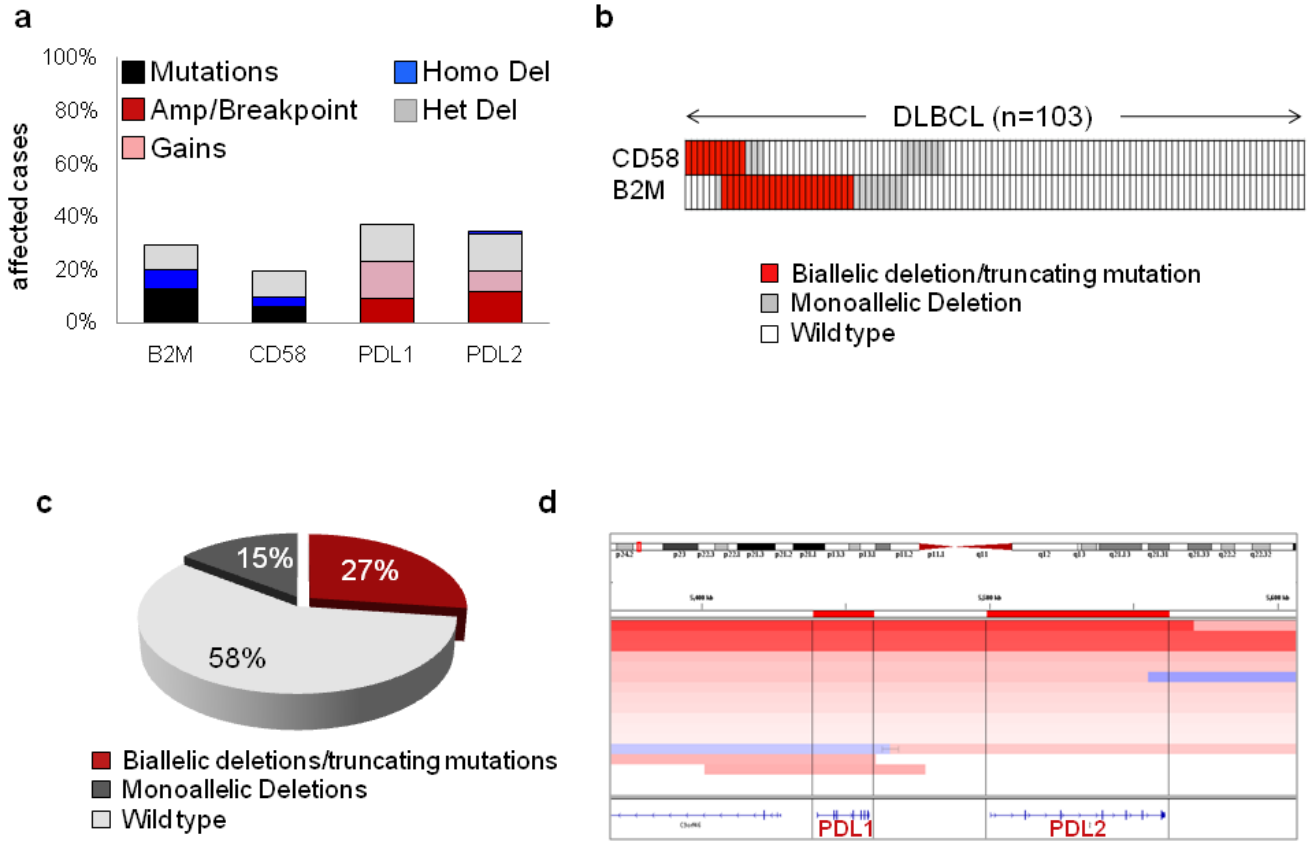
Nature Genetics: doi:10.1038/ng.892

Supplementary Figure 3



Supplementary Figure 3. Top 50 genes affected by copy number losses (top) and gains (bottom) in DLBCL, based on the ComFocal algorithm. Genes are ranked from highest to lowest, according to their significance score (see Supplementary Notes). Light blue denotes common fragile sites. Note that gains/amplifications encompassing only a subset of gene exons were also considered in this analysis, although the functional consequences of these lesions are less clear.

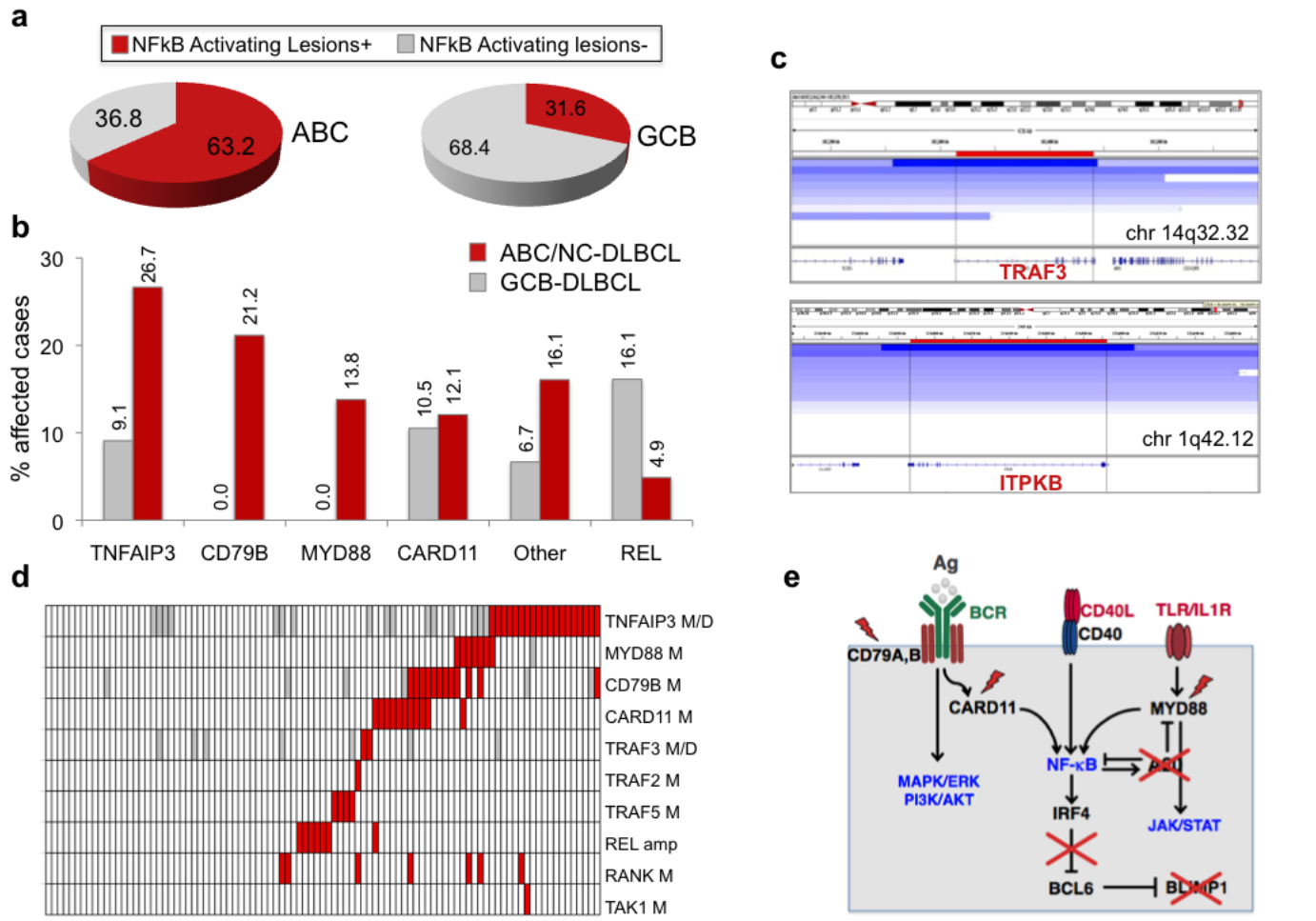
Supplementary Figure 4



Supplementary Figure 4. Genetic lesions of immunomodulatory genes in DLBCL. a, Frequency of genomic aberrations at four genes involved in T cell-B cell recognition. **b,** Relationship between inactivating mutations of *B2M* and *CD58*. Each column in the heatmap represents a patient, and the status of the two genomic loci is indicated by the color. **c,** Overall proportion of DLBCL patients harboring genetic lesions in *CD58* and *B2M*. **d,** Graphic display of copy number data from 15 patients harboring amplifications and rearrangements involving the *PDL1* and *PDL2* loci. Segmentation data were visualized using the IGV tool, where each track represents one sample, white denotes a normal (diploid) copy number, red indicates a region of copy number gain and blue indicates a region of copy number loss. Note that, due to the presence of non-tumor cells infiltrating the biopsy, the inferred copy number, and the corresponding color intensity, may vary across samples. Individual genes in the region are aligned in the bottom panel, and the boxed areas (defined by the red bars at the top) highlight the two *PDL* genes.

Nature Genetics: doi:10.1038/ng.892

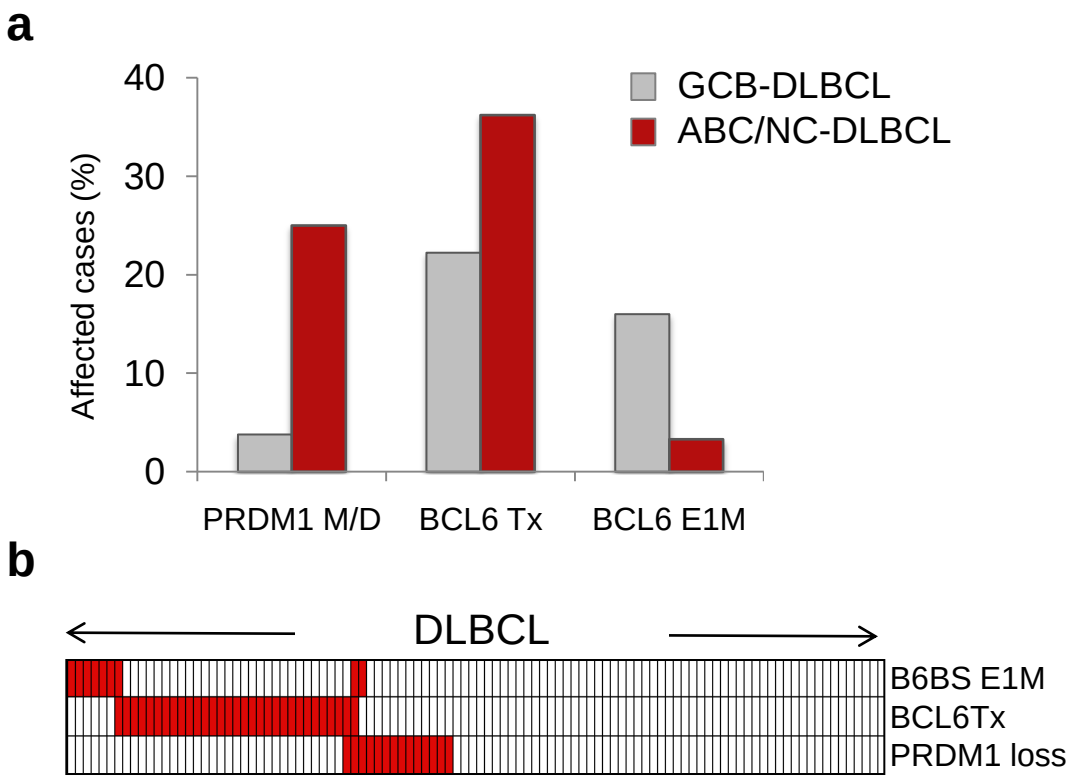
Supplementary Figure 5



Supplementary Figure 5. DLBCL-associated genetic lesions leading to activation of NF-κB. a, Overall prevalence of cases carrying structural alterations that converge on the activation of the NF-κB signaling pathway (mutation and copy number data, combined). **b,** Frequency of genetic lesions affecting components of the NF-kB, BCR and TLR pathway in major DLBCL phenotypic subtypes. The following genes are included in the category designated as "other": *TRAF2*, *TRAF3*, *TRAF5*, *TAK1* and *TNFRSF11A* (also known as *RANK*). For *REL*, only high copy number amplifications are scored. **c,** Inferred log2 copy number data from DLBCL primary cases carrying loss of *TRAF3* (upper panel) and *ITPKB* (lower panel), shown using the IGV visualization tool. Each track represents a patient. Blue lines indicate the extent of the deletions identified in the DLBCL samples, with the darker blue segment corresponding to a homozygous loss. **d,** Relative distribution of genetic lesions affecting the indicated genes in individual DLBCLs (primary cases only). Each column represents one patient, with color codes indicating the presence or absence of the corresponding feature, as follows: red, structural alteration; white, wild type; light grey, monoallelic deletions (*TNFAIP3* and *TRAF3)* or missense mutations whose functional significance is unknown (*CD79B* and *MYD88*). Only samples with complete information about the indicated lesions are considered in the plot. **e,** Schematics of the NF-κB-IRF4-BCL6-PRDM1 pathway; only some of the upstream activating signals are shown in the figure. Most common targets of genetic alterations are indicated.

Nature Genetics: doi:10.1038/ng.892

Supplementary Figure 6



Supplementary Figure 6. Mutually exclusive genetic alterations of BCL6 and PRDM1. **a**, Frequency of *PRDM1* biallelic loss (mutations and/or deletions), *BCL6* translocations (Tx), and mutations affecting two *BCL6* binding sites within the gene noncoding exon 1 in two major phenotypic sybtypes of DLBCL (primary biopsies only). **b**, Relative distribution of the indicated lesions in individual DLBCL cases. Only samples with complete information about the three types of lesion are considered in the plot.

