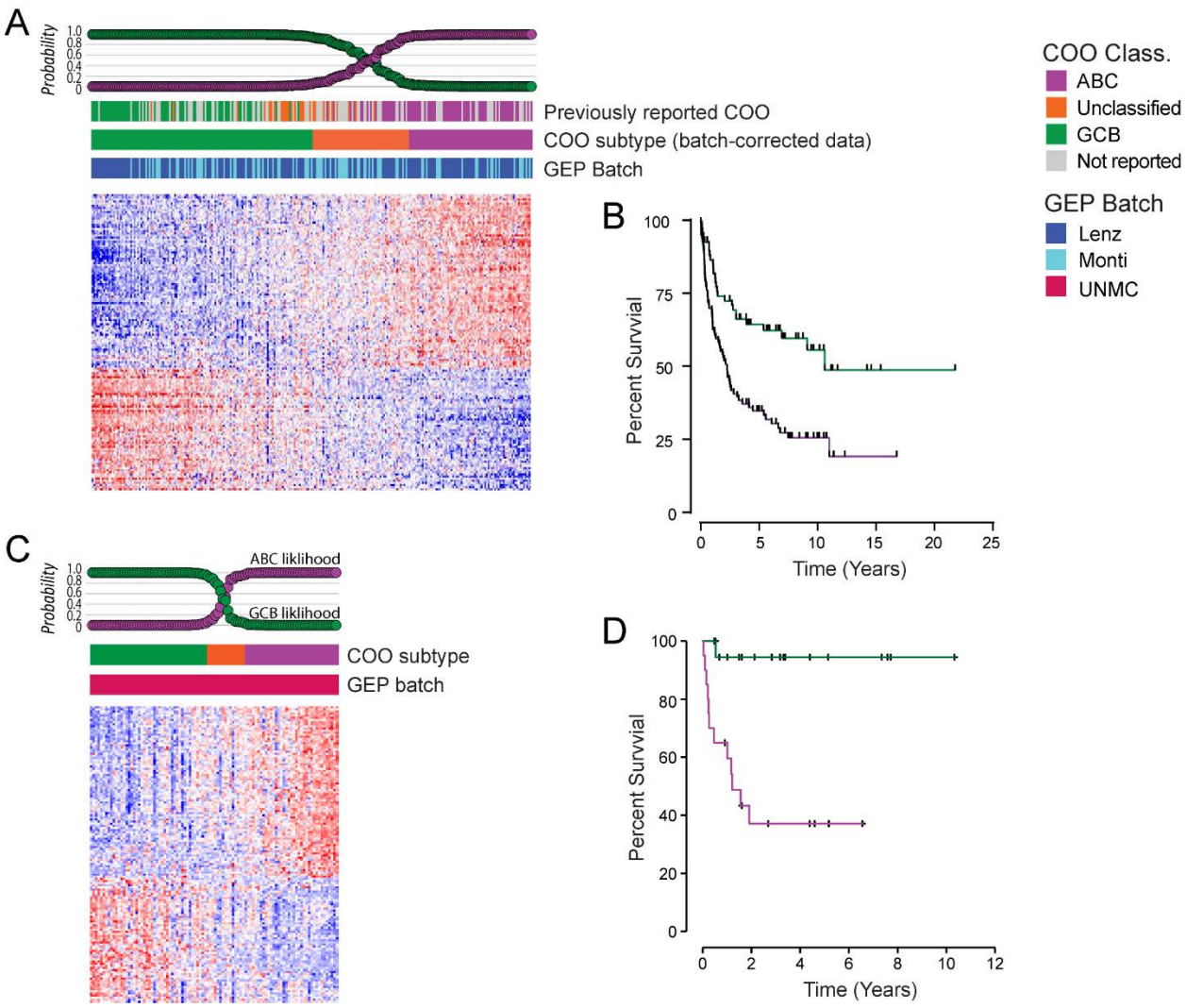

Figure S1: Validation of cell of origin subtyping. A) Gene expression microarray data from 249 tumors from 2 previous studies was batch corrected using COMBAT and cell of origin subtyping performed using a Bayesian classifier with the 140 gene signature reported by Wright et al. Subtypes are shown relative to the previously-reported COO subtype for these samples. **B)** Overall survival data for GCB-like (green) and ABC-like (purple) tumors from (A) is shown. There is a highly significant difference in overall survival (log rank $P < 0.001$) thereby validating the success of COO subtyping. **C)** COO subtyping from gene expression microarray data of an additional 98 tumors that could not be batch-corrected with the other cohorts. **D)** Overall survival is shown for GCB-like (green) and ABC-like (purple) tumors with available data. There is a highly significant difference in overall survival (log rank $P < 0.001$) thereby validating the success of COO subtyping.

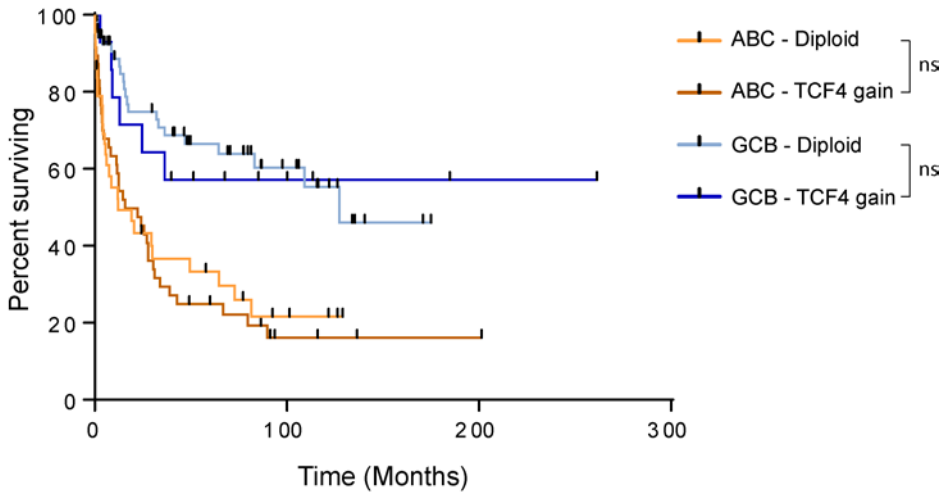

Figure S2: Comparison of outcome in CHOP-treated DLBCL patients stratified by cell-of-origin subtype and 18q21 gain. The association between *TCF4* gain and outcome was not independent of COO subtype, as shown by the non-significant (ns) difference between GCB or ABC patients with (TCF4 gain) or without (Diploid) 18q21 DNA copy number gains.

