SUPPLEMENTARY TABLES

Table S1. List of antibodies used in flow cytometry.

Surface molecule	Fluorochrome	Clone	Manufacturer
Anti-mouse antibodies			
B220	PerCP	RA3-6B2	BD
CD4	Alexa700	L3T4	eBioscience
CD38	Alexa700	90	eBioscience
CD83	Biotin	Michel19	Biolegend
CD86	Alexa488*	GL1	eBioscience
CXCR4	PE	2B11	eBioscience
CXCR5	FITC	2G8	BD
FAS	PE-Cy7	Jo2	BD
GL-7	FITC	GL-7	BD
Igκ	FITC	187.1	BD
Igλ$_{1-3}$	FITC	R26-46	BD
MHC II (I-Ab)	FITC	AF6-120.1	BD
Anti-human antibodies			
CCR6	PE	29-2L17	Biolegend
CD3	FITC	UCHT1	Beckman Coulter
CD19	PE-Cy7	SJ25C1	BD
CD23	FITC	EBVCS2	eBioscience
CD27	FITC	M-T271	BD
CD38	PerCP-Cy5.5	HIT2	Biolegend
CD44	FITC	IM7	eBioscience
CD69	FITC	FN50	BD
CD77	FITC	38-13	AbD Serotec
CD83	Biotin	HB15e	Biolegend
CD86	Pacific Blue	IT2.2	Biolegend
CXCR4	PE	12G5	Biolegend
CXCR5	Alexa488	RF8B2	Biolegend
IgD	FITC	IA6-2	BD
IgD	V450	IA6-2	BD
Igκ	FITC	G20-193	BD
Igλ	FITC	1-155-2	eBioscience
MHC-II (HLA-DR)	APC-Cy7	L243	BD
Secondary reagents			
Streptavidin	APC	---	BD

* conjugated to fluorochrome in our laboratory

Table S2. List of antibodies used in immunofluorescence microscopy.

Surface molecule	Fluorochrome	Clone	Manufacturer
Anti-mouse antibodies			
AID	Purified	mAID-2	eBioscience
Bcl6 (N3)	Purified	Rabbit polyclonal	Santa Cruz Biotechnology
Mouse IgH+L	Biotin	Horse polyclonal	Vector Laboratories
Anti-human antibodies			
AID	Purified	mAID-2	eBioscience
CD23	Alexa488*	1B12	Novocastra
CD83	Purified	1H4b	Abcam
CD86	APC	IT2.2	Biolegend
CXCR4	PE	12G5	Biolegend
IgD	Biotin	IA6-2	BD
IgD	FITC	IA6-2	BD
Secondary reagents			
Streptavidin	Alexa647	---	Invitrogen
Rat IgG	Biotin	Donkey polyclonal	Southern Biotech
Streptavidin	Cy3	---	Jackson Immunoresearch
Streptavidin	Alexa647	---	Invitrogen

* conjugated to fluorochrome in our laboratory using Zenon kit (Invitrogen)

Table S3. Sequencing of V regions from human light and dark zone cells.

	No. of Sequences	Mutations/ Base Pair	Mutations/ V-region	p-value*	Replacement/ Silent Ratio	p-value*
LZ pooled	37	0.0439	13.63	0.444	3.11	0.397
DZ pooled	36	0.0496	14.60		2.73	
LZ Pt. 1	13	0.0385	16.33	0.852	4.07	0.088
DZ Pt. 1	12	0.0407	12.00		3.13	
LZ Pt. 2	15	0.0373	10.93	0.303	2.81	0.820
DZ Pt. 2	14	0.0501	14.71		2.68	
LZ Pt. 3	9	0.0627	18.44	0.787	2.77	0.556
DZ Pt. 3	10	0.0588	17.30		2.53	

* Student's T-test; Pt: patient

Table S4. Full list of genes included in each of the gene signatures constructed.

(Microsoft Excel spreadsheet)

Table S5. Overlap between selected gene sets from the GSEA database and the common human/mouse signatures light and dark zone signatures.

