Supplemental Methods

IRF8 is a transcriptional activator of CD37 expression in diffuse large B-cell lymphoma

Suraya Elfrink[1], Martin ter Beest[1,#], Luuk Janssen[1,2,#], Marijke P. Baltissen[3], Pascal W. T. C. Jansen[3], Angelique N. Kenyon[1], Raymond M. Steen[1], Daynelys de Windt[1], Philipp M. Hagemann[1], Corine Hess[4], Dick-Johan van Spronsen[4], Brigiet Hoevenaars[5], Ellen van der Spek[6], Zijun Y. Xu-Monette[7], Ken H. Young[7], Charlotte Kaffa[8], Sander Bervoets[8], Jolien van Heek[9], Eva Hesius[9], Charlotte M. de Winde[1,10], Michiel Vermeulen[3], Michiel van den Brand[2,11], Blanca Scheijen[2], Annemiek B. van Spriel[1,*]

Sequencing of *CD37* promoter region

Genomic DNA was extracted from CD37-negative (n=12) formalin-fixed paraffin-embedded or frozen diffuse large B-cell lymphoma tumor sections. CD37 protein expression in the tumor samples was determined previously (Elfrink et al. *Blood* 2019). The region approximately 1.5kb upstream of *CD37* (chr19:49,837,166-49,838,841 (GRCh37/hg19)) was amplified with AmpliTaq Gold® DNA Polymerase (Applied Biosystems/ThermoScientific) in six or seven polymerase chain reactions (PCRs), using the primers described in the table below. All products allowed for Sanger sequencing using an M13 primer. Sequences were analyzed using SnapGene version 3.2.1 (GSL Biotech).

Primers for amplification and sequencing of the *CD37* upstream locus in DLBCL tumor samples.

	Forward primer sequence (5' -> 3')	**Reverse primer sequence (5' -> 3')**
#1	TGTAAAACGACGGCCAGTTGACTGTCAGGATCCCAGG	CAGGAAACAGCTATGACCTTTCTGGGCTCAAGTGATCC
#2*	TGTAAAACGACGGCCAGTGAAAAGAAGACAGGCTGGGC	CAGGAAACAGCTATGACCGAACTCCTGGACTCAAGCGA
#2a	TGTAAAACGACGGCCAGTGAAAAGAAGACAGGCTGGGC	GTCCTTCCTCCTTTCTCTCTTCA
#2b	TGTAAAACGACGGCCAGTCTCGAGCCTGGGAAGGTTG	CAGGAAACAGCTATGACCACTCCTGGACTCAAGCGACC
#3	TGTAAAACGACGGCCAGTAAGGGGCCAGAAGCAGAGAG	CAGGAAACAGCTATGACCGAGTACGGTGGGGAAAACAG
#4	TGTAAAACGACGGCCAGTCTCAGAAGGCAGAGGTGGG	CAGGAAACAGCTATGACCAGCTGTGTCTCCCATTTCCT
#5	TGTAAAACGACGGCCAGTTAAGGGGACAGAGATGGGGA	CAGGAAACAGCTATGACCGGACACATGCTCTGGTACCT
#6	TGTAAAACGACGGCCAGTGACTGCCCCATGAGTCCTT	CAGGAAACAGCTATGACCAGGAAGTACTTGATGAGGCTGA
M13 sequencing primers	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC

* Amplification of this locus was done by either one (#2) or two (#2a and #2b) polymerase chain reactions.

***CD37* exon sequencing**

Genomic DNA of cell lines was isolated using the ISOLATE II Genomic DNA Kit (Bioline) according to manufacturer's protocol. *CD37* exons were amplified from the genomic DNA of OCI-Ly19 and SU-DHL-6 cells and subsequently Sanger sequenced using the primers indicated in Supplemental Table S1. Sequences were analyzed using SnapGene version 3.2.1 (GSL Biotech).

Targeted DNA methylation analysis

DNA methylation was assessed using bisulfite sequencing PCR (BSP) as follows: genomic DNA isolated from cell lines and DLBCL tumor samples were converted using EZ DNA Methylation-Gold™ Kit (Zymo Research) according to manufacturer's protocol. Converted DNA was amplified using AmpliTaq Gold™ DNA Polymerase (Applied Biosystems/Thermo Fisher Scientific) and the primers stated in the table below, designed using Methyl Primer Express software v1.0 (Applied Biosystems/Thermo Fisher Scientific). For region A and C a subsequent nested PCR was performed to increase the yield. Using the A-overhang created by the DNA polymerase, PCR products were ligated into pGEM-T Easy vector (Promega). Single clones were Sanger sequenced by means of M13 primers. Sequences were analyzed using SnapGene version 3.2.1 (GSL Biotech).

Primers for bisulfite sequencing PCR.

	Forward primer sequence (5' -> 3')	**Reverse primer sequence (5' -> 3')**
Region A	GGAGTTTTGATTTTATGATTAGATAGATGGTGGGG	AACTCRAAATAAAAAAATCCCTAAAAATAAAAAAAAC
Region A nested	TTGATTTTATGATTAGATAGATGGTGGGGTTGTTT	AAATAAAAAAATCCCTAAAAATAAAAAAAACRCTAAACC
Region B	TGAGTGGATYGTTTATTTTATTAGGTGAAG	CCCAACCCCCCTCACCCACAAAAAAC
Region C	AAGGATTTTGGGTTTGTTTTYGATTT	TATAAACRTAATCCAAATTTCTACAC
Region C nested	TTTTGGGTTTGTTTTYGATTTTATTT	CRTAATCCAAATTTCTACACAAAAAT
M13 sequencing primers	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC

Quantitative RT-PCR

Total RNA was isolated from frozen sections (10µm) of human DLBCL tumors and tonsil of a healthy donor, using TRIzol reagent (ThermoScientific) followed by phenol/chloroform extraction using

standard procedure. After DNase I (amplification grade; Invitrogen) treatment, equal amounts of RNA were used for cDNA preparation using Moloney murine leukemia virus (M-MLV) reverse transcriptase (Thermo Fisher Scientific) with random hexamer primers (Sigma-Alderich). cDNA was analyzed by qRT-PCR analysis for expression of *IRF8* and *CD37* with PowerUp Sybr Green Master Mix (ThermoScientific) using Quantstudio 3 PCR system (ThermoScientific). For each sample technical replicates were taken along and results were normalized to GAPDH expression. Primer sequences are listed in Supplemental Table S7.

