SUPPLEMENTAL TABLE LEGENDS

Table S1. Glycopeptide analysis information.

Table S2. Crystallographic data and refinement statistics for Fab L14.

Table S3. Distribution, amino acid motif and location of the N-glycosylation sites acquired in the tumor IGHV-IGHD-IGHJ rearrangements of DLBCL primary samples.

Table S4. Location pattern of the acquired N-glycosylation sites in the tumor IGHV-IGHD-IGHJ rearrangements and mutational profile of the DLBCL primary samples.

Table S5. Distribution and prevalence of individual genetic lesions in the CDR+ve and CDR-ve EZB subtype of GCB-DLBCL.

Table S6. Characteristics of the DLBCL primary samples used for association with a *BCL2* translocation and/or functional studies.

Table S7. Characteristics of DLBCL cell lines.

Table S8. Progression-free survival at 9 years of EZB GCB-DLBCL defined by the acquisition (CDR+ve) or not (CDR-ve) of N-glycosylation sites in the CDR or by acquisition of a MYC translocation (MYC+ve) or not (MYC-ve).

Table S9. Differentially expressed genes between CDR+ve and CDR-ve EZB MYC-ve DLBCL.

Table S10. Pathways of interest overexpressed in CDR+ve versus CDR-ve EZB MYC-ve DLBCL by Gene Set Enrichment Analysis.

Table S11. Immunoglobulin gene sequence analysis of Raji and Raji/DC-SIGN cell lines.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Frequency of AGS by COO and genetic subtypes.

(A) Frequency of rearrangements with AGS in the CDR (blue bars) or the FR only (white bars) relative to the total number of each genetic subtype according to the LymphGen algorithm in all DLBCL. Using the Fisher exact test, the frequency of rearrangements with AGS in the CDR in the EZB subtype was significantly higher than in any other genetic subtype ($p<0.01$) except N1 ($p=0.07$). The frequency of AGS in the FR only of the EZB subtype was not significantly different from any other genetic subtype ($p>0.05$) except BN2 ($p=0.0279$). **(B)** Distribution of genetic subtypes (according to LymphGen algorithm) within GCB-DLBCL that have not acquired sites in the CDR**.**

Figure S2. Genetic aberrations associated with the acquisition of glycosylation sites, double-hit status, and genetic subtype in GCB-DLBCL cases.

An oncoplot showing the frequency of the genetic variants associated with the acquisition of glycosylation sites and double-hit (DHIT) status in the EZB subtype **(A)** and all the GCB-DLBCL **(B)**. Genetic variants include those associated with the EZB, BN2, ST2, and A53 genetic subtypes according to the LymphGen algorithm, ordered first by *BCL2* translocation status, then by frequency of genetic alteration in the cohort.

Figure S3. *IGHV-IGHD-IGHJ* and *IGKV-IGKJ* amino acid sequence of L14 and L29.

IMGT/V-QUEST alignment of the tumor *IG* heavy and light chain rearranged transcripts from L14 and L29 lymphomas. FRs and CDRs with numbering according to IMGT are represented. Dashes indicate homology to germline; upper cases indicate amino acid replacement. The acquired N-glycosylation sites are highlighted in yellow. The location of those acquired N-glycosylation sites and relevant amino acid residues which were highlighted in Figure 2 are also represented. In the L14 *IGH* rearrangement, the germline unoccupied glycosylation site NHS is at codons 57-59.

Figure S4. Glycosylation pattern by immunoblot of the surface Ig variable region from GCB-DLBCL cell lines.

Cell surface proteins of CDR+ve (WSU-FSCCL, NU-DHL1, SU-DHL6, OCI-Ly19, WSU-NHL) and CDR-ve (WSU-DLCL2, 1 AGS in FR only) GCB-DLBCL cell lines, were isolated by surface biotinylation assay. The glycosylation pattern of the sIg was analyzed by immunoblotting after glycan digestion with Endo H or PNGase F. Anti-μ and anti-γ primary antibodies were used depending on sIg isotype expression.Also refer to Table S7 for AGS location in each cell line.

Figure S5. Binding of DC-SIGN is specific for sIg-Mann+ve tumor cells.

DC-SIGN binding was determined by flow cytometry in lymphocyte subsets from PBMCs of healthy donors and in germinal center (GC) B cells from reactive tonsils and lymph nodes **(A)**, or GCB-DLBCL cell lines with AGS **(B-C)**. Endo H treatment was performed on the sIg-Mann+ve WSU-FSCCL and NU-DHL1 cell lines prior to DC-SIGN binding assay, to verify specific binding to oligomannose-type glycans **(B)**. Red lines indicate treatment with recombinant DC-SIGN-Fc and staining with secondary antibody FITC-conjugated anti-Fc. Black lines indicate staining with secondary FITC-conjugated anti-Fc antibody only, as control. In **(B)**, blue lines indicate treatment with recombinant DC-SIGN-Fc and staining with secondary antibody FITC-conjugated anti-Fc following treatment with Endo H of the lymphoma cells at 1:10 dilution at 37°C for 2 hours before DC-SIGN binding; grey lines indicate staining with secondary FITC-conjugated anti-Fc antibody only following incubation with Endo H, as control. Also, refer to Table S7 for the intensity of DC-SIGN binding to the cell lines.

Figure S6. Signaling analysis in DLBCL primary samples.

(A) Gating strategy for the selection of the tumor population in signaling experiments in two samples (16-TB0006, upper plots, and 16-TB0014, lower plots). Tumor cells were selected as CD20+/BCL2^{hi} events. IC, isotype control antibody. **(B)** Three repeats of DC-SIGN and anti-Ig-mediated intracellular signaling (pSYK) by phosflow assay in the sIg-Mann+ve cell line NU-DHL1. **(C)** Flow cytometry histograms for intracellular signaling (pSYK) in the samples and cell line shown in Figure 5. **(D)** Soluble recombinant DC-SIGN was incubated with 500nM hIgG1-D1 before the treatment of the primary GCB-DLBCL sample 16-TB0014 for 5 minutes. Phosphorylation of SYK at Y525/526 was measured in the CD20+/BCL2^{hi} tumor population by flow cytometry. Also, refer to Tables S6-S7 for characteristics of primary samples and cell lines.

Figure S7. DC-SIGN expression in Raji, Raji/DC-SIGN, or MoDCs.

(A) DC-SIGN (CD209) expression was determined by flow cytometry in Raji and Raji/DC-SIGN cells. **(B)** Phenotypic characterization of MoDCs showing positivity for the DC markers CD209, CD1a, and CD206 and negativity for macrophage markers CD14 and CD163. IC, isotype control antibody. Ab, antibody.

Figure S8. Inhibition of clustering by hIgG1-D1.

(A) Clustering assay of the sIg-Mann-ve HBL1 cell line with Raji/DC-SIGN (upper panels) or Raji (lower panel) was performed by inverted fluoresce microscopy. The microscopy images show no DC-SIGN-specific interaction between the two cell populations. **(B)** Clustering between sIg-Mann+ve (WSU-FSCCL and NU-DHL1) or sIg-Mann-ve (HBL1) DLBCL cell lines and Raji (dark grey bars) or Raji/DC-SIGN (black bars) was measured by flow cytometry. Raji/DC-SIGN were treated with 10nM hIgG1-D1 (light grey bars) or left untreated before co-culture with DLBCL cell lines.

Figure S9. Clonally related lymphoma cells collected from independent tissue sites at diagnosis of FL and transformation into DLBCL preserve structure and function of Ig-Mann.

Biopsy of a renal hilar lymph node at the time of diagnosis of grade 1 FL (specimen 13-TB0265), and biopsy of a supraclavicular lymph node at the time of transformation into DLBCL (specimen 16-TB0013). **(A)** IMGT/V-QUEST alignment of the *IGHV-IGHD-IGHJ* rearrangements of the FL and transformed DLBCL sample. The dominant *IGHV-IGHD-IGHJ* rearrangements from the FL and DLBCL biopsies are clonally related. FRs and CDRs with numbering according to IMGT are represented. Dashes indicate homology; upper cases indicate amino acid replacement. The AGS in the CDR3 (highlighted in yellow) is preserved at the time of transformation into DLBCL. **(B)** Hematoxylin and eosin staining of the FL and DLBCL biopsy specimens. Both FL 13-TB0265 and DLBCL 16-TB0013 specimens are shown at 200X magnification. **(C)** DC-SIGN binding to the sIg-Mann expressed on the FL and DLBCL samples was determined by flow cytometry. Red lines: tumor cells were stained with soluble recombinant DC-SIGN-Fc and subsequently with FITC-conjugated anti-Fc. Black lines: tumor cells were not stained with soluble recombinant DC-SIGN-Fc and were subsequently stained with FITC-conjugated anti-Fc. Grey lines: autofluorescence of unstained cells. **(D)** SYK phosphorylation at Y525/526 in the FL and DLBCL samples was measured following 15 minutes incubation with soluble DC-SIGN (red line) or anti-Ig (blue line), or no treatment (NT, black line). **(E)** Raji/DC-SIGN cells were treated with 10nM hIgG1-D1 or left untreated (NT) for 30 minutes, and subsequently co-cultured with FL or DLBCL cells. Clustering was determined by inverted fluorescence microscopy.

