

Supplemental Methods

Burkitt lymphoma samples

All clinical samples were studied according to an IRB protocol approved by the National Cancer Institute. Pre-treatment sBL biopsy samples (n=28) used for RNA-seq were previously analyzed by gene expression profiling¹. In addition, RNA-seq was performed on 13 BL cell lines. For Sanger sequencing analysis of *CCND3*, *TCF3*, and *ID3*, sBL was represented by primary biopsy samples (n=78) and EBV-negative BL cell lines derived from sBL (n=11). Clinical characteristics of Burkitt lymphoma and diffuse large B cell lymphoma cases used for RNA-seq or Sanger sequencing are summarized in Supplemental Table 6. hivBL samples (n=9) were from patients treated at the National Cancer Institute². eBL was represented by early passage cultures derived from primary eBL biopsy samples (n=35)³, by long-term EBV-positive cell lines derived from eBL cases (n=5) and primary biopsies of endemic Burkitt lymphoma (n=14). Formalin fixed paraffin embedded primary biopsies of endemic Burkitt lymphoma were obtained from St. Mary's Hospital Lacor as part of the EMBLEM (Epidemiology of Burkitt lymphoma in East African children and minors) study. FFPE tissue sections were evaluated by three pathologists confirming BL diagnosis.

Cell lines

Cell lines were cultured in RPMI 1640 medium supplemented with penicillin/streptomycin and 10% fetal bovine serum or, for the OCI series of cell lines, Iscove's medium with 20% fresh human plasma. Cells were maintained in a humidified, 5% CO₂ incubator at 37°C. All cell lines were engineered to express an ecotropic retroviral receptor and the bacterial tetracycline repressor as previously described⁴.

Sample preparation for RNA-sequencing

RNA was extracted using the AllPrep kit (QIAGEN) following the manufactures instructions. Sequencing libraries were prepared using the TruSeq RNA sample Prep kit v2 (Illumina) following the manufactures instructions. Paired end 108 bp read sequencing was performed on a HiSeq 2000 system (Illumina).

RNA-Seq alignment and analysis

Paired-end reads were mapped to the RefSeq database (National Center for Biotechnology Information (NCBI build 37) using the Burrows-Wheeler Aligner (BWA) software⁵ with default parameters. Reads that failed to map to RefSeq were mapped to the Ensembl database, which includes additional transcripts and pseudogenes⁶, and the remaining unmapped reads were mapped to the human genome assembly (NCBI build 37). These alignments from paired-end RNA-seq were utilized to generate single nucleotide variation (SNV) calls. Reads that mapped to the same starting position (redundant reads) and mismatches from reads with Phred quality scores of less than 20 were discarded. The number of mutated versus non-mutated reads was aggregated based on the three steps of RNA-seq alignment. Initial SNVs were recorded if there was evidence for more than three mutant reads from both the forward and reverse strand and the ratio of mutant reads vs. total coverage was greater than 20%. Based on initial discovered SNVs, we then extracted all of the original reads that covered those SNV locations. If these spanned exon boundaries, the extracted reads were split based on known exon structures before aligning them to the human reference genome (NCBI build 37) using Bowtie⁷ and NovoAlign (<http://www.novocraft.com>). The new list of SNVs was generated based on the mismatches that were detected in these alignments and matched to the

initial list of SNVs. With this approach, we removed the majority of false positive SNVs. SNVs that corresponded to SNPs in dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>, build#132) or the 1000 genomes database (<http://www.1000genomes.org/>; May 2011 release) were excluded. Functional effects of SNVs were predicted using SIFT (Sorting Intolerant From Tolerant)⁸ and Polyphen-2 (Polymorphism Phenotyping) (<http://genetics.bwh.harvard.edu/pph2/>) using three protein data sets for prediction (Blink/Swiss-Prot/NCBI). Gene expression was calculated in reads per kilobase per million mapped sequence reads (RPKM)⁹. RefSeq transcriptome definition was used to derive digital gene expression. For genes with single transcript isoforms, full length transcripts were utilized to calculate RPKM. For genes with multiple isoforms, the length of gene was calculated as all bps covered by any exon of this gene. Sequencing data has been deposited in NCBI Sequence Read Archive (SRA048058).

Calculation of significance of BL specific mutations

In calculating the p-value for the mutation prevalence between subtypes, we felt it important to take into account the fact that the different subtypes may have a different level of coverage for a given gene, and so that a reduction in observed prevalence may actually represent a difference in detection. We observed that for a mutation with high coverage, an average of 40% of the reads were found to be mutant. Thus, given a mutation with coverage c , the probability that it would satisfy our requirement of 3 or more mutant reads could be calculated from the binomial distribution as

$$P(\text{Mutant Reads} \geq 3 | \text{Mutant sample}) = 1 - \sum_{k=0}^2 \binom{c}{k} (0.4)^k (0.6)^{c-k}$$

To calculate this value for individual cases, we used the average coverage of a sample and a continuous version of the binomial so that the detection probability of a given sample with average coverage c_i was defined as

$$d_i \equiv 1 - \sum_{k=0}^2 \frac{\Gamma(c_i + k + 1)}{\Gamma(c_i + 1)\Gamma(k + 1)} (0.4)^k (0.6)^{c_i - k}$$

If $c_i < 3$ then this was modified to assume a coverage of 3 over a portion of the gene, so that

$$d_i \equiv \frac{c_i}{3} (0.4)^3$$

From this we derive

$$P(\text{Observing mutation in sample } i) = d_i P(\text{sample } i \text{ is a Mutant})$$

or

$$P(\text{sample } i \text{ is a Mutant}) = \frac{P(\text{Observing mutation in sample } i)}{d_i}$$