Retroviral transduction of IRF8

IRF8 was subcloned into pBabe-puro[1] using standard cloning methods. 293FT cells were transfected with pBabe-puro-IRF8, pCMV-VSVG and pMD-MLVogp using Lipofectamine 3000 (ThermoFisher) to generate retrovirus. After 24 hours, supernatant of the transfected 293FT cells was collected, filtered and added to the OCI-Ly19 cells together with polybrene (Sigma/Merck) at 8 µg/mL. Fresh medium was added to the transfected 293FT cells and the previous step was repeated. After 24 hours incubation, the transduced OCI-Ly19 cells were incubated with selection medium (0.2 µg/mL of puromycin (Invivogen), and cells were analyzed after two weeks of incubation.

Western blot

Cells were lysed in NP40 lysis buffer (200mM NaCl, 100 mM NaF, 20mM Tris-HCl pH8.0, 0.5% NP-40, 5mM EDTA, 1mM Na3VO4, 1mM PMSF and complete protease inhibitor), centrifuged and supplemented with Laemmli loading buffer. Samples were boiled at 95°C for 5 minutes, separated on a 10% SDS-PAGE gel, and blotted onto a polyvinylidene fluoride (PVDF) membrane. Membranes were incubated with mouse anti-IRF8 (clone E-9, Santa Cruz Biotechnology) and mouse anti-γ-tubulin (GTU-88, Sigma-Aldrich) antibody sequentially, both directly followed by goat-anti-mouse-AlexaFluor790 (Invitrogen) and imaged on an Odyssey CLx (LI-COR). Western blots were analyzed with the Image Studio software (version 5.0, LI-COR). For detection of PU.1, cell lysates were prepared in 1% SDS and

10 mM Tris-HCl (pH 6.8) supplemented with protease inhibitors (Roche). Protein concentration was determined and 30 µg was loaded on a 10% SDS-PAA gel. After transfer on PVDF, the blot was blocked with Odyssey Intercept TBS blocking buffer. Blots were probed overnight with rabbit-anti-PU.1 (Cell Signaling Technology) diluted 1:1000 in Intercept blocking buffer supplemented with 0.1% Tween-20. After extensive washing, the blots were probed with goat-anti rabbit IRDye800 W (Li-Cor) and scanned with either Odyssey CLx or a Typhoon (GE) scanner. Blots were reprobed with rat-anti-alpha-Tubulin antibody (YOL1/34, Novus Biologicals) at 1:2000 dilution followed by detection with goat-anti-rat IgG AlexaFluor680 (ThermoScientific).

Correlation analysis of IRF8 protein with *CD37* mRNA expression levels

The 20Q4 release of the Cancer Dependency Map (https://depmap.org/portal/) dataset was used to investigate IRF8 protein and *CD37* mRNA expression in non-Hodgkin lymphoma cell lines.[2,3] The dataset included a proteome dataset and a dataset with information about TPM from RNA seq studies.[4] Data was analysed with R and Pearson Correlation Coefficient was calculated with the *cor.test* function from the stats package version 4.1.1.

***IRF8* mutation analysis in correlation with *CD37* expression levels**

From a public available RNA-Seq dataset (GOYA DLBCL clinical trial; accession: GSE125966), 33 SRA files were downloaded from the SRA repository with ID SRP183071, using prefetch (version 2.9.2), and converted to fastq files with fastq-dump (version 2.9.2). These fastq files were analyzed for variants using the CalliNGS-NF pipeline (commit version 9e7ebf0). References were gathered from the GATK resource bundle for human genome assembly 38. After variant calling the variants were annotated using Gvanno (version 1.4.3) with their grch38 data bundle. Based on the severity of the impact, high impact variants with frameshift or truncating mutations were considered mutated, where the detected variants samples were grouped into mutated and non-mutated samples. This resulted in 9 mutated samples with a variant allele frequency of minimal 15% of the reads and 24 non-mutated samples.

With DESeq2 (version 1.22.2) counts were normalized and a linear model was fit with stats (version 3.6.2).

Gene expression profiling case selection

The 498 cases with gene expression profiling data deposited in GSE31312 were from the 560 cases of the discovery cohort (Xu-Monette *et al. Blood* 2016) for which *CD37* and *IRF8* GEP data were available.

Statistics

Statistical analyses were performed as indicated in the respective figure legends. (un)paired t-tests and one- or two-way ANOVA with multiple comparison post-hoc correction were used to compare two or multiple groups of samples with continuous data, respectively. Fisher's exact test and Chi-squared test were applied to compare two or multiple groups of categorical data, respectively. Linear correlations were determined using the Pearson correlation coefficient.

Visual abstract

Visual abstract was created with BioRender (Biorender.com).