Nature Genetics: doi:10.1038/ng.892

Supplementary Table 1. Features of the 6 DLBCL discovery cases analyzed by Whole Exome Sequencing

Variable	Patient ID					
	2112	**2203**	**2204**	**2205**	**2209**	**2210**
DLBCL subtype	NC	ABC	ABC	ABC	non-GCB	GCB
Age	55	62	43	42	79	73
Sex	F	F	M	F	F	M
Tumor site	lymph node	lymph node	lymph node	lymph node	spleen	lymph node
IgHV	IGHV3-66*01	IGHV3-30*03	IGHV4-34*01	IGHV1-18*01	IGHV4-34*02	IGHV4-34*01
IgHD	IGHD3-3*01	IGHD3-9*01	IGHD3-3*01	IGHD1-26*01	IGHD5-12*01	IGHD6-6*01
IgHJ	IGHJ4*02	IGHJ6*02	IGHJ4*02	IGHJ6*02	IGHJ6*02	IGHJ5*02
% Homology	89.12	87.61	92.28	83.33	86.67	96.12
MYC Translocation	nd	Neg	Neg	Neg	Neg	Neg
BCL2-IGH Translocation	nd	Neg	Neg	Neg	Neg	Neg
BCL6 Translocation	Neg	Neg	Pos	Neg	Pos	Neg
6q21 deletion	Neg	Neg	Pos	Neg	Neg	Pos
TP53 deletion	Neg	Neg	Pos	Neg	Neg	Neg

NC, unclassified; ABC, activated B cell-like; GCB, germinal center B cell-like

Nature Genetics: doi:10.1038/ng.892

Supplementary Table 2. Results of Titanium 454 sequencing and mapping after whole exome capture

Sample ID	Target Region Coverage (%)			Mean	N of mapped	Unique reads in	N of variant reads (HC)		
	1X	≥5X	HCC^	Depth^^	reads (%)	region (%)	Total	Known	Novel*
2112N	95.8	76.8	70.2	10.0	2465148 (99.12%)	1655515 (70.4%)	10758	10161	597
2112T	96.1	78.6	71.1	9.5	2502590 (99.10%)	1706903 (71.6%)	10902	10193	709
2203N	96.0	77.2	69.7	9.0	2274722 (98.91%)	2038393 (88.6%)	9730	9116	614
2203T	95.4	75.5	68.3	8.9	2192985 (98.78%)	1981190 (89.2%)	9423	8683	740
2204N	95.7	79.4	72.7	10.0	2489172 (99.01%)	2282059 (90.8%)	9859	9141	718
2204T	96.0	79.4	74.3	12.8	4219862 (98.72%)	3332313 (77.9%)	10247	9354	893
2205N	96.5	81.4	74.3	9.7	2341274 (98.98%)	1658253 (74.3%)	11478	10663	815
2205T	96.3	79.2	72.1	9.4	2297582 (98.98%)	1619798 (74.1%)	11349	10585	764
2209N	96.2	79.4	72.3	9.4	2479111 (99.19%)	1721161 (72.3%)	10626	10012	614
2209T	94.7	70.8	64.7	9.0	2344386 (99.27%)	1593773 (70.9%)	8997	8464	533
2210N	96.3	78.9	72.1	9.6	2376481 (99.07%)	1722250 (75.4%)	10741	10064	677
2210T	95.6	75.7	70.1	10.2	2388566 (98.84%)	1739238 (75.8%)	10570	9881	689

^ High Confidence Coverage, i.e. % of the target region where coverage depth was estimated to be sufficient for detecting both alleles of heterozygous variants. This value can be taken as an estimate of the mutation-detection sensitivity (see methods)

^^ Mean number of sequence reads covering the target exome

* Number of variant reads (High Confidence) not reported in the NCBI dbSNP database, before cross-comparison with paired normal

Nature Genetics: doi:10.1038/ng.892

Supplementary Table 3. Validated somatic mutations identified by whole exome sequencing in the DLBCL discovery panel

Sample ID	Gene	Exon#	Chr	Start^	End^	Ref nt	Var nt	%	CovT	CovN	Mutation Type	AA change	SIFT prediction	Polyphen prediction	COSMIC database^^	CGC database**	CCDS ID
2112	ADCK1	10	14	77469480	77469480	C	T	23	13	-	Missense	P522L	Damaging (low confidence)	Probably damaging	No	No	CCDS9869.1
2112	BRSK1	10	19	60506038	60506038	G	A	50	8	-	Missense	R340H	Tolerated	Unknown	Yes (4)	No	CCDS12921.1
2112	GCFC1	8	21	33049459	33049459	C	T	33	10	-	Missense	R487Q	Damaging	Probably damaging	No	No	CCDS33541.1
2112	CARD11	4	7	2950673	2950673	G	A	50	6	8	Missense	T128M	Damaging	Probably damaging	Yes (3)	Yes	CCDS5336.2
2112	CD79B	4	17	59360531	59360531	A	T	82	11	-	Missense	Y196N	Damaging	Probably damaging	No	Yes	CCDS11655.1
2112	CPNE7	16	16	88189480	88189480	G	A	44	9	-	Missense	V578I	Damaging	Probably damaging	No	No	CCDS10980.1
2112	SPECC1	1	17	19940675	19940675	C	G	57	8	-	Missense	S40C	Damaging (low confidence)	Probably damaging	No	No	CCDS32590.1
2112	DENND3	9	8	142242636	142242636	T	G	25	12	8	Missense	Y379D	Tolerated	Probably damaging	No	No	CCDS34947.1
2112	DUSP27	5	1	165363521	165363521	G	A	60	10	17	Missense	M843I	Tolerated	Benign	No	No	CCDS30932.1
2112	GIMAP5	2	7	150070209	150070209	T	C	43	9	5	Missense	S17P	Tolerated	Benign	Yes (1)	No	CCDS5907.1
2112	IGSF5	5	21	40073073	40073073	G	A	56	10	-	Missense	R302H	Damaging (low confidence)	Probably damaging	No	No	CCDS33562.1
2112	OFD1	16	X	13688504	13688504	G	T	62	13	-	Missense	K668N	Damaging (low confidence)	Possibly damaging	No	No	CCDS14157.1
2112	PDCD4	10	10	112645760	112645760	T	A	30	10	-	Missense	V425E	Damaging	Benign	Yes (2)	No	CCDS7567.1
2112	PPP2R5A	1	1	210526077	210526077	T	A	50	6	8	Missense	M1K	Damaging (low confidence)	Possibly damaging	Yes (1)	No	CCDS1503.1
2112	PRDM1	1	6	106643019	106643019	T	G	73	11	8	Splice site	E97fs	Damaging	Damaging	Yes (1)	Yes	CCDS5054.1
2112	PRX	4	19	45594857	45594857	C	G	35	17	-	Missense	K414N	Tolerated	Probably damaging	Yes (2)	No	CCDS33028.1
2112	SCNN1A	1	12	6353892	6353892	C	T	55	11	-	Missense	G107R	Tolerated	Benign	No	No	CCDS8543.1
2112	SDCCAG1	17	14	49351269	49351269	C	T	31	16	-	Missense	R545Q	Damaging	Probably damaging	No	No	CCDS9694.1
2112	SERPINA6	1	14	93850239	93850239	C	T	58	12	-	Missense	S167N	Tolerated	Possibly damaging	Yes (3)	No	CCDS9924.1
2112	STOML2	2	9	35092777	35092777	G	C	30	11	11	Missense	S30C	Damaging (low confidence)	Probably damaging	No	No	CCDS6577.1
2112	STRADA	2	17	59157777	59157777	G	A	68	19	-	Missense	S16L	Damaging (low confidence)	Possibly damaging	Yes (1)	No	CCDS32703.1
2112	UBE2A	2	X	118592926	118592926	G	C	30	10	-	Missense	A27P	Damaging	Possibly damaging	No	No	CCDS14581.1
2112	WDR88	4	19	38330426	38330426	C	G	80	11	-	Missense	A171G	Damaging	Probably damaging	Yes (3)	No	CCDS12429.1
2203	ADAMTSL3	15	15	82373042	82373042	G	A	67	7	15	Missense	R632Q	Tolerated	Benign	Yes (18)	No	CCDS10326.1
2203	ARHGEF2	17	1	154188192	154188192	C	G	62	10	12	Splice site	D725fs	Damaging	Damaging	Yes (1)	No	CCDS1125.1
2203	BCL2L10	1	15	50191929	50191929	C	T	45	13	2	Missense	R96K	Tolerated	Benign	No	No	CCDS10148.1
2203	C12orf35	1	12	32028772	32028772	C	T	46	17	8	Nonsense	Q1206*	Damaging	Damaging	Yes (1)	No	CCDS8725.2
2203	CAMTA1	9	1	7646677	7646677	G	A	67	7	5	Missense	G495S	Damaging (low confidence)	Probably damaging	Yes (9)	No	CCDS30576.1
2203	CD36	4	7	80130402	80130402	C	A	42	14	9	Missense	T197K	Tolerated	Benign	Yes (1)	No	CCDS34673.1
2203	CD74	2	5	149766917	149766917	G	A	86	7	12	Missense	L97F	Damaging	Probably damaging	Yes (1)	Yes	CCDS34276.1
2203	CDC42EP1	2	22	36294321	36294321	G	A	30	10	9	Missense	A242T	Damaging (low confidence)	Unknown	No	No	CCDS13949.1
2203	CXorf48	2	X	134131250	134131250	T	G	27	11	-	Missense	K71N	Tolerated	Possibly damaging	No	No	CCDS14647.1
2203	ETF1	9	5	137872286	137872286	T	G	37	8	16	Missense	Q401P	Damaging	Possibly damaging	Yes (2)	No	CCDS4207.1
2203	FBXO31	8	16	85925216	85925216	G	A	43	8	2	Missense	R392C	Damaging (low confidence)	Benign	Yes (1)	No	CCDS32501.1
2203	GPR112	3	X	135256002	135256002	C	G	62	21	-	Missense	T824S	Damaging (low confidence)	Benign	Yes (11)	No	CCDS35409.1
2203	HLA-DMB	3	6	33012974	33012979	CAGGTG	-	43	7	17	Frameshift del	C192fs	Damaging	Damaging	No	No	CCDS4760.1
2203	HNF1B	2	17	33173723	33173723	G	T	44	10	3	Missense	A122D	Damaging	Probably damaging	Yes (3)	No	CCDS11324.1
2203	LRP1B	4	2	141728635	141728635	G	A	100	3	10	Missense	T130I	Tolerated	Benign	Yes (34)	No	CCDS2182.1
2203	MDN1	69	6	90453413	90453413	G	A	36	13	15	Missense	S3834F	Damaging	Probably damaging	Yes (8)	No	CCDS5024.1
2203	MEGF10	8	5	126774052	126774052	-	AGG	43	9	16	In frame ins	G330+	Damaging	Damaging	Yes (4)	No	CCDS4142.1
2203	MTMR8	9	X	63473881	63473881	C	A	27	11	-	Nonsense	G365*	Damaging	Damaging	Yes (4)	No	CCDS14379.1
2203	MYOM2	1	8	1986364	1986364	G	A	29	17	8	Missense	R26Q	Tolerated	Probably damaging	Yes (5)	No	CCDS5957.1
2203	PDE6A	2	5	149294380	149294380	G	A	37	8	13	Missense	A190V	Damaging	Probably damaging	Yes (2)	No	CCDS4299.1
2203	RASGEF1A	5	10	43015668	43015668	C	T	50	8	21	Missense	V242I	Tolerated	Benign	No	No	CCDS7202.2
2203	RYR1	82	19	43718491	43718506	ATGCCTTTGAGAGACA	-	54	15	8	Frameshift del	N3844fs	Damaging	Damaging	Yes (10)	No	CCDS33011.1
2203	SERPINA1	1	14	93918765	93918765	C	T	45	12	10	Missense	G188D	Tolerated	Benign	Yes (1)	No	CCDS9925.1
2203	TSC22D1	1	13	44045498	44045498	C	T	33	16	16	Missense	A905T	Damaging (low confidence)	Possibly damaging	No	No	CCDS31966.1
2203	TSPAN7	4	X	38418509	38418509	C	T	46	13	-	Missense	R146C	Damaging	Possibly damaging	Yes (1)	No	CCDS14248.1
2203	ZFHX3	8	16	71386682	71386682	G	A	40	11	10	Missense	P2467L	Not scored	Unknown	Yes (2)	No	CCDS10908.1
2203	ZNF311	5	6	29074501	29074501	C	A	67	6	4	Missense	C135F	Damaging	Probably damaging	No	No	CCDS34357.1
2203	ZNF521	3	18	21059298	21059298	A	G	31	20	10	Missense	L861S	Tolerated	Benign	Yes (7)	Yes	CCDS32806.1
2203	ZWILCH	1	15	64584741	64584741	G	A	44	9	4	Missense	R4Q	Damaging (low confidence)	Benign	Yes (1)	No	CCDS10219.1
2204	NCEH1	2	3	173848463	173848463	C	T	42	16	14	Missense	V124I	Tolerated	Benign	No	No	CCDS33893.1
2204	ABCC9	36	12	21851680	21851680	T	C	38	16	14	Missense	D1439G	Damaging	Benign	Yes (4)	No	CCDS8693.1
2204	DPF2	10	11	64872979	64872979	G	T	36	14	11	Splice site	G381fs	Damaging	Damaging	Yes (1)	No	CCDS8100.1
2204	MYD88	5	3	38157645	38157645	T	C	20	17	4	Missense	L265P	Not scored	Probably damaging	Yes (93)	Yes	CCDS2674.2