From this we define

$$Q_1 \equiv \frac{\text{\#Mutants observed in subset 1}}{\sum_{i \in \text{Subset 1}} d_i}$$

$$Q_2 \equiv \frac{\text{\#Mutants observed in subset 2}}{\sum_{i \in \text{Subset 2}} d_i}$$

$$Q_T \equiv \frac{\text{Total \#Mutants observed in both subsets}}{\sum d_i}$$

To be the true prevalence of mutations within subset 1, subset 2, or the combined data taking coverage into account. Under a null hypothesis in which Q_T is the true prevalence for both subgroups, we find by application of the central limit theorem that

$$Q_1 - Q_2 \sim N\left(0, \frac{\sum_{i \in \text{Subset 1}} (Q_T d_i)(1 - Q_T d_i)}{(\sum_{i \in \text{Subset 1}} d_i)^2} + \frac{\sum_{i \in \text{Subset 2}} (Q_T d_i)(1 - Q_T d_i)}{(\sum_{i \in \text{Subset 2}} d_i)^2}\right)$$

From which we can derive a p-value for the difference of mutation prevalence in the two groups adjusted for coverage differences. This adjusted p-value relies on the assumption that the coverage is uniform across the gene, and so may produce anomalous significant results particularly in genes with extremely low coverage. Therefore to be conservative we also applied a Fisher's exact test on the number of observed mutations ignoring difference in coverage, and reported the larger of the two p-values.

Sanger re-sequencing of pSNVs

Genomic DNA containing putative SNVs was PCR amplified using primers shown in Supplemental Table 7. PCR products were bidirectionally sequenced using an ABI 3730 Genetic Analyzer (Applied Biosystems). Sequence electropherograms were manually reviewed.

Gene copy number analysis of CDKN2A

DLBCL samples were previously analyzed by array CGH¹⁰. Additional BL and DLBCL samples were analyzed by TaqMan Copy Number Assay (Applied Biosystems).

Retroviral vectors and retroviral transduction

Retroviral transductions and shRNA toxicity assays were done as previously described¹¹. The shRNA sequences used in individual experiments are listed in Supplemental Table 8.

Expression vectors and cDNA mutagenesis:

The expression vector, vLyt2-CCND3 was created by inserting PCR-generated *CCND3* cDNA (Origene) into the pBMN-IRES-Lyt2 vector (provided by G. Nolan), using the following PCR primers: 5'-

ACGCGGCCGCGGATCCGAAATCACCATGGAGCTGCTGTGTTGCGAAGGCACC -3'

and 5'-GTGCGGCCGCGCCTCGAGTTATCACTACAGGTGTATGGCTGTGACATCTGTAG-3'.

The vLyt2-EGFP-CCND3 expression vector, encoding amino terminus EGFP-tagged CCND3 was constructed by ligation of two step PCR-generated CCND3 and EGFP fusion product into the pBMN-IRES-Lyt2 vector. CCND3, EGFP or EGFP-CCND3 fusion PCR products were generated using the following primer pairs:

5'- ACGCGGCCGCGGATCCGAAATCACCATGGTGAGCAAGGGCGAGGAGCTGTTAC -3' (GFP-for), 5'-

GTGCCGGGTGCCTTCGCAACACAGCAGCTCCTTGTACAGCTCGTCCATGCCGAGAG TG -3' (EGFP/CCND3-rev), 5'-

CACTCTCGGCATGGACGAGCTGTACAAGGAGCTGCTGTGTTGCGAAGGCACCCGGC AC -3' (EGFP/CCND3-for), 5'-

GTGCGGCCGCGCCTCGAGTTATCACTACAGGTGTATGGCTGTGACATCTGTAG -3' (CCND3-rev).

CCND3 mutants were created with the QuikChange Site-Directed Mutagenesis Kit (Agilent), using either vLyt2-EGFP-CCND3 or vLyt2-CCND3 vector as templates. All cDNA inserts from PCR cloning and site-directed mutagenesis were verified by sequencing. The CCND3 mutagenesis primers used were the following:

Q276*-for 5'- GGCTCCAGCAGCTAAGGGCCCAGCCAGAC -3' and Q276*-rev 5'-
GTCTGGCTGGGCCCTTAGCTGCTGGAGCC-3'. T283A-for 5'-
CAGCCAGACCAGCGCTCCTACAGATGTC -3' and T283A-rev 5'-
GACATCTGTAGGAGCGCTGGTCTGGCTG-3'. P284L-for 5'-
CAGCCAGACCAGCACTCTTACAGATGTCACAG -3' and P284L-rev 5'-
CTGTGACATCTGTAAGAGTGCTGGTCTGGCTG-3'. I290R-for 5'-
GATGTCACAGCCAGACACCTGTAGTG-3' and I290R-rev 5'-
CACTACAGGTGTCTGGCTGTGACATC-3'.

ID3 expression vectors were created by inserting PCR-generated *ID3* cDNAs (Origene)
into the pRetroCMV/TO/PG vector, using the PCR primers: 5'-
CTGGATCCGCCACCATGAAGGCGCTGAGCCCGGTGC -3' and 5'-
GACTCGAGTTATCAGTGGCAAAAGCTCCTTTTGTCG -3'. pRetroCMV/TO/PG FLAG-
ID3 encoding amino terminus FLAG-tagged *ID3* were created similarly using primers 5'-
CTGGATCCGCCACCATGGACTACAAGGATGACGATGACAAGAAGGCGCTGAGCCC
GGTGC -3' and 5'-
TGCACTCGAGTTATCATTTATCGTCATCGTCTTTGTAGTCGTGGCAAAAGCTCCTTTT
GTCGTTG -3'.