References

1. Morgenstern JP, Land H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* 1990;18(12):3587.
2. Depmap Broad: Depmap achilles 21q4 public. Cambridge, MA: Broad Institute; 2021.
3. Tsherniak A, Vazquez F, Montgomery PG, et al. Defining a Cancer Dependency Map. *Cell*. 2017;170(3):564-576.e16.
4. Nusinow DP, Szpyt J, Ghandi M, et al. Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell*. 2020;180(2):387-402.e16.

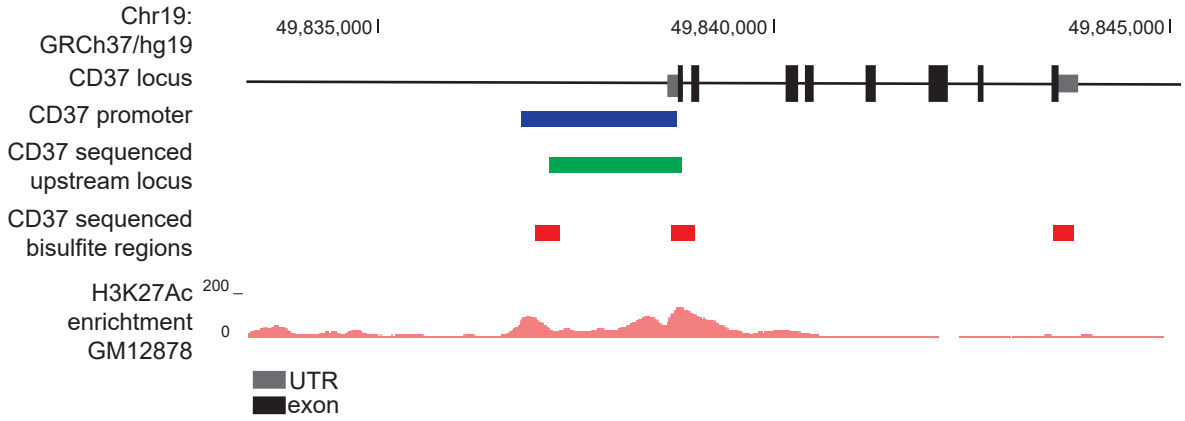

Figure S1. Schematic representation of the *CD37* gene and upstream sequence. The blue bar indicates the locus of the *CD37* promoter as used in the promoter vector (Main text; Figure 1). The green bar indicates the sequenced upstream locus in n=12 CD37 phenotypically negative but genetically wild-type DLBCL tumors. The red bars from left to right indicate the sequenced regions A, B and C after bisulfite conversion. H3K27Ac enrichment is shown in the lymphoblastoid GM12878 cell line (as obtained via UCSC Genome Browser, Integrated Regulation from ENCODE tracks, data from the Bernstein Lab at the Broad Institute). For primer sequences see Supplemental Methods.