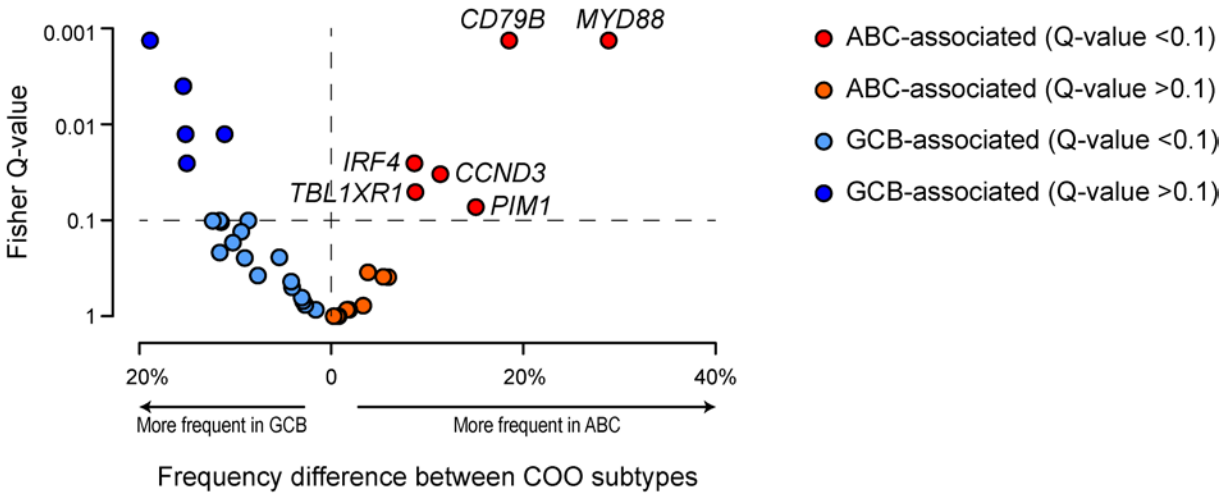

Figure S3: Recurrently mutated genes enriched in the ABC-like DLBCL subtype. A volcano plot show the frequency difference and statistical significance of 40 recurrently mutated genes from 199 DLBCL tumors with matched NGS, DNA copy number and COO subtype data. Mutations in *MYD88* were most significantly associated with the ABC-like subtype, motivating their selection as the seed feature for REVEALER analysis. Other mutations that were significantly associated with ABC-like DLBCL (red) are annotated and included in Figure 1E for comparison to 18q21 DNA copy number gains.

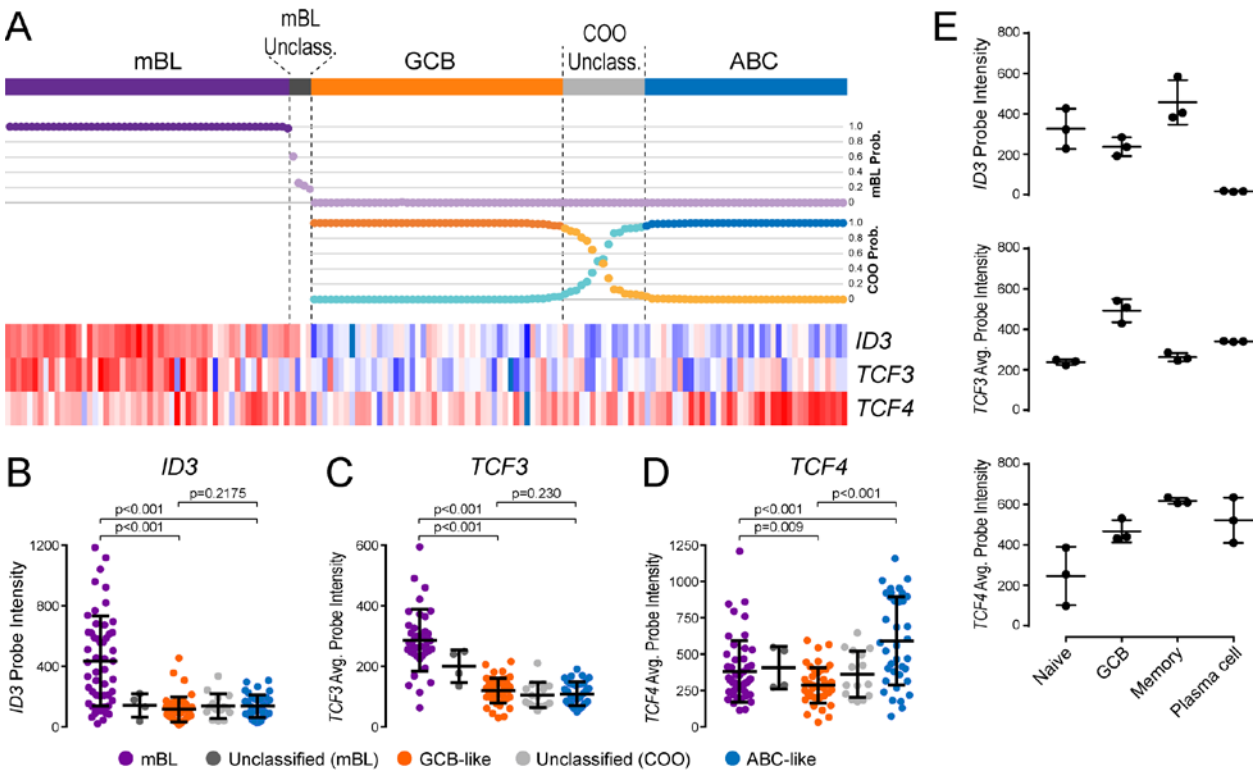

Figure S4: Expression patterns of TCF3 and TCF4 in normal and malignant B-cells. A) Gene expression microarray data from 154 high-grade B-cell lymphomas were used to classify tumors into molecular Burkitt lymphoma (mBL, n=52), unclassified (n=4) or non-mBL (n=98). The non-mBL were then further classified by cell of origin subtype into germinal center B-cell-like (GCB, n=46), unclassified (n=15), or activated B-cell-like (ABC, n=37). The average probe intensity for TCF3, ID3 and TCF4 are shown for each tumor, arranged by classification scores. **B)** *ID3* expression is significantly higher in mBL than all other molecular subtypes. **C)** *TCF3* expression is significantly higher in mBL than all other molecular subtypes. **D)** TCF4 is variably expressed in mBL and ABC-like DLBCL, but is significantly higher in ABC-like DLBCL compared to all other molecular subtypes. **E)** The expression of *ID3*, *TCF3* and *TCF4* is for different subsets of normal human B-cells. The pattern of *TCF3* and *TCF4* is observably different, with *TCF3* (middle) expression elevated in germinal center B-cells (GCB) and declining in subsequent stages of development, whereas *TCF4* expression is elevated in GCB cells and then equal or higher in subsequent stages of development.

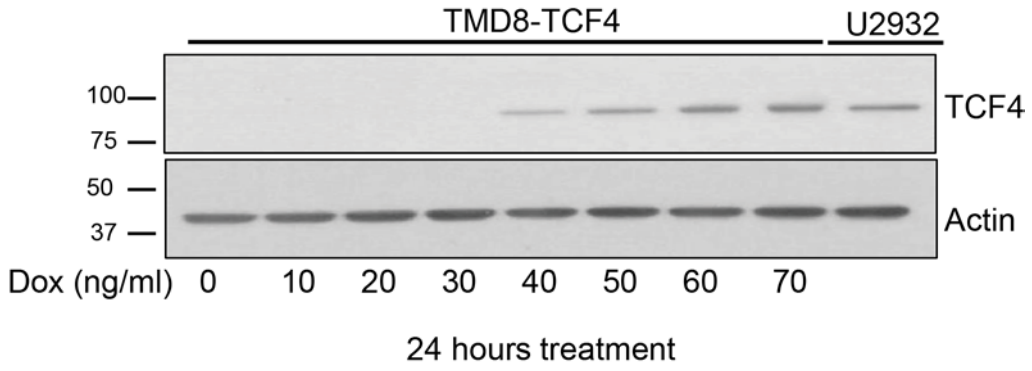

Figure S5: Tetracycline induction of TCF4 expression. A titration was used to identify the dose of tetracycline required to induce equivalent expression of TCF4 to that observed in the U2932 cell line with TCF4 DNA copy number gain. The dose of 60ng/mL was selected.