Figure S10. Progression-free survival and gene expression profile of EZB MYC+ve lymphomas acquiring N-glycosylation sites in the CDR

(A) Progression-free survival in CDR+ve and CDR-ve EZB GCB-DLBCL with MYC translocation (EZB MYC+ve). The number of patients at risk are indicated in blue (CDR-ve) or in red (CDR+ve) at each time point (years). **(B)** Differential gene expression in CDR+ve compared to CDR-ve EZB MYC+ve DLBCL. Each point represents a gene and the fold change and P-value for differential expression in CDR+ve vs CDR-ve EZB MYC+ve DLBCL. Positive $\log_2$ fold change genes were expressed higher in CDR+ve while negative $\log_2$ fold change genes were expressed higher in CDR-ve. Grey points were not differentially expressed genes. Light blue/pink points were differentially expressed at P=0.05. No differentially expressed genes were identified at FDR=0.05 after controlling for multiple testing in the EZB MYC+ve group.

SUPPLEMENTAL METHODS

Primary samples

Primary DLBCL samples were collected from fresh tumor lymph nodes at the Department of Pathology, University of Rochester Medical Center, and at the University of Southampton in the Mature lymphoid malignancies observational study (NIHR/UKCRN Portfolio ID: 31076). Control samples were collected from fresh reactive tonsils and lymph nodes at the University of Southampton. Cell suspensions were cryopreserved in FBS + 10% dimethyl sulfoxide. Before each assay, cells were thawed and viable cells were separated by Lymphoprep (Stemcell Technologies, Cambridge, UK). Diagnosis of GCB-DLBCL vs ABC-DLBCL was performed by immunohistochemistry using Hans algorithm.[1] *BCL2* translocation (MBR, 3'MBR, and mcr breakpoints) was determined by PCR using the BIOMED-2 primers and procedures.[2] Peripheral blood mononuclear cells (PBMCs) were isolated from healthy volunteers by Lymphoprep separation.

Cell lines and culture conditions

WSU-FSCCL cell line was obtained from DSMZ. HBL1, TMD8, SU-DHL6, OCI-Ly1, OCI-Ly7, OCI-Ly19, NU-DHL1, WSU-NHL, and WSU-DLCL2 were provided by Riccardo Dalla-Favera (Columbia University, NY). The Raji cell line transduced with DC-SIGN (Raji/DC-SIGN) and the parental Raji cells were generated and provided by Teunis Geijtenbeek (University of Amsterdam). Raji/DC-SIGN carries a TAC>TAG mutation at codon 103 of the IGKV FR3, leading to a non-productive IG sequence (Table S11), and lack Ig protein expression or DC-SIGN binding on the cell surface. WSU-FSCCL cells were cultured in RPMI 1640 medium (Sigma-Aldrich) supplemented with 7% fetal bovine serum (FBS; Biosera), 0.05mM 2-mercaptoethanol, 2mM glutamine, and 1% penicillin/streptomycin; NU-DHL1 were cultured in RPMI 1640 supplemented with 10% FBS, 0.05mM 2-mercaptoethanol, 2mM glutamine and 1% penicillin/streptomycin; TMD8, Raji and Raji/DC-SIGN were cultured in RPMI 1640 supplemented with 10% FBS, 2mM glutamine and 1% penicillin/streptomycin; HBL1, SU-DHL6, OCI-Ly1, OCI-Ly7, OCI-Ly19, and WSU-DLCL2 were cultured in IMDM medium (Gibco) supplemented with 10% FBS, 2mM glutamine and 1% penicillin/streptomycin.

Phenotypic analyses of primary samples and cell lines

Cells from DLBCL primary samples and lines were stained for 30 minutes on ice with the following antibodies: APC-conjugated anti-CD10, PerCP-Cy5.5-conjugated anti-CD20 (Biolegend, London, UK), PE-conjugated $F(ab')_2$ anti-IgM and FITC-conjugated $F(ab')_2$ anti-IgG (Dako), PE-conjugated anti-kappa and FITC-conjugated anti-lambda (BD Biosciences), or corresponding polyclonal or monoclonal control antibodies. Phenotype of normal PBMCs, reactive tonsils or lymph nodes was performed using the following antibodies: Pacific Blue-conjugated anti-CD19, PerCP-Cy5.5-conjugated anti-CD5, APC-conjugated or PE-conjugated or FITC-conjugated anti-IgD, APC-Cy7-conjugated anti-CD27, APC-conjugated anti-CD10, PerCP-Cy5.5-conjugated anti-CD20, FITC-conjugated anti-CD77 or corresponding control antibodies (Biolegend). Data were acquired on a FACS Canto II (BD Biosciences) and analyzed with FlowJo software (LLC, Ashland, USA).

Surface protein biotinylation

Cell surface proteins were isolated using the Cell Surface Protein Isolation kit (Pierce, Thermo Fisher Scientific), and the glycosylation pattern of the sIg was analyzed as previously described.[3,4] Briefly, surface proteins were biotinylated, isolated from the cell lysate by incubation with Neutravidine agarose slurry and digested by treatment with Endo H (high-mannose specific) or PNGase F (removes all glycans regardless of the composition) (New England Biolabs) for 3 hours at 37°C. Biotinylated proteins were then separated by sodium-dodecyl-sulphate 10% polyacrylamide gel electrophoresis (SDS-PAGE) and analyzed by immunoblotting using a primary rabbit anti-human μ or γ chain antibody (Jackson ImmunoResearch Laboratories).

Immunoglobulin gene analysis from public RNAseq data

The tumor *IGHV-IGHD-IGHJ-IGHC* rearranged transcript sequences from RNAseq data were generated using an analytical pipeline from Blachly *et al*,[5] modified to improve the speed of computation without affecting quality. In particular, data were downloaded from the National Cancer Institute Genomic Data Commons (GDC) portal at https://portal.gdc.cancer.gov/ in the form of aligned BAM files (accession phs001444.v1.p1).[6] Data were converted into FASTQ format using SAMtools v1.6.[7] To reduce the computational resources needed for *de novo* transcriptome assembly, data filtering was performed by alignment of the FASTQ files to the hg38 reference genome using HISAT2,[8] and only reads that were either unmapped or aligned to the *IG* loci as primary or secondary alignments were retained. *De novo* transcriptome assembly was performed using Trinity v2.4.0,[9] with a minimum contig length of 500 bp and without *in silico* normalization. *IG* loci-derived assembled transcripts were identified using BLASTN against *IGH* reference IMGT/LIGM-DB database,[10] and quantified using Kallisto v0.43.1.[11] The transcripts with the 5 highest transcripts per million (TPM) were aligned to IMGT/LIGM-DB using IMGT/V-QUEST alignment tool http://www.imgt.org/IMGT_vquest/input. Those sequences that did not align with any human *IGH* sequence were discarded. Of the remaining sequences, the tumor-derived *IGHV-IGHD-IGHJ-IGHC* rearranged transcript was identified by complying with all of the following criteria in hierarchical order: i) full transcript sequence (from codon 1 in FR1 to codon 129 in FR4 included), ii) 'productive' V-DOMAIN functionality call ('no rearrangement found', 'unproductive', 'no results' calls were discarded), iii) the most abundant (dominant) *IGHV-IGHD-IGHJ-IGHC* rearranged transcript as defined by the highest estimated read count (est. count) and iv) the dominant sequence had to have at least 5-fold higher est. count than additional rearrangements identified, v) identifiable IGHC class and subclass, vi) homology to germline *IGHV*<98%. In 17 independent primary mature B-cell tumor samples (51-92% tumor infiltration by flow cytometry) investigated by RNAseq and Sanger sequencing in Southampton, *IGHV*, *IGHD*, *IGHJ* allele concordance, CDR3 identity, and full 100% *IGHV-IGHD-IGHJ* nucleotide sequence identity between the two methods was observed in 16/17 samples (94% concordance), validating the 94%-100% specificity of the original method by Blachly et al.[5] while RNAseq also confidently identified the tumor IGHC isotype. By this hierarchical approach, the full tumor *IGHV-IGHD-IGHJ-IGHC* rearranged transcripts were confidently identified from 308 primary DLBCL. The mean est. count was 282,789 ± 15,174 SEM and the mean TPM was 518,250 ± 14,847 SEM. Cell of origin and genetic subtype associated with each case were obtained from ref.[6,12]

Immunoglobulin gene analysis from primary lymphoma samples or cell lines

The *IG* heavy and light chain rearrangements from primary lymphoma samples or cell lines were obtained by sequencing following RNA extraction, cDNA preparation, and polymerase-chain-reaction as described previously.[13,14]

IGHV, *IGHD,* and *IGHJ* use, and *IGV* homology to the closest germline sequence were determined using the IMGT database and IMGT/V-QUEST alignment tool as above. The deduced amino acid sequences were scanned to identify the acquired N-glycosylation sites (NxS/T; x $\neq$ P),[15,16] and the position of the sites was defined according to the IMGT unique Lefranc numbering at http://www.imgt.org/IMGTScientificChart/Numbering/IMGT-Kabat_part1.html. Those sites which were across the FR/CDR border were assigned to the CDR.