The *ID3* mutants were generated with the QuikChange Site-Directed Mutagenesis Kit
using the following mutagenesis primers: C47*-for 5'-
GACATGAACCACTGATACTCCCGCCTGCG-3' and C47*-rev 5'-
CGCAGGCGGGAGTATCAGTGGTTCATGTC-3'. P56S-for 5'-
GCGGGAAGTGGTATCCGGAGTCCCGAG-3' and P56S-rev 5'-

CTCGGGACTCCGGATACCAGTTCCCGC-3'. Q63FS-for 5'-
GTCCCGAGAGGCACTCACTTAGCCAGGTGGAAATC-3' and Q63FS-rev 5'-
GATTTCCACCTGGCTAAGTGAGTGCCTCTCGGGAC-3'. L64F-for 5'-
GAGAGGCACTCAGTTTAGCCAGGTGG-3' and L64F-rev 5'-
CCACCTGGCTAAACTGAGTGCCTCTC-3'. E68*-for 5'-
CAGCTTAGCCAGGTGTAAATCCTACAGCGCGTC-3' and E68*-rev 5'-
GACGCGCTGTAGGATTTACACCTGGCTAAGCTG-3'. ID3 splice site mutants were
generated by PCR cDNA from BL2 cell line using the same primers used for generation of ID3
wild type expression vector.

E47 expression vectors were created by inserting PCR-generated *E47* cDNAs (Origene)
into the pRetroCMV/TO/PG vector, using the PCR primers: 5'-
CTGGATCCGCCACCATGAACCAGCCGCAGAGGATGGC - 3' and 5'-
GACTCGAGTTATCACATGTGCCCGGCGGGGTTGTG - 3'. E47 mutants were created with
the QuikChange Site-Directed Mutagenesis Kit, using pRetroCMV/TO/PG-E47 vector as
templates. *E47* mutants were generated using the following mutagenesis primers: V557E-for:
5'- GCATGGCCAATAACGCGCGGGAGCGGGAGCGCGTGCGGGATATTAACGAGG - 3'
and V557E-rev: 5'-
CCTCGTTAATATCCCGCACGCGCTCCCGCTCCCGCGCGTTATTGGCCATGC - 3'.
N551K-for: 5'- GAGAGGCGCATGGCCAAGAACGCGCGGGAGCG - 3' and N551K-rev:
5'- CGCTCCCGCGCGTTCTTGGCCATGCGCCTCTC - 3'. D561E-for: 5'-
GAGCGGGTGCGCGTGCGGGAAATTAACGAGGCCTTCCGGGAG - 3' and D561E-rev:
5'- CTCCCGGAAGGCCTCGTTAATTTCCCGCACGCGCACCCGCTC - 3'..

The PTPN6 expression vector was created by inserting RT-PCR-generated *PTPN6* cDNA into the pRetroCMV/TO/PG vector, using the PCR primers 5'-TCGGATCCAAGCTTGAAATCACCATGGTGAGGTGGTTTCACCGAGACCTC -3' and 5'-GACTCGAGTTATCACTTCCTCTTGAGGGAACCCTTGC -3'.

All coding sequences were verified by full length re-sequencing

Proliferation analysis of mutant and wild type cyclin D3:

Cells were transduced with retroviral vectors co-expressing an inducible shRNAs targeting the 3' UTR of *CCND3* and a Puromycin resistance gene. Following puromycin selection, cells were superinfected with retroviruses (vLyt2-EGFP-*CCND3*) expressing either wild-type or mutant *CCND3* coding regions fused to GFP. At this point shRNA expression was induced and the cell fraction positive for GFP and Lyt2 was monitored over time by flow cytometry.

Apoptosis assay:

Cells were fixed using 1.5 % paraformaldehyde (Electron Microscopy Sciences) and ice-cold methanol, stained with anti-Active Caspase-3-PE (Cat. # 51-68655X) (BD Pharmingen) and with anti-Cleaved PARP-Alexa647 (F21-852) (BD Pharmingen) antibodies and analyzed by FACS.

Cell cycle analysis:

Paraformaldehyde-methanol-fixed cells were re-suspended in PBS containing 50 µg/ml Propidium Iodide and 0.5 mg/ml RNaseA followed by incubation at 37°C for 30 min and FACS analysis.

Pulse chase analysis of wild type and mutant cyclin D3

Gumbus cells were transduced with GFP-tagged wild type or mutant CCND3 (vLyt2-EGFP-CCND3). 10^7 cells per time-point were incubated in cysteine / methionine free media for 30 minutes, pulse labeled with 0.3mCi/ml EasyTag EXPRESS³⁵S protein labeling mix (Perkin Elmer) for 45 minutes and then chased with RPMI with unlabeled cysteine and methionine (2.5mM). At indicated time-points cells were lysed in RIPA lysis buffer and GFP-CCND3 fusion protein was immunoprecipitated with GFP-Trap_A beads (Chromotek). SDS-PAGE gels were dried and bands corresponding to GFP-CCND3 were quantified using a phosphorimager (Storm 860 Molecular Imager) and analyzed with ImageQuant software.

Tumor model and therapy study using PD 0332991:

A murine model of human Burkitt lymphoma was established by subcutaneous injection of 1×10^7 Gumbus cells which stably express luciferase (Gumbus/luc) into the right flank of female nonobese diabetic/severe combined immunodeficient (NOD/SCID) mice (Jackson Laboratory). The tumor growth was monitored by both bioluminescence imaging and measuring tumor size in two orthogonal dimensions. Bioluminescence images were obtained using an *in-vivo* imaging system and the images were analyzed using Living Imaging software (Xenogen Corporation). Tumor volume was calculated by using the formula $\frac{1}{2}(\text{long dimension})(\text{short dimension})^2$. Eleven days after injection of the tumor cells, the average tumor volume reached

100mm³ and drug therapy was started. The Gumbus/luc tumor bearing NOD/SCID mice were divided into two groups, with comparable tumor burden between groups as evaluated by bioluminescence imaging and tumor volume. PD 0332991 (Chemietek) was dissolved in sodium lactate buffer (50 mM, pH 4) and given daily at 150 mg/kg by oral gavage for 10 days. The control mice received the same amount of sodium lactate buffer by oral gavage. Tumor volume was monitored during this time. At day 10 after initiation of the therapy, all of the mice were euthanized. All animal experiments were approved by the National Cancer Institute Animal Care and Use Committee (NCI ACUC) and were performed in accordance with NCI ACUC guidelines

ID3 overexpression toxicity assay:

pRetroCMV/TO/PG-ID3 vector that co-expresses ID3 coding sequence and GFP was transduced into lymphoma cell lines. Two days after retroviral transduction, doxycycline was added to induce ID3 gene expression and toxicity was evaluated by measuring the frequency of GFP⁺ cells relative to the first time point.