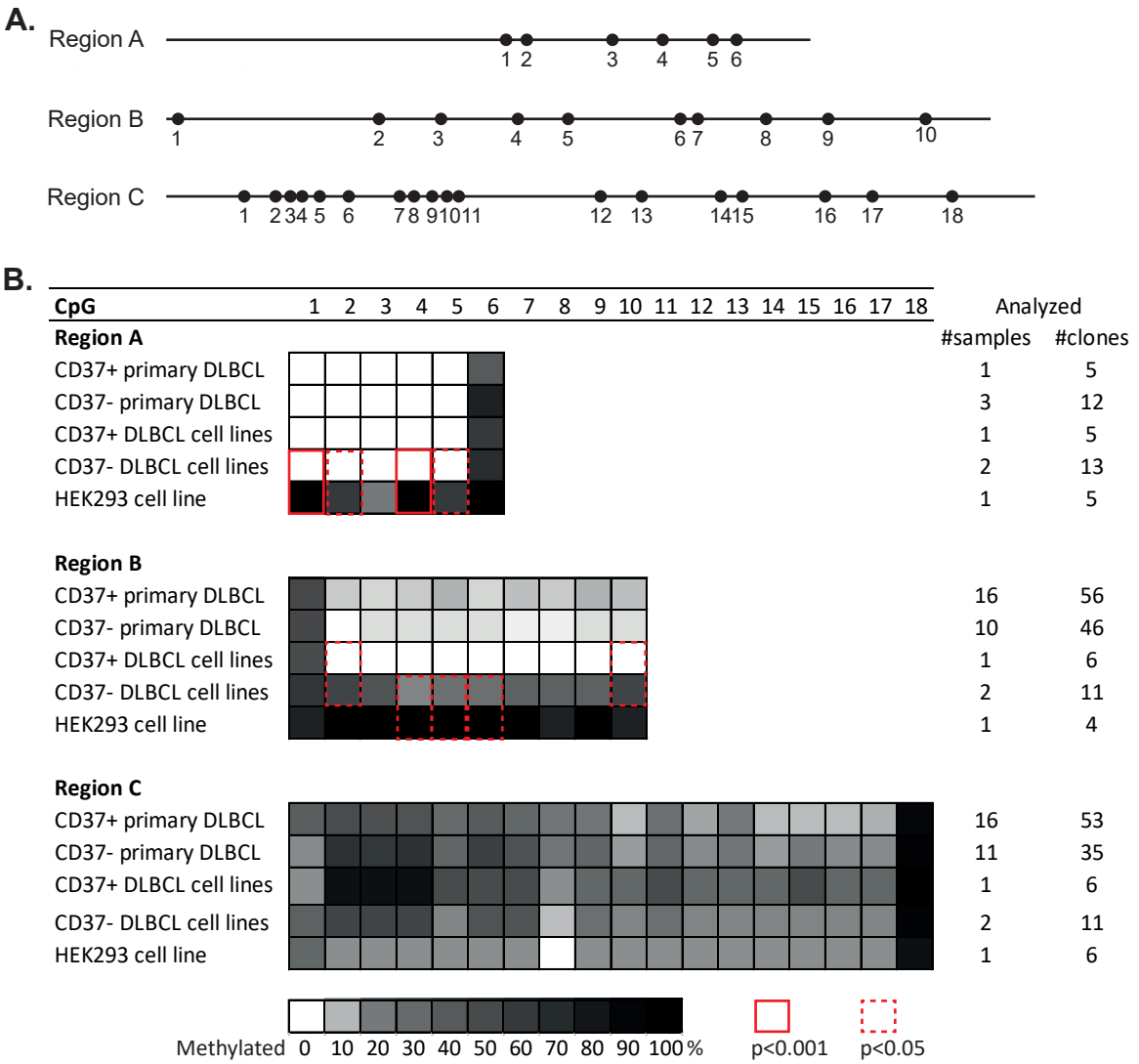

Figure S2. Targeted DNA methylation analysis. A) Schematic scaled overview of the three analyzed bisulfite sequencing PCR regions. Each dot represents the location of a CpG. The location of the regions is indicated in Figure S1. B) Schematic overview of the percentage methylation per CpG in three regions of the *CD37* gene. Genomic DNA of each sample was bisulfite converted, amplified for each region, cloned and sequenced. CD37+: CD37-positive DLBCL sample. CD37-: CD37-negative DLBCL sample. The CD37 protein expression was determined in cell lines by flow cytometry (main Figure 1A) and in patient samples using immunohistochemistry staining for CD37 as described in Methods. We compared the methylation levels per CpG in CD37+ to CD37- DLBCL cell lines, in CD37- DLBCL cell lines compared to the non-hematopoietic control HEK293 cell line, and in CD37+ compared to CD37- DLBCL tumor samples, using Fisher's exact test on the absolute number clones per CpG that are methylated (Supplemental Table S8-13).

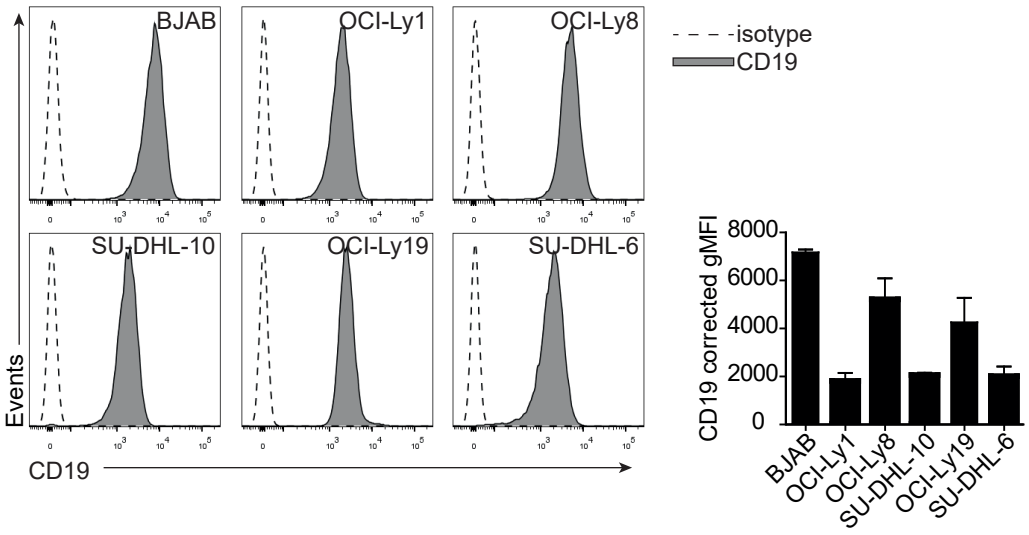

Figure S3. CD19 membrane expression on DLBCL cell lines. Representative histograms (left) and quantification (right) of flow cytometry analysis of CD19 membrane expression on DLBCL cell lines. Corrected gMFI: geometric mean fluorescence intensity corrected for gMFI of isotype control. Data represents mean ±SEM of three independent experiments.

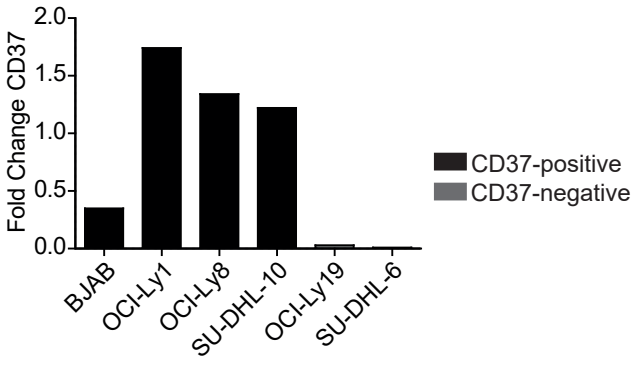

Figure S4. *CD37* mRNA expression in DLBCL cell lines. qPCR analysis of *CD37* mRNA levels in DLBCL cell lines. Ct-values were corrected for Ct-value of the housekeeping gene GAPDH. Fold change of mRNA expression was normalized to the average expression in CD37-positive cell lines BJAB, OCI-Ly1, OCI-Ly8, and SU-DHL-10. Data represents n=1.