2204	OR8H3	1	11	55646778	55646778	G	A	27	50	22	Missense	M118I	Damaging	Probably damaging	Yes (2)	No	CCDS31519.1
2205	AFF2	11	X	147845734	147845734	A	G	67	7	12	Missense	H820R	Tolerated	Benign	Yes (5)	No	CCDS14684.1
2205	DCHS1	2	11	6611935	6611935	G	A	43	14	10	Missense	A659V	Damaging	Probably damaging	Yes (3)	No	CCDS7771.1
2205	DHX15	4	4	24167097	24167097	T	C	25	12	10	Missense	M246V	Damaging	Probably damaging	Yes (1)	No	CCDS33966.1
2205	DPYD	19	1	97473123	97473123	G	A	33	12	4	Missense	P772L	Damaging	Probably damaging	Yes (5)	No	CCDS30777.1
2205	DPYD	6	1	97937522	97937522	T	G	33	9	4	Missense	E218A	Damaging	Probably damaging	Yes (5)	No	CCDS30777.1
2205	DSC3	11	18	26842010	26842010	A	G	37	8	4	Missense	Y545H	Tolerated	Possibly damaging	Yes (2)	No	CCDS32810.1
2205	DUSP9	3	X	152568717	152568717	G	A	43	8	11	Missense	M306I	Damaging	Probably damaging	Yes (2)	No	CCDS14724.1
2205	ENAM	8	4	71728945	71728948	TCAA	-	36	12	10	Frameshift del	S980fs	Damaging	Damaging	Yes (3)	No	CCDS3544.1
2205	GBP6	10	1	89623413	89623415	AAG	-	45	12	13	In frame del	ΔK567	nd	nd	Yes (2)	No	CCDS723.1
2205	HMGB1	3	13	29934772	29934772	A	T	50	8	11	Missense	V125D	Damaging	Probably damaging	Yes (2)	No	CCDS9335.1
2205	HMGB1	3	13	29934833	29934833	A	G	43	7	10	Missense	F105L	Damaging	Possibly damaging	Yes (2)	No	CCDS9335.1
2205	KLF2	3	19	16298726	16298726	G	A	50	11	14	Missense	E318K	Damaging	Probably damaging	No	No	CCDS12343.1
2205	MAGEC3	1	X	140812782	140812782	C	A	25	12	21	Missense	F524L	Tolerated	Possibly damaging	Yes (3)	No	CCDS14676.1
2205	OFD1	1	X	13663116	13663116	T	C	83	7	4	Missense	M2T	Damaging (low confidence)	Possibly damaging	No	No	CCDS14157.1
2205	PCDHB10	1	5	140553576	140553576	G	A	24	26	21	Missense	V423I	Tolerated	Possibly damaging	Yes (1)	No	CCDS4252.1
2205	PHOX2A	2	11	71629883	71629883	C	G	42	13	4	Missense	E106Q	Damaging	Possibly damaging	No	No	CCDS8214.1
2205	TLL2	14	10	98136730	98136730	C	T	25	16	17	Missense	E608K	Damaging	Probably damaging	Yes (2)	No	CCDS7449.1
2205	ZNF394	3	7	98929662	98929662	G	A	30	11	11	Missense	S371F	Damaging	Probably damaging	No	No	CCDS5666.1
2209	AKAP8	14	19	15326900	15326901	AT	-	25	13	13	Frameshift del	H635fs	Damaging	Damaging	Yes (2)	No	CCDS12329.1
2209	B2M	1	15	42791105	42791105	T	A	47	18	12	Splice site	R23fs	Damaging	Damaging	Yes (2)	No	CCDS10113.1
2209	CCND3	5	6	42011666	42011666	A	T	43	7	6	Missense	I290K	Damaging	Benign	No	Yes	CCDS4863.1
2209	CREBBP	2	16	3840874	3840874	G	A	75	9	6	Nonsense	R75*	Damaging	Damaging	Yes (10)	Yes	CCDS10509.1
2209	EXTL2	4	1	101112475	101112475	A	T	37	8	11	Missense	S202T	Tolerated	Benign	No	No	CCDS775.1
2209	KCNA5	1	12	5024215	5024215	G	A	37	8	14	Missense	R214H	Damaging	Probably damaging	Yes (4)	No	CCDS8536.1
2209	MYH7	28	14	22957362	22957362	C	T	37	8	3	Missense	E1356K	Damaging	Possibly damaging	Yes (3)	No	CCDS9601.1
2209	MYO1G	8	7	44977063	44977063	G	C	27	11	7	Missense	L323V	Tolerated	Benign	Yes (3)	No	CCDS34629.1
2209	PABPC5	1	X	90577701	90577701	G	A	40	15	12	Missense	A157T	Damaging	Probably damaging	Yes (3)	No	CCDS14460.1
2209	PCDHB3	1	5	140462098	140462098	G	A	75	9	7	Missense	V561M	Damaging	Probably damaging	Yes (2)	No	CCDS4245.1
2209	PMS1	1	2	190364797	190364797	C	T	75	4	8	Missense	A6V	Damaging	Possibly damaging	Yes (4)	Yes	CCDS2302.1
2209	SMARCA1	8	X	128467832	128467832	C	T	37	8	10	Missense	R335H	Damaging	Probably damaging	Yes (5)	No	CCDS14612.1
2209	TMEM30A	2	6	76034110	76034110	A	T	40	10	15	Nonsense	C104*	Damaging	Damaging	No	No	CCDS4983.1
2209	TNFAIP3	1	6	138234353	138234357	GTAAG	-	37	8	10	Splice site	G99fs	Damaging	Damaging	Yes (117)	Yes	CCDS5187.1
2210	COL4A6	3	X	107440641	107440641	C	T	30	9	8	Missense	G47E	Damaging	Unknown	Yes (8)	No	CCDS14541.1
2210	DCC	11	18	48988045	48988045	A	G	19	17	7	Splice site	N575fs	Damaging	Damaging	Yes (9)	No	CCDS11952.1
2210	FKBP9	7	7	33002401	33002401	G	A	23	14	4	Missense	D381N	Tolerated	Benign	Yes (14)	No	CCDS5439.1
2210	MED12L	15	3	152550628	152550630	AGA	-	37	8	6	In frame del	ΔK746	nd	nd	Yes (5)	No	CCDS33876.1
2210	OR51D1	1	11	4618421	4618421	T	A	28	18	12	Missense	H275Q	Damaging	Probably damaging	No	No	CCDS31357.1
2210	RGAG1	1	X	109582172	109582172	G	A	23	29	21	Missense	M557I	Tolerated	Benign	Yes (3)	No	CCDS14552.1
2210	SYNE1	58	6	152736050	152736050	A	G	67	6	3	Splice site	E3116fs	Damaging	Damaging	Yes (23)	No	CCDS5235.1

In CCDS Reference Sequence

^ Numbering according to the Human Genome hg18 assembly

^^ For genes annotated in the COSMIC database, the number of mutations that have been reported is given in brackets

**Cancer Gene Census database, March 2011 (http://www.sanger.ac.uk/genetics/CGP/Census)

Abbreviations: Refnt, reference nucleotide; Varnt, variant nucleotide; CovT, coverage tumor; CovN, coverage normal; %, fraction of sequence reads showing the variant nucelotide; AA, amino acid; fs, frameshift; Δ, deletion; +, insertion; nd, not determined (The PolyPhen-2 algorithm predicts only the impact of amino acid substitutions)

Nature Genetics: doi:10.1038/ng.892

Supplementary Table 4. Germline mutations identified by Whole Exome Sequencing and not annotated in the dbSNP database (Release 132)

Sample ID	Gene	Exon#	Chr	Start^	End^	Ref nt	Var nt	%	CovT	CovN	Mutation Type	AA change
2112	ADAMTS14	11	10	72168715	72168715	G	A	54	15	-	Missense	G571R
2112	DNAH3	37	16	20949939	20949939	T	A	55	13	-	Missense	I1790F
2112	FBL	2	19	45023207	45023207	C	T	100	4	-	Missense	R24H
2112	KLHL25	1	15	84113993	84113993	A	G	50	10	-	Missense	M18T
2112	LILRA1	5	19	59799035	59799039	GGAGA	TGGGG*	40	11	-	Missense	GE261-262WG*
2112	MAP3K11	7	11	65130086	65130086	G	A	75	8	-	Missense	P549L
2112	MTMR15	1	15	28984876	28984876	G	A	33	15	-	Missense	E240K
2112	MYLK	11	3	124910250	124910250	C	T	80	5	6	Missense	V640M
2112	NLRP12	3	19	59005086	59005086	C	T	36	11	-	Missense	E547K
2112	NLRP7	3	19	60143347	60143347	T	G	44	18	-	Missense	M218L
2112	NSMCE4A	3	10	123720467	123720467	T	C	60	5	-	Missense	M160V
2112	P2RX6	1	22	19699548	19699548	A	T	87	9	-	Nonsense	K29*
2112	PRKAR1B	2	7	686808	686808	G	C	75	4	9	Missense	P87A
2112	RBMXL1	1	1	89221825	89221825	T	A	23	13	9	Missense	R91S
2112	SIPA1L3	7	19	43302348	43302348	G	A	60	7	-	Missense	V952M
2112	TCHH	2	1	150350837	150350837	-	ins 14bp	25	17	4	Frameshift ins	E495fs
2112	TRPM5	23	11	2383341	2383341	C	T	100	3	-	Missense	G1128D
2203	CLMN	9	14	94739127	94739127	T	G	67	8	4	Missense	E771R
2203	FAM46A	1	6	82518509	82518509	-	ins 15bp	75	5	2	In frame ins	L23+
2203	FBXO17	1	19	44132510	44132510	G	C	60	5	1	Missense	A97G
2203	LRIG2	15	1	113458629	113458629	C	T	43	9	10	Missense	A713V
2203	LRRC16B	34	14	23604756	23604756	G	T	100	6	2	Missense	G1161V
2203	NARFL	11	16	720576	720576	C	T	50	8	4	Missense	G425S
2203	NLGN2	6	17	7259834	7259834	G	A	36	18	6	Missense	D440N
2203	OAS2	10	12	111931327	111931327	C	T	57	8	3	Missense	R650C
2203	SERPINA6	1	14	93850627	93850627	C	A	73	17	7	Missense	G38C
2203	SLC16A9	5	10	61082565	61082565	A	C	50	11	2	Missense	F501V
2203	SPSB3	2	16	1768452	1768452	C	T	57	8	5	Missense	G97R
2203	SPTB	25	14	64309317	64309317	T	C	31	19	4	Missense	S1763G
2203	TIA1	11	2	70295066	70295066	T	C	56	10	11	Missense	DN314
2203	TIAM1	13	21	31459179	31459179	A	T	79	23	8	Missense	Q307R
2203	TTC39B	14	9	15178134	15178134	T	A	40	18	12	Splice site	A412fs
2203	UCN2	1	3	48575260	48575260	G	T	60	5	2	Missense	T101N
2203	ULK1	16	12	130964854	130964854	G	A	36	12	12	Missense	A418T
2203	UTRN	42	6	144895974	144895974	G	T	83	6	9	Missense	D2040Y
2203	WDR81	1	17	1578578	1578578	C	T	60	5	5	Missense	T1192M
2203	ZNF532	3	18	54757736	54757736	T	C	58	16	6	Missense	Y870H
2204	AGBL5	11	2	27143910	27143912	ATC	-	50	11	17	In frame del	ΔH712
2204	ARFGEF1	31	8	68292819	68292819	C	T	56	19	7	Missense	A1483S
2204	C10orf90	3	10	128183115	128183115	C	T	75	20	6	Missense	G215E
2204	CCDC66	10	3	56625096	56625098	-	TCT	38	20	18	In frame ins	S490+
2204	DAO	6	12	107812255	107812255	C	T	67	10	3	Missense	R199W
2204	DST	23	6	56593355	56593355	T	C	40	21	7	Missense	M1146V

2204	DUOX2	21	15	43180766	43180766	T	C	42	19	5	Splice site	G951fs
2204	F11R	4	1	159237177	159237177	G	A	67	16	15	Missense	R86W
2204	F2RL2	2	5	75949683	75949683	C	T	50	10	11	Missense	R202Q
2204	FBXL13	18	7	102241087	102241090	TAGA	-	71	8	6	Frameshift del	K716fs
2204	HPS4	10	22	25189975	25189975	T	C	50	13	6	Missense	M541V
2204	LRRC55	1	11	56706205	56706205	C	T	67	15	4	Missense	R88C
2204	NR1D2	8	3	23993918	23993918	C	T	70	23	3	Splice site	P579fs
2204	OR2T4	1	1	246591561	246591561	T	G	80	12	17	Missense	M19R
2204	SLC7A14	1	3	171727302	171727302	C	T	50	9	11	Missense	G40R
2204	TACC2	3	10	123835139	123835139	C	T	60	5	4	Missense	P1045L
2204	TCF3	18	19	1562736	1562736	G	C	80	7	8	Missense	S645R
2204	TSHZ3	2	19	36461133	36461133	T	C	50	8	9	Missense	E469G
2204	VRK2	4	2	58167059	58167059	G	A	45	18	7	Missense	G113E
2205	C15orf27	6	15	74250450	74250450	A	G	44	9	8	Missense	N192S
2205	C16orf84	14	16	87309113	87309113	C	T	37	8	3	Missense	P481L
2205	CDSN	2	6	31191924	31191924	C	T	100	7	4	Missense	A483T
2205	CLEC18B	9	16	73002364	73002364	G	A	25	12	12	Missense	R352C
2205	CTU2	14	16	87309113	87309113	C	T	37	8	3	Missense	P481L
2205	E2F8	10	11	19203863	19203863	G	C	33	12	10	Missense	P673R
2205	GIGYF2	20	2	233405971	233405971	G	A	80	5	2	Missense	R897H
2205	GPLD1	5	6	24583389	24583389	G	A	50	8	3	Missense	A127V
2205	GPRIN2	1	10	46419215	46419215	G	A	38	13	11	Missense	R110Q
2205	GRP	3	18	55048655	55048658	GGAA	-	50	18	10	Frameshift del	R141fs
2205	HBEGF	2	5	139705773	139705773	G	A	75	4	8	Missense	P43S
2205	MAMDC4	5	9	138868103	138868122	TGGCAGAGCACAGGGCCCTG	-	43	8	3	Frameshift del	W170fs
2205	PADI3	5	1	17465819	17465819	C	T	44	9	8	Missense	P143S
2205	PRG2	3	11	56912693	56912693	C	T	68	20	4	Missense	R144Q
2205	PTK2B	26	8	27366590	27366590	C	T	60	5	-	Missense	A864V
2205	RBP3	4	10	48001988	48001988	G	A	100	4	4	Missense	L1223F
2205	SIM2	10	21	37039088	37039088	G	A	44	19	10	Missense	V453I
2209	AARS2	11	6	44380814	44380814	C	G	50	7	5	Missense	D512H
2209	ABCG8	4	2	43932394	43932394	C	T	50	9	2	Nonsense	R164*
2209	IPO9	5	1	200087888	200087888	C	A	60	5	9	Missense	P183Q
2209	LRRC8E	2	19	7870859	7870859	C	A	55	11	8	Missense	D484E
2209	OR4M2	1	15	19870180	19870180	-	TAAA	53	20	16	Frameshift ins	M81fs
2209	PDHX	11	11	34973068	34973068	G	A	37	8	6	Missense	E427K
2209	PES1	11	22	29306109	29306109	T	A	67	12	5	Missense	D367V
2209	PUSL1	7	1	1236201	1236201	C	T	100	6	6	Missense	A250V
2209	RNF113B	1	13	97627389	97627389	T	G	75	4	-	Missense	K35Q
2209	RTN1	3	14	59264032	59264032	G	A	75	4	2	Missense	R375W
2209	SMEK1	14	14	90997657	90997657	C	T	50	6	3	Missense	G725R
2209	TMEM180	3	10	104220467	104220467	G	A	50	8	-	Missense	G103S
2210	C2orf3	3	2	75782889	75782889	C	T	44	9	7	Missense	R188K
2210	C5orf15	2	5	133323316	133323316	T	C	64	11	4	Missense	N145S
2210	C9orf43	9	9	115227487	115227487	-	GCA	44	10	11	In frame ins	Q303+
2210	CHD1	28	5	98234400	98234400	G	A	67	6	5	Missense	P1290L