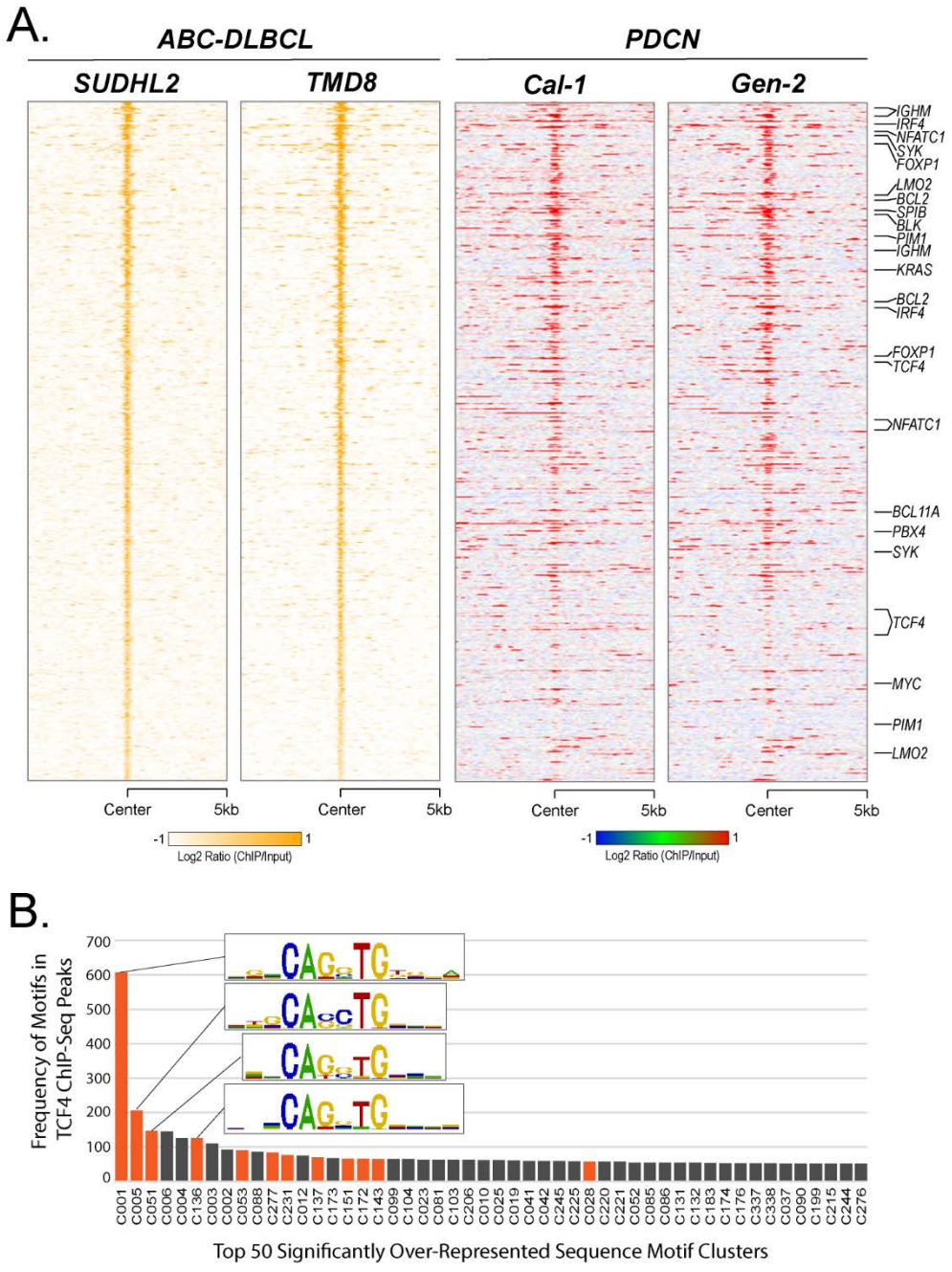

Figure S6: ChIP-seq peaks for TCF4. A) The TCF4 ChIP-seq peaks from Figure 3B are shown and aligned with TCF4 ChIP-seq signal from plasmacytoid dendritic cell neoplasm (PDCN) cell lines that expression TCF4 (Ceribelli et al., GEO Accession GSE76147). Many of the regions that are bound by TCF4 in ABC-like DLBCL cell lines can also be seen to be bound by TCF4 in PDCN cell lines. **B)** The frequency of the 50 most significantly enriched DNA sequence motif clusters from TCF4 ChIP-seq peaks are shown. Motif clusters with a dominant E-box (CANNTG) are colored in orange, with the motif logos shown for the top 4 E-box motif clusters.