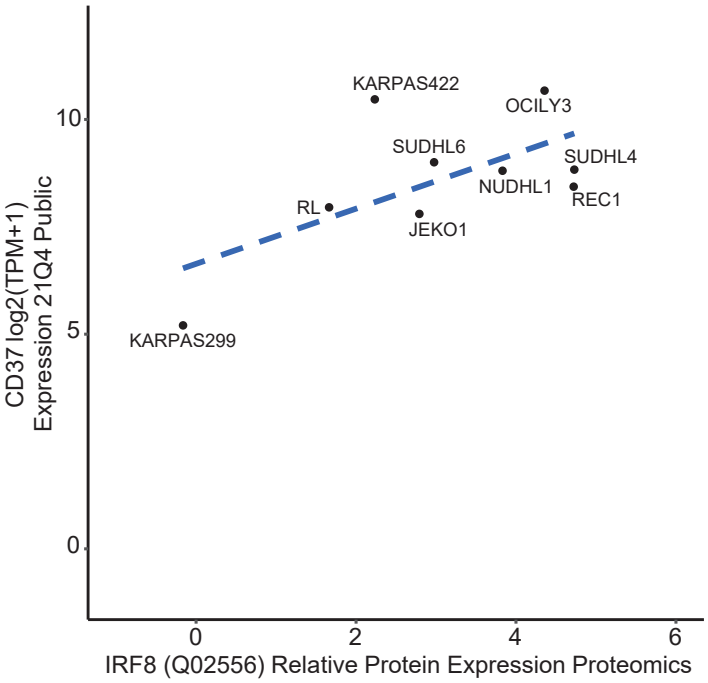

Figure S5. Correlation analysis of IRF8 protein with CD37 mRNA expression levels in non-Hodgkin lymphoma cell lines. The Cancer Dependency Map (https://depmap.org/portal/) dataset was used to investigate IRF8 protein and CD37 mRNA expression in non-Hodgkin lymphoma cell lines. Pearson correlation r=0.65, p=0.06.

> chr19:49,836,812-49,838,766 (GRCh37/hg19)

GTTTCAAAAAGACCCCTCTGGGGCTATGTGGGGGACAGACTGGAGGGAGAAACTAGAAGCCAGGGAGAAGC
CTAAAGTCAGGGTCCAGGTAGGAGAGCAGGAGCCCAGAGCCTGCCTGAGCCCTCAGGACAGAGAAGAGGGG
TGGAGTGGAGACTCAGCAGGCAGGAGGGATCTGGCTAAGAGACTGATGCGCTGTTGGGGAAGAGGGGACA
AGATGGAGCTTTGACCTCATGACTAGATAGATGGTGGGGCTGCCTCAGATGAGGAAACCAGGTGAGAGAATT
TGGGGACATCAGAGGAAACTGCAGAGGCCACTGGGTGTGCACCTGCAACCAGGCACTGCAGACCCAGATCTG
ACTGTCAGGATCCCAGGCTCAGTGCCTCCTCCTCTGGGGCCTCCGCTCCTCGGGCCCCACTGCAGAGGCAGGA
AGTACCGCCTGATGCTCAGCTTCCGCTGGGGGTGTTGGCACGGCCCAGCGCCCTCCCCACTCCCAGGGACCTT
CCCACCTCGAGCCTGTAGCACTTTCACTTCCCCAGGTACCCAGCCCCTGCCCCCAGCACCTGGGTGCTCCCTGGT
CCAGAACAGGCTGCAGACTCTGAGAGGCACAAAAACAAGTGGGGAAGGGGGAGACAGAGAAAAGAAGACA
GGCTGGGCACAGTGGCTCACGCCCATAATCCCAGCACTTTAGGAGGCCAAGGCAGGAGGATCACTTGAGCCC
AGAAATTTGAGACCAGCCTAGGCAACACAGTGAGACTCCCATCTCTACAAAAAATGAAGAAAATTAGCCAGGC
ATGGTGGTGTGCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGTGCAAGGATTGCTCGAGCCTGGGAAGGT
TGGGGCTGCAGTGAGCTATGATCGCACTACTGCATTCCAGCCTGGGCGACCCTGTCTCAAAAACAAAAACAAA
AGAGAGATGAAGAGAGAAAGGAGGAAGGACCCAAAGAAGGGGCCAGAAGCAGAGAGAGGCCAGGCACAG
TGGCTCACGCCTGTAATCCCAACACTTTGGGAGGCCAAGGCAGGAGGGTCGCTTGAGTCCAGGAGTTCCAGA
CCAGCCTGGGCAACCTAGTGAGACCCTATCTCTGTTTTTTTTTAATTAAATATAAAAAATACAACACTCCCACTC
CATAGCAAGACCCCATTTCTACAAAAAATACAAAAATTAGCCAGGCATGGTAGTGTGCACCTGTAATCCCAGCT
ACTCAGAAGGCAGAGGTGGGAGGATAGCTTGAGCCCAGGCTACAGTGAGCTGTTTTCCCCACCGTACTCCAGC
CTGGGTGAAAAGTGAGACCCTGTCTCAAAAAGAAACAAATATATAAATAAAATAAAGAAACACAGAGAGAGG
GAGGATAGAGACCCCAGGAGGCAGGAGGACAGAGACCCAGAGGAGGGAGAGACTCTGCTGAGATAAGGGG
ACAGAGATGGGGAGTAGCAGGAAGGTGGGGACAAAGACCCAGAGAGGAGGAACAAGAGACCCAGGGCAG
GAAATGGGAGACACAGCTGGATGCAGAGATGGGAGAGCTTAGCAGGCCATGAGGTCAGAGTGTCTGTGAGG
TCAGCTGCTGTCCAGGGTCAGCCCTGAGGTCCCCATGCCCAGAGGTGGGTTCCAGTAGATCAGGGGTCCCTTT
AGATCCCCCAGCGCCTGAGACATGGAGACCTGGGAGGGACTGCCCCATGAGTCCTTGGTGCCCACAAATTCCA
AGGCCCCTCAGGTACCAGAGCATGTGTCCCAGTGGAGCAGGTGAGGGAGACAGGGACAGAGAGAGAGACAG
AAAGCCACCTTACATGAAGCGGGAGTGGGTGGAGCCCCACCCCTGGAACTTCCTCTTTTGGGGTTCTTCCTTTC
TCTCTCAGCTCTCCGTCTCTCTTTCTCTCTCAGCCTCTTTCTTTCTCCCTGTCTCCCCCACTGTCAGCACCTCTTCTG
TGTG

Red: ETS-IRF composite element (EICE)

Blue: Start exon 1 *CD37* (untranslated)

TCGAGCCTGTAGCACTTTCACTTCCCCAGGTACCCAGCCC Oligo bait sequence for DNA pull down

Underlined Amplified regions (2) in ChIP followed by qPCR.