Genetic variant analysis

Published DLBCL genetic variant data were downloaded from the National Cancer Institute Genomic Data Commons at https://portal.gdc.cancer.gov/ (accession phs001444.v1.p1) in the form of filtered and annotated MAF files and tables of summarized translocations, fusions, and DHIT lymphoma status generated as per Wright *et al*.[12] Downloaded files were processed using custom UNIX and R code and genetic variant plots were produced using GenVisR.[17]

RNAseq data analysis

RNAseq data were downloaded in the form of aligned bam files from the National Cancer Institute Genomic Data Commons (accession phs001175), converted to fastq files using SAMtools,[7] quality checked using FastQC (Babraham Bioinformatics, Cambridge, UK) then aligned to the hg38 reference genome using HISAT2.[8] Aligned data was used to produce tables of read counts using HTseq-count against Ensembl GRCh38 v94 gene annotations.[18,19] Tables of raw counts were analyzed for differential expression using EdgeR v3.30 and custom R code.[20,21] Gene set enrichment analysis was performed with fGSEA using the ranked $\log_2$ fold change (positive to negative) gene list from CDR+ve vs CDR-ve EZB MYC-ve (within GCB-DLBCL) differential expression output and gene sets from MSigDB v7.2.[22,23]

Fab generation

Plasmids encoding lymphoma IG heavy and light chain, with the heavy chain containing a C-terminal His-tag, were transiently co-transfected at a ratio of 1:1 in HEK293F cells, known to reproduce the Ig mannosylation pattern of primary cells, [24] using OptiPro SFM and FreeStyle Max reagent (Invitrogen). The cells were transfected at a density of $1x10^6$ cells/ml and incubated for 5 days at 37°C with 8% CO_2 and 125 rpm shaking using a New Brunswick™ S41i (Eppendorf) incubator. To specifically boost the expression levels of L29, MEXi293E cells (IBA Lifesciences) were used instead of 293F. Plasmids encoding lymphoma IG heavy and light chain, with the heavy chain containing a C-terminal His-tag, were transiently co-transfected at a ratio of 1:1 using 1.25µg/ml of each plasmid, and 7.5µg/ml of PEI MAX transfection reagent and the cells were incubated for 7 days at 37°C with 8% CO2 and 125rpm shaking.

For glycopeptide mass spectrometry, Fabs were denatured, reduced, and alkylated at room temperature in 50mM Tris/HCl, pH 8.0 containing 6M urea and 5mM dithiothreitol

for 1 hour, followed by 20mM iodacetamide addition for another hour in the dark, and then DTT (20mM). Fabs were buffer-exchanged into 50mM Tris/HCl, pH 8.0 using Vivaspin columns, and made available for the spectrometry.

Nickel affinity purification

After harvesting, the cells were spun down at 3220g for 10 minutes and the supernatant applied to a 250ml Stericup-HV sterile 0.45µm filter (MilliPore). His-tagged lymphoma Fabs were purified using nickel affinity 5ml HisTrap FF columns (GE Healthcare). Imidazole was added to the filtered supernatant to a final concentration of 20 mM to minimize non-specific interactions. The column was first washed with 10 column volumes of washing buffer (50mM sodium phosphate, 300mM NaCl) at pH 7 before loading the sample onto the column. The column was then washed with washing buffer (10 column volumes) and the Fabs eluted in 5 column volumes of elution buffer (300mM imidazole in washing buffer). The elution was concentrated by a Vivaspin column (10 kDa cut-off) to a volume of 1ml.

Size exclusion chromatography (SEC)

Lymphoma Fabs were further purified using SEC on an Akta Start system (GE Healthcare). A Superdex 75pg 16/600 column was washed with 10mM Tris (pH 8.0) and 150mM NaCl at a rate of 1ml/minute, then 1ml of the nickel-affinity purified material was injected onto the column. Fractions separated by SEC were pooled according to their corresponding peaks on the size exclusion spectra. Both samples were then analyzed using a SDS-PAGE gradient gel (4-12%).

In-gel digestion of N-glycans

The SDS-PAGE gel bands corresponding to the lymphoma Fabs were excised and washed with 100% acetonitrile and water three times. The N-glycans ion mobility mass spectrometry analyses were then enzymatically released by within-gel digestion by PNGase F (New England BioLabs) for 16 hours at 37°C.

Determination of L14 Fab crystal structure

Purified IgG L14 Fab fragment was concentrated to 9.5 mg/ml and crystallized at room temperature using the sitting drop vapor diffusion method using a 1:1 ratio of precipitant (100nl) to protein (100nl). Crystals appeared after 238 days in a precipitant containing 25% (w/v) polyethylene glycol 3350, 0.2M magnesium chloride, and 0.1M bis-Tris pH 5.5, and were flash-frozen in liquid nitrogen with a cryoprotectant containing 25% glycerol (v/v) plus precipitant. X-ray diffraction data were recorded at beamline I04 at Diamond Light Source (Oxfordshire, UK). These data were indexed, integrated, and scaled with XIA2.[25] The structure of L14 Fab was solved by molecular replacement with PHASER[26] using the structure of 5d3 Fab (PDB ID: 5NIV) as a search model. Prior to iterative model building and refinement in COOT,[27] and Refmac5,[28] respectively, an initial model was built automatically in ARP/wARP.[29] MolProbity was used to validate the model and to allow continuous connectivity of the protein chain.[30]

Monocyte isolation and dendritic cell differentiation

Monocytes were isolated from PBMCs using the $CD14^+$ Microbeads kit (Miltenyi Biotec) and cultured for 6 days in RPMI1640 with 10% FBS, 800UI/ml GM-CSF and 500UI/ml IL-4 (R&D Systems) to allow differentiation into dendritic cells (MoDCs), as previously described.[31] MoDC phenotype was verified by flow cytometry prior to each assay by staining with the following antibodies: APC-conjugated anti-CD209, Brilliant Violet 421-conjugated anti-CD14, PE-conjugated anti-CD1a, FITC-conjugated anti-CD206, FITC-conjugated anti-CD163 (all from Biolegend) and with corresponding isotype control antibodies.

Intracellular signaling analysis

DLBCL primary cells were plated in U-bottom 96-well plates and left for 30 minutes on ice either with 20µg/ml DC-SIGN-Fc or with soluble polyclonal goat $F(ab')_2$ anti-human IgM or IgG (Sothern Biotech) or untreated. Cells were then incubated at 37°C for 1, 5 or 15 minutes to allow intracellular signaling. When specified, DC-SIGN-Fc was pre-incubated with 500nM hIgG1-D1 for 30 minutes on ice before cell treatment. For Phosflow studies, cells were fixed with paraformaldehyde (final concentration 1.6%) for 5 minutes at room temperature, washed in PBS and permeabilized with ice-cold 90% methanol for 10 minutes on ice. Cells were washed in PBS + 1% BSA and stained with Alexa Fluor 488-conjugated anti-pSYK (Y525/526) antibody (clone C87C1) (Cell Signaling Technology), PerCP-Cy5.5-conjugated anti-CD20 (cloneH1/FB1) antibody and PE-conjugated anti-Bcl-2 antibody (BD Biosciences). Data were acquired on a FACS Canto II.

For western blot analysis, cells were washed in ice-cold PBS and lysed in RIPA buffer (150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCL, pH 8.0) supplemented with 1X protease inhibitor and phosphatase inhibitors (Sigma-Aldrich). Protein concentration was measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). After incubation for 5 minutes at 95°C in Red Loading buffer and dithiothreitol (DTT; Cell Signaling Technology), equal amounts of proteins were resolved by SDS-PAGE and transferred to nitrocellulose membranes (GE Healthcare). Membranes were blocked with 5% BSA in Tris-buffered saline, before incubation with the following primary antibodies: anti-phosphoAKT (Ser473), anti-AKT, anti-phosphoERK (Thr202/Tyr2014), anti-ERK (all from Cell Signaling Technology), and anti-GAPDH (6C5) (Ambion). Subsequently, membranes were washed and stained with horseradish peroxidase (HRP)-conjugated secondary antibodies (Dako), and images were acquired after exposure to chemiluminescence reagents (Pierce SuperSignal West Pico, Thermo Fisher Scientific) using the ChemiDoc-It imaging system (UVP).

Statistical analysis

Categorical variables were compared with Fisher's exact test. Continuous variables were compared by the Mann-Whitney non-parametric tests. For clinical correlation studies of CDR+ve and CDR-ve DLBCL, the clinical data associated with the National Cancer Institute Genomic Data Commons project were obtained from Wright *et al*.[12] MYC status in the EZB subtype was obtained from Schmitz et al.[6] which classified as MYC+ve any EZB lymphoma which had a MYC translocation by FISH break-apart probe, a MYC mutation, a MYC fusion transcript (by FusionCatcher), or a MYC amplification. Progression-free survival was calculated by Kaplan-Meier method using log-rank statistics. Analyses were performed with GraphPad Prism software version

8.2.1 or the Statistical Package for the Social Sciences (SPSS) software v.27.0. All statistical tests were 2-sided. Statistical significance was defined as *P*-value <0.05.