2210	CLK1	9	2	201428051	201428051	A	G	67	6	-	Missense	L354P
2210	DYNC1LI1	2	3	32586842	32586842	-	ins 18bp	78	10	9	In frame ins	G74+
2210	DYSF	30	2	71654858	71654858	-	CGGAGG	60	5	8	In frame ins	G1066+
2210	EVPL	22	17	71517924	71517924	C	T	75	4	7	Missense	R986Q
2210	ARHGEF40	3	14	20612981	20612981	C	T	71	7	3	Missense	P418S
2210	GLIS1	7	1	53747441	53747441	C	T	60	6	7	Missense	G549S
2210	KBTBD6	1	13	40604613	40604613	C	T	62	8	7	Missense	R12H
2210	PDE4DIP	3	1	143727234	143727234	G	T	21	14	12	Missense	L142I
2210	PIH1D1	4	19	54643069	54643069	C	A	100	4	2	Splice site	Q132fs
2210	PLEKHO1	6	1	148398233	148398233	A	G	80	6	3	Missense	Q374R
2210	SLC2A4	5	17	7128023	7128023	C	T	50	6	6	Missense	T152I
2210	SLFN11	3	17	30704621	30704621	C	T	50	14	10	Missense	R590H
2210	TAF6	14	7	99542828	99542828	C	T	100	5	4	Missense	G671S
2210	TDRD6	1	6	46768854	46768854	T	C	64	12	-	Missense	I1677T
2210	TEX15	1	8	30824989	30824989	T	C	50	15	5	Missense	M363V
2210	USP21	6	1	159399997	159399997	A	G	37	9	5	Missense	K360E
2210	USP31	16	16	22988154	22988154	G	C	62	9	8	Missense	P925A

In CCDS Reference Sequence

^ Numbering according to the Human Genome hg18 assembly

* Unclear interpretation on Sanger sequencing

Abbreviations: Refnt, reference nucleotide; Varnt, variant nucleotide; CovT, coverage tumor; CovN, coverage normal; %, fraction of sequence reads showing the variant nucleotide; AA, amino acid; fs, frameshift; +, insertion

Supplementary Table 5. Segments (regions) of tumor-acquired copy number alterations identified in the DLBCL discovery panel

Sample ID	Chr	Cytoband	Start^	End^	Copy number	Segment size (Kb)	N of genes	First 10 genes in segment	N of miRs
2112	1	1p21.3	97436118	97610158	2.83	174.040	1	DPYD	0
2112	1	1q42.12	224371678	225060761	1.42	689.083	6	ACBD3 MIXL1 LIN9 PARP1 C1orf95 ITPKB	0
2112	2	2q11.2	97697953	97890694	1.37	192.741	2	ZAP70 TMEM131	0
2112	3	3q26.32	178281694	178722001	1.48	440.307	1	TBL1XR1	0
2112	3	3q27.3 - 3q28	188988884	190112563	1.40	1123.679	2	LPP FLJ42393	1
2112	6	6p25.3 - 6q12	94649	65204103	3.41	65109.454	658	DUSP22 IRF4 EXOC2 HUS1B LOC285768 FOXQ1 FOXF2 FOXC1 GMDS C6orf195	8
2112	9	9p24.3 - 9p21.3	30910	21966858	2.60	21935.948	95	FOXD4 CBWD1 LOC642313 C9orf66 DOCK8 KANK1 DMRT1 DMRT3 DMRT2 SMARCA2	3
2112	9	9p21.3	21967182	22001380	0.91	34.198	3	CDKN2A CDKN2BAS CDKN2B	0
2112	9	9p21.3 - 9q34.3	22001477	140211203	2.63	118209.726	772	CDKN2BAS DMRTA1 ELAVL2 TUSC1 C9orf82 PLAA IFT74 LRRC19 TEK NCRNA00032	20
2112	12	12q24.33	131746103	132209971	1.53	463.868	10	POLE PXMP2 PGAM5 ANKLE2 GOLGA3 CHFR ZNF605 ZNF26 ZNF84 ZNF140	0
2112	14	14q11.1 - 14q32.33	18072112	106356482	2.66	88284.370	728	OR11H12 POTEG P704P OR4Q3 OR4M1 OR4N2 OR4K2 OR4K5 OR4K1 OR4K15	62
2112	15	15q22.2	58105293	58137917	0.84	32.624	0		0
2112	18	18q12.3	37296957	37315890	0.93	18.933	1	KC6	0
2112	19	19p13.3	41898	1683588	3.12	1641.690	63	OR4F17 FLJ45445 PPAP2C MIER2 THEG C2CD4C SHC2 ODF3L2 MADCAM1 C19orf20	0
2203	1	1q21.2;q23.2	148116963	158679606	2.65	10562.643	312	HIST2H2BE HIST2H2AC HIST2H2AB BOLA1 SV2A SF3B4 MTMR11 OTUD7B VPS45 PLEKHO1	6
2203	1	1q23.2;q23.3	158679714	159170764	0.72	491.050	8	SLAMF6 CD84 SLAMF1 CD48 SLAMF7 LY9 CD244 ITLN1	0
2203	1	1q23.3 - 1q41	159176740	212727697	2.57	53550.958	416	ITLN2 F11R TSTD1 USF1 ARHGAP30 PVRL4 KLHDC9 PFDN2 NIT1 DEDD	16
2203	2	2p14	65115199	68866670	1.49	3751.471	16	CEP68 RAB1A ACTR2 SPRED2 MEIS1 ETAA1 C1D WDR92 PNO1 PPP3R1	0
2203	3	3p21.31	48098697	49510404	1.37	1411.707	51	MAP4 CDC25A CAMP ZNF589 NME6 SPINK8 FBXW12 PLXNB1 CCDC51 CCDC72	2
2203	3	3p14.2	60959632	61067664	1.30	108.032	1	FHIT	0
2203	3	3q13.31 - q13.33	115205834	123389343	1.31	8183.509	46	KIAA1407 QTRTD1 DRD3 ZNF80 TIGIT MIR568 ZBTB20 GAP43 LSAMP LOC285194	2
2203	6	6p21.33	31350674	31797988	1.41	447.314	31	HLA-B MICA HCP5 HCG26 MICB MCCD1 BAT1 SNORD117 SNORD84 ATP6V1G2	0
2203	6	6q22.32 - 6q24.1	126956390	142041811	1.31	15085.421	79	RSPO3 RNF146 ECHDC1 KIAA0408 C6orf174 C6orf58 THEMIS PTPRK LAMA2 ARHGAP18	1
2203	6	6q26	161463399	161529006	1.33	65.607	2	AGPAT4 NCRNA00241	0
2203	7	7q31.1	110569932	110591894	1.16	21.962	1	IMMP2L	0
2203	7	7q31.2 - q33	115342402	132932923	1.30	17590.521	105	TFEC TES CAV2 CAV1 MET CAPZA2 ST7OT1 ST7 ST7OT4 ST7OT2	9
2203	10	10p13 - 10p11.21	13693396	39116215	1.35	25422.819	116	PRPF18 FRMD4A MIR1265 FAM107B CDNF HSPA14 SUV39H2 DCLRE1C MEIG1 OLAH	7
2203	10	10q25.2	113778727	113806357	1.27	27.630	0		0
2203	11	11q13.1 - 11q13.4	66010921	71284357	1.35	5273.436	85	DPP3 BBS1 ZDHHC24 ACTN3 CTSF CCDC87 CCS RBM14 RBM4 RBM4B	1
2203	14	14q24.3;q31.1	78235108	78511599	1.24	276.491	1	NRXN3	0
2203	17	17q23.3	59663266	59835377	3.21	172.111	2	TEX2 PECAM1	0
2203	19	19p13.3	4293341	4401376	1.36	108.035	4	MPND SH3GL1 CHAF1A UBXN6	0
2203	19	19p13.12	15241237	15261430	1.19	20.193	1	BRD4	0
2203	19	19p13.12;p13.11	15991604	16144389	1.47	152.785	6	LOC126536 FLJ25328 TPM4 RAB8A HSH2D CIB3	0
2203	X	Xq28	152270993	154835536	2.54	2564.543	76	ZNF275 ZFP92 TREX2 HAUS7 BGN ATP2B3 FAM58A DUSP9 PNCK SLC6A8	1
2204	1	1p13.1	116883520	116981713	1.03	98.193	2	CD58 IGSF3	0
2204	1	1q22 - 1q23.3	154291720	161645271	2.41	7353.551	162	ROBLD3 RAB25 MEX3A LMNA SEMA4A SLC25A44 PMF1 BGLAP PAQR6 SMG5	3
2204	1	1q23.3 - 1q41	161645856	220481486	3.58	58835.630	394	PBX1 LMX1A RXRG LOC400794 LRRC52 MGST3 ALDH9A1 TMCO1 UCK2 FAM78B	18
2204	1	1q41	220496286	221283671	1.29	787.385	7	HHIPL2 TAF1A MIA3 AIDA C1orf58 FAM177B DISP1	0
2204	2	2p23.3	23995166	26531052	1.39	2535.886	30	ATAD2B UBXN2A MFSD2B C2orf44 FKBP1B SF3B14 TP53I3 PFN4 LOC375190 C2orf84	1
2204	2	2q11.2	96895331	100237267	1.35	3341.936	27	SEMA4C FAM178B FAHD2B ANKRD36 ANKRD36B COX5B ACTR1B ZAP70 TMEM131 VWA3B	0
2204	2	2q11.2	101427700	101812077	1.35	384.377	2	RFX8 MAP4K4	0
2204	3	3p26.3 - 3p22.1	35333	42177166	2.31	42141.833	209	CHL1 CNTN6 CNTN4 IL5RA TRNT1 CRBN LRRN1 SETMAR SUMF1 ITPR1	4
2204	3	3p21.31 - 3p11.1	50603537	90026778	2.69	39423.241	153	CISH MAPKAPK3 DOCK3 MANF RBM15B VPRBP RAD54L2 TEX264 GRM2 IQCF6	3
2204	4	4p16.3	2269	1599701	1.37	1597.432	29	ZNF595 ZNF718 ZNF876P ZNF732 ZNF141 ABCA11P ZNF721 PIGG PDE6B ATP5I	1
2204	4	4p11 - 4q12	48777028	52423397	1.42	3646.369	1	DCUN1D4	0
2204	4	4q32.1 - 4q32.3	156566057	166919176	1.27	10353.119	36	GUCY1A3 GUCY1B3 ACCN5 TDO2 CTSO PDGFC GLRB GRIA2 LOC340017 FAM198B	2
2204	5	5p15.33 - 5q15	68520	93937512	2.85	93868.992	352	PLEKHG4B LRRC14B CCDC127 SDHA PDCD6 AHRR LOC100310782 C5orf55 EXOC3 LOC25845	10
2204	5	5q23.2 - 5q35.3	123706589	180002458	2.32	56295.869	484	ZNF608 GRAMD3 ALDH7A1 PHAX C5orf48 LMNB1 MARCH3 FLJ44606 MEGF10 PRRC1	17
2204	6	6p25.1 - 6p21.32	5348925	33520286	2.36	28171.361	382	FARS2 NRN1 F13A1 LOC285780 LY86 RREB1 SSR1 CAGE1 RIOK1 DSP	4
2204	6	6p21.32 - 6q27	33520753	170892918	1.82	137372.165	698	SYNGAP1 ZBTB9 BAK1 GGNBP1 C6orf227 ITPR3 C6orf125 IP6K3 LEMD2 MLN	13
2204	7	7q31.2 - 7q32.1	114488159	127932090	1.29	13443.931	60	TFEC TES CAV2 CAV1 MET CAPZA2 ST7OT1 ST7 ST7OT4 ST7OT2	3
2204	7	7q33 - 7q36.2	134215887	153251193	1.40	19035.306	176	CALD1 AGBL3 TMEM140 C7orf49 WDR91 STRA8 CNOT4 NUP205 PL-5283 SLC13A4	4
2204	8	8q23.1 - 8q24.12	109970014	119848365	1.26	9878.351	18	TRHR NUDCD1 ENY2 PKHD1L1 EBAG9 SYBU KCNV1 CSMD3 MIR2053 TRPS1	1
2204	8	8q24.12 - 8q24.13	122256149	126953719	1.32	4697.570	28	HAS2 HAS2AS ZHX2 DERL1 WDR67 FAM83A LOC100131726 C8orf76 ZHX1 ATAD2	1
2204	8	8q24.13 - 8q24.21	127027548	129424604	2.42	2397.056	10	FAM84B POU5F1B LOC727677 MYC PVT1 MIR1204 MIR1205 MIR1206 MIR1207 MIR1208	5
2204	8	8q24.21 - 8q24.22	130109887	135486109	0.82	5376.222	18	GSDMC FAM49B ASAP1 ASAP1IT1 ADCY8 EFR3A OC90 HHLA1 KCNQ3 HPYR1	0
2204	8	8q24.22 - 8q24.23	135487437	139168319	1.26	3680.882	6	ZFAT ZFATAS MIR30B MIR30D LOC286094 KHDRBS3	2
2204	10	10q23.33 - 10q26.3	94801170	135356682	2.92	40555.512	325	EXOC6 CYP26C1 CYP26A1 MYOF CEP55 GPR120 RBP4 PDE6C FRA10AC1 LGI1	10