Gene Sets	Description	Up in	k/K	p-value (overlap)§*	NES (Hs/Mm)	Nominal p-value (Hs/Mm)*	FDR q-value (Hs/Mm)*
				Overlap analysis (*Hs* and *Mm* combined)		**GSEA analysis** (individually by species)	
C2.CGP							
BASSO_CD40_SIGNALING_UP	Gene upregulated by CD40 signaling in Ramos cells (EBV-negative Burkitt's lymphoma)	LZ	0.31	0	2.92/2.81	0/0	0/0
DIRMEIER_LMP1_RESPONSE_EARLY	Genes upregulated in B2264-19/3 cells (primary B lymphocytes) within 30-60 min after activation of LMP1	LZ	0.39	0	2.77/2.67	0/0	0/0
SCHUHMACHER_MYC_TARGETS_UP	Genes upregulated in P493-6 cells (Burkitt's lymphoma) induced to express MYC	LZ	0.19	8.02E-13	2.65/1.58	0/0.010	0/0.071
BROCKE_APOPTOSIS_REVERSED_BY_IL6	Genes upregulated in INA-6 cells (multiple myeloma, MM) by re-addition of IL6 after its initial withdrawal for 12h.	LZ	0.11	1.07E-12	1.69/2.09	0/0	0.044/0.001
C3.TFT							
V$NFKAPPAB_01	Motif GGGAMTTYCC, matches annotation for NFKB/RELA	LZ	0.10	7.46E-08	1.75/2.11	0/0	0.025/0.001
V$SRF_01	Motif ATGCCCATATATGGWNNT, matches annotation for SRF	LZ	0.11	0.004	1.81/1.48	0/0.009	0.011/0.199
V$STAT_01 (STAT6)	Motif TTCCCGKAA; matches annotation for STAT6	LZ	0.06	0.005	1.45/1.76	0.001/0	0.248/0.014
C5.BP							
CELL_CYCLE_PROCESS	Gene Ontology GO:0022402	DZ	0.18	0	2.89/2.10	0/0	0/0.009
MITOTIC_CELL_CYCLE	Gene Ontology GO:0000278	DZ	0.20	0	2.75/2.05	0/0	0/0.008
MICROTUBULE_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	Gene Ontology GO:0000226	DZ	0.24	2.50E-06	2.65/1.83	0/0.002	0/0.044
CYTOKINESIS	Gene Ontology GO:0000910	DZ	0.32	7.46E-06	2.48/2.04	0/0	0/0.007