SUPPLEMENTAL REFERENCES

1. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 2004;103:275-82.
2. van Dongen JJ, Langerak AW, Bruggemann M, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 2003;17:2257-317.
3. Coelho V, Krysov S, Ghaemmaghami AM, et al. Glycosylation of surface Ig creates a functional bridge between human follicular lymphoma and microenvironmental lectins. Proc Natl Acad Sci U S A 2010;107:18587-92.
4. Drennan S, Chiodin G, D'Avola A, et al. Ibrutinib therapy releases leukemic surface IgM from antigen drive in chronic lymphocytic leukemia patients. Clin Cancer Res 2018.
5. Blachly JS, Ruppert AS, Zhao W, et al. Immunoglobulin transcript sequence and somatic hypermutation computation from unselected RNA-seq reads in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 2015;112:4322-7.
6. Schmitz R, Wright GW, Huang DW, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N Engl J Med 2018;378:1396-407.
7. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
8. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37:907-15.
9. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644-52.
10. Giudicelli V, Duroux P, Ginestoux C, et al. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. Nucleic Acids Res 2006;34:D781-4.
11. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 2016;34:525-7.
12. Wright GW, Huang DW, Phelan JD, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. Cancer Cell 2020;37:551-68 e14.
13. Forconi F, Sahota SS, Raspadori D, Mockridge CI, Lauria F, Stevenson FK. Tumor cells of hairy cell leukemia express multiple clonally related immunoglobulin isotypes via RNA splicing. Blood 2001;98:1174-81.
14. Forconi F, Sozzi E, Rossi D, et al. Selective influences in the expressed immunoglobulin heavy and light chain gene repertoire in hairy cell leukemia. Haematologica 2008;93:697-705.
15. Forconi F, Capello D, Berra E, et al. Incidence of novel N-glycosylation sites in the B-cell receptor of lymphomas associated with immunodeficiency. Br J Haematol 2004;124:604-9.
16. Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK. Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. Blood 2002;99:2562-8.
17. Skidmore ZL, Wagner AH, Lesurf R, et al. GenVisR: Genomic Visualizations in R. Bioinformatics 2016;32:3012-4.
18. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166-9.
19. Hunt SE, McLaren W, Gil L, et al. Ensembl variation resources. Database (Oxford) 2018;2018.
20. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 2012;40:4288-97.

21. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.
22. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
23. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003;34:267-73.
24. McCann KJ, Ottensmeier CH, Callard A, et al. Remarkable selective glycosylation of the immunoglobulin variable region in follicular lymphoma. Mol Immunol 2008;45:1567-72.
25. Winter G. xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography 2010;43:186-90.
26. McCoy AJ, Gross-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst 2007;40:658-74.
27. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126-32.
28. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240-55.
29. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. ARP/wARP and molecular replacement. Acta Crystallogr D Biol Crystallogr 2001;57:1445-50.
30. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:W375-83.
31. Gringhuis SI, den Dunnen J, Litjens M, van der Vlist M, Geijtenbeek TB. Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to Mycobacterium tuberculosis, HIV-1 and Helicobacter pylori. Nat Immunol 2009;10:1081-8.

Figure S1

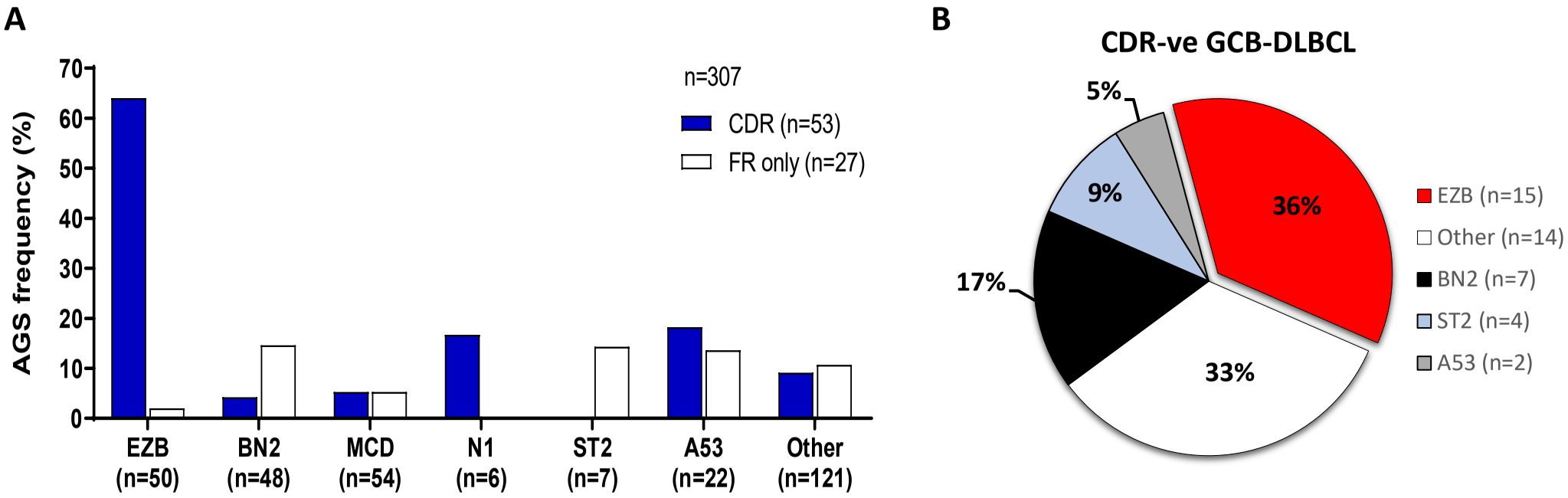
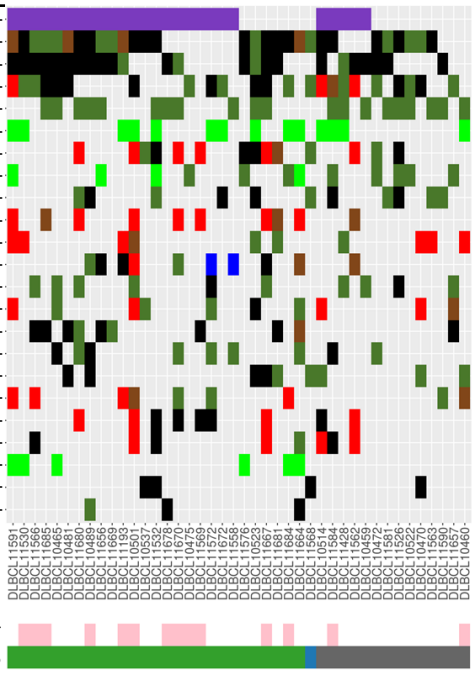

A
AGS frequency (%)
70
60
50
40
30
20
10
0
n=307
CDR (n=53)
FR only (n=27)
EZB (n=50)
BN2 (n=48)
MCD (n=54)
N1 (n=6)
ST2 (n=7)
A53 (n=22)
Other (n=121)
B
CDR-ve GCB-DLBCL
5%
9%
36%
17%
33%
EZB (n=15)
Other (n=14)
BN2 (n=7)
ST2 (n=4)
A53 (n=2)

Figure S2

A

Mutation frequency (%)			Gene
total	CDR+ve	CDR-ve	
61.9	77.8	33.3	BCL2 fusion
69.0	74.1	60.0	TNFRSF14
57.1	63.0	46.7	KMT2D
54.8	48.1	66.7	CREBBP
47.6	40.7	60.0	EZH2
33.3	37.0	26.7	REL
33.3	37.0	26.7	FAS
28.6	25.9	33.3	STAT6
26.2	18.5	40.0	IRF8
23.8	33.3	6.7	PTEN
23.8	22.2	26.7	MAP2K1
23.8	33.3	6.7	GNA13
23.8	22.2	26.7	EP300
23.8	25.9	20.0	CIITA
23.8	33.3	6.7	ARID1A
21.4	25.9	13.3	MEF2B
21.4	18.5	26.7	HLA-DMB
21.4	25.9	13.3	FBXO11
21.4	25.9	13.3	C10orf12
19.0	18.5	20.0	EBF1
14.3	22.2	0.0	MIR17HG
9.5	7.4	13.3	S1PR2
7.1	11.1	0.0	BCL7A