2204	13	13q11 - 13q34	17924937	114126487	2.87	96201.550	364	LOC284232 LOC348021 DKFZp686A1627 TUBA3C LOC100101938 TPTE2 MPHOSPH8 PSPC1 ZMYM5 ZMYM:	14
2204	17	17p13.3 - 17p11.2	514	18065802	1.35	18065.288	318	RPH3AL C17orf97 FAM101B VPS53 FAM57A GEMIN4 ELP2P GLOD4 RNMTL1 NXN	11
2204	17	17p11.2	18065982	18384390	2.54	318.408	11	LLGL1 FLII SMCR7 TOP3A SMCR8 SHMT1 EVPLL LOC339240 LGALS9C LOC220594	0
2204	17	17q23.3 - 17q25.3	59342527	78643088	2.41	19300.561	266	GH1 CD79B SCN4A C17orf72 ICAM2 ERN1 SNORD104 SNORA76 TEX2 PECAM1	7
2204	18	18p11.32 - 18q23	1543	76116029	2.72	76114.486	292	USP14 THOC1 COLEC12 CETN1 CLUL1 C18orf56 TYMS ENOSF1 YES1 ADCYAP1	9
2204	21	21p11.2 - 21p11.1	9758730	10197771	1.18	439.041	6	TPTE BAGE5 BAGE3 BAGE4 BAGE2 BAGE	0
2204	Y	Yq11.21 - Yq11.23	14079303	57728335	1.15	43649.032	64	UTY TMSB4Y VCY1B VCY NLGN4Y FAM41AY2 FAM41AY1 NCRNA00230B NCRNA00230A XKRY	0
2205	1	1p36.33 - 1p35.2	51586	33172955	1.56	33121.369	493	OR4F5 LOC100132062 LOC100133331 LOC100132287 OR4F3 OR4F29 OR4F16 NCRNA00115 LOC643837	9
2205	1	1q21.3 - 1q25.1	152859696	172135734	1.71	19276.038	305	ADAR KCNN3 PMVK PBXIP1 PYGO2 SHC1 CKS1B FLAD1 LENEP ZBTB7B	11
2205	1	1q25.2	175729961	175762917	1.33	32.956	0		0
2205	1	1q31.1 - 1q44	185283605	247190999	2.48	61907.394	436	FAM5C RGS18 RGS21 RGS1 RGS13 RGS2 UCHL5 TROVE2 GLRX2 CDC73	14
2205	2	2p11.1;q11.1	91638457	94691602	0.88	3053.145	0	(centromeric region)	0
2205	2	2q21.3 - 2q24.3	136652914	168352732	1.36	31699.818	80		1
2205	3	3p12.3	77638126	77695070	1.40	56.944	1	ROBO2	0
2205	3	3q27.3;q28	187955510	189858646	1.44	1903.136	18	EIF4A2 SNORD2 MIR1248 SNORA81 SNORA63 SNORA4 RFC4 ADIPOQ ST6GAL1 RPL39L	1
2205	4	4q13.1	65863932	65938754	1.40	74.822	1	EPHA5	0
2205	6	6q23.2;q23.3	135174467	137363938	1.45	2189.471	15	ALDH8A1 HBS1L MYB AHI1 MIR548H4 C6orf217 PDE7B FAM54A BCLAF1 MAP7	1
2205	6	6q23.3	137367796	138585301	0.95	1217.505	8	IL20RA IL22RA2 IFNGR1 OLIG3 TNFAIP3 PERP KIAA1244 PBOV1	0
2205	6	6q23.3 - q25.2	138585800	155045608	1.45	16459.808	75	KIAA1244 HEBP2 NHSL1 CCDC28A ECT2L REPS1 C6orf115 HECA TXLNB CITED2	0
2205	13	13q21.31	61299128	61322787	1.08	23.659	0		0
2205	19	19p12 - 19q12	24167277	32657343	1.29	8490.066	0		0
2209	6	6q13 - 6q25.3	73505939	159545634	1.36	86039.695	377	KCNQ5 KHDC1L KHDC1 C6orf147 DPPA5 C6orf221 OOEP DDX43 C6orf150 MTO1	6
2210	1	1q21.1 - 1q31.2	142693888	191475047	2.16	48781.159	639	PDE4DIP	26
2210	1	1q31.2	191475560	191488526	4.58	12.966	1	CDC73	0
2210	1	1q31.2 - 1q44	191491706	247190999	2.23	55699.293	1273	PTBP2 DPYD MIR137 SNX7 LPPR5 LOC100129620 LPPR4 PALMD FRRS1 AGL	62
2210	3	3p26.3 - 3q29	35333	199380503	2.27	199345.170	1189	CHL1 CNTN6 CNTN4 IL5RA TRNT1 CRBN LRRN1 SETMAR SUMF1 ITPR1	67
2210	5	5p15.33 - 5q35.3	68520	180722914	2.27	180654.394	981	CCDC125 TAF9 RAD17 PLEKHG4B MARVELD2 LRRC14B CCDC127 SDHA OCLN PDCD6	62
2210	6	6p25.3 - 6p11.2	94649	57758722	2.27	57664.073	673	TRIM27 ZNF311 DUSP22 OR2W1 LOC100129636 LOC401242 OR2B3 OR2J3 IRF4 OR2J2	22
2210	6	6p11.2	57759523	58179324	1.52	419.801	0		0
2210	6	6p11.2 - 6q11.1	58190043	62067209	2.30	3877.166	1	GUSBP4	0
2210	6	6q11.1 - 6q14.1	62067505	76120636	1.56	14053.131	36	KHDRBS2 LGSN PTP4A1 PHF3 EYS MCART3P BAI3 LMBRD1 COL19A1 COL9A1	2
2210	6	6q14.1 - q616.1	76121163	92554936	2.27	16433.773	64	FILIP1 SENP6 MYO6 IMPG1 HTR1B IRAK1BP1 PHIP HMGN3 LCA5 SH3BGRL2	3
2210	6	6q16.1 - 6q16.3	92556135	103840867	1.54	11284.732	23	EPHA7 TSG1 MANEA FUT9 KIAA0776 FHL5 GPR63 NDUFAF4 KLHL32 C6orf167	1
2210	6	6q16.3	103841326	103868730	2.25	27.404	0		0
2210	6	6q16.3 - 6q21	103868742	111038725	1.58	7169.983	41	HACE1 LIN28B BVES POPDC3 C6orf112 PREP PRDM1 ATG5 AIM1 RTN4IP1	1
2210	6	6q21	111041147	111687254	2.31	646.107	6	CDK19 AMD1 GTF3C6 RPF2 GSTM2P1 SLC16A10	0
2210	6	6q21 - 6q24.2	111690279	143202785	1.73	31512.506	148	KIAA1919 REV3L TRAF3IP2 FYN WISP3 TUBE1 C6orf225 LAMA4 RFPL4B MARCKS	3
2210	6	6q24.2 - 6q25.1	143204925	152424193	2.27	9219.268	52	AIG1 ADAT2 PEX3 FUCA2 LOC285740 PHACTR2 LTV1 C6orf94 PLAGL1 HYMAI	0
2210	6	6q25.1 - 6q25.3	152426493	157695987	1.58	5269.494	16	ESR1 SYNE1 MYCT1 VIP FBXO5 MTRF1L RGS17 OPRM1 IPCEF1 CNKSR3	3
2210	6	6q26 - 6q27	161247281	165933532	1.53	4686.251	9	MAP3K4 AGPAT4 NCRNA00241 PARK2 PACRG LOC285796 QKI C6orf118 PDE10A	0
2210	6	6q27	165933750	170892918	2.15	4959.168	40	PDE10A C6orf176 LOC441177 T PRR18 SFT2D1 BRP44L RPS6KA2 MIR1913 RNASET2	3
2210	7	7p22.3 - 7q31.1	52899	110759081	2.27	110706.182	719	FAM20C PDGFA FLJ44511 PRKAR1B HEATR2 SUN1 GET4 ADAP1 COX19 CYP2W1	45
2210	8	8q13.3 - 8q24.3	72955205	146268947	2.27	73313.742	372	TRPA1 LOC100132891 LOC392232 KCNB2 TERF1 C8orf84 LOC100130301 RPL7 RDH10 STAU2	37
2210	13	13q11 - 13q34	17924937	114126487	2.25	96201.550	384	LOC284232 LOC348021 DKFZp686A1627 TUBA3C LOC100101938 TPTE2 MPHOSPH8 PSPC1 ZMYM5 ZMYM:	31
2210	15	15q26.3	98242463	100217560	1.52	1975.097	15	ADAMTS17 FLJ42289 LASS3 LINS1 ASB7 ALDH1A3 LRRK1 CHSY1 SELS SNRPA1	0
2210	19	19q13.42	60859780	61226466	1.58	366.686	9	U2AF2 EPN1 NLRP9 RFPL4A NLRP11 NLRP4 NLRP13 NLRP8 NLRP5	0
2210	X	Xp22.33 - Xq28	108465	154632859	2.68	154524.394	944	PLCXD1 PPP2R3B SHOX CRLF2 CSF2RA IL3RA SLC25A6 NCRNA00105 ASMTL P2RY8	76

^ Numbering according to the Human Genome hg18 assembly

Note: segments highlighted in grey represent subclonal lesions, as confirmed by FISH analysis in selected chromosomes (e.g., trisomy of chr. 3, observed in 10% of the nuclei). These segments were not computed in the calculation of the total load of genetic lesions presented in Figure 3

Supplementary Table 6. Enriched functional categories in DLBCL altered genes (DAVID analysis)

Category	Enrichment Score	P-value	Benjamini
Regulation of transcription	5.336	1.05E-08	2.45E-05
Lymphocyte activation/differentiation	2.835	5.23E-07	6.08E-04
Regulation of T cell activation/differentiation	2.288	5.91E-06	0.002
Cyclin-dependent protein kinase regulator activity	1.143	3.41E-05	0.008
Chromatin modification	3.354	8.26E-05	0.008
T cell proliferation/activation	1.675	1.02E-04	0.009
Nuclear component	2.466	2.41E-05	0.009
Myeloid dendritic cell activation/differentiation	1.919	1.15E-04	0.010
DNA methylation	2.493	2.61E-04	0.049
Cell adhesion	2.218	0.002	0.063
Cell cycle checkpoint	1.576	0.002	0.089
Antigen processing and presentation	1.809	7.48E-04	0.092
Regulation of cell projection	0.903	0.003	0.096
B cell activation/differentiation	1.449	0.003	0.107
Protein binding	1.756	0.001	0.110
Regulation of apoptosis	1.481	0.004	0.125
GTPase binding	2.230	0.003	0.159
Plasma membrane component	1.134	0.004	0.179
Regulation of cell growth	0.963	0.010	0.228
Regulation of protein kinase activity	1.171	0.011	0.248
Tube development	1.348	0.012	0.258
Response to radiation	0.726	0.013	0.258
MHC protein binding	0.658	0.008	0.265
Regulation of immune response	1.068	0.013	0.267
Regulation of cAMP biosynthetic process	0.803	0.020	0.350
Tumor necrosis factor binding	1.125	0.017	0.351
Nucleotide binding	1.623	0.012	0.355
JNK MAP cascade	1.009	0.025	0.402
Phosphatase activity	1.040	0.028	0.429
Histone deacetylase activity	0.997	0.016	0.430
RNA localization	0.888	0.030	0.440
Peptide transport/secretion	1.074	0.030	0.440
Response to hypoxia	0.258	0.030	0.443
Ear development	0.854	0.034	0.456
Nuclear/Mitochondrial envelope	0.292	0.015	0.478
Sensory perception	0.592	0.037	0.479
Cell morphogenesis	0.828	0.041	0.499
Regulation of DNA binding	1.022	0.042	0.503
Regulation of DNA binding	0.645	0.042	0.503
Cell cycle	0.479	0.044	0.505
Lymphocyte proliferation	0.993	0.046	0.519
Transforming growth factor beta binding	1.005	0.039	0.522
Mitochondrial transport	0.658	0.047	0.528
Ion channel activity	0.690	0.042	0.531
Ion binding	1.020	0.046	0.549

Supplementary Table 7. Significant regions of deletion in DLBCL according to the GISTIC algorithm

Cytoband	q value	Residual q value	Wide peak boundaries	Total N of genes	Gene names (first 20 genes in region)
1p36.31	2.10E-05	0.00015508	chr1:5736742-6359635	9	RPL22 KCNAB2 ACOT7 ICMT CHD5 NPHP4 GPR153 RNF207 HES3
1p13.1	7.45E-08	7.45E-08	chr1:116878071-116919472	2	CD58 IGSF3
1q42.12	0.013866	0.041044	chr1:224370620-225066589	6	PARP1 ITPKB ACBD3 MIXL1 LIN9 C1orf95
2q22.3	0.079791	0.079791	chr2:142281763-148092099	5	KYNU ZEB2 LRP1B ARHGAP15 GTDC1
3p14.2	1.03E-10	1.03E-10	chr3:60416800-60452452	1	FHIT
4p15.1	0.00068122	0.00068122	chr4:29660618-31420982	1	PCDH7
4q34.1	0.00097534	0.057584	chr4:172327318-172866785	1	GALNT17
6q21	1.24E-13	1.24E-13	chr6:106420940-110855391	34	hsa-mir-587 AIM1 PRDM1 FOXO3 GPR6 SMPD2 NR2E1 DDO SNX3 CD164 WASF1 ATG5 ZBTB24 FIG4 SEC63 SESN1 OSTM1 C6orf203 CDC40 SOBP
6q23.3	2.34E-13	0.030288	chr6:137447314-138401397	4	TNFAIP3 OLIG3 IL22RA IFNGR1
7q32.1	0.090495	0.090495	chr7:127206505-129956043	39	hsa-mir-335 hsa-mir-183 hsa-mir-129-1 hsa-mir-593 OPN1SW CALU CPA1 CPA2 FLNC IMPDH1 IRF5 LEP MEST NRF1 SMO UBE2H ATP6V1F KIAA0265 AHCYL2 TNPO3
8p23.3	9.12E-08	9.12E-08	chr8:591044-2318161	7	hsa-mir-596 CLN8 MYOM2 DLGAP2 ARHGEF10 KBTBD11 ERICH1
9p21.3	1.19E-22	1.19E-22	chr9:21966859-21986983	1	CDKN2A
10q24.32	0.14377	0.14377	chr10:102841044-114442753	73	hsa-mir-609 hsa-mir-936 hsa-mir-146b ADD3 ADRA2A ARL3 COL17A1 CYP17A1 DUSP5 FGF8 TLX1 MXI1 NFKB2 PITX3 PSD FBXW4 TAF5 TECTB XPNPEP1 SHOC2
12q24.33	0.22126	0.22126	chr12:131799881-131902313	3	GOLGA3 ANKLE2 PGAM5
13q14.3	0.0044568	0.0044568	chr13:48174250-49867923	15	hsa-mir-15a MLNR KPNA3 TRIM13 FNDC3A PHF11 RCBTB1 CYSLTR2 C13orf1 CDADC1 CAB39L SETDB2 EBPL ARL11 KCNRG
14q32.32	0.18402	0.18402	chr14:102081628-102445346	2	TRAF3 RCOR1
15q21.1	1.50E-16	1.5E-16	chr15:42769393-42798306	1	B2M
16q12.2	0.038012	0.038012	chr16:52334966-55602461	32	hsa-mir-138-2 AMFR BBS2 CES1 CETP GNAO1 MMP2 MT1A MT1B MT1E MT1F MT1G MT1H MT1M MT1X MT2A MT3 SLC6A2 SLC12A3 NUP93
17p13.1	3.76E-09	3.76E-09	chr17:7295122-7524765	17	ATP1B2 CD68 CHRNB1 EIF4A1 POLR2A SHBG SOX15 TP53 TNFSF13 TNFSF12 FXR2 MPDU1 SENP3 ZBTB4 SAT2 TNFSF12-TNFSF13 AMAC1L3
18q22.3	0.013866	0.013866	chr18:65346422-69222105	6	SOCS6 CD226 RTTN NETO1 CBLN2 DOK6
19p13.3	0.00068122	0.00068122	chr19:6470458-6533436	1	TNFSF9

Supplementary Table 8. Significant regions of amplification in DLBCL according to the GISTIC algorithm