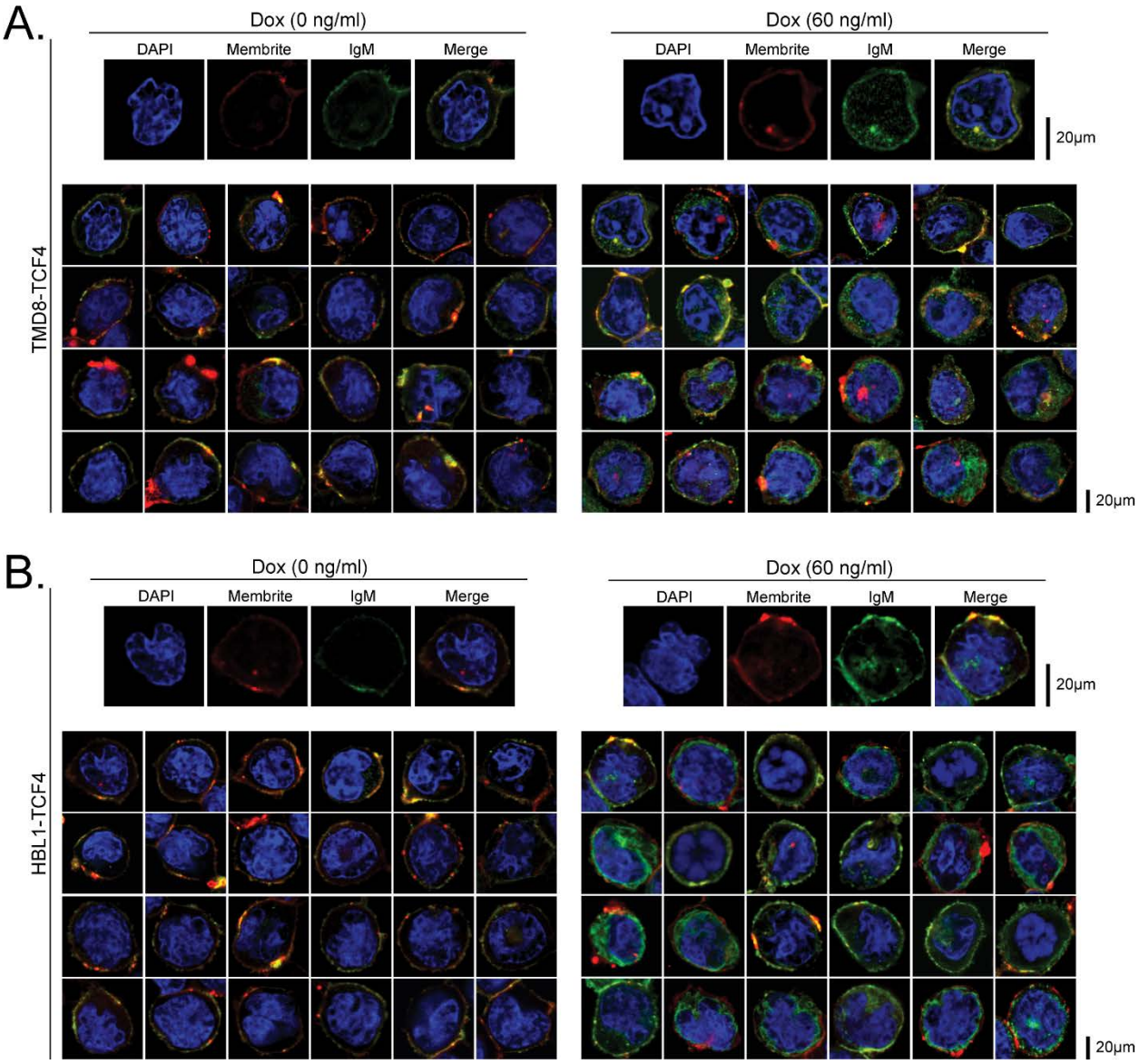

Figure S7: Cellular localization of TCF4-induced IgM. The expression of TCF4 was induced in ABC-like DLBCL cell lines with low *TCF4* DNA copy number and expression, TMD8 (**A**) and HBL1 (**B**). The nucleus is stained by DAPI (blue), the cell membrane by Membrite (red), and IgM by mouse anti-IgM antibody with Alexa-555-conjugated goat anti-mouse secondary (green). An example of each individual stain and its merged image is shown above, and merged images for 24 cells shown below. The large images above correspond to the top left image in the set of 24. It can be seen that the induction of TCF4 leads to a marked increase of IgM on the cell surface, as well as in intracellular puncta, in both cell lines. Images are representative of 3 replicate experiments. Scale bar=20μm.

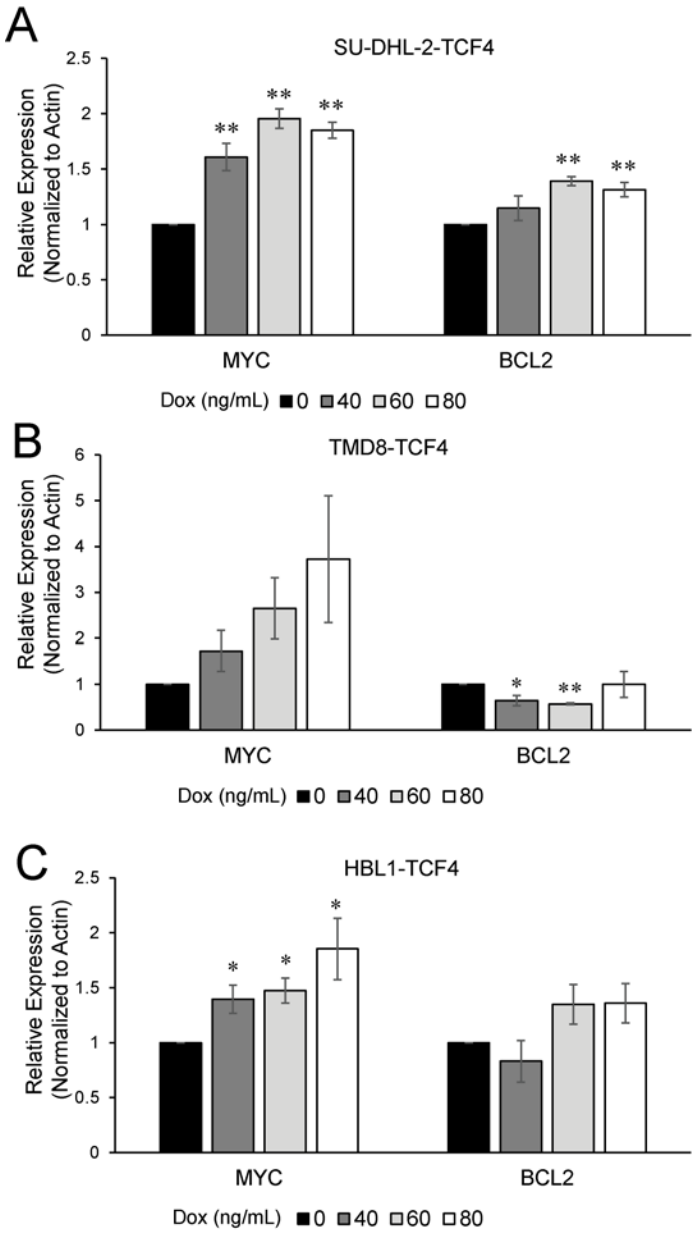

Figure S8: Quantification of TCF4-induced MYC and BCL2 protein expression. Bar graphs show the means +/- S.E.M for the quantification of triplicate western blots for MYC and BCL2 expression following tetracycline-inducible expression of TCF4 in SUDHL2 (A) TMD8 (B) and HBL1 (C). (*P<0.05, **P<0.01)

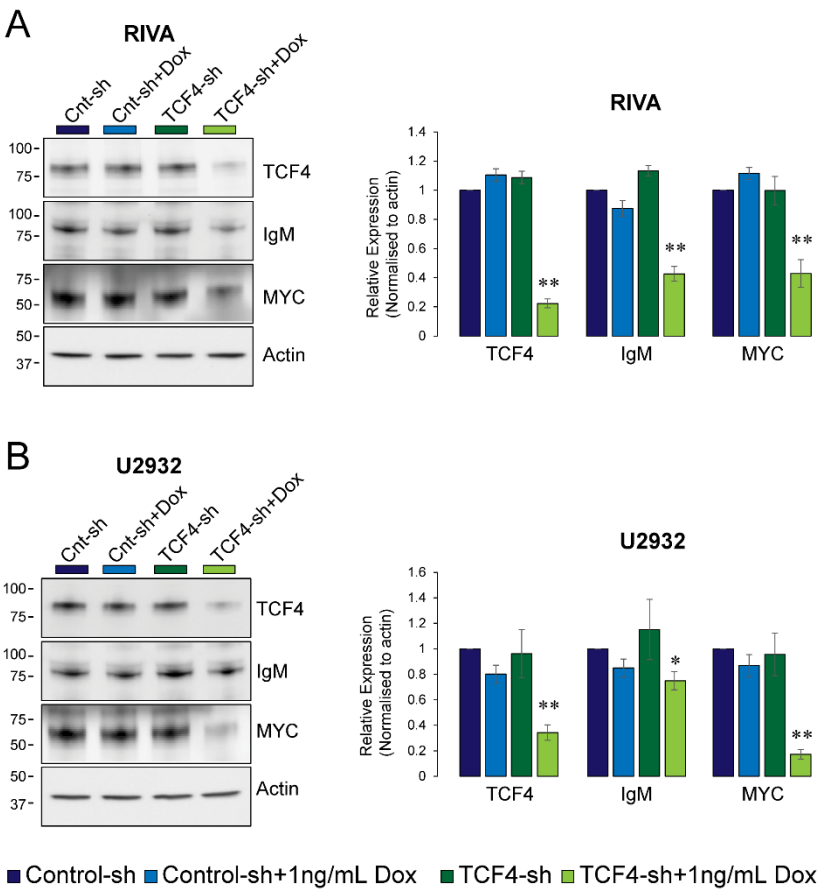

Figure S9: shRNA-mediated knock-down of TCF4. Knock-down of TCF4 using tetracycline-inducible shRNA expression in RIVA (A) and U2932 (B) significantly reduced IgM and MYC expression compared to a control shRNA targeting luciferase. Bars represent the mean +/- S.E.M. of three independent experiments, with the difference between dox-treated control and TCF4 shRNA conditions assessed using a Student's T-test, *p<0.05, **p<0.01.