k/K: fraction of the total gene set contained within the common human/mouse signature; NES: Normalized enrichment score; FDR: False discovery rate; CGP: chemical and genetic perturbations (knowledge-based); TFT: transcription factor motifs (computational); BP: Gene Ontology, biological processes; Mm: *Mus musculus*; Hs: *Homo sapiens*; LZ: light zone; DZ: dark zone.
* "0" indicates p-value below the minimum calculated by the algorithm.
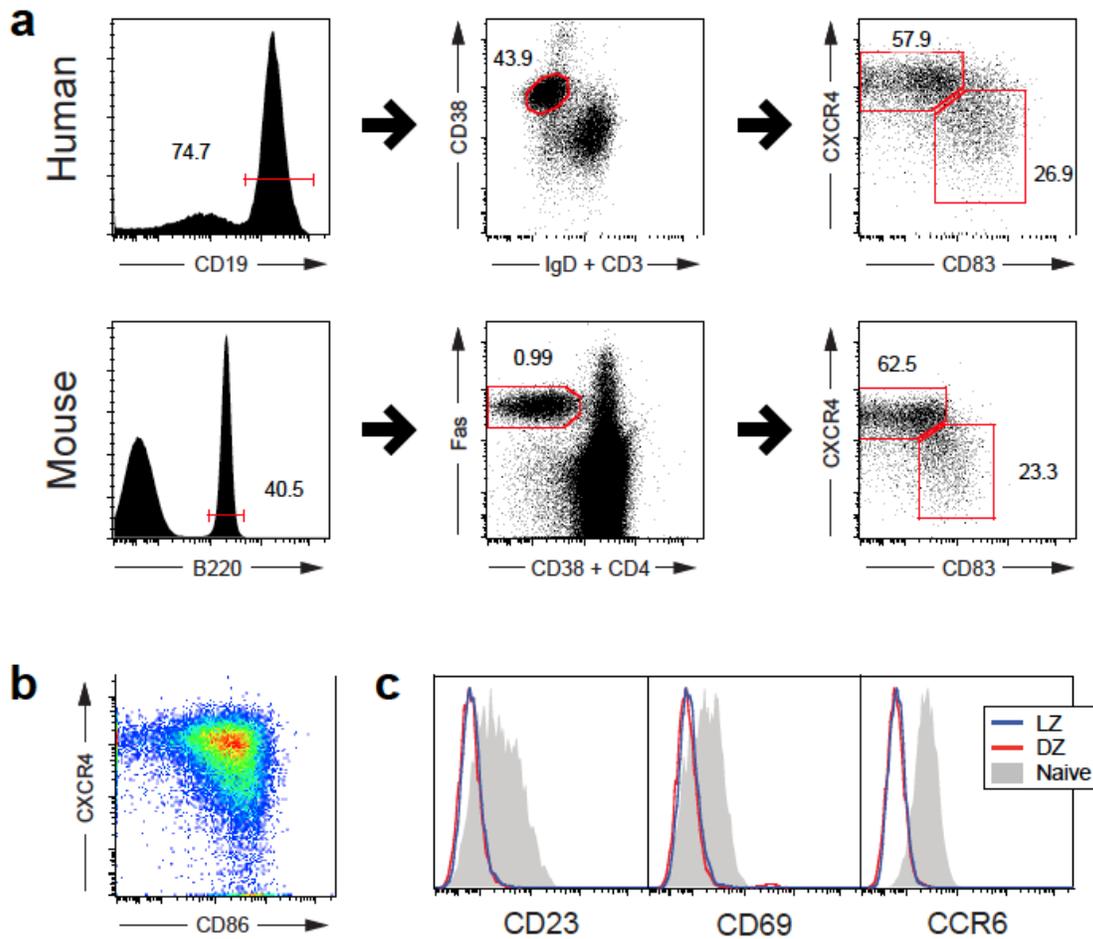
Figure S1. Flow cytometry analysis of human and mouse light zone/dark zone B cells. (A) Gating strategy for identifying LZ and DZ cells in GCs from human tonsil (top) and mouse lymph nodes and spleen (bottom). (B) Distribution of human GC B cells according to expression of CXCR4 and CD86. (C) Low expression of CD23, CD69, and CCR6 by human GC B cells.