Gene expression analysis of cells with TCF3 knock down, ID3 overexpression or

Rapamycin treatment:

For TCF3 knockdown, BL cell lines were transduced with either pRSMX-P-shTCF3 or pRSMX-P-ctrl. For studying gene expression changes by overexpression of ID3 BL cells were transduced with pRetroCMV/TO/PG-ID3-WT or pRetroCMV/TO/PG-empty. Transduced cells were selected using puromycin (2 µg/ml) for 3-7 days and shRNA or gene expression was induced with doxocyclin (50 ng/ml) for indicated time points. For gene expression changes by

Rapamycin, cells were treated with either 100 pM of Rapamycin (EMD Chemicals) or DMSO for 3, 6, 12 and 24 hours. Gene expression profiling was performed using two-color human Agilent 4x44K gene expression arrays following the manufacturer's protocol comparing signal from DMSO or dox-treated control cells (Cy3) to experimentally manipulated cells (Cy5). Array elements were filtered for those above confidence thresholds for spot size, architecture, and level above local background. These criteria are a feature of the Agilent gene expression software package for Agilent 4x44k arrays. Gene expression data has been deposited in GEO.

Rapamycin and TCF3 signature analysis

A “rapamycin down-regulated” signature was defined as those genes that decreased in expression by 0.4 log₂ in at least 3 of the 4 time points in both BL cell lines. A “rapamycin up-regulated” signature was defined as those genes that increased in expression by 0.4 log₂ in at least 3 of the 4 time points in both BL cell lines. A “TCF3 up-regulated” gene signature was defined as those genes that decreased by >0.33 log₂ in >70% of samples with ID3 overexpression or in >70% of samples with TCF3 knock down. A “TCF3 up-regulated” gene signature was defined as those genes that increased by >0.33 log₂ in >70% of samples with ID3 overexpression or in >70% of samples or with TCF3 knock down. To test the statistical significance of the TCF3 regulated genes, we performed a one-sample Random Variance T-test¹² on the 18 arrays representing TCF3 knockdown or ID3 over expression. A similar test was done on the 8 Rapamycin arrays.

We wished to test the hypotheses that the rapamycin down-regulated signature genes were more highly expressed in BL than in GCB DLBCL and that the rapamycin up-regulated signature genes were more highly expressed in GCB DLBCL than in BL. We calculated the t-

test between BL and GCB DLBCL samples for all genes on the U133+ array, and calculated the Kolmogorov-Smirnov (KS) statistic for the difference in the distribution of these statistics within the rapamycin up- and down-regulated signatures and for the entire gene set. We randomly permuted the BL and GCB DLBCL class labels, and repeated the same procedure to obtain a KS statistic for these genes on effectively random data. This was repeated 10000 times. A permutation p-value was calculated based on the proportion of times that the permuted KS statistic was greater in absolute value than the value obtained for the unpermuted labels. A similar test was performed to demonstrate the enrichment of genes highly expressed in BL versus GCB or ABC among the “TCF3 up-regulated” gene signature and genes highly expressed in GCB or ABC versus BL among the “TCF3 down-regulated” gene signature.

Validation of gene expression changes in cells with TCF3 knock down by qPCR:

Indicated BL cell lines were transduced with either TCF3 or control shRNAs and RNA was extracted, DNase digested and retro transcribed using the Omniscript RT kit (QIAGEN). QPCR was performed using either Taqman probes or SybrGreen reagent (Applied Biosystems) using the primers: ID3-for: 5'- CTTAGCCAGGTGGAAATCCTACAG – 3' and ID3-rev 5'- TTGTCGTTGGAGATGACAAGTTCC – 3'. IGKC-for: 5'- TCTTCCCGCCATCTGATGAGCAGTTG– 3' and IGKC-rev: 5'- TTGCTGTCCTGCTCTGTGACACTCTC– 3'. IGHM-for: 5'- AGCAGAATGCGTCCTCCATGTGTG– 3' and IGHM-rev: 5'- GTCATAGGTGGTCAGGTCTGTGAC– 3'.

Downregulation of BCR in shTCF3 knockdown cells:

BL cell lines were transduced with either shRNA targeting TCF3 or control shRNAs, induced for 1 day with doxycycline (50 ng/ml), stained with anti-CD79B-Alexa647 antibody (SN8) (BD Pharmingen) and analyzed by FACS.

Chromatin immunoprecipitation:

For each chromatin preparation, 1×10^8 BL cell line cells were collected, resuspended in RPMI without FBS and cross-linked with 1% formaldehyde for 5' at RT'. Cross-linking was quenched by addition of 125 mM Glycine for 5' at RT. Cells were washed with ice-cold PBS and nuclei were isolated by rocking for 10' at 4°C in NP-40 buffer.

Isolated nuclei were resuspended in shearing buffer (0.1% SDS, 10mM EDTA and 50 mM Tris-HCl pH 8.0) and DNA was fragmented to a 200-500bp range by a combination of Covaris S220 (Covaris) and Misonix XL (Qsonica) sonicators. Following sonication, chromatin preparations were diluted with RIPA buffer.