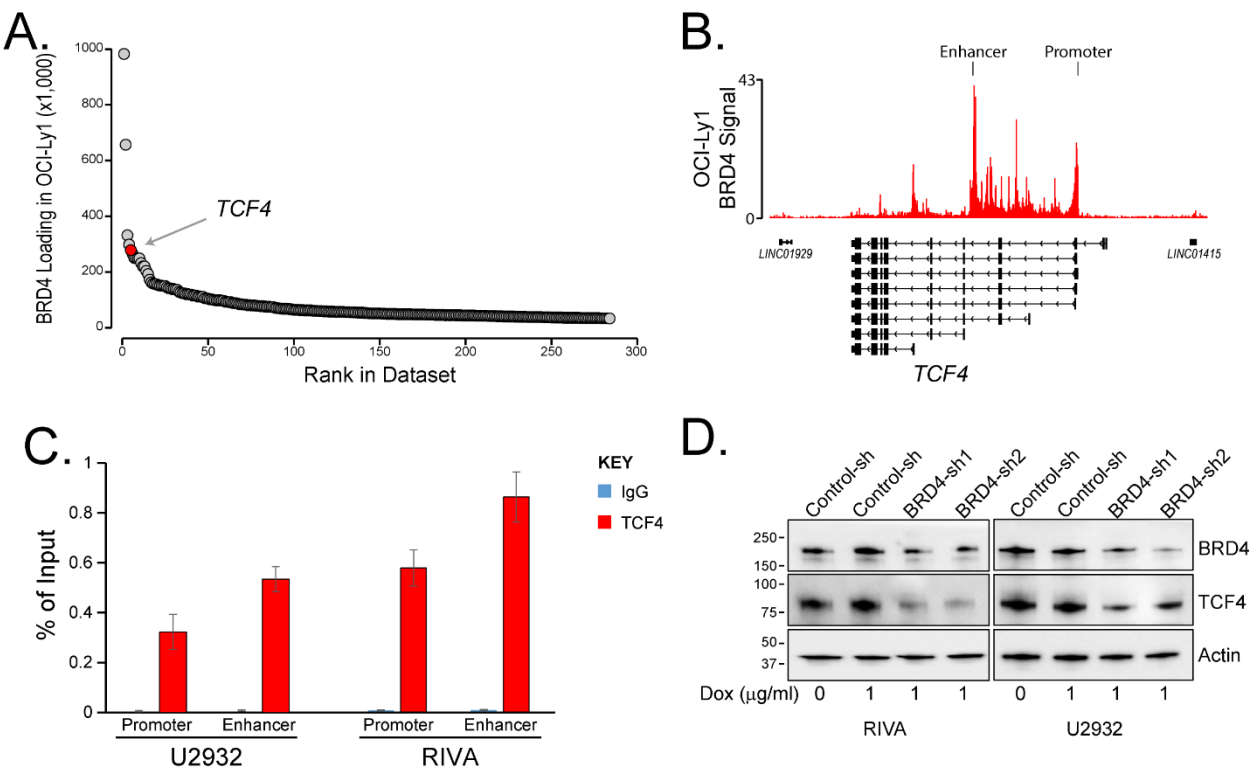

Figure S10: BRD4 regulates TCF4 in DLBCL cell lines. A) Previously reported BRD4 ChIP-seq data from the OCI-Ly1 cell line shows that TCF4 is the 5th most highly BRD4-loaded gene in the genome. **B)** A wiggle plot shows strong BRD4 signal over the TCF4 gene in OCI-Ly1, including peaks that correspond to promoter and enhancer regions. **C)** ChIP-qPCR for BRD4 confirmed binding at the promoter and enhancer regions (shown in B) in two ABC-like DLBCL cell lines with DNA copy gain of the TCF4 locus, U2932 and RIVA. **D)** Tetracycline-inducible shRNA knock-down of BRD4 using two unique hairpins led to a significant reduction of TCF4 in the RIVA and U2932 cell lines, confirming a direct role for BRD4 in regulating *TCF4* expression in this context.