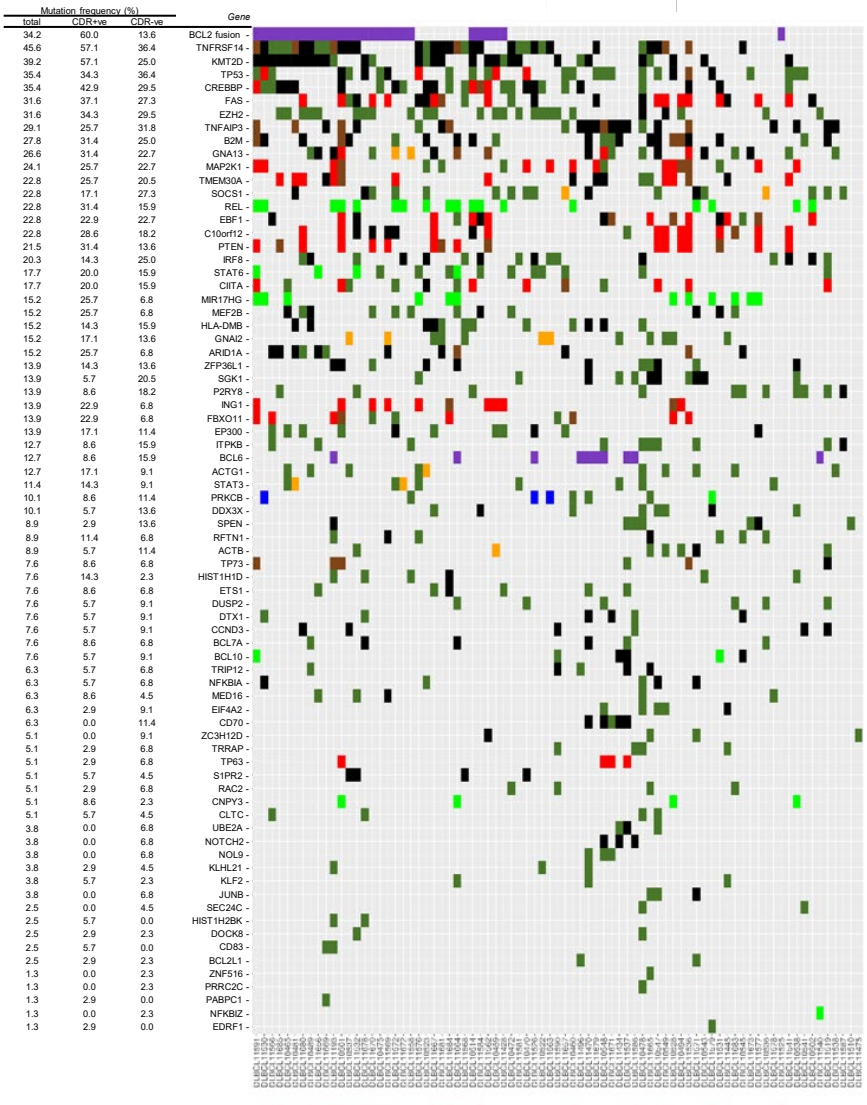
Mutation type
FUSION
TRUNC
MISSENSE
HOMDEL
HETLOSS
AMP
INFRAME

DHIT status
NEG (n=28)
POS (n=14)

AGS status
CDR (n=27)
FR_only (n=1)
No_AGS (n=14)

Sample n=42

B



Mutation type
FUSION
TRUNC
MISSENSE
HOMDEL
HETLOSS
AMP
GAIN
INFRAME

DHIT status
NEG (n=61)
POS (n=18)

AGS status
CDR (n=35)
FR_only (n=9)
No_AGS (n=35)

Subtype
EZB (n=42)
BN2 (n=8)
ST2(n=4)
A53 (n=4)
Other (n=21)

Sample n=79

Figure S3

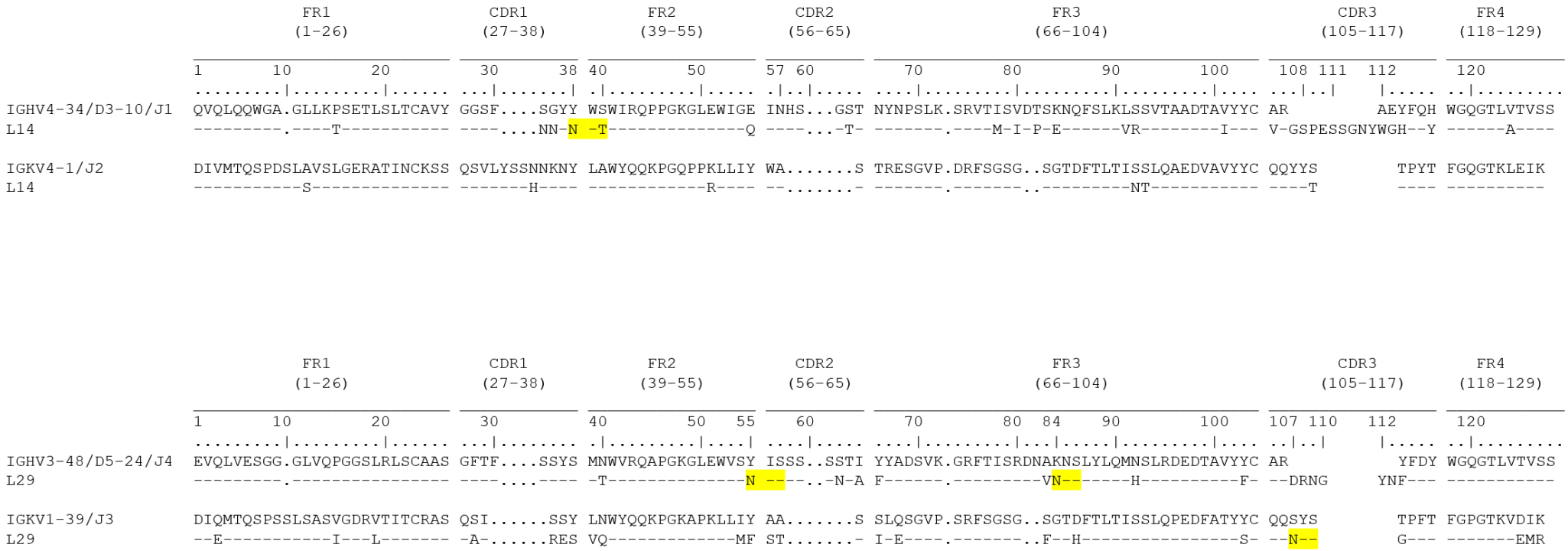

FR1 (1-26)
CDR1 (27-38)
FR2 (39-55)
CDR2 (56-65)
FR3 (66-104)
CDR3 (105-117)
FR4 (118-129)
IGHV4-34/D3-10/J1
L14
IGKV4-1/J2
L14
IGHV3-48/D5-24/J4
L29
IGKV1-39/J3
L29

Figure S4

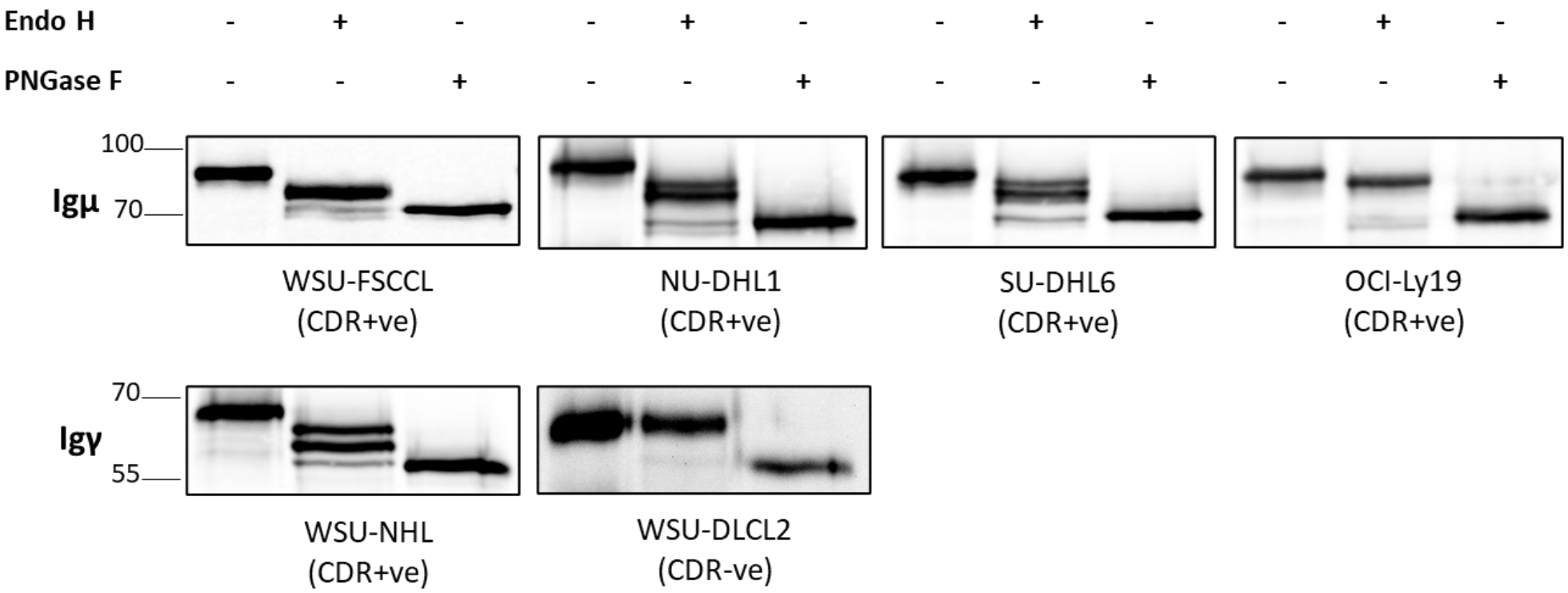

Endo H - + - - + - - + - - + -
PNGase F - - + - - + - - + - - +
100
Igμ
70
WSU-FSCCL (CDR+ve)
NU-DHL1 (CDR+ve)
SU-DHL6 (CDR+ve)
OCI-Ly19 (CDR+ve)
70
Igγ
55
WSU-NHL (CDR+ve)
WSU-DLCL2 (CDR-ve)