Figure S6. CD37 upstream sequence.

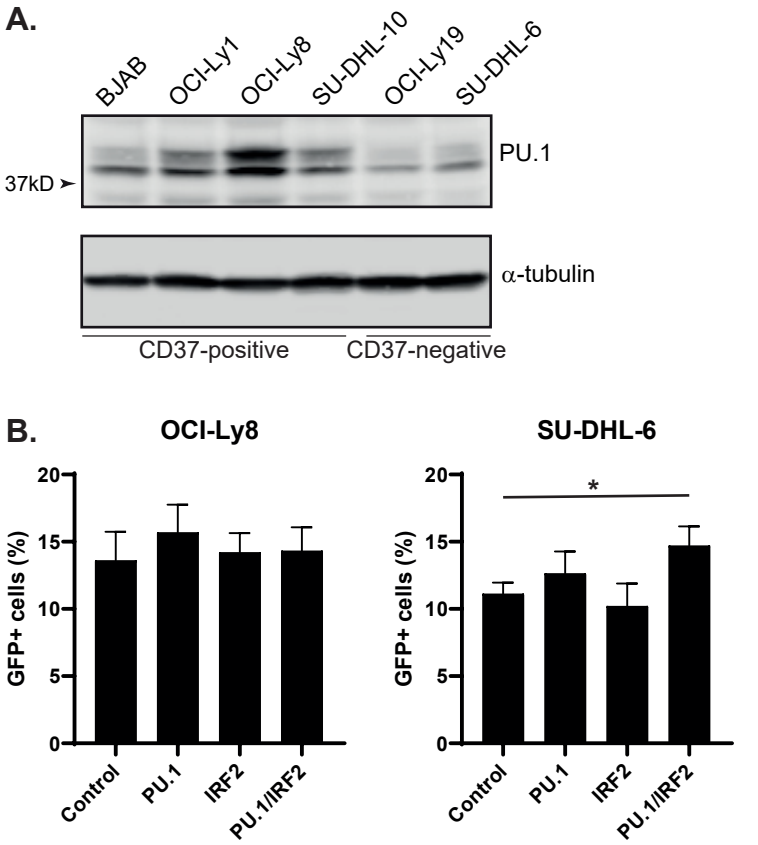

Figure S7. PU.1 expression and *CD37* promoter activity in DLBCL cell lines. A) Representative western blot stained for PU.1 and α-tubulin as loading control in indicated CD37-positive and CD37-negative DLBCL cell lines. B) Percentage GFP-positive cells in a CD37-positive (OCI-Ly8) and a CD37-negative DLBCL cell line (SU-DHL-6) transfected with the CD37 promoter-GFP construct with or without (control) indicated IRF8 cofactors. Repeated measures one-way ANOVA analysis was followed by Dunnett's multiple comparisons test, *p<0.05. Data represents mean+SEM of four independent experiments.

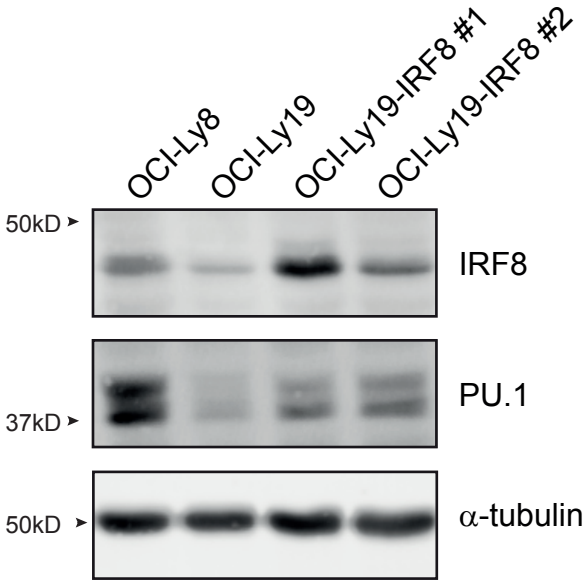

Figure S8. IRF8 and PU.1 expression after IRF8 overexpression via viral transduction. OCI-Ly19 cells were transduced to overexpress IRF8. Western blot stained for IRF8, PU.1 and γ-tubulin as loading control in transduced OCI-Ly19 cells (2 independent transductions). IRF8-positive cell line OCI-Ly8 and non-transduced OCI-Ly19 cells are shown as positive and negative control, respectively.