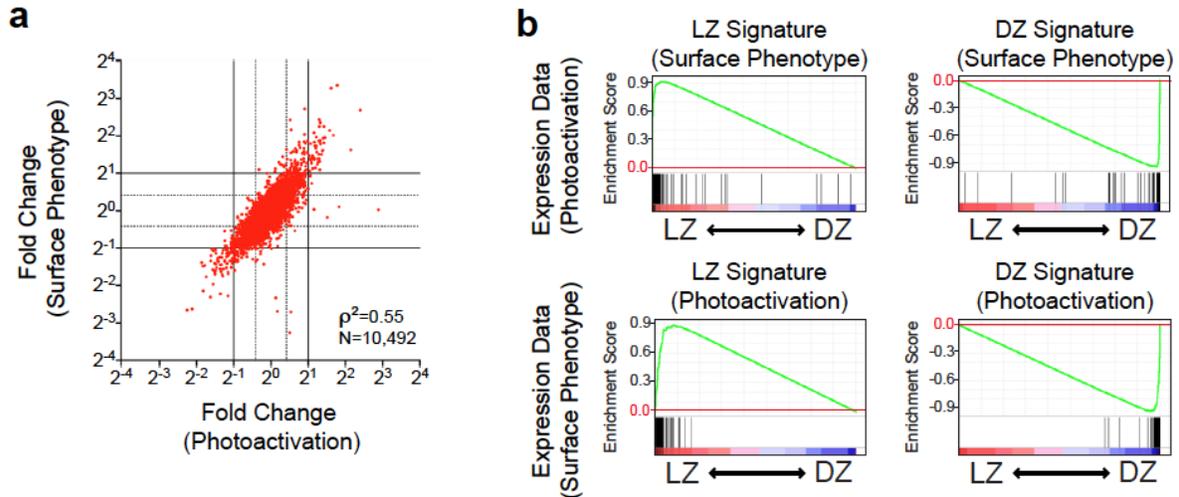
Figure S2. Comparison of LZ/DZ profiles obtained by flow cytometry and *in situ* photoactivation. (A) Correlation of fold differences between LZ and DZ cells sorted based on anatomical position (photoactivation, X-axis; data from Victora et al (2010)) and on expression of CXCR4 and CD83 (Surface Phenotype, Y-axis). To minimize noise from probes with trace or no signal, only the ~10 thousand probes with mean raw expression >100 in both datasets are plotted. (B) Gene set enrichment analysis of the expression of LZ and DZ gene signatures obtained by photoactivation or by sorting in the reciprocal datasets. Nominal and adjusted P-values and false discovery rates are all below the detection level (<0.001) for all comparisons.
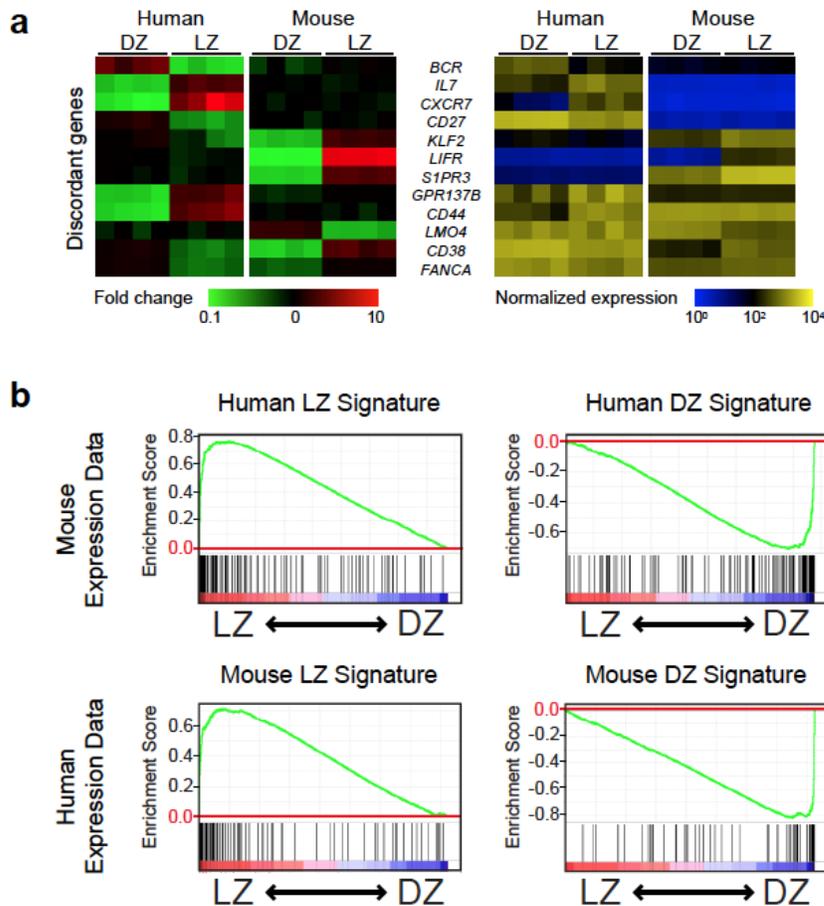