For each immunoprecipitation, 5×10^7 chromatin cell equivalents were incubated over night with 10 µg of the antibodies anti-E2A (Santa Cruz Biotechnology; Cat# SC-349X) or anti-Flag-M2 (Sigma) and subsequently incubated with 50 µl of Protein G magnetic beads (Invitrogen) for 4 h at 4°C. Protein G bound complexes were washed 4 times with RIPA Buffer, once with LiCl Buffer (10 mM Tris-HCl pH8, 250 mM LiCl, 0.5% NP40, 0.5% Sodium Deoxycholate, 1 mM EDTA), once with TE pH 8.0 and resuspended in 100 µl TE pH8 containing RNase A (0.2 µg/µl). Unprocessed chromatin samples were processed in parallel as input reference. Reverse cross-link was performed over night at 65°C, followed by treatment with 20 µg Proteinase K (Invitrogen) for 2 h at 50°C and DNA purification using QIAquick PCR Purification columns (QIAGEN).

ChIP samples and corresponding input DNA were analyzed by real-time PCR after a 10 fold dilution, in triplicates reactions using SybrGreen PCR master mix on an ABI7500 Taqman instrument. Primers are listed in Supplemental Table 9.

Flag-BIOTIN Chromatin Immunoprecipitation

Vectors containing *BirA* (pEF1 α BirAV5) and the double tag Flag-BIOTIN (pEF1 α FLBIO) constructs were kindly provided by Dr. Orkin, SH (Dana-Farber Cancer Institute). pRCMV/TO/ Flag-BIOTIN/E47 construct was obtained by insertion of cDNAs encoding E47 (WT, N551K, V557E and D561E). BL41 and Namalwa cells stably expressing the prokaryotic biotinylating enzyme *BirA* were transduced with either empty vector or pRCMV/TO/Flag-BIOTIN E47 (WT, N551K, V557E and D561E) and selected via puromycin. Flag-BIOTIN-E47 expression was induced using 20 ng/ml Doxycycline 16h before cell harvest. Fixation and sonication were performed as described above. 2.5×10^7 chromatin cell equivalents were incubated with 50 μ l of streptavidin magnetic beads (Invitrogen) overnight at 4°C. Streptavidin bound complex were washed 4 times with RIPA Buffer, once with LiCl Buffer, and once with TE pH8, before resuspension in 150 μ l TE pH8 containing RNase A (0.2 μ g/ μ l final). Unprocessed chromatin samples were processed in parallel as input reference. Reverse cross-linking was performed over night at 65°C. Samples were treated with Proteinase K (Invitrogen) for 4h at 50°C and DNA was purified with QIAquick PCR Purification columns (QIAGEN). ChIP DNA was used to generate ChIP-Seq libraries with the ChIP-Seq Sample Preparation Kit (Illumina), according to manufacturer's instructions. Sequencing was performed on a GA2x (Illumina) sequencer, with single reads (SR) of 36 bp length.

ChIP-Seq Peak calling

36 bp sequence tags were aligned to human genome build 36 (hg18) with ELAND_EXTENDED software. Redundant reads were removed and reads uniquely mapping to reference genome were used for further analysis. A maximum of two mismatches was allowed for each read.

To perform “peak calling”, the genome was divided into “bins” or 25 bp. Each sequencing tag was associated with the bin of its start site and an additional extension of 7 bins (for a total of 200bp) along the direction of its read to match the average library size. For a given experiment, the number of “hits” in a bin was defined as the number of extended tags associated with that bin. Bins with fewer than 3 hits were removed. Within the BL41 ChIP-seq experiments, we achieved the greatest proportion of bins within 2 kb of gene transcription start sites (TSS) if we restricted ourselves to bins that had 12 or more hits, and we used this as a cutoff for a bin having significant binding of TCF3 in BL41. “Peaks” were formed by merging consecutive significant bins. The height of the peak was defined as the largest number of hits in any of the bins forming the peak. The “apex” of the peak was defined as the bin that had this largest number of hits. In the case of ties, the earliest such peak was chosen as the apex. Peaks for TCF3 binding Namalwa cells were defined similarly, with 13 hits found to be the optimal choice for significance. We also performed a pair of control experiments based on an empty vector controls in the BL41 and Namalwa cell line, to remove artifactual binding effects. Those peaks which contained a bin with 8 or more hits from the BL41 control or 9 or more hits from the Namalwa control (which had a greater number of overall tags) were excluded from the table. A peak was defined to be associated with a RefSeq accession if its apex was either within the body of the gene or was upstream of the gene by no more than 10kb. All ChIP-Seq data are available under the SRA accession SRA052618.

ChIP Seq Statistical analysis

To identify genomic regions that showed significantly different binding in between E47 WT and the N551K mutant, in both BL41 and Namalwa BL cell lines, the following strategy was adopted. For each cell line, we assumed that the number of hits in a bin for a given experiment was distributed according to a Poisson distribution, and that under the null hypothesis the expected number of hits in a given bin will be equal in the two experiments up to a possible experiment dependent scaling factor that will be the same for all bins. Under this formulation the maximum likelihood test for the difference in numbers of hits reduces to a binomial test for the number of wild type hits observed in a bin relative to total number of combined wild type and N551K hits. The null probability of a wild type hit for the test depends on the proper scaling factor for each group. We estimated this by considering all the bins that had 3 or more hits for both the TCF3 and N551K experiments, and which had no hits in the empty vector control library. This resulted in a WT probability of 0.446 in the BL41 experiment, and 0.442 in the Namalwa experiment. We used these estimates to generate for each bin in each experiment a binomial p-value for the E47 WT vs. E47 N551K differential binding. Only those bins for which this p-value was less than 0.001 in both the BL41 and N551K, and for which the direction of the effect was the same (larger number of hits in both comparisons WT versus N551K, or reduced number of hits in both comparisons WT versus N551K) were defined significantly differential bins. As before, sets of one or more significant 25-bp bins were merged to identify differentially bound genomic regions. In doing so, we weakened the previously described requirement of continuity and allowed for gaps of 3 or fewer non-significantly different bins to be ignored when merging bins to generate a “differential peak”. For each “differential peak” we calculated the product of the BL41 and Namalwa p-values of each bin, and defined the Apex of the peak as the

bin that had the lowest P-value product. As before, a peak was associated with a RefSeq if its apex was either within the body of the gene or was upstream of the gene by no more than 10kb.