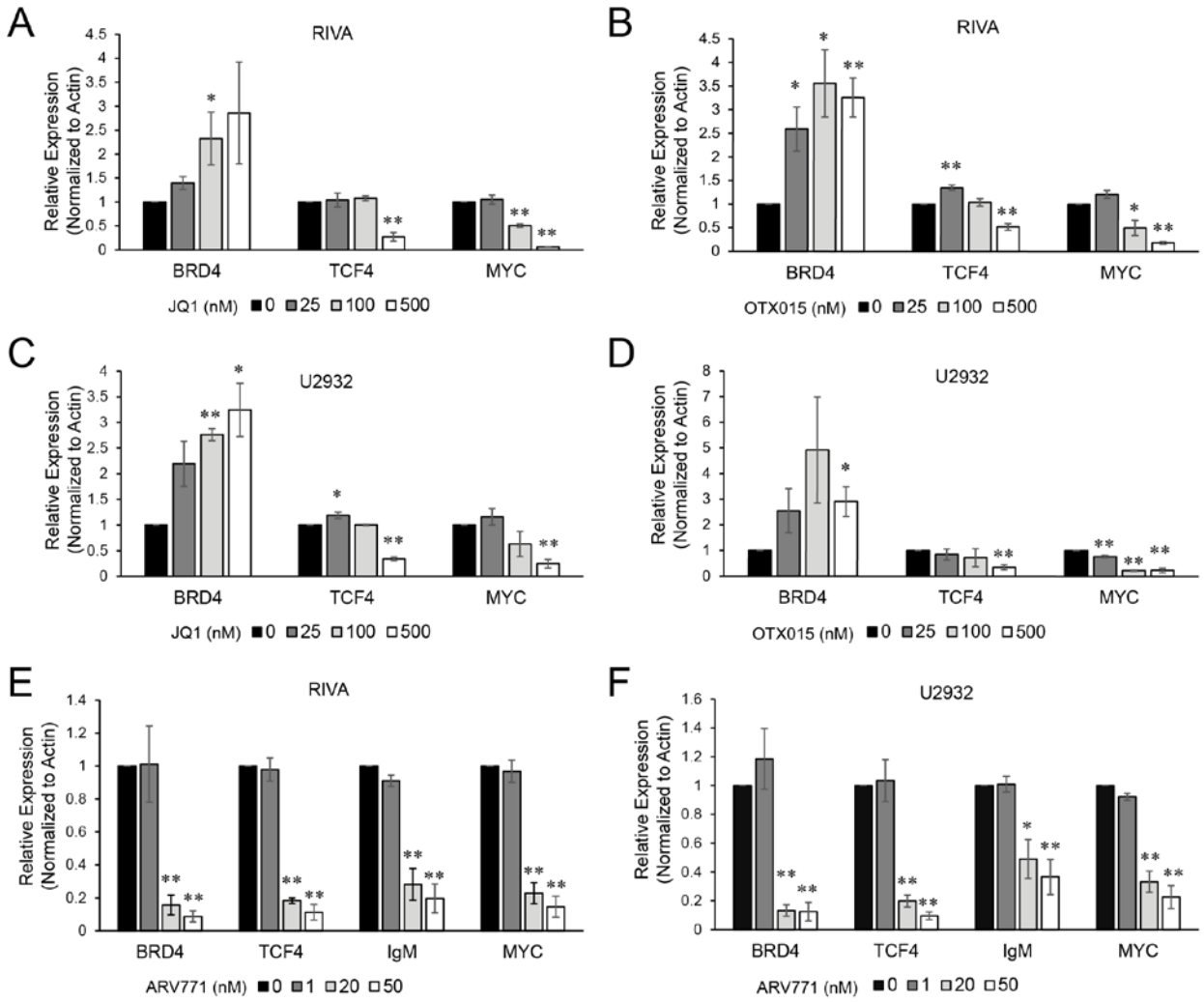

Figure S11: Western blot quantification of triplicate BET inhibitor experiments. The quantification of BRD4, TCF4 and MYC protein abundance following treatment with JQ1 and OTX015 is shown for RIVA (A-B) and U2932 (C-D). The quantification of these targets in addition to IgM following treatment with ARV771 in RIVA (E) and U2932 (F).

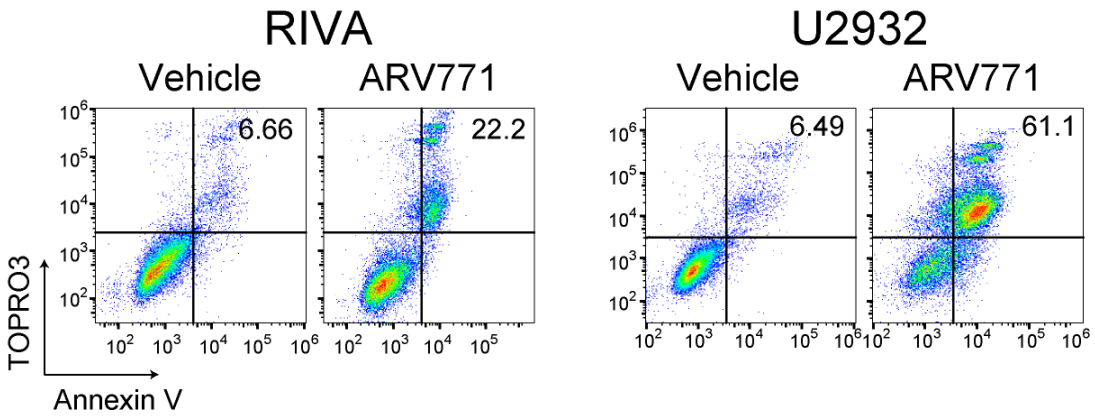

Figure S12: ARV771 induces apoptosis in RIVA and U2932 cell lines. Cells were treated with 50nM of ARV771 or vehicle control for 72h and analyzed by flow cytometry using Annexin-V/TOPRO3 staining. This confirmed that ARV711 was able to induce apoptosis in these cell lines, as previously observed with other small molecule BET inhibitors.

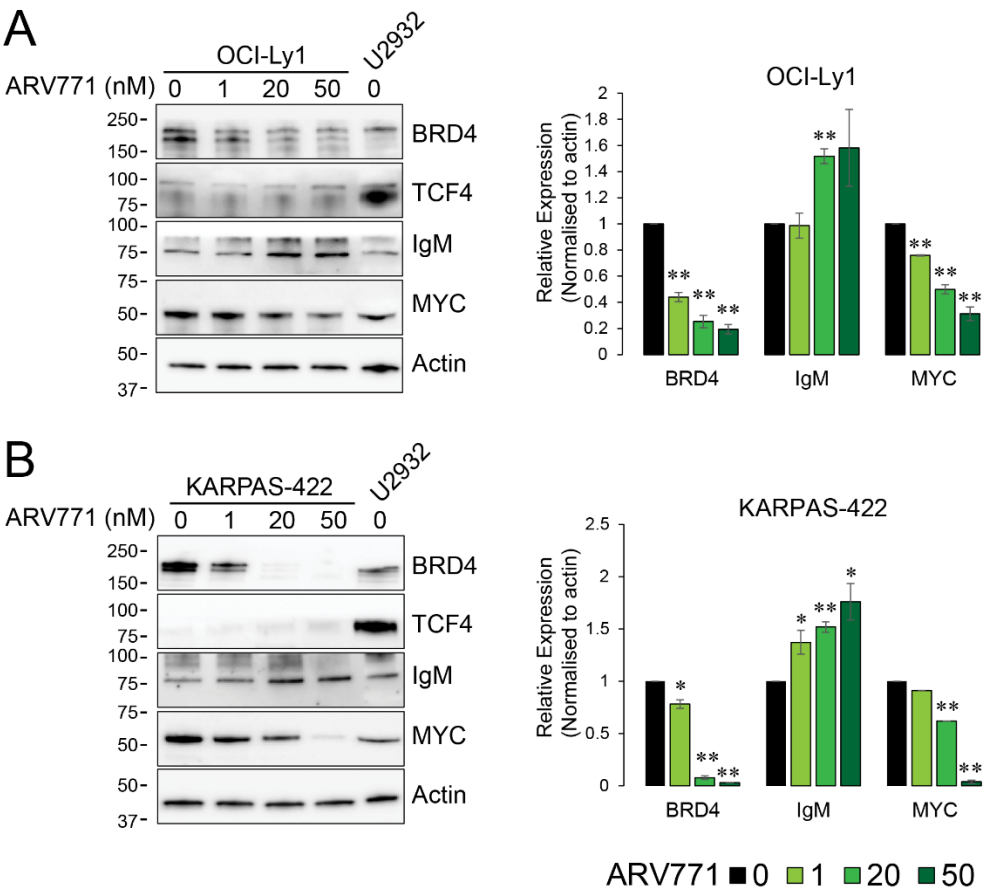

Figure S13: Effect of ARV-771 on IgM expression in GCB-like DLBCL cell lines. Two GCB-like DLBCL cell lines with low (OCI-Ly1, A) or undetectable (KARPAS-422, B) levels of TCF4 were exposed to increasing concentrations of ARV-771. The expression of BRD4, TCF4, IgM and MYC was evaluated by western blot (left) and quantified over triplicate experiments with reference to Actin (right). Visible and significant reduction of BRD4 and MYC can be observed in both cell lines. TCF4 expression was too low to reliably quantify. Unexpectedly, we observed a modest, but significant and reproducible, increase in IgM expression following ARV-771 treatment.

SUPPLEMENTARY MATERIALS AND METHODS

Targeted next generation sequencing (NGS) and variant calling.