Cytoband	q value	Residual q value	Wide peak boundaries	Total N of genes	Gene names (first 20 genes in region)
1q25.1	1.94E-05	1.94E-05	chr1:172017518-175588211	17	hsa-mir-488 ASTN1 SERPINC1 TNR GPR52 RABGAP1L CACYBP KLHL20 DARS2 FAM5B PAPPA2 TNN MRPS14 RFWD2 ZBTB37 CENPL RC3H1
2p16.1	2.38E-09	2.38E-09	chr2:60854354-61001391	2	REL PAPOLG
3q12.3	0.0027524	0.0027524	chr3:102748375-126619117	126	hsa-mir-198 hsa-mir-568 hsa-mir-567 ADCY5 ADPRH ALCAM ATP6V1A CASR CBLB CD80 CD86 CD47 CSTA DRD3 GAP43 GOLGB1 GSK3B GTF2E1 HCLS1 HGD
5p13.2	0.084633	0.084633	chr5:34945647-39658596	28	hsa-mir-580 C9 DAB2 FYB GDNF IL7R LIFR PRLR RAD1 SKP2 SLC1A3 OSMR NUP155 NIPBL WDR70 BXDC2 AGXT2 C5orf42 SPEF2 LMBRD2
6p21.32	0.01695	0.01695	chr6:31641586-32562240	58	AGER AIF1 CFB C2 C4A C4B CLIC1 CREBL1 CSNK2B CYP21A2 DOM3Z HLA-DRA HSPA1A HSPA1B HSPA1L LTA LTB MSH5 NEU1 NOTCH4
7q21.13	0.0015045	0.0015045	chr7:83319325-99776465	113	hsa-mir-106b hsa-mir-591 hsa-mir-489 ASNS AZGP1 CALCR KRIT1 CDK6 COL1A2 CYP3A7 CYP3A4 CYP3A5 CYP51A1 DLX5 DLX6 DYNC1I1 GNG11 GNGT1 GRM3 MCM7
8q24.21	0.0094423	0.0094423	chr8:127021989-128834542	2	MYC FAM84B
9p24.1	0.00017135	0.00017135	chr9:5486919-5566409	1	PDCD1LG2
9p13.2	0.0012577	0.057507	chr9:35988727-37238610	9	CCIN CLTA PAX5 RECK MELK GNE ZCCHC7 RNF38 C9orf19
11q23.3	0.0012577	0.0012577	chr11:116789150-117413537	6	FXYD2 IL10RA CEP164 FXYD6 DSCAML1 TMPRSS13
12q14.1	0.002096	0.002096	chr12:58692629-68346744	45	hsa-mir-548c hsa-let-7i AVPR1A CPM GNS IFNG LYZ MDM2 RAP1B YEATS4 HMGA2 DYRK2 USP15 TMEM5 CCT2 FRS2 CPSF6 WIF1 IRAK3 XPOT
13q32.3	0.22085	0.22085	chr13:90219486-99997053	31	hsa-mir-623 hsa-mir-92a-1 DCT EBI2 GPC5 GPR18 RANBP5 PCCA DNAJC3 RAP2A SLC15A1 ZIC2 STK24 CLDN10 TM9SF2 GPC6 MBNL2 FARP1 ABCC4 SOX21
18q21.2	1.55E-08	0.12602	chr18:50839879-51607473	1	TCF4
18q21.33	2.44E-09	2.44E-09	chr18:58944230-59229769	3	BCL2 KDSR VPS4B
19q13.33	0.15813	0.15813	chr19:54014687-62347297	267	hsa-mir-935 hsa-mir-373 hsa-mir-519a-2 hsa-mir-643 hsa-mir-125a hsa-mir-150 AP2A1 KLK3 BAX CD33 SIGLEC6 CD37 CGB ETFB FCAR FCGRT FLT3LG FPR1 FPRL1 FPRL2

Supplementary Table 9. Somatic mutations in recurrently mutated WES and MCR genes

This table does not include previously identified genes as well as *MLL2*, *B2M*, *CD58* and *MEF2B*, which are reported separately

Gene Symbol	Sample ID	Subtype	Exon	Nucleotide change*	AA change*	PolyPhen-2 prediction (score)	Affected functional domain	mRNA RefSeq
ADAMTSL3	2134	GCB	E13	T1519C	V432A	Benign (1.261)	TSP type-1 2	NM_207517.2
	2203^	ABC	E16	G2119A	R632Q	Benign (0.001)	inter TSP type 4 and 5	
	2032	GCB	E21	A3834G	I1204V	Benign (0.971)	none	
	2032	GCB	E21	G3603A	E1127K	Possibly damaging (1.671)	none	
AKAP8	2110	ABC	E4	C267G	A69G	Benign (0.168)	none	NM_005858.2
	2027	GCB	E14	C1890T	T610M	Benign (0.634)	none	
	2209^	ABC	E14	ΔAT(1965-1966)	H635fs	Damaging	none	
C12orf35	2072	ABC/NC	E4	T2360G	F649C	Probably damaging (0.914)	none	NM_018169.3
	2026	ABC	E4	A3237C	K941N	Benign (0.02)	none	
	2026	ABC	E4	A3259C	N949H	Possibly damaging (0.553)	none	
	2203^	ABC	E4	C4030T	Q1206X	Damaging	none	
CCND3	2163^	ABC	E5	C1015T	P284S	Possibly damaging (1.8)	none	NM_001760.3
	2075	GCB	E5	C943T	Q260X	Damaging	none	
	2209^	ABC	E6	T1024A	I290K	Benign (0.028)	none	
CD36	2016	ABC	E7	T778C	F135L	Benign (0.5)	Extracellular domain	NM_001001547.2
	2158	GCB	E7	T778C	F135L	Benign (0.5)	Extracellular domain	
	2203^	ABC	E8	C965A	T197K	Benign (0.005)	Extracellular domain	
	2126	GCB	E8	G943T	V190F	Possibly damaging (0.480)	Extracellular domain	
	2043	ABC	E11	A1154G	K260R	Probably damaging (0.997)	Extracellular domain	
	2012	GCB	E12	T1364C	I330T	Possibly damaging (1.887)	Extracellular domain	
	2032	GCB	E12	A1320C	E315D	Benign (0.274)	Extracellular domain	
DCHS1	2091	ABC	E2	G1734A	V442I	Probably damaging (0.994)	Cadherin 4 domain	NM_003737.2
	2205^	ABC	E3	C2386T	A659V	Probably damaging (0.999)	Cadherin 6 domain	
	2097	ABC	E20	C8841T	P2811S	Probably damaging (0.999)	Cadherin 26 domain	
DSC3	2020	GCB	E7	A1017C	T288P	Possibly damaging (1.623)	Extracellular, Cadherin 2	NM_001941.3
	2205^	ABC	E11	T1788C	Y545H	Possibly damaging (0.443)	Extracellular, Cadherin 4	
DUSP27	2102	GCB	E5	A2614A	K872E	Benign (1.272)	none	NM_001080426.1
	2033	ABC	E5	A1969G	I657V	Probably damaging (0.965)	Ser-rich	
	2112^	NC	E5	G2529A	M843I	Benign (0.377)	Ser-rich	
DPYD	2060	ABC	E10	A1060T	D308V	Possibly damaging (0.22)	none	NM_000110.3
	2205^	ABC	E20	C2452T	P772L	Probably damaging (1)	none	
HNF1B	2060	ABC	E1	G294T	V25L	Possibly damaging (0.555)	Dimerization	NM_000458.2
	2016	ABC	E1	T403G	V61G	Benign (0.000)	none	
	2203^	ABC	E2	C586A	A122D	Probably damaging (0.999)	none	
	2027	GCB	E3	G879C	D220H	Benign (0.033)	none	
	2016	ABC	E7	T1594G	V458G	Probably damaging (0.961)	none	
KLF2	2108	ABC	E2	G482C	G133A	Benign#	Poly-Gly	NM_016270.2
	2108	ABC	E2	G568A	G162S	Benign#	Interaction with WWP1	
	2205^	ABC	E3	G1036A	E318K	Probably damaging (0.993)	Zinc finger C2H2-type 2	

MAGEC3	2025	ABC	E10c	T1757C	L586S	Probably damaging (0.988)	MAGE 2 (456 – 643)	NM_138702.1
	2157	GCB	E10b	T1570G	F524V	Benign (0.008)	MAGE 2 (456 – 643)	
	2205^	ABC	E10b	C1572A	F524L	Possibly damaging (0.562)	MAGE 2 (456 – 643)	
MED12L	2137^	GCB	E31	C4572T	R1512W	Probably damaging (0.995)	none	NM_053002.3
	2210	ABC	E15	ΔAGA (2274-2276)	ΔK746	Unknown	none	
MYOM2	2108	ABC	E12	C1504T	R456X	Damaging	Fibronectin type-III 1	NM_003970.2
	2110	ABC	E12	C1504T	R456X	Damaging	Fibronectin type-III 1	
	2057	GCB	E15	G1870T	G578W	Probably damaging (1)	Fibronectin type-III 2	
	2026	ABC	E29	C(-1)T	K1151fs	Damaging	Ig-like C2-type 4	
	2093	ABC	E17	G2216C	S693T	Probably damaging (0.997)	Fibronectin type-III 3	
	2044	ABC	E18	C2291T	T718M	Probably damaging (1)	Fibronectin type-III 4	
	2019	ABC	E36	G4166A	R1343H	Probably damaging (1)	none	
	2034	ABC	E37	C4343T	T1404I	Possibly damaging (0.707)	Ig-like C2-type 5	
	2203^	ABC	E2	G215A	R26Q	Probably damaging (0.997)	none	
MTMR8	2017	ABC	E14	G1789A	G574D	Benign (0.001)	none	NM_017677.2
	2203^	ABC	E9	G1161T	G365X	Damaging	Myotubularin phosphatase	
NOTCH1	2132	ABC	E34	ΔCT (7544-7545)	P2515fs	Damaging	PEST domain	NM_017617.2
	2100	GCB	E34	C7249T	Q2417X	Damaging	PEST domain	
PMS1	2093	ABC	E9	T2321A	L598I	Benign (0.072)	DNA binding, HMG box	NM_000534.4
	2209^	non-GCB	E2	C546T	A6V	Possibly damaging (0.484)	none	
TLL2	2075	GCB	E2	G488A	V88M	Benign (0.000)	Propeptide	NM_012465.2
	2137^	GCB	E4	G649T	K141N	Benign (0.001)	Propeptide	
	2205^	ABC	E14	G2048A	E608K	Probably damaging (0.999)	EGF-like 1; calcium-binding	
TRAF3	2147	GCB	E10	C1281T	R310X	Damaging	none	NM_145726.1
	2174	ABC	E5	C705T	R118W	Probably damaging (1)	TRAF like domain	
TSC22D1	2148	GCB	E5	C2776T	S762L	Probably damaging (0.794)	Gln-rich	NM_183422.2
	2203^	ABC	E5	G3204A	A905T	Possibly damaging (0.940)	none	
TMEM30A	2127	ABC	E3	C690T	R138X	Damaging	none	NM_018247.3
	2209^	non-GCB	E2	T590A	C104X	Damaging	none	
	2057	GCB	E7	A1266T	I330F	Probably damaging (0.999)	none	
ANKLE2	2034	ABC	E8	G1583A	V506I	Benign (0.001)	none	NM_015114.1
	2065	ABC	E11	C2139T	P691L	Benign (0.000)	none	
	2148	GCB	E11	C2139T	P691L	Benign (0.000)	none	
	2160	ABC/NC	E11	C2139T	P691L	Benign (0.000)	none	
KDM2B	2156^	GCB	E4	dupl CTGAA (429-432)	P121fs	Damaging	none	NM_032590.3
	2094	ABC/NC	E9	C1075G	N334K	Probably damaging (1.000)	JmjC domain	
	2027	GCB	E14	G2102C	E677Q	Benign (0.015)	Zinc finger PHD type	
	2137^	GCB	E23	C4010T	Q1313X	Damaging	Repeat LRR 7	

* Numbering according to the corresponding mRNA and protein RefSeq

^ For these samples, paired normal DNA was available and confirmed the somatic origin of the mutation

According to the SIFT prediction algorithm (this mutation was scored as unknown by PolyPhen-2)

Abbreviations: ABC, activated B cell-type; GCB, germinal center B cell type; NC, unclassified; non-GCB, non-Germinal Center type (IHC classification); Δ, deletion; fs, frameshift; dupl, duplication

Supplementary Table 10. Mutations of the *MLL2* gene in DLBCL

Sample ID	DLBCL Subtype	Exon	Nucleotide change*	AA change*	Affected Domain^^
Missense Mutations					
2089	GCB-DLBCL	E14	G4163T (+/+)	R1388L (+/+)	none
2171	ABC-DLBCL	E17	C4517G	T1506S	5 PHD domain
DB	GCB-DLBCL	E23	C5386T (+/+)	R1796W (+/+)	none
2057	GCB-DLBCL	E31	T7829C	L2610P	none
2136	ABC-DLBCL	E31	G7904A	R2635Q	none
Ly4	GCB-DLBCL	E39	A11849G (+/+)	Q3950R (+/+)	none
2147	GCB-DLBCL	E42	G13976C	R4659P	none
2125	GCB-DLBCL	E46	G14395A	V4799M	none
2041	NC-DLBCL	E48	T15167C	L5056P	none
2027	GCB-DLBCL	E48	T15274A	C5092S	none
SUDHL8	GCB-DLBCL	E51	A16310G	N5437S	SET domain
Nonsense Mutations					
SUDHL6	GCB-DLBCL	E5	C631T	Q211X	5 PHD, HMG, PHD, FYRN, FYRC, SET domain
2078	GCB-DLBCL	E14	T4212A	Y1404X	3 PHD, HMG, PHD, FYRN, FYRC, SET domain
2162	ABC/NC-DLBCL	E20	G4999T	E1667X	HMG, PHD, FYRN, FYRC, SET domain
Ly1	GCB-DLBCL	E26	C5707T	R1903X	HMG, PHD, FYRN, FYRC, SET domain
SUDHL10	GCB-DLBCL	E31	C7228T	R2410X	PHD, FYRN, FYRC, SET domain
2069^	ABC/NC-DLBCL	E34	C8458T	Q2820X	PHD, FYRN, FYRC, SET domain
2192	GCB-DLBCL	E39	C11911T	Q3971X	PHD, FYRN, FYRC, SET domain
2152^	ABC-DLBCL	E39	C11962T	Q3988X	PHD, FYRN, FYRC, SET domain
SUDHL6	GCB-DLBCL	E44	G14134T	E4712X	PHD, FYRN, FYRC, SET domain
2195	ABC-DLBCL	E48	T15618A	Y5206X	FYRN, FYRC, SET domain
Frameshift/Splice site Mutations					
2110	ABC-DLBCL	E6	Δ14bp (777-790)	L260fs	4 PHD, HMG, PHD, FYRN, FYRC, SET domain
2106	GCB-DLBCL	E7	insCACTC (935_936)	H313fs	4 PHD, HMG, PHD, FYRN, FYRC, SET domain
Ly18	GCB-DLBCL	E7	T(+2)GT	A319fs	4 PHD, HMG, PHD, FYRN, FYRC, SET domain
Ly18	GCB-DLBCL	E10	ΔC1940	P658fs	3 PHD, HMG, PHD, FYRN, FYRC, SET domain
2067	GCB-DLBCL	E13	ΔTC (-4)	V1340fs	3 PHD, HMG, PHD, FYRN, FYRC, SET domain
2168	ABC-DLBCL	E35	Δ10bp (10304-10313)	I3435fs	PHD, FYRN, FYRC, SET domain
2127	ABC-DLBCL	E36	ΔCT (10369-10370)	L3457fs	PHD, FYRN, FYRC, SET domain
2157^	GCB-DLBCL	E36	insG (10394_10395)	P3466fs	PHD, FYRN, FYRC, SET domain
Ly1	GCB-DLBCL	E48	ΔC14782	P4929fs	PHD, FYRN, FYRC, SET domain
2160	GCB-DLBCL	E48	ΔG15133	P5046fs	PHD, FYRN, FYRC, SET domain
2027	GCB-DLBCL	E48	ΔC15546	G5182fs	FYRN, FYRC, SET domain
SUDHL8	GCB-DLBCL	E48	ΔC15546	G5182fs	FYRN, FYRC, SET domain