Supplemental Table 14 shows differential peaks that had at least one bin with a P-Value of less than 10^{-10} in both experiments, and that at the same time satisfied, in both experiments, a fold change cutoff of at least 4 fold (up or down) between the bin with the most WT hits and the one with the most N551K hits. The 139 genomic regions showing N551K binding greater than the WT, and the 212 genomic regions showing WT binding greater than the N551K were used as input sequences to perform motif analysis with the software MEME¹³. A ± 200 bp window centered on the peak apex was used for this analysis.

E47 rescue experiment:

BL cell lines were transduced with pRetroCMV/TO/P-E47-WT or pRetroCMV/TO/P-E47-mutant retroviral constructs co-expressing a puromycin resistance gene and E47 wild type or E47 mutant coding sequences. Selected cells were subsequently transduced with a construct expressing GFP reporter gene and inducibly expressing a shRNA targeting the 3'UTR of TCF3. GFP expressing cells were monitored over indicated time.

Phosflow Analysis:

Cell lines were first transduced with virus and shRNA expression was induced with the addition of 50 ng/ml doxycycline for 2 days. At this point, cells were fixed in 1.5% paraformaldehyde (Electron Microscopy Sciences) for 10 minutes at room temperature to preserve phosphorylation, spun down and permeabilized with cold methanol overnight. Cells were washed 2-times in FACS buffer (PBS+0.5% FBS) and stained with p-Akt (S473)-

Alexa647 (D9E) or isotype control (DA1E) (Cell Signaling Technology) for 20 minutes at room temperature. Cells were washed and resuspended in FACS buffer for analysis on a FACSCalibur (BD Biosciences).

Mutant ID3 and mutant E47 Western Blot Analysis:

All lysates were extracted in NP40 buffer, incubated on ice for 10' and vortexed at 4 °C for 2'. Total extracts were quantified with the Bradford reagent and equal amounts were used either for direct western blots or to set up TCF3 immunoprecipitation reactions. Samples were separated on 12% SDS PAGE and blotted to a 0.22 µm Nitrocellulose membrane. The following antibodies were used for protein detection: ID3 (Cat. # SC-13046) (Santa Cruz Biotechnology), TCF3-E12/E47 (Cat. # SC-349X) (Santa Cruz Biotechnology), TCF3-E47 (G127-32) (BD Pharmingen), Histone H3, unmodified (Cat. # ab1791-100) (Abcam).

pRetroCMV/TO/PG-ID3-WT (2.5 µg) was expressed in HEK293T cells together with either an empty vector control or in the presence of increasing amounts (0.1 µg, 0.5 µg, 2.5 µg) of pRetroCMV/TO/PG-E47-WT or pRetroCMV/TO/PG-E12-WT constructs. Western blot analysis was used to detect ID3 and TCF3 in whole cell lysates followed by quantization of scanned images using ImageJ software.

For the ID3 experiments, pRetroCMV/TO/PG-ID3-WT and pRetroCMV/TO/PG-ID3-mutant were transduced into the ID3-negative cell line Namalwa. Following puromycin selection and 2 days of induction with doxycycline (20 ng/ml), cells were collected for analysis. For the E47 experiments, pRetroCMV/TO/PG-E47-WT and pRetroCMV/TO/PG-E47-mutant were co-transfected together with pRetroCMV/TO/PG-ID3-WT construct into HEK293T cells. Two days after transfection, cells were collected for analysis.

Other Western Blot Analysis:

Cell lines were treated with either DMSO vehicle control or 25 μ M LY294002 (Sigma) for 1 hour under normal culture conditions. 10^6 cells were washed and resuspended in 1X Laemmli sample buffer and boiled. Samples were separated on 10% SDS PAGE gel and transferred to a PVDF membrane. Proteins were detected with the following antibodies: p-AKT (S473) (D9E), AKT (5G3), p-p70S6K (T389) (108D2), and p70S6K (Cat. #9202) (Cell Signaling Technology). BL cell lines were transduced with pRetroCMV/TO/PG-PTPN6 or pRSMX-P-shTCF3 or empty vector and selected for 2 days using puromycin. PTPN6 and TCF3 shRNA expression was induced by addition of doxycycline (20 ng/ml) for 2 days, respectively. Proteins were detected with the following antibodies: p-AKT (S473) (D9E), AKT (5G3), SHP-1 (Cat # sc-287) (Santa Cruz Biotechnology) or TCF3-E12/E47 (Cat. # SC-349X) (Santa Cruz Biotechnology). shRNA mediated knock down was verified using the antibodies: anti-CDK4 (Cat # SC-260) (Santa Cruz Biotechnology), anti-CCND3 (G107-565), anti-CCND2 (Cat # SC-181) (Santa Cruz Biotechnology).

MTS Cell Viability Assays:

BL cell lines were plated at a density of 2×10^5 cells per well of a 96-well plate with the indicated concentration of drugs. Cells were grown for 3 days and cell viability was assessed using the CellTiter 96 Aqueous Non-Radioactive Cell Proliferation Assay (Promega) following the manufacturers directions.