Figure S5

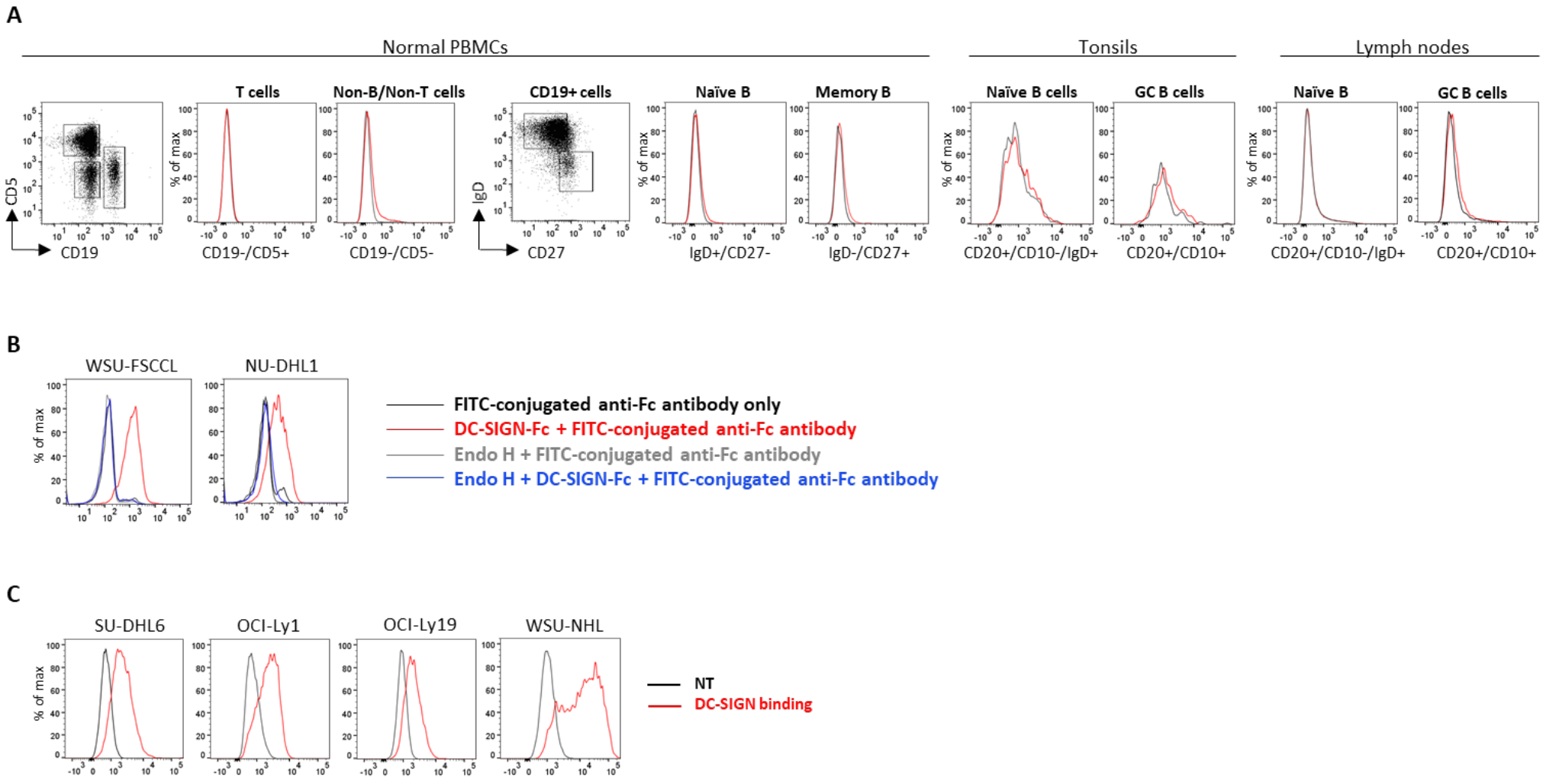
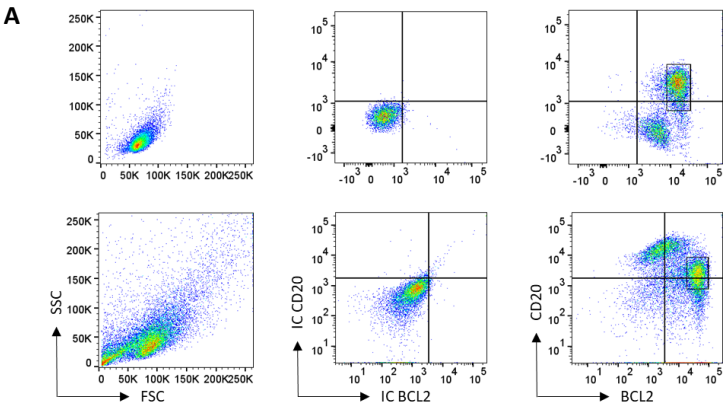
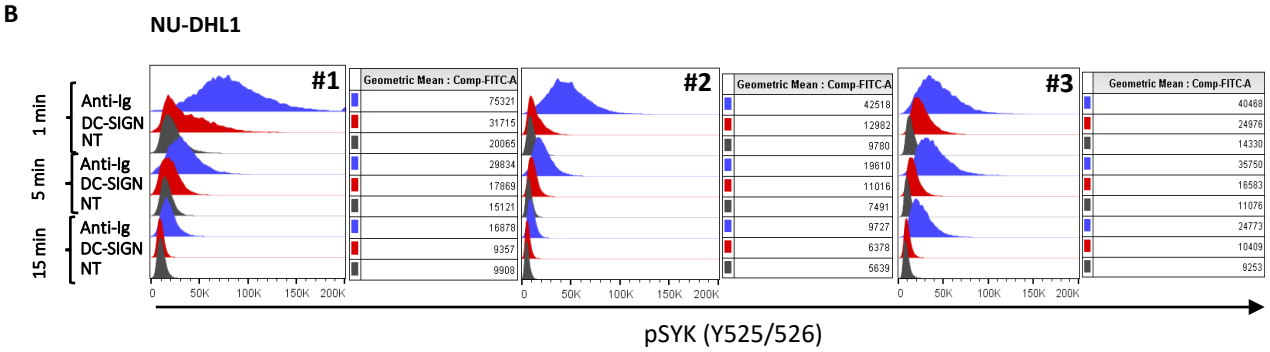
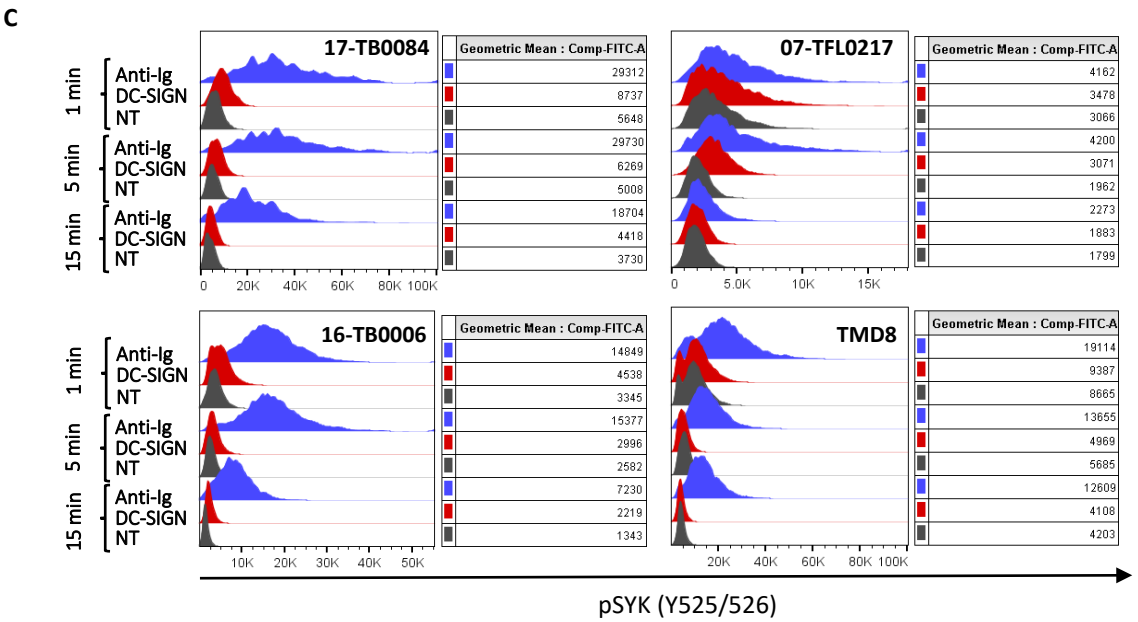