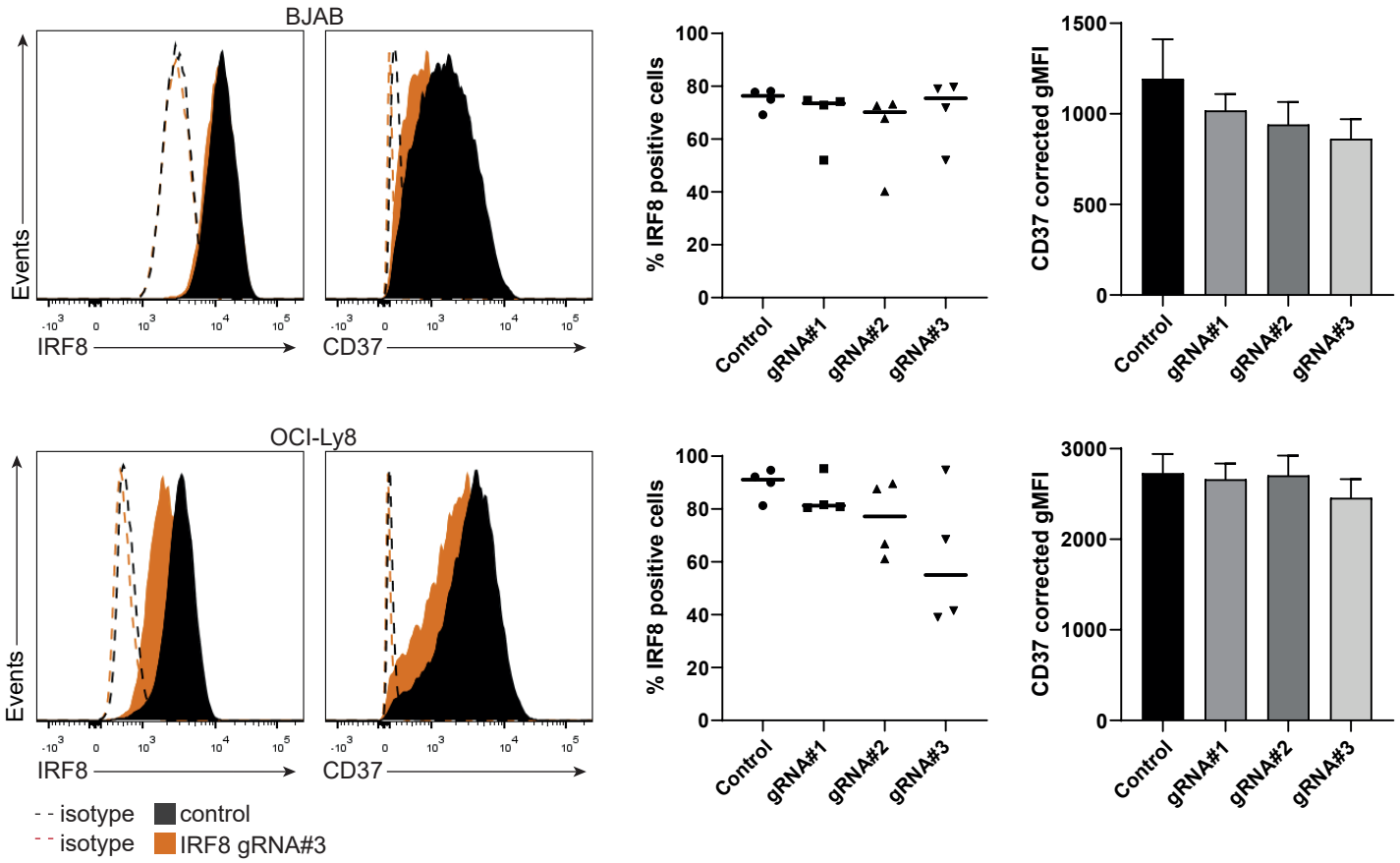

Figure S9. CD37 expression in IRF8 KO cells. Representative histograms (left) and quantification (right) of flow cytometry analysis of BJAB (top) and OCI-Ly8 (bottom) control (empty vector transfected) and IRF8 knock-out (gRNA) cells obtained using CRISPR gRNA set #2 (see Supplemental Table S6). The percentage of IRF8 positive cells was determined as percentage of cells with IRF8 signal above isotype level. Corrected gMFI: corrected for isotype control. Data represents mean+SEM of four independent experiments.