Figure S3. A common set of genes distinguishes LZ and DZ B cells in mouse and human. (A) *Right*: Heatmaps showing differential expression in LZ and DZ of genes whose changes in expression are not conserved between species, or do not follow the same trend in both mouse and human. Colors indicate fold change between one zone and the opposite zone from the same sample (human) or pool (mouse). *Left*: Absolute expression of the same genes in mouse and human. (B) Gene set enrichment analysis overlaying the mouse and human LZ and DZ gene signatures (as shown in Fig. 3) on datasets derived from the opposite species. Nominal and adjusted P-values and false discovery rate are all below the detection level (<0.001) for all comparisons.
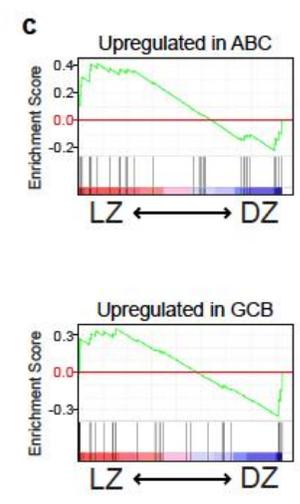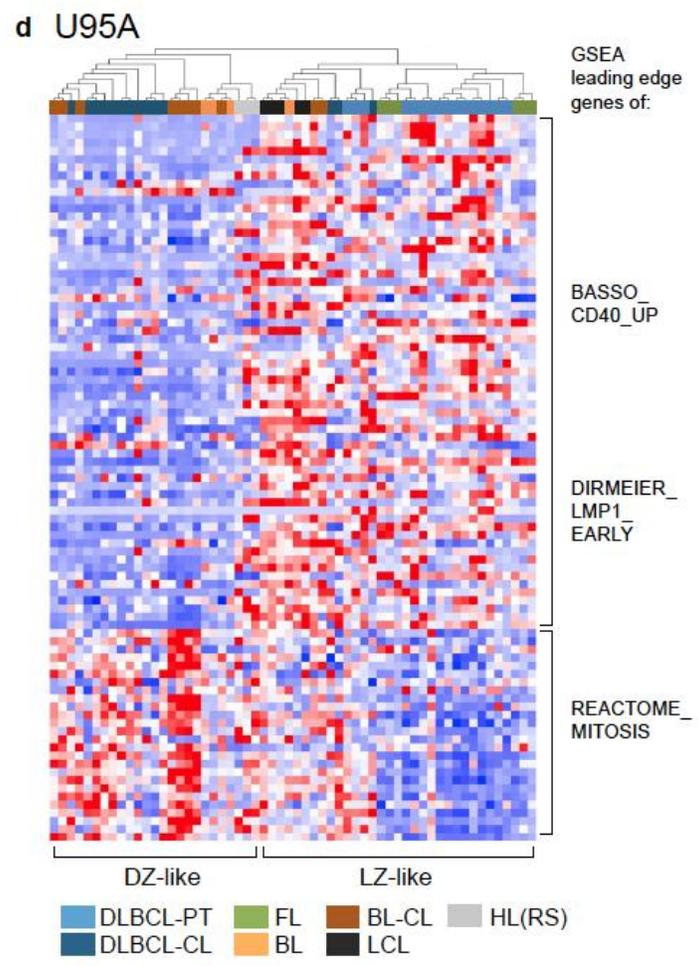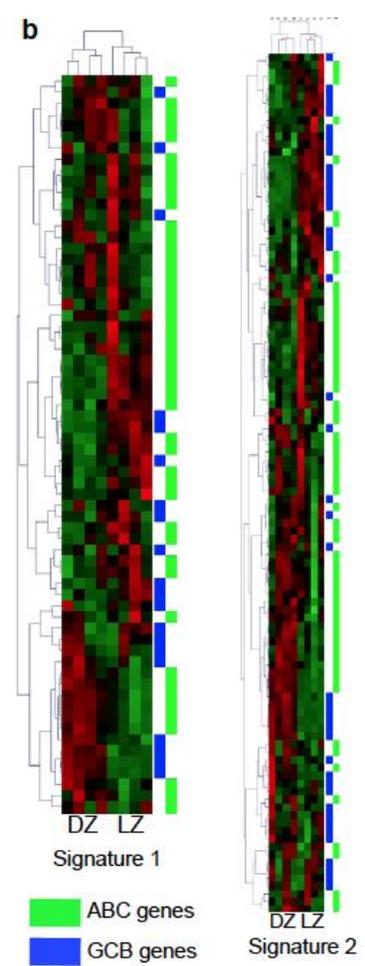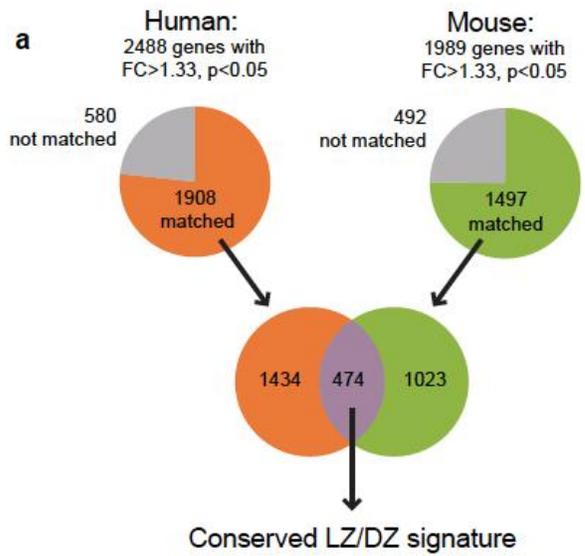
a