Supplemental Figure Legends

Supplemental Figure 1. Single nucleotide variants in BL observed by RNA-seq. a,

Frequencies of putative SNVs in AID hotspots (DGYW or WRCH motifs within 2kb from the transcriptional start site) detected by RNA sequencing. **b,** *TP53* mutations in BL. Amino acid positions 1 – 142 are shown according to protein accession NP_001119589. **c,** *MYC* mutations in BL. Amino acid positions 1 – 326 are shown according to protein accession NP_002458. *TSS:* *transcriptional start site*

Supplemental Figure 2. Cell cycle regulators in lymphomas. a,

An RNA interference screen identified shRNAs targeting CDK6 as toxic for BL cell lines. Indicated lymphoma cell lines were transduced with an shRNA library targeting 1051 genes as described⁴. After 21 days in culture, shRNA abundance was compared between cell populations in which shRNA expression was induced or uninduced. Error bars are s.e.m. (n=4). **b,** Increase of cells in G1 phase by knockdown of CCND3. BL cell lines were transduced with constructs co-expressing a puromycin resistance gene and an inducible shRNA targeting CCND3. Puromycin selected cells were induced for two days, fixed with ethanol and DNA content was analyzed by FACS using Propidium Iodide staining. **c,** Selective toxicity of CCND1, CDK4 and CCND2 shRNAs in Mantle cell lymphoma (MCL) and ABC DLBCL cell lines, respectively. Shown is the fraction of GFP⁺, shRNA-expressing cells relative to the GFP⁻, shRNA-negative fraction at the indicated times, normalized to the day 0 values. Data represent at least three independent experiments. **d,** Treatment of BL xenograft model with PD 0332991 results in profound reduction of tumor volume. Gumbus cells stably expressing luciferase were transplanted into NOD/SCID mice. PD 0332991 (150 mg/kg) or vehicle (sodium lactate 50 mmol/L, pH 4) was daily administered orally for 10 days. Tumor burden was evaluated by bioluminescence imaging. Values indicate average relative luminescence units at indicated time points. **e,** Mutant CCND3 proteins are not destabilized in response to phosphatase inhibition. BL41a cells expressing GFP-CCND3 proteins were treated for 30 min with 750 nM pan-phosphatase inhibitor okadaic acid and analyzed by FACS.

Supplemental Figure 3. Mutations affecting the ID3-TCF3 transcription factor module reveal an important role for TCF3 in BL pathogenesis.

a, The majority of BL with mutant ID3 cases harbor more than one mutation in *ID3*, indicating frequent biallelic inactivation. Frequencies of ID3 mutations in different lymphoma subtypes subdivided into cases with one mutation or more than one mutation in *ID3*. 40% of the cases with more than one ID3 mutation were analyzed by PCR cloning. In all cases tested, both alleles were mutated. See Supplemental Table 4 for details. **b,** The E47 isoform of TCF3 is highly expressed in BL. Shown are ratios of reads per kilobase per million mapped sequence reads (RPKM) of E47 and E12 isoforms. **c,** Mutations in *ID3* splice donor site result in a deletion of 57 base pairs mediated by the usage of an alternative, exonic splice donor site. **d,** An RNA interference screen revealed oncogenic addiction of TCF3 in BL cell lines. Indicated lymphoma cell lines were transduced with an shRNA library targeting 1051 genes as described⁴. After 21 days in culture, shRNA abundance was compared between cell populations in which shRNA expression was induced or uninduced. Error bars are s.e.m. (n=4). **e,** Rescue of shTCF3 mediated toxicity by E47 wild type and E47 mutants. Indicated cell lines were transduced with retroviral construct co-expressing a puromycin resistance gene and the coding sequence of E47 wild type or E47 mutants. Selected cells were subsequently transduced with a construct expressing GFP reporter gene and inducibly expressing an shRNA targeting the 3'UTR of TCF3. GFP expressing cells were monitored over the indicated time by FACS. **f,** TCF3 shRNA have selective toxicity for BL cell lines. Shown is the fraction of GFP+, shRNA-expressing cells relative to the GFP-, shRNA-negative fraction at the indicated times, normalized to the day 0 values. Data represent at least four independent experiments. **g,** Expression of wild type ID3 protein causes a time-dependent toxicity in BL cell lines. Shown is the fraction of cells co-expressing GFP relative to the GFP-positive fraction on day 0 of induction. Data represent at least three independent experiments. **h.** WT ID3 is stabilized by E47 and E12. Wild type ID3 was co-transfected in HEK293T cells with increasing amounts of WT E47 or WT E12 constructs or empty vector control. Western blot analysis indicated low expression of WT ID3 protein in the absence of E47 or E12 proteins, while strong increase of ID3 levels occurred when TCF3 isoforms were co-expressed with ID3 **i,** Validation of TCF3 target genes by qPCR. Cells were transduced with a retroviral construct co-expressing a puromycin resistance gene and an inducible shRNA targeting TCF3. Puromycin selected cells were analyzed by qPCR for the indicated genes. **j,** Chromatin immunoprecipitation (ChIP)

confirms binding of TCF3 at regulatory regions of TCF3 regulated genes. TCF3 ChIP were performed in three TCF3 dependent BL cell lines: BL41, Namalwa and Thomas. For each cell line, enriched DNA was used to perform qPCR for the indicated genomic regions. TCF3 signals were reported as % of the starting chromatin (Input).

Supplemental Figure 4. A TCF3-regulatory network contributes substantially to the BL phenotype .

a, Changes in gene expression were profiled in ID3-mutant BL cell lines following TCF3 knock down or wild type ID3 overexpression. Shown are genes that were downregulated by at least 0.33 log₂ (left panel) and genes that were upregulated by at least 0.4 log₂ (right panel) by TCF3 knockdown or by ID3 overexpression in at least 70 % of samples. To test the statistical significance of the TCF3 regulated genes, a one-sample Random Variance T-test on the 18 arrays representing TCF3 knockdown or ID3 over expression. 96% of the signature genes showed significant change in expression at $p < 0.01$ with a median p-value of 3.25×10^{-5} .