A
Normal PBMCs
Tonsils
Lymph nodes
T cells
Non-B/Non-T cells
CD19+ cells
Naïve B
Memory B
Naïve B cells
GC B cells
Naïve B
GC B cells
CD5
CD19
% of max
CD19-/CD5+
CD19-/CD5-
IgD
CD27
IgD+/CD27-
IgD-/CD27+
CD20+/CD10-/IgD+
CD20+/CD10+
CD20+/CD10-/IgD+
CD20+/CD10+
B
WSU-FSCCL
NU-DHL1
% of max
FITC-conjugated anti-Fc antibody only
DC-SIGN-Fc + FITC-conjugated anti-Fc antibody
Endo H + FITC-conjugated anti-Fc antibody
Endo H + DC-SIGN-Fc + FITC-conjugated anti-Fc antibody
C
SU-DHL6
OCI-Ly1
OCI-Ly19
WSU-NHL
% of max
NT
DC-SIGN binding

Figure S6


A
SSC
FSC
IC CD20
IC BCL2
CD20
BCL2


B
NU-DHL1
#1
#2
#3
Geometric Mean : Comp-FITC-A
1 min
5 min
15 min
Anti-Ig
DC-SIGN
NT
pSYK (Y525/526)


C
17-TB0084
07-TFL0217
16-TB0006
TMD8
Geometric Mean : Comp-FITC-A
1 min
5 min
15 min
Anti-Ig
DC-SIGN
NT
pSYK (Y525/526)

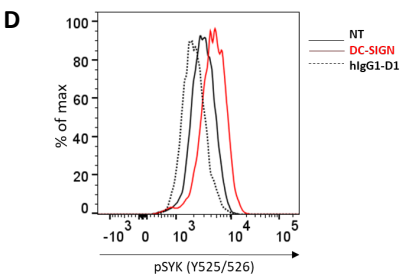

D
% of max
NT
DC-SIGN
hIgG1-D1
pSYK (Y525/526)

Figure S7

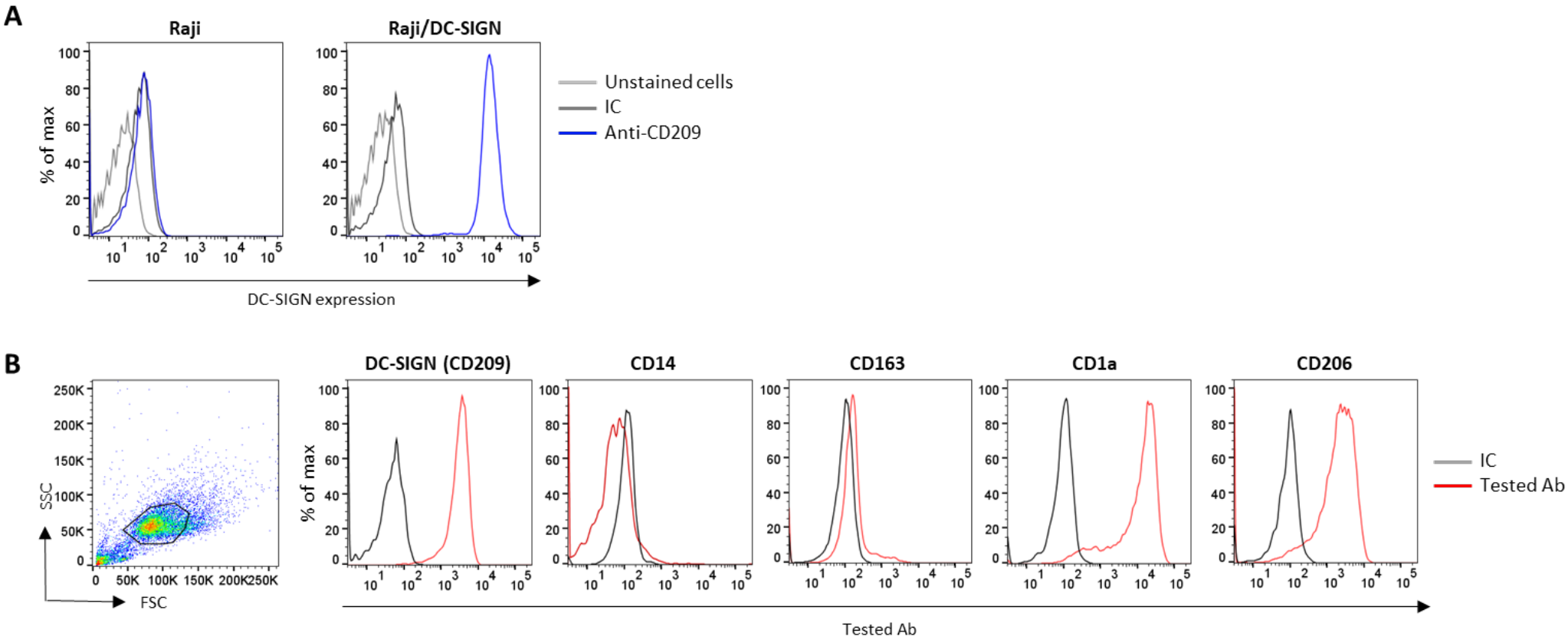

A
Raji
Raji/DC-SIGN
% of max
DC-SIGN expression
Unstained cells
IC
Anti-CD209
B
SSC
FSC
DC-SIGN (CD209)
CD14
CD163
CD1a
CD206
% of max
Tested Ab
IC
Tested Ab

Figure S8

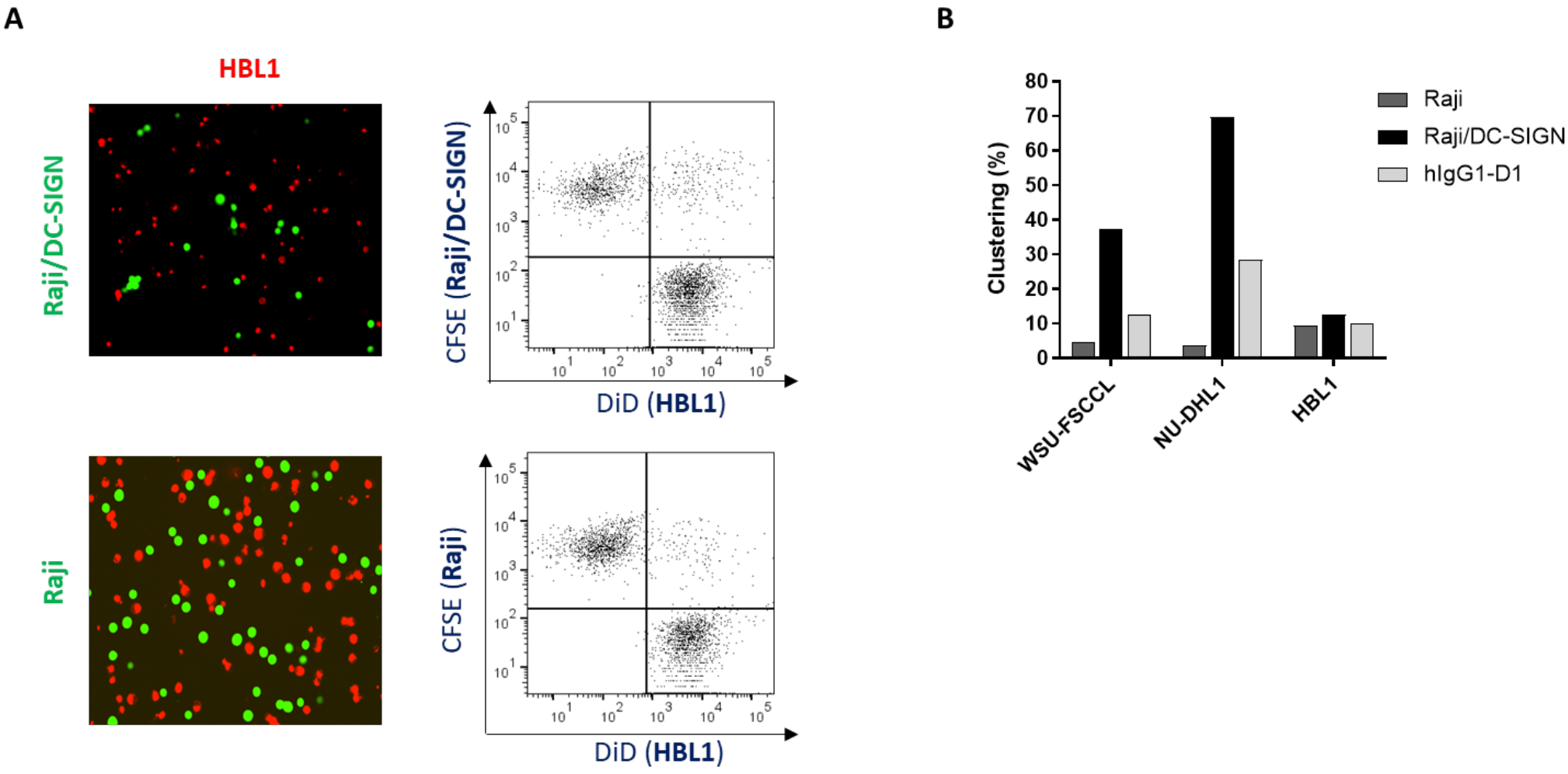

A
HBL1
Raji/DC-SIGN
Raji
CFSE (Raji/DC-SIGN)
CFSE (Raji)
DiD (HBL1)
DiD (HBL1)
B
Clustering (%)
80
70
60
50
40
30
20
10
0
Raji
Raji/DC-SIGN
hIgG1-D1
WSU-FSCCL
NU-DHL1
HBL1