Human:
2488 genes with
FC>1.33, p<0.05

Mouse:
1989 genes with
FC>1.33, p<0.05

580
not matched

1908
matched

492
not matched

1497
matched

1434   474   1023

Conserved LZ/DZ signature

b

DZ   LZ
Signature 1

DZ LZ
Signature 2

ABC genes
GCB genes

d U95A

GSEA
leading edge
genes of:

BASSO_
CD40_UP

DIRMEIER_
LMP1_
EARLY

REACTOME_
MITOSIS

DZ-like   LZ-like

DLBCL-PT   FL   BL-CL   HL(RS)
DLBCL-CL   BL   LCL

c

Upregulated in ABC

Enrichment Score

0.4
0.2
0.0
-0.2

LZ ←→ DZ

Upregulated in GCB

Enrichment Score

0.3
0.0
-0.3

LZ ←→ DZ

Figure S4. Definition of a common LZ/DZ gene signature for mouse and human and its relationship to different B cell lymphomas. (A) Genes in the low-stringency (1.33-fold cutoff) mouse and human LZ/DZ signatures were first matched to their homologues in the opposite species by Official Gene Symbol. Genes that could not be matched (the fraction in grey in the charts) were discarded. The overlap between the two gene sets thus obtained was defined as the common human/mouse signature. A smaller, more stringent version of this signature (1.5-fold cutoff) was used for clustering analysis. See Table S2 for the full signatures. (B) Expression pattern of ABC and GCB signature genes/probesets (Alizadeh et al., 2000; Rosenwald et al., 2002)) across human LZ and DZ samples. (C) GSEA enrichment plots for ABC or GCB gene signatures in LZ/DZ gene expression profiles (Normalized enrichment scores, 1.25 and 1.03, respectively. P-values, 0.15 and 0.41). (D) Hierarchical clustering of GC-derived B-NHL and various lymphoid cell lines based on the expression of genes in a "compound pathway signature." In this case, the samples analyzed correspond to tumoral B cells purified from primary lymphoma cases by magnetic bead isolation (all samples have been previously reported in GSE2350, Gene