*Numbering according to GenBank accession No. NM_003482.3 (mRNA) and NP_003473.3 (protein) respectively
Mutations affecting intronic sequences within consensus splice sites are indicated by the distance from the corresponding exon
^In this sample, analysis of paired normal DNA confirmed the somatic origin of the mutation
^^Note that none of the missense mutations could be successfully scored by the SIFT and Polyphen-2 algorithms
Abbreviations: ABC, activated B cell-type; GCB, germinal center B cell-type; NC, unclassified; Δ, deletion; ins, insertion; fs, frameshift; +/+, hemizygous or homozygous mutation

Supplementary Table 11. Copy number losses and genomic breaks affecting histone methylation genes in DLBCL

Gene Symbol	Sample ID	Chr	Cytoband	Start position^	End position^	Seg.mean	Seg.observed CN	Size (Kb)	Aberration	N of Genes	Genes
MLL3	2201	7	7q36.1	151478879	151980320	-0.5497	1.37	501.441	loss	2	MLL3 XRCC2
	2124	7	7q36.1 - 7q36.3	148747000	158820000	-0.2331	1.70	10073.000	loss (subclonal)	68	MLL3 and 67 other genes
	2204	7	7q33 - 7q36.2	134215887	153251193	-0.52327	1.40	19035.306	loss	176	MLL3 and 175 other genes
	2168	7	7q32.3 - 7q36.3	130669532	158819753	-0.3754	1.54	28150.221	loss	190	MLL3 and 189 other genes
	2151	7	7q32.1 - 7q36.1	127457093	151504219	0.4188	2.67	24047.126	gain/breakpoint	204	MLL3 and >200 other genes
MLL5	2025	7	7q22.2	104431254	104805720	-0.6521	1.27	374.466	loss	2	MLL5 SRPK2
	2168	7	7p13 - 7q22.2	44647000	105575241	-0.3865	1.53	60928.241	loss	318	MLL5 and >200 other genes
MLL	2126	11	11p15.5 - 11q23.3	188510	117880221	0.6330	3.10	117691.711	gain/breakpoint	1076	MLL and >200 other genes
KDM2B	2157	12	12q24.31	120430813	120550171	-0.6175	1.30	119.358	loss	2	KDM2B ORAI1
	2180	12	12q24.31	120267091	120754295	-0.6723	1.26	487.204	loss	7	KDM2B and 6 other genes
	2176	12	12q24.31	119623368	122061907	-0.5864	1.33	2438.539	loss	42	KDM2B and 41 other genes
	2194	12	12q24.31	119223646	121827176	-0.5786	1.34	2603.530	loss	48	KDM2B and 47 other genes
	2135	12	12q23.2 - 12q24.33	102104106	129246565	-0.8973	1.07	27142.459	loss	214	KDM2B and >200 other genes
	2196	12	12q23.1 - 12q24.33	94841355	132288250	-0.2269	1.71	37446.895	loss (subclonal)	279	KDM2B and >200 other genes
	2162	12	12p13.33 - 12q24.33	20691	132288250	-0.2418	1.69	132267.559	loss (subclonal)	1085	KDM2B and >200 other genes

^ Numbering according to the Human Genome hg18 assembly

Supplementary Table 12. Mutations of the *B2M* gene in DLBCL

Sample ID	Subtype	Allele**	Status	Exon	Nucleotide change*	AA change*	Functional consequences	PolyPhen-2 prediction (score)
2038	GCB-DLBCL	A	M	E1	T62A	M1K	No translation initiation	Possibly damaging (0.564)
		B	M	E1	T80C	L7S	Unknown	Benign (0.076)
2075	GCB-DLBCL	A	M	E1	T62G	M1R	No translation initiation	Possibly damaging (0.733)
		B	G	na	-	-	-	-
2100	GCB-DLBCL	A	M	E1	T62A	M1K	No translation initiation	Possibly damaging (0.564)
		B	unknown	unknown				
2138	GCB-DLBCL	A	M	E1	T62G	M1R	No translation initiation	Possibly damaging (0.733)
		B	G	na	-	-	-	-
2156^	GCB-DLBCL	A	M	E1	G82A, T309G	A8T, Y83X	Truncated protein	nd
		B	M	E2	G(+1)A	R23fs	Truncated protein	nd
2171	ABC-DLBCL	A	M	E1	ΔGCGCTACTCTC (91-101)	L11fs	Truncated protein	nd
		B	M	E2	A245C	N62T	Unknown	Probably damaging (0.997)
2183	ABC-DLBCL	A	M	E1	ΔCT (97-98)	L15fs	Truncated protein	nd
		B	unknown	unknown				
2110	ABC-DLBCL	A	M	E1	ΔCT (103-104)	L14fs	Truncated protein	nd
		B	G	na	-	-		-
2209^	non-GCB	A	M	E1	T(+2)A	R23fs	Truncated protein	nd
		B	G	na	-	-		-
Ly1	GCB-DLBCL	A	M	E1	T62C	M1T	No translation initiation	Possibly damaging (0.564)
		B	M	E2	T352A	Y98N		Probably damaging (1.0)
2033	ABC-DLBCL	A	M	E2	G299A	W80X	Truncated protein	nd
		B	D	E1-3	na	na	Allelic loss	
2089	GCB-DLBCL	A	M	E2	Δ42bp (239-280)	L60fs	Truncated protein	nd
		B	D	na	na	na	Allelic loss	
2148	GCB-DLBCL	A	M	E2	C318G, C271A, T304G	H71N, F82V, Y86X	Truncated protein	nd
		B	G	na	-	-	-	-
WSU	GCB-DLBCL	A	M	E1	ΔGGCCGAGATG (54-63)	ΔM1	No translation initiation	nd
		B	D	E1-3	na	na	Allelic loss	
2152	ABC-DLBCL	A	M	E2	insA (345-346)	D96fs	Truncated protein	nd
		B	unknown	unknown				
2161	GCB-DLBCL	A	M	E2	insT (239-240)	K61fs	Truncated protein	nd
		B	G	na	-	-		
SUDHL2	ABC-DLBCL	A	M	E2	T193G	C45G	Unknown	Probably damaging (0.967)
		B	D	E1-3	na	na	Allelic Loss	

* Numbering according to GenBank accession No. NM_004048 (mRNA) and NP_004039.1 (protein) respectively

^ For these samples, paired normal DNA was available and confirmed the somatic origin of the mutation

** In cases carrying multiple sequence variants, the allelic distribution of the mutations was formally proven by cloning and sequencing PCR products encompassing the various changes; in primary cases where no copy number data were available, the status of the second allele is given as unknown

Abbreviations: ABC, activated B cell-type; GCB, germinal center B cell-type; Δ, deletion; fs, frameshift; ins, insertion; na, not applicable; M, mutated; D, deleted; G, germline; nd, not determined (The PolyPhen-2 algorithm predicts only the impact of amino acid substitutions; note that in cases carrying multiple mutations within the same allele, only the nonsense mutation is considered)

Supplementary Table 13. Copy number losses of the *B2M* gene in DLBCL

Sample ID	Chr	Cytoband	Start position^	End position^	Seg.mean	Seg.observed CN	Size (Kb)	Aberration	N of Genes	Genes
2126	15	15q21.1	42771402	42801338	-2.7261	0.30	29.936	loss homozygous	1	B2M
2019	15	15q21.1	42761080	42824209	-0.8982	1.07	63.129	loss	2	B2M TRIM69
2094	15	15q21.1	42717188	42796158	-1.5882	0.67	78.970	loss homozygous	2	B2M SPG11
2101	15	15q21.1	42736553	42792673	-1.3642	0.78	56.120	loss homozygous	2	B2M SPG11
2101	15	15q21.1 - 15q21.2	42792952	48353000	-0.6927	1.24	5560.048	loss	34	B2M and 33 other genes
2165	15	15q21.1	42789067	42843504	-1.6319	0.65	54.437	loss homozygous	2	B2M TRIM69
2180	15	15q15.3 - 15q21.1	42695657	42807609	-1.8679	0.55	111.952	loss homozygous	2	B2M SPG11
2147	15	15q21.1	42718127	42849292	-1.5344	0.69	131.165	loss homozygous	3	B2M SPG11 TRIM69
2176	15	15q15.3 - 15q21.1	42663754	42831204	-2.0990	0.47	167.450	loss homozygous	3	B2M SPG11 TRIM69
2033	15	15q15.3 - 15q21.1	42635102	42831204	-1.0865	0.94	196.102	loss	4	B2M EIF3J SPG11 TRIM69
2195	15	15q15.3 - 15q21.1	41938355	42814441	-0.8182	1.13	876.086	loss	7	B2M and 6 other genes
2089	15	15q15.1 - 15q21.1	40127344	44171662	-0.6393	1.28	4044.318	loss	57	B2M and 56 other genes
2168	15	15q15.1 - 15q21.1	38888416	43088136	-1.3992	0.76	4199.720	loss homozygous	70	B2M and 69 other genes
2184	15	15q14 - 15q21.2	36854077	48291699	-0.7589	1.18	11437.622	loss	125	B2M and 124 other genes
2194	15	15q12 - 15q21.1	25114232	46089975	-0.6471	1.28	20975.743	loss	156	B2M and 155 other genes
2034	15	15q11.2 - 15q21.2	20303872	50060564	-0.5851	1.33	29756.692	loss	203	B2M and >200 other genes
2124	15	15q11.2 - 15q22.2	20232602	58570000	-0.2416	1.69	38337.398	loss (long arm, subclonal)	247	B2M and >200 other genes
2122	15	15q11.2 - 15q26.3	20224751	100286551	-0.2423	1.69	80061.800	loss (long arm, subclonal)	539	B2M and >200 other genes
2192	15	15q11.2 - 15q26.3	20073110	100286551	-0.2116	1.73	80213.441	loss (long arm, subclonal)	539	B2M and >200 other genes
2196	15	15q11.2 - 15q26.3	20224751	100286551	-0.2066	1.73	80061.800	loss (long arm, subclonal)	539	B2M and >200 other genes
2025	15	15q11.1 - 15q26.3	18276329	100286551	-0.6675	1.26	82010.222	loss (long arm)	727	B2M and >200 other genes

^ Numbering according to the Human Genome hg18 assembly

Supplementary Table 14. Mutations of the *CD58* gene in DLBCL

Sample ID	Subtype	Allele	Status**	Exon	Nucleotide change*	AA change*	Functional consequences
2138^	GCB-DLBCL	A	M	E3	G611T	E164X	Truncated protein
		B	G	na	-	-	-
2026	ABC-DLBCL	A	M	E2	ΔTAT (476-478)	ΔY119	Unknown
		B	unknown	unknown			
2072	NC-DLBCL	A	M	E2	ΔTTCT (467-470)	F116fs	Truncated protein
		B	unknown	unknown			
2038	GCB-DLBCL	A	M	E2	ΔTG (376-377)	V86fs	Truncated protein
		B	unknown	unknown			
2156^	GCB-DLBCL	A	M	E2	ΔGGTTTATTTAGACACTGTGT (361-380)	R80fs	Truncated protein
		B	M	E3	insA (519-520)	C134fs	Truncated protein
2108	ABC-DLBCL	A	M	E3	insG (535-536)	E139fs	Truncated protein
		B	unknown	unknown			

* Numbering according to GenBank accession No. NM_001779.2 (mRNA) and NP_001770.1 (protein) respectively

^ For these samples, paired normal DNA was available and confirmed the somatic origin of the mutation

** In cases carrying multiple sequence variants, the allelic distribution of the mutations was formally proven by cloning and sequencing PCR products encompassing the various changes; in cases where no copy number data were available, the status of the second allele is given as unknown

Abbreviations: ABC, activated B cell-type; GCB, germinal center B cell-type; NC, unclassified; Δ, deletion; fs, frameshift; ins, insertion

Supplementary Table 15. Copy number losses of the *CD58* gene in DLBCL

Sample ID	Chr	Cytoband	Start position^	End position^	Seg.mean	Seg.observed CN	Size (Kb)	Aberration	N of Genes	Genes
2025	1	1p13.1	116864186	116914082	-2.6735	0.31	49.896	loss homozygous	1	CD58
2163	1	1p13.1	116844399	116889302	-0.7329	1.20	44.903	loss	1	CD58
2204	1	1p13.1	116883520	116981713	-1.164	1.03	98.193	loss	2	CD58 IGSF3
2141	1	1p13.1	116868616	116931567	-2.1374	0.45	62.951	loss homozygous	2	CD58 IGSF3
2141	1	1p13.1	116478246	116864186	-0.5539	1.36	385.940	loss	3	ATP1A1 C1orf161 CD58
2176	1	1p13.1	116879075	116988908	-1.9606	0.51	109.833	loss homozygous	2	CD58 IGSF3
2176	1	1p13.2 - 1p13.1	111686452	116878070	-0.7087	1.22	5191.618	loss	48	CD58 and 47 other genes
2136	1	1p13.1	116197766	117071909	-0.5349	1.38	874.143	loss	5	ATP1A1 C1orf161 CD58 IGSF3 SLC22A15
2165	1	1p13.2 - 1p13.1	115016588	117187099	-0.6783	1.25	2170.511	loss	17	CD58 and 16 other genes
2015	1	1p13.2 - 1p12	115801000	118589000	-0.3382	1.58	2788.000	loss	19	CD58 and 18 other genes
2034	1	1p13.3 - 1p11.2	109562000	121041748	-0.4626	1.45	11479.748	loss	115	CD58 and 114 other genes
2168	1	1p36.11 - 1q21.1	24087544	143553903	-0.3839	1.53	119466.359	loss	726	CD58 and >200 other genes
2135	1	1p13.2 - 1q44	114498544	247190999	-0.4290	1.49	132692.455	loss	974	CD58 and >200 other genes
2032	1	1p36.33 - 1p11.2	218557	120832474	-0.3981	1.52	120613.917	loss	1019	CD58 and >200 other genes
2033	1	1p36.33 - 1p11.2	51586	121045307	-0.3335	1.59	120993.721	loss	1020	CD58 and >200 other genes