Figure S9

A

```
                   FR1                          CDR1          FR2              CDR2       FR3                                            CDR3                FR4
                   (1-26)                       (27-38)       (39-55)          (56-65)    (66-104)                                       (105-117)           (118-129)
                   1        10        20        30            40        50     60         70        80        90        100              111   112           120
IGHV3-23/D3-3/J4   EVQLLESGG.GLVQPGGSLRLSCAAS   GFTF....SSYA  MSWVRQAPGKGLEWVSA ISGS..GGST YYADSVK.GRFTISRDNSKNTLYLQMNSLRAEDTAVYYC AK            YFDY  WGQGTLVTVSS
FL  (13-265)       D---M----.-------------IS-   T---....---V  -T---------------  --SH..-LA- ------R.---V--------V----H---LD------- TTNATDFDFWSGNG--H   ----------
DLBCL(16-13)       D---M----.D------------IS-   T---....---V  -T--------------T  --SH..-LA- ------R.---V--------V----H---LD------- TTNATDFDFWSGNG--H   ----------
```

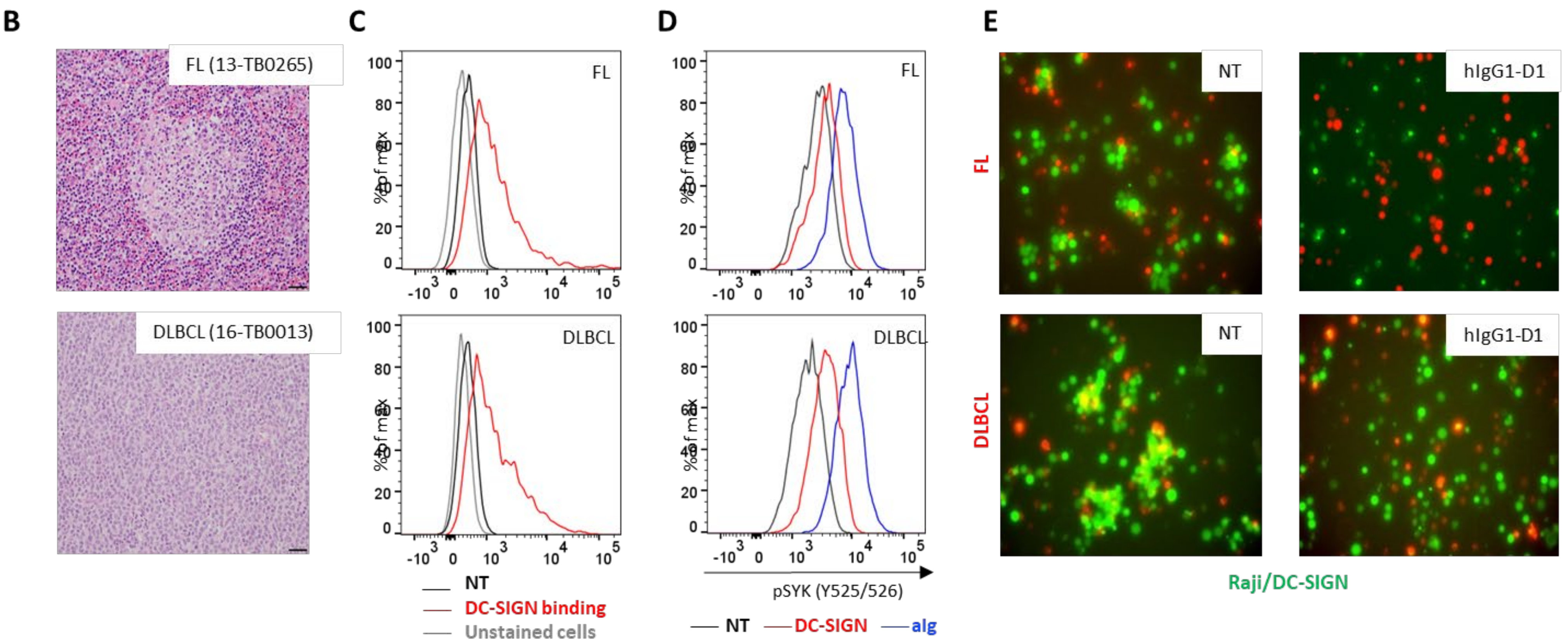
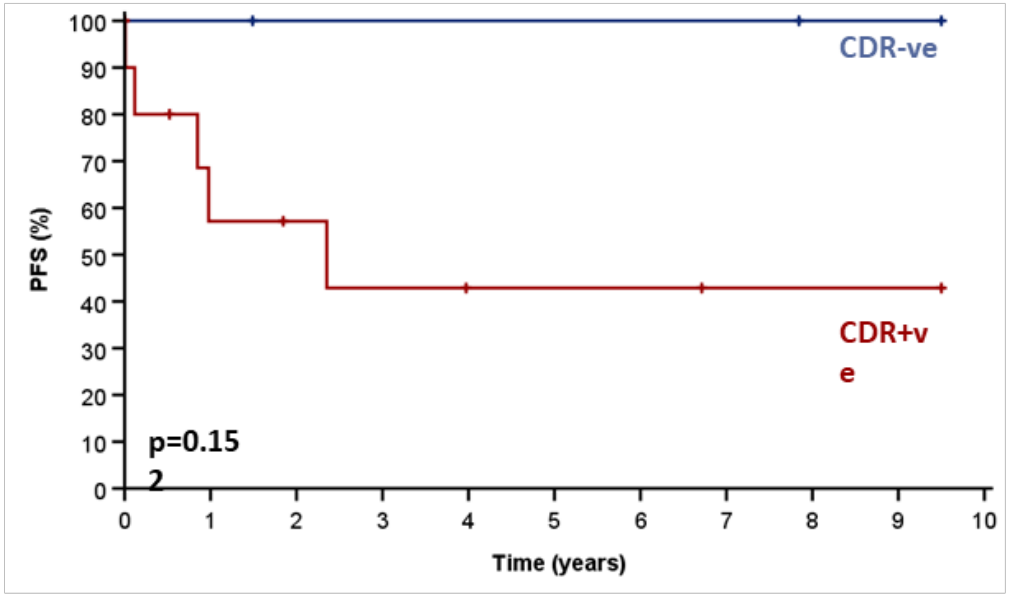
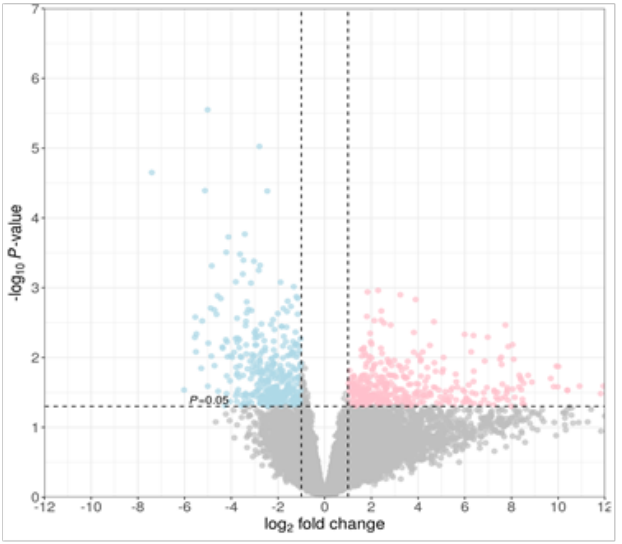


Figure S10

A



B



AGS	Total	Progressed (%)	Median PFS (years)	SE
CDR-ve	3	0 (0)	Not reached	-
CDR+ve	12	5 (41.7)	2.35	1.68
Overall	15	5 (33.3)	Not reached	-

Differentially expressed genes in CDR+ve vs CDR-ve EZB MYC+ve	
Upregulated	0
Downregulated	0
Not significant	20674