Expression Omnibus database). Cases are color-coded. DLBCL, Diffuse Large B-cell Lymphoma. BL, Burkitt's Lymphoma. FL, Follicular Lymphoma, HL, Hodgkin Lymphoma. (PT, primary tumors. CL, cell lines). LCL, lymphoblastoid cell lines. All Hodgkin lymphoma samples correspond to established cell lines.
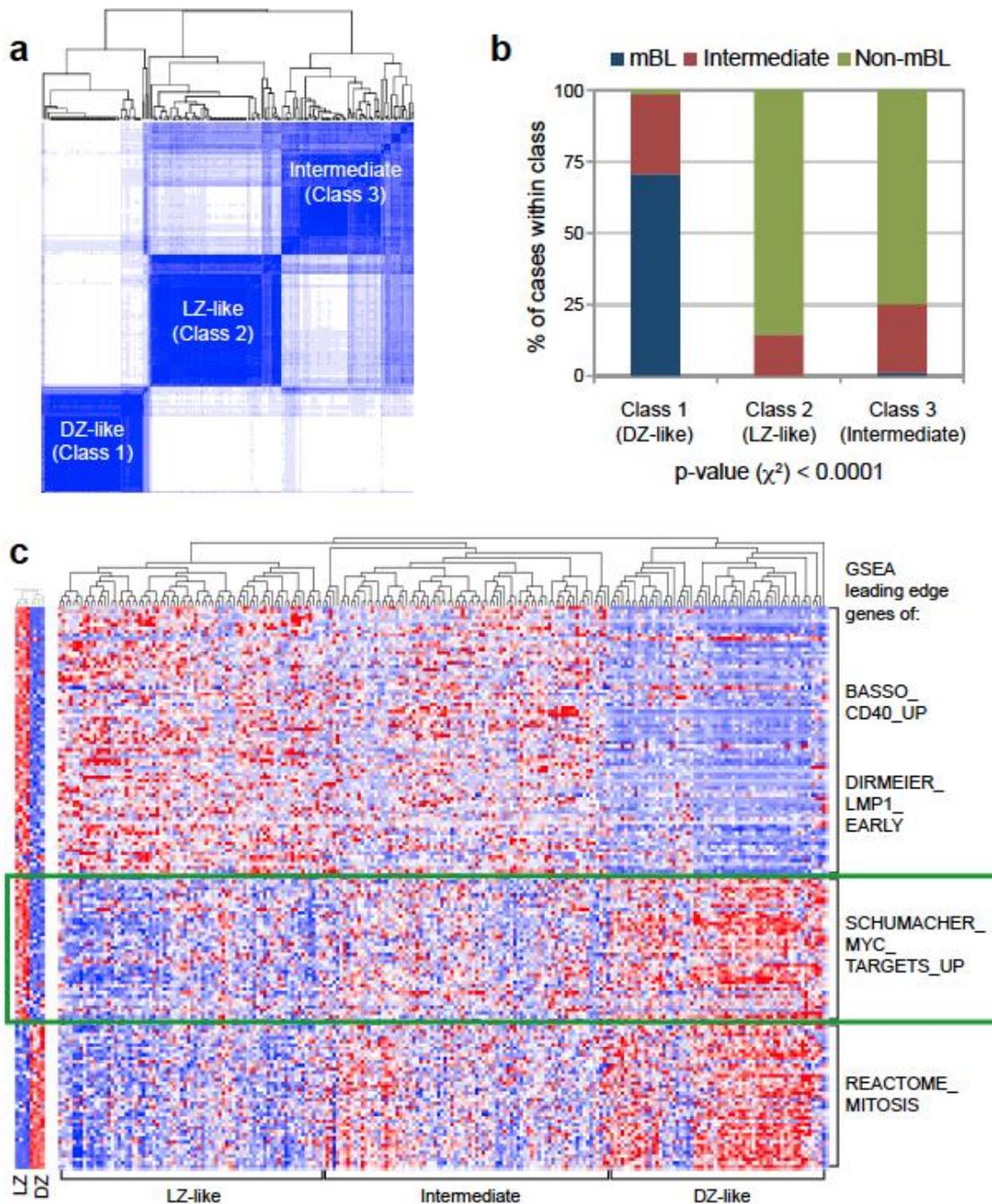
Figure S5: Pathway-based Hierarchical Clustering identifies 3 subgroups of aggressive mature B-NHL based on their resemblance to LZ/DZ GC B-cells.

(A) Consensus clustering (200 bootstraps) of the aggressive mature B-NHL case series described by Hummel et al (2006), according to the expression pattern of a "compound pathway signature" differentiating LZ and DZ GC B-cells. Shown is

the clustering image with k=3 (no significant improvement of the CDF was observed with higher k values). (B) Distribution of molecular Burkitt Lymphomas (mBL) and non-molecular (non-mBL) cases, as defined by Hummel et al (2006), among the 3 different subgroups identified by the consensus clustering analysis depicted in (A). The p-value shown refers to the significance of the distribution of the three subclasses in Hummel et al (2006) (mBL, intermediate, non-mBL) among the three consensus clustering subgroups (Chi-squared test). (C) Hierarchical clustering of mature B-NHL lymphoma cases described by Hummel et al (2006) (see also Fig. 6). Clusters are highlighted for clarity, and are built based on the expression pattern of all genes included in the "compound pathway signature." Because of the relevance of c-Myc-activated gene programs in the biology of Burkitt's lymphoma and their enrichment in LZ GC B-cells, an additional signature (described in Table S3) was included in the analysis. Note the discordant behavior of this signature in normal DZ B cells and DZ-like aggressive B-NHL (= molecular BL).