^ Numbering according to the Human Genome hg18 assembly

Supplementary Table 16. Copy number losses of the *TNFSF9* gene in DLBCL

Sample ID	Chr	Cytoband	Start position^	End position^	Seg.mean	Seg.observed CN	Size (Kb)	Aberration	N of Genes	Genes
2101	19	19p13.3	6475406	6529028	-1.3442	0.79	53.622	loss homozygous	1	TNFSF9
2025	19	19p13.3	6378364	6529028	-2.2145	0.43	150.664	loss homozygous	6	TNFSF9 and 5 other genes
2032	19	19p13.3	6473782	6720475	-0.9985	1.00	246.693	loss	7	TNFSF9 and 6 other genes
2194	19	19p13.3 - 19p13.2	6089027	7978919	-0.5320	1.38	1889.892	loss	49	TNFSF9 and 48 other genes
2168	19	19p13.3 - 19p13.2	6150971	8115572	-1.3811	0.77	1964.601	loss homozygous	50	TNFSF9 and 49 other genes
2182	19	19p13.3 - 19p13.2	6353648	9507586	-0.6219	1.30	3153.938	loss	76	TNFSF9 and 75 other genes
2034	19	19p13.3 - 19q12	4064312	32458000	-0.3557	1.56	28393.688	loss	442	TNFSF9 and >200 other genes
2019	19	19p13.3 - 19p12	41898	24252646	-0.2231	1.71	24210.748	loss	578	TNFSF9 and >200 other genes
2046	19	19p13.3 - 19q13.31	214395	48344501	-0.2759	1.65	48130.106	loss	835	TNFSF9 and >200 other genes
2135	19	19p13.3 - 19q13.43	41898	63789654	-0.3707	1.55	63747.756	loss (entire chromosome)	1518	TNFSF9 and >200 other genes

^ Numbering according to the Human Genome hg18 assembly

Supplementary Table 17. Copy number gains affecting genes involved in B cell differentiation in DLBCL

Gene Symbol	Sample ID	Chr	Cytoband	Start position^	End position^	Seg.mean	Seg.observed CN	Size (Kb)	Aberration	N of Genes	Genes
PAX5	2147	9	9p13.2	36837139	37027976	0.5906	3.01	190.837	gain	1	PAX5
	2182	9	9p13.3 - 9p13.2	35989466	37630824	1.0339	4.10	1641.358	amp	14	PAX5 and 13 other genes
	2020	9	9p13.3 - 9p13.2	35124375	37891883	1.0493	4.14	2767.508	amp	42	PAX5 and 41 other genes
	2151	9	9p21.1 - 9p13.2	32426000	37237137	0.2153	2.32	4811.137	gain	80	PAX5 and 79 other genes
	2171	9	9p21.3 - 9p13.1	22103895	39332563	0.4742	2.78	17228.668	gain	107	PAX5 and 106 other genes
	2168	9	9p21.3 - 9p12	22016992	40572694	0.4481	2.73	18555.702	gain	110	PAX5 and 109 other genes
	2019	9	9p21.3 - 9q13	22474702	70301266	0.2572	2.39	47826.564	gain	128	PAX5 and 127 other genes
	2034	9	9p21.3 - 9q13	21989228	70133638	1.7990	6.96	48144.41	amp	129	PAX5 and 128 other genes
	2023	9	9p24.1 - 9p11.2	5571140	44190837	0.6546	3.15	38619.697	gain	177	PAX5 and 176 other genes
	2162	9	9p24.3 - 9p12	30910	41942504	0.2459	2.37	41911.594	gain	203	PAX5 and >200 other genes
	2089	9	9p24.3 - 9p11.2	30910	43054830	1.3602	5.13	43023.92	amp	207	PAX5 and >200 other genes
	2172	9	9p24.3 - 9q22.31	30910	93978334	0.5258	2.88	93947.424	gain	302	PAX5 and >200 other genes
	2094	9	9p24.3 - 9q33.2	30910	123808552	0.4341	2.70	123777.642	gain	483	PAX5 and >200 other genes
	2135	9	9p21.2 - 9q34.3	26148371	140194638	0.4591	2.75	114046.267	gain	657	PAX5 and >200 other genes
	2101	9	9p21.3 - 9q34.3	22072569	140211203	0.6971	3.24	118138.634	gain	660	PAX5 and >200 other genes
	2112	9	9p21.3 - 9q34.3	22001477	140211203	0.3944	2.63	118209.726	gain	660	PAX5 and >200 other genes
	2165	9	9p24.1 - 9q34.3	5805583	140211203	0.4802	2.79	134405.62	gain	717	PAX5 and >200 other genes
	2181	9	9p24.3 - 9q34.3	30910	140211203	0.2655	2.40	140180.293	gain (entire chromosome)	746	PAX5 and >200 other genes
IRF4	2110	6	6p25.3	327897	418248	0.4673	2.77	90.351	gain	1	IRF4
	2124	6	6p25.3 - 6p24.1	94649	11698000	0.3057	2.47	11603.351	gain	62	IRF4 and 61 other genes
	2163	6	6p25.3 - 6p23	94649	15355372	0.5957	3.02	15260.723	gain	76	IRF4 and 75 other genes
	2192	6	6p25.3 - 6p22.2	94649	25962000	0.3023	2.47	25867.351	gain	116	IRF4 and 115 other genes
	2017	6	6p25.3 - 6p21.33	94649	30897673	0.4510	2.73	30803.024	gain	252	IRF4 and >200 other genes
	2157	6	6p25.3 - 6p21.32	94649	32968730	0.5719	2.97	32874.081	gain	344	IRF4 and >200 other genes
	2162	6	6p25.3 - 6p21.1	94649	41778611	0.3228	2.50	41683.962	gain	456	IRF4 and >200 other genes
	2172	6	6p25.3 - 6p21.1	94649	42113427	1.0594	4.17	42018.778	amp	464	IRF4 and >200 other genes
	2144	6	6p25.3 - 6p12.3	94649	45521325	1.0511	4.14	45426.676	amp	521	IRF4 and >200 other genes
	2173	6	6p25.3 - 6p12.3	94649	46237643	0.2112	2.32	46142.994	gain	524	IRF4 and >200 other genes
	2196	6	6p25.3 - 6p12.1	94649	56967000	0.1714	2.25	56872.351	gain	582	IRF4 and >200 other genes
	2112	6	6p25.3 - 6q12	94649	65204103	0.7713	3.41	65109.454	gain	591	IRF4 and >200 other genes
	2041	6	6p25.3 - 6q14.1	247294	77300412	0.4619	2.75	77053.118	gain	621	IRF4 and >200 other genes
ETS1	2196	11	11q24.2 - 11q24.3	127277784	128406451	0.8661	3.65	1128.667	gain	7	ETS1 and 6 other genes
	2127	11	11q23.3 - 11q24.3	118011767	127836821	0.2576	2.39	9825.054	gain	110	ETS1 and 109 other genes
	2025	11	11q23.3 - 11q25	117413538	134449982	0.6231	3.08	17036.444	gain	154	ETS1 and 153 other genes
	2094	11	11q22.3 - 11q24.3	107643213	129097547	0.4949	2.82	21454.334	gain	204	ETS1 and >200 other genes
	2168	11	11q23.1 - 11q25	110198401	134449982	0.6088	3.05	24251.581	gain	215	ETS1 and >200 other genes
	2180	11	11q22.3 - 11q24.3	105812443	129117874	0.9738	3.93	23305.431	gain	215	ETS1 and >200 other genes
	2034	11	11q22.3 - 11q25	107654049	134449982	0.2933	2.45	26795.933	gain	225	ETS1 and >200 other genes
	2195	11	11q22.1 - 11q25	100154156	134449982	0.8872	3.70	34295.826	gain	269	ETS1 and >200 other genes
	2141	11	11q14.2 - 11q24.3	87496474	128435255	0.5980	3.03	40938.781	gain	288	ETS1 and >200 other genes
	2032	11	11q14.3 - 11q25	89597951	134449982	0.3041	2.47	44852.031	gain	299	ETS1 and >200 other genes
	2046	11	11q14.2 - 11q25	87309810	134449982	0.4809	2.79	47140.172	gain	310	ETS1 and >200 other genes
	2097	11	11q14.2 - 11q25	86466000	134449982	0.2314	2.35	47983.982	gain	311	ETS1 and >200 other genes
	2201	11	11q14.2 - 11q25	86490476	134429213	0.4104	2.66	47938.737	gain	311	ETS1 and >200 other genes
	2015	11	11q14.1 - 11q25	81644000	134449982	0.2622	2.40	52805.982	gain	332	ETS1 and >200 other genes
	2144	11	11q13.4 - 11q25	74640094	134449982	0.4708	2.77	59809.888	gain	367	ETS1 and >200 other genes
	2122	11	11q13.3 - 11q25	70230704	134449982	0.5345	2.90	64219.278	gain	419	ETS1 and >200 other genes
	2043	11	11q13.2 - 11q25	67734671	134449982	0.4338	2.70	66715.311	gain	442	ETS1 and >200 other genes
	2075	11	11p15.5 - 11q25	188510	134449982	0.4288	2.69	134261.472	gain (entire chromosome)	1218	ETS1 and >200 other genes
	2140	11	11p15.5 - 11q25	188510	134449982	0.5916	3.01	134261.472	gain (entire chromosome)	1218	ETS1 and >200 other genes
	2182	11	11p15.5 - 11q25	188510	134449982	0.5038	2.84	134261.472	gain (entire chromosome)	1218	ETS1 and >200 other genes

^ Numbering according to the Human Genome hg18 assembly

Supplementary Table 18. Mutations of the *MEF2B* gene in DLBCL

Sample ID	Subtype	Exon	Nucleotide change*	AA change*	Affected Domain	PolyPhen-2 prediction (score)
2026	ABC-DLBCL	E3	G343A	E77K	MEF2 domain	Possibily damaging (0.849)
2034	ABC-DLBCL	E3	C348G	S78R	MEF2 domain	Benign (0.39)
2125	GCB-DLBCL	E3	T275C	L54P	MADS box	Probably damaging (1.0)
2144	GCB-DLBCL	E3	C226A	L38I	MADS box	Probably damaging (1.0)
2127	ABC-DLBCL	E5	C625T	R171X	TAD2, C-term	nd
2073	GCB-DLBCL	E6	G734A	R207Q	TAD1	Probably damaging (0.999)
2162	ABC-DLBCL	E7	G851A	G246E	none	unknown**
2109	GCB-DLBCL	E8	Δ(908-938)	P267fs	TAD2, C-term	nd
Ly4	GCB-DLBCL	E4	G428A	G105E	none	Probably damaging (1.0)
SUDHL4	GCB-DLBCL	E3	A362T	D83V	MEF2 domain	Probably damaging (0.999)

* Numbering according to GenBank accession No. NM_001145785.1 (mRNA) and NP_001139257.1 (protein), corresponding to the MEF2B isoform a.
** The algorithm could not compute a score for this change, in both isoforms of the MEF2B protein
Abbreviations: ABC, activated B cell-type; GCB, germinal center B cell-type; Δ, deletion; fs, frameshift; nd, not determined (the PolyPhen-2 algorithm predicts only the impact of amino acid substitutions)

Nature Genetics: doi:10.1038/ng.892

Supplementary Table 19. Mutation analysis of common targets of genetic lesions in DLBCL cell lines

Sample ID	Subtype	Gene Symbol												
		TNFAIP3^	MYD88	CD79B	CARD11	TRAF2	CREBBP^	EP300^	EZH2	MLL2	MEF2B	CD58	PRDM1^	B2M
DB	GCB	WT	WT	WT	WT	WT	WT	m	WT	m	WT	WT	WT	WT
LY18	GCB	ND	WT	WT	ND	ND	WT	WT	WT	M/M	WT	WT	WT	WT
LY1	GCB	WT	WT	WT	WT	WT	M	WT	m	M/M	WT	WT	WT	m/m
LY4	GCB	WT	WT	WT	WT	WT	WT	WT	WT	m	m	WT	WT	WT
LY7	GCB	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
LY8	GCB	D	WT	WT	ND	ND	m	WT	WT	WT	WT	WT	WT	WT
SUDHL4	GCB	WT	WT	WT	WT	WT	WT	WT	m	WT	m	WT	WT	WT
SUDHL5	GCB	D	WT	WT	WT	m	D	WT	WT	WT	WT	WT	D	WT
SUDHL6	GCB	WT	m	WT	WT	WT	M	m	m	M/M	WT	WT	WT	WT
SUDHL7	GCB	WT	WT	WT	m	WT	m/m	WT	WT	WT	WT	WT	WT	WT
SUDHL8	GCB	WT	WT	WT	WT	WT	M	M/m	WT	M/m	WT	WT	WT	WT
SUDHL10	GCB	WT	WT	WT	WT	WT	M	m	WT	M	WT	WT	WT	WT
VAL	GCB	m	WT	m	WT	WT	M	WT	m	WT	WT	WT	WT	WT
WSU	GCB	WT	WT	WT	m	m	m	WT	WT	WT	WT	WT	WT	M/D
RCK8	ABC/NC	M/M	WT	WT	WT	WT	WT	D*	WT	WT	WT	M/D	WT	WT
HBL1	ABC	WT	m	m	WT	WT	WT	WT	WT	WT	WT	D/D	WT	WT
LY10	ABC	D	m	WT	WT	WT	WT	WT	WT	WT	WT	WT	M/D	WT
LY3	ABC	D	m	WT	m	WT	WT	WT	WT	WT	WT	WT	M/D	WT
RIVA	ABC	D	WT	WT	WT	WT	M	WT	WT	WT	WT	WT	D	WT
SUDHL2	ABC	M/M	m	WT	WT	WT	WT	M/D	WT	WT	WT	WT	m/D	m/D
U2932	ABC	D	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	D	WT

Abbreviations: M, truncating mutation; m, missense mutation; M/M or m/m, biallelic mutation; D, hemizygous deletion; D/D, homozygous deletion; ND, not determined

^This gene was also studied by FISH analysis. Data for SUDHL5, SUDHL6, SUDHL8, SUDHL10 and DB are from the Tumorscape database (http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf)

*Data from Garbati et al, Cancer Letter 2011