In brief, 500-1,000ng of genomic DNA was sheared using a Covaris S2 instrument and libraries prepared using KAPA Hyper Prep kits (Roche) and Illumina TruSeq Adapters (Bioo Scientific) according to the manufacturer's protocol. A maximum of 6 cycles of PCR was used for library preparation. Samples were 12-plexed, subjected to hybrid capture with a 5.3Mbp Nimblegen SeqCap custom reagent (Roche), and amplified by 8 cycles of PCR. Each pool was sequenced on a single lane of an Illumina HiSeq 2500 instrument in high-output mode using 100bp paired-end reads at the Hudson Alpha Institute for Biotechnology Genome Sequencing Laboratory. Raw sequencing reads were aligned to the human genome (hg19) using a BWA-Mem (*1*), realigned around InDels using GATK(*2*), sorted and deduplicated using Piccard tools, and variants were called according to a consensus between VarScan2(*3*) and GATK Unified Genotyper (*2*). This approach has been validated to have a specificity of 92.9% and a sensitivity of 86.7% (*4*). Average on-target rate for this dataset was 88% and average depth of coverage 623X (min = 122X, max = 1396X). Raw FASTQ files for the targeted NGS of previously published DLBCLs (n=119; European Nucleotide Archive Accession ERP021212) (*5*) and Burkitt's lymphomas (*6*) were also analyzed with the same pipeline, and the results integrated. The DNA copy number of UNMC DLBCL and Burkitt lymphoma cohorts was determined using CopyWriteR (*7*) with 200kB windows.

ChIP-sequencing of TCF4

Five million of cells were fixed with 1% formaldehyde for 10 min, quenched by addition of 125 mM Glycine for 5 min at RT, washed with ice-cold PBS then resuspended and incubated in ice-cold ChIP buffer (10mM Tris-HCl pH 8.0, 6.0 mM EDTA, 0.5% SDS and protease inhibitor) for 1hour. In the same time, 5µg of antibodies (TCF4(*8*)) or control rabbit IgG (Cell Signaling; 2729) were allowed to bind to dynabeads Protein-G (Invitrogen; 10003D) in binding buffer (0.2% BSA, 0.1% Tween-20 in PBS) for 2 hours. Chromatin was sheared using Covaris M220. Sonicated lysates were diluted in dilution buffer (10mM Tris-HCl pH8, 140 mM NaCl, 1mM EDTA pH 8, 0.5 mM EGTA, 1% Triton X-100, and 0.1% Sodium Deoxycholate) and added to antibody bound Protein-G beads for immunoprecipitation overnight at 4ºC. Note; for ChIP normalization, spike in chromatin/antibody was added to sonicated lysates (53083/61686; Active Motif). Next day, beads bound complexes were washed 5 times with RIPA buffer (1% NP40, 0.1% SDS, and 0.5% Sodium Deoxycholate in PBS), 2 times with LiCl buffer (10 mM Tris-HCl pH8, 250 mM LiCl, 0.5% NP40, 0.5% Sodium Deoxycholate, 1 mM EDTA), once with TE buffer pH 8.0 and finally resuspended in 50µl of TE buffer containing 20µg of proteinase K and RNase A (0.2 µg/µl). TE buffer, RNase A and proteinase K mixture was also added in total chromatin samples in parallel as input reference. Reverse cross-linking was performed at thermal cycler (4 hours 37°C, 4 hours 50°C, and overnight 65°C). DNA purification was performed with SPRIselect beads (Beckman Coulter; B23317) and further processed for library generation with KAPA HyperPrep kit (KK8502) according to the kit protocol.

Libraries were 6-plexed and sequenced on a single land of an Illumina HiSeq 4000 with 100bp paired-end reads at the MD Anderson Sequencing and microarray facility. Sequencing reads were aligned to the human genome (hg19) using BWA-Mem (*1*), realigned around InDels using GATK (*2*), sorted and deduplicated using Piccard tools. Peaks were called in TCF4 ChIP samples compared to their input control using EaSeq with global thresholding. Peaks were annotated according to the transcription start site of the nearest RefSeq gene and filtered based upon FDR (<0.1), log2ratio of TCF4 ChIP vs. isotype control (≥2.0), peaks that overlapped between TMD8 and SUDHL2, and peaks corresponding to genes with differential expression between ABC-like DLBCL tumors with or without TCF4 DNA copy number gain. Peaks within 2kbp of the transcription start site were defined as 'promoter' peaks, those outside of the promoter region but within the coding region of the gene were defined as 'intragenic' peaks, and those outside of these regions but within 50kbp of the transcription start site were defined as distant 'enhancer' peaks. For visualization, files were converted to wiggle format and viewed using the Integrative Genomics Viewer(*9*). The wiggle file for H3K27Ac ChIP-seq for CD20+ B-cells was downloaded directly from UCSC Genome Browser (https://genome.ucsc.edu/ENCODE/downloads.html). Significantly over-represented DNA sequence motifs (FDR<0.05) were identified in TCF4 ChIP-seq peaks compared to the reference genome (hg19) using CisFinder (*10*) with the default settings. Motifs with 75% homology were collapsed to motif clusters.

Confocal Microscopy for IgM

For IgM staining, TMD8-TCF4, or HBL1-TCF4 cell lines were untreated or treated with doxycycline (60ng/ml) for 24 hours. Cell were stained with membrane staining dye (Membrite™ Fix 660/680; 30098-T) for 5 min at 37°C according to recommended protocol. Cells were washed twice with PBS and cytospun (ThermoFisher; A78300003) on glass slides at 1000 RPM for 5 min. Cells on slides were fixed with cold methanol for 15 min at -20°C, washed three times with cold PBS, followed by blocking with 5% BSA/0.1% Triton X-100 in PBS for two hour at room temperature. IgM antibody (ab200541) was diluted with 1% BSA in PBS at 1:100 dilution and incubated overnight at 4°C in a humidified chamber. Next day, slides were washed 5 times with 0.1% Triton X-100 followed by secondary antibody incubation (Alexa Fluor-555; A32727) diluted in 1% BSA in PBS at 1:250 dilution for two hours at room temperature. Slides were washed 5-6 times followed by mounting with ProLongTM Gold Antifade Mountant (ThermoFisher; P36930). Images were acquired on Leica Confocal microscope (Leica TCS SP8). Background was subtracted using cells stained with only secondary antibody (without IgM antibody incubation) before taking images. For control, cells were also stained with isotype antibody (ab170190; 1:100 dilution) followed by secondary antibody staining.

BCR stimulation and western blots

TMD8-TCF4 or SU-DHL-2-TCF4 cell lines were treated with doxycycline (60ng/ml) for 24 hours to induce TCF4 expression. Cells were washed once and re-suspended in complete RPMI-1640 media and then stimulated with soluble anti-IgM (5µg/ml; AffiniPure

F(ab')$_2$ Fragment Goat Anti-Human IgM) for 5 minutes at 37°C. Cells after stimulation were immediately lysed on ice with RIPA buffer for further western blots and probed for antibodies: TCF4 (SAB1404449), IgM (ab200541), Btk (8547), pY223-Btk (5082), BLNK (36438), pY96-BLNK (3601), and actin (12620).