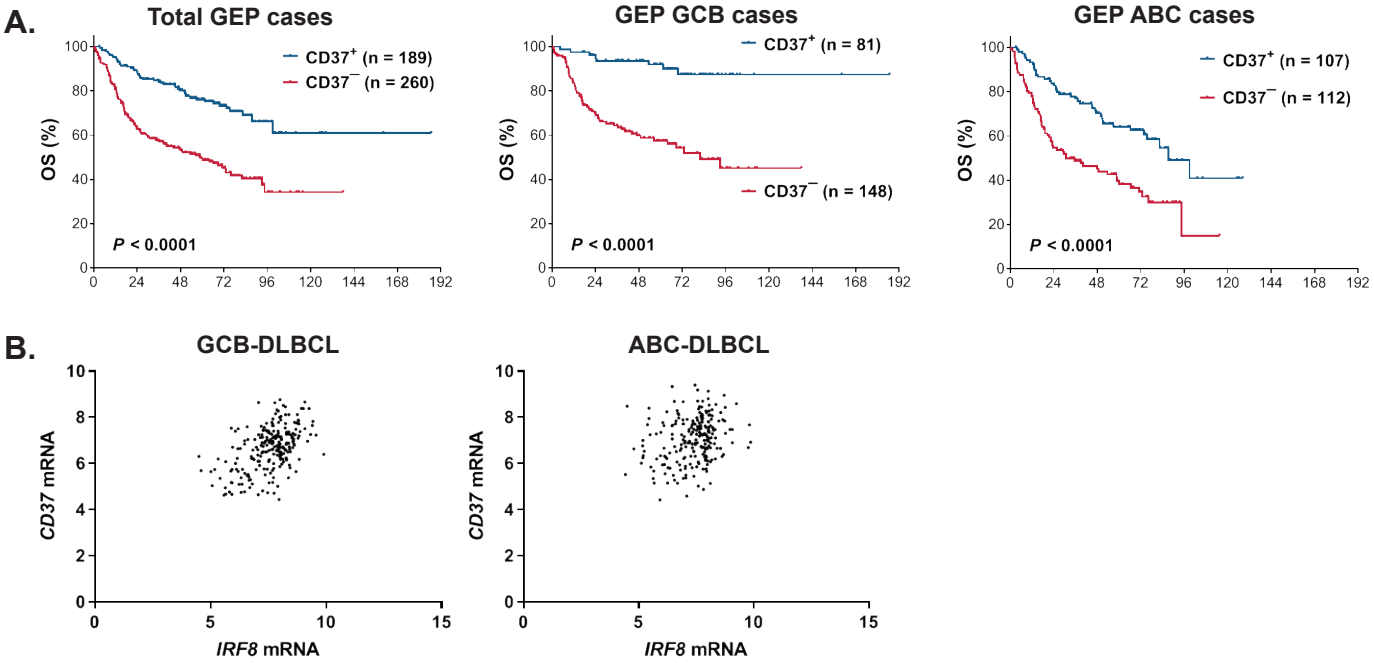

Figure S10. Additional analyses of the gene expression profiling cohort. A) Survival analysis of the gene expression profiling cohort. Prognostic analysis of the GEP cases in Figure 5A for which CD37 immunohistochemistry and survival data was available. B) Analysis of IRF8 and CD37 mRNA levels in primary GCB- and ABC-DLBCL separately. GCB: Pearson correlation r=0.5037, 95% confidence interval 0.4034-0.5920, R^2=0.2537, ****p<0.0001. ABC: Pearson correlation r=0.2469, 95% confidence interval 0.1211-0.3648, R^2=0.0609, ***p=0.0002.

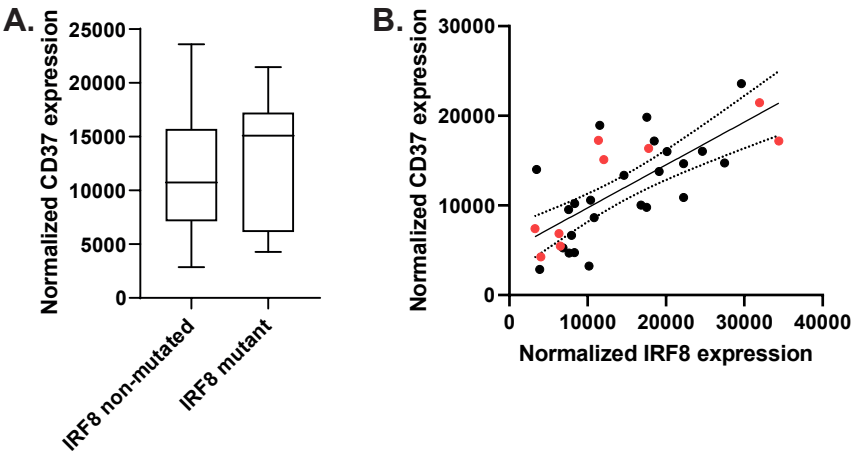

Figure S11. Impact of IRF8 mutations on *CD37* expression levels in primary DLBCL patient samples.
A) Comparison of normalized *CD37* expression levels between high impact IRF8 mutated DLBCL patient samples (n=9) and non-mutated DLBCL samples (n=24). B) Correlation plot of normalized *IRF8* expression levels versus normalized *CD37* expression levels in DLBCL (n=33). Colored dots indicate samples with an IRF8 mutation. Linear regression including 95% confidence interval, R2=0.5306. Spearman rank correlation r=0.743, ****p<0.0001. The data of both A) and B) were based on public available RNA-Seq datasets (GSE125966).

A.

	CD37-negative (%)	CD37-positive (%)	Total	P value
Non-GCB	42 (49)	43 (41)	85	.3046
GCB	43 (51)	61 (59)	104	
Total	85	104	189	

B.

	IRF8-low (%)	IRF8-high (%)	Total	P value
Non-GCB	60 (47)	25 (41)	85	.5320
GCB	68 (53)	36 (59)	104	
Total	128	61	189	

C.

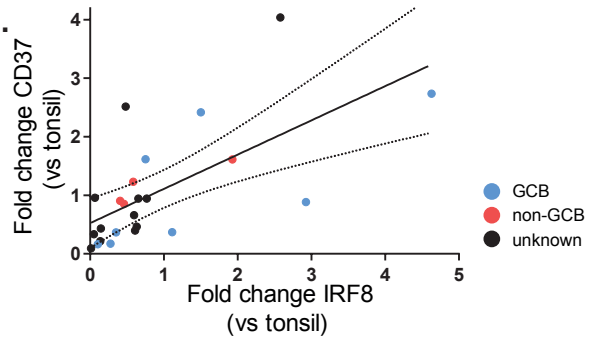

Figure S12. GCB/non-GCB status of the DLBCL cohort and *IRF8* and *CD37* mRNA expression of a subcohort. Sufficient data about CD37, IRF8 and GCB/non-GCB status were available for 189 out of the 206 analyzed patients. An overview of the GCB/non-GCB status in CD37-positive and CD37-negative samples (A) and IRF8-low and IRF8-high samples (B) is shown. CD37 status was determined negative if the IHC score was <10%. IRF8 status was determined low if the IHC score was <60%. P value shows the statistical significance obtained using Fisher's exact test. C) qPCR analysis of *IRF8* and *CD37* mRNA levels in primary DLBCL (n=24/206), normalized to expression in tonsil tissue. Linear regression including 95% confidence interval, R^2=0.4268. Pearson correlation r=0.6533, ***p=0.0005.