Previously published gene expression profiling datasets from BL, GCB DLBCL, and ABC DLBCL were used¹. **b,** To test the enrichment of genes highly expressed in BL versus GCB DLBCL or ABC DLBCL, a t-test between BL and GCB DLBCL or ABC DLBCL samples and Kolmogorov-Smirnov (KS) statistic for the difference in the distribution of these statistics within the TCF3 signature and for the entire gene set were calculated (see methods). In no case was a permutation KS statistic found to be more significant than the KS statistic observed on the unpermuted data. Therefore, we conclude that the gene lists are significantly enriched in those genes associated with expression in BL with $p < 0.001$. The curves indicate the cumulative distribution of GCB DLBCL or ABC DLBCL vs. BL t-statistics for the entire set of genes profiled. Genes expressed at high levels in BL tumors relative to GCB DLBCL or ABC DLBCL tumors are at the left and those at low relative levels in BL tumors are at the right. Genes in the TCF3 down-regulated signature and the TCF3 up-regulated signature are highlighted with vertical green hash marks indicating their ranking with respect to the other genes on the array. **c,** Enrichment analysis of TCF3 signature genes in BL harboring mutant ID3 and/or TCF3 and BL with wild type ID3 and TCF3 using t-test and Kolmogorov-Smirnov (KS) statistics.

Supplemental Figure 5. a, An RNA interference screen identified shRNAs targeting CD79A, but not CARD11, as toxic for most BL cell lines analyzed. Indicated lymphoma cell lines were transduced with an shRNA library targeting 1051 genes as described⁴. Error bars are s.e.m. (n=4). See text for details. **b,** Overexpression of ID3 reduces PI(3) kinase activity. The indicated BL cell lines were transduced with a control vector or ID3 overexpression vector. ID3 overexpression was induced for 18 hours, and cells were analyzed by FACS for phospho-S473-AKT staining as a measure of PI(3) kinase activity. **c,** BL cell lines do not depend on NF- κ B activation. Shown are viable cells, as quantified by an MTS assay, in cultures of BL and DLBCL lines treated with indicated concentrations of the I κ B kinase (IKK) β inhibitor MLN120B. **d,** *MIR17HG* is highly expressed in BL. Shown are RNA-seq reads per kilobase per million mapped sequence reads (RPKM) of *MIR17HG*. **e,** Enrichment analysis of rapamycin signature genes in BL and GCB DLBCL biopsies using t-test and Kolmogorov-Smirnov (KS) statistics. The red curve indicates the cumulative distribution of GCB DLBCL vs. BL t-statistics for the entire set of genes on the array. Genes expressed at high levels in BL tumors relative to GCB DLBCL tumors are at the left and those at low relative levels in BL tumors are at the right. Genes in the rapamycin down-regulated signature (top panel) and the rapamycin up-regulated signature (bottom panel) are highlighted with vertical green hash marks indicating their ranking with respect to the other genes on the array.

Supplemental Figure 6. a, shRNA control experiments for shCCND3 and shCDK6 toxicity assays (Figure 1H). Cells were infected with a negative control shRNA (left panel) or a positive control shRNA targeting *RPL6* encoding the ribosomal protein L6 (right panel). Shown is the fraction of GFP⁺, shRNA-expressing cells relative to the GFP⁻, shRNA-negative fraction at the indicated times, normalized to the day 0 values. **b,** shRNA control experiments for shCCND2, shCCND1 and shCDK4 toxicity assays (Supplemental Figure 2C); performed as described above **c,** shRNA mediated knockdown of CCND3 in OCI-Ly19 cells. Cells were transduced with an shRNA targeting *CCND3* or a control shRNA, selected with puromycin for 2 days, shRNA expression was induced using doxycycline and analyzed by Western Blot. **d,** shRNA mediated knockdown of CCND2 in HBL1 cells. **e,** shRNA mediated knockdown of CDK4 in OCI-Ly19 cells. **f,** shRNA mediated knockdown of CDK6 in Thomas cells and CCND1 in Jeko cells. Cells were transduced with an shRNA targeting *CDK6* or *CCND1* or a control shRNA, selected with

puromycin for 2 days, shRNA expression was induced using doxycycline and analyzed by qPCR. Shown is the relative mRNA expression normalized to control shRNA cells. **g**, Pulse chase experiments show increased half-life of mutant cyclin D3 isoforms. Gumbus cells were transduced with GFP-tagged wild type or mutant CCND3 and pulse labeled with 0.3mCi/ml EasyTag EXPRESS³⁵S protein labeling mix for 45 minutes. GFP-cyclinD3 fusion proteins were immunoprecipitated, immunoblotted and quantified using a phosphorimager at indicated time points.

Supplemental Figure 7. a, shRNA control experiments for shTCF3 toxicity assays (Figure 2F and Supplemental Figure 3F). Performed as described for Supplemental Figure 6A. **b**, shRNA mediated knockdown of TCF3 in Thomas cells. Cells were transduced with indicated shRNAs targeting *TCF3* or a control shRNA, selected with puromycin for 2 days, shRNA expression was induced using doxycycline and analyzed by Western Blot. **c**, Expression of wild type ID3 protein causes a time-dependent toxicity in most BL cell lines. Shown is the fraction of cells co-expressing GFP relative to the GFP-positive fraction on day 0 of induction. Data represent at least three independent experiments. **d**, ID3 is proteasomally degraded in the absence of TCF3. Wild type ID3 was transfected in 293T. Cells were treated for 4h with 100 nM, 200 nM or 400 nM of Bortezomib (PS-341) or DMSO and ID3 level assessed by Western Blot. **e**, ID3 mRNA levels of Namalwa cells transduced with mutant and wildtype ID3 shown in Figure 2I. Indicated ID3 isoforms were induced in Namalwa cells for 2 days and analyzed for ID3 transcripts by qPCR. Shown are relative ID3 expression levels normalized to ID3 expression from cells transduced with wild type ID3.

Supplemental Figure 8. ChIP-seq binding of mutant and wild type E47

a, Shown are representative ChIP-