Cloning and tetracycline inducible expression of TCF4 dominant-negative construct

Two dominant-negative mutants of *TCF4* gene (TCF4dn: TCF4^{R582P} or TCF4$^{\Delta BR}$) were generated using site directed mutagenesis kit (NEB; E0554S) with overlapping PCR method and were cloned into sleeping beauty vector (pSBtet-GP; Addgene 60495) by replacing luciferase gene to respective TCF4dn mutants. DLBCL cell lines (U2932, RIVA, OCI-LY10, TMD8, HBL1, SUDHL2, SUDHL4, OCI-LY1, and K-422) expressing TCF4dn genes were generated by co-transfecting transposase expressing vector (pCMV-CATT7-SB100; Addgene 34879) and sleeping beauty vector expressing TCF4dn genes or empty vector (EV) using neon transfection system according to the manufacturer instructions (Neon transfection system; Invitrogen, CA, USA). Transfected cells were selected for stable cell line generation with puromycin (1µg/ml) for one week and maintained in 10% tetracycline negative FBS (Corning; MT35075CV) containing RPMI media.

For immunoblots and ChIP-qPCR, cells were treated with doxycycline (60ng/ml) for 48 hours to induce TCF4 expression. Immunoblots and ChIP-qPCR were performed as described in “Materials and Methods”.

For cell competition assay, equal number of parental cells (non-transfected) were mixed with equal fraction of GFP+ cells possessing the *TCF4^{R582P} or TCF4$^{\Delta BR}$* tetracycline-

inducible dominant-negative construct. Cells were exposed to doxycycline (60ng/ml; every alternative day) and the GFP+ fraction of cells were measured every 2-3 days for 10 days using flow cytometry. The viability of TCF4dn expressing DLBCL cell lines was evaluated by measuring the frequency of GFP+ cells at indicated days relative to the first time point (day 0).

TCF4 overexpression and ARV771 treatment

Tetracycline inducible wild-type TCF4 expressing DLBCL cell lines (U2932 and RIVA) were generated by co-transfecting sleeping beauty vector expressing TCF4 gene with transposase expressing vector using neon transfection system as described in “Materials and Methods”. These stable lines were treated with doxycycline (60ng/ml) in order to induce exogenous TCF4 expression for 12 hours. Doxycycline induced cells were then treated with ARV771 (50nM) for 24 hours (for western blots) or 48 hours (for apoptosis assay by Annexin V/To-PRO-3 staining).

Lentiviral shRNA mediated BRD4 knockdown

Lentiviral tetracycline based inducible shRNA (TRIPZ plasmids) for BRD4 (sh1: V3THS_378004, sh2: V3THS_326487) and non-targeting control (RHS4743) were purchased from Dharmacon. Lentivirus were generated using the Lenti-X Packaging Single Shots (VSV-G) (Clontech; 631276) transfection system according to the manufacturer instructions. These lentivirus were used to transduce DLBCL cell lines (U2932 and RIVA) by spinoculation method. Cells after transduction were selected for stable cell line generation with puromycin (1µg/ml) and maintained in 10% tetracycline

negative FBS (Corning; MT35075CV) containing RPMI media. Expression of shRNA was induced by treating cells with doxycycline (1000ng/ml) for three days, lysed for immunoblots with RIPA buffer and probed for indicated antibodies.

Murine Xenograft Experiments

Reagents and antibodies. ARV-771 was kindly provided by Arvinas, Inc. (New Haven, CT) D-Luciferin (potassium salt) was obtained from Gold Biotechnology, Inc. (St Louis, MO). BD Matrigel Matrix High Concentration was obtained from BD Biosciences (Franklin Lakes, NJ) (Catalog number 354248).

Cell lines. Luciferase-expressing RIVA and U2932 cells were created by transducing cells with Luc-ZSGreen. pHIV-Luc-ZsGreen was a gift from Bryan Welm (Addgene plasmid # 39196). High GFP-expressing cells were isolated by flow sorting for GFP expression in the M. D. Anderson Flow Cytometry and Cellular Imaging Core Facility (FCCICF), a shared resource partially funded by NCI Cancer Center Support Grant P30CA16672.

In vivo studies. All animal studies were performed under a protocol approved by the IACUC at M.D. Anderson Cancer Center, an AAALAC-accredited institution. Five million RIVA or U2932 cells (mixed with Matrigel at a volume ratio of 1:1) were subcutaneously injected in the left flank of male athymic nude mice (nu/nu) (n = 8 per group). Tumor volume was calculated by the ½(length x width^2) method. Treatment was initiated when the mean tumor volumes reached ~150 mm3. Mice were treated with vehicle (10% [1:1 solutol: ethanol] and 90% D5-water, s.c. daily x 5 days per week) or ARV-771 (30 mg/kg, s.c., daily x 5 per week). The RIVA mouse model was treated for two weeks. Due to

slower tumor growth, the U2932 mouse model was treated for three weeks. For bioluminescent imaging, mice were IP-injected with 100 µL of 75 mg/kg D-Luciferin potassium salt (reconstituted in 1X PBS and sterile-filtered through a 0.2 um filter) incubated for 5 minutes, anesthetized with isoflurane and imaged once per week utilizing a Xenogen IVIS-200 imaging system (PerkinElmer) to monitor disease status and treatment efficacy. One mouse from each cohort was euthanized after three weeks of treatment for biomarker analysis. Mice bearing tumors greater than 1500 mm^3 were removed from study and humanely euthanized (carbon dioxide inhalation and cervical dislocation) according to the IACUC-approved protocol. Veterinarians and veterinary staff assisting in determining when euthanasia was required were blinded to the experimental conditions of the study. Tumor size was compared among cohorts by unpaired t-test. The survival of the mice is represented by a Kaplan Meier plot. Differences in survival were calculated by a Mantel-Cox log-rank test. P values less than 0.05 were considered significant. For immunoblotting, tumors from respective vehicle and ARV771 treated xenografts were extracted. Tumor lysates were prepared using RIPA lysis buffer as described in "Materials and Methods" section and probed for indicated primary antibodies.

Supplementary References

1. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **1303.3997**, (2013).
2. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).
3. D. C. Koboldt *et al.*, VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* **22**, 568-576 (2012).

4. M. R. Green *et al.*, Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E1116-1125 (2015).
5. K. Krysiak *et al.*, Recurrent somatic mutations affecting B-cell receptor signaling pathway genes in follicular lymphoma. *Blood* **129**, 473-483 (2017).
6. A. Bouska *et al.*, Adult High Grade B-cell Lymphoma with Burkitt Lymphoma Signature: Genomic features and Potential Therapeutic Targets. *Blood*, (2017).
7. T. Kuilman *et al.*, CopywriteR: DNA copy number detection from off-target sequence data. *Genome biology* **16**, 49 (2015).
8. M. Ceribelli *et al.*, A Druggable TCF4- and BRD4-Dependent Transcriptional Network Sustains Malignancy in Blastic Plasmacytoid Dendritic Cell Neoplasm. *Cancer cell* **30**, 764-778 (2016).
9. J. T. Robinson *et al.*, Integrative genomics viewer. *Nature biotechnology* **29**, 24-26 (2011).
10. A. A. Sharov, D. B. Dudekula, M. S. Ko, CisView: a browser and database of cis-regulatory modules predicted in the mouse genome. *DNA research : an international journal for rapid publication of reports on genes and genomes* **13**, 123-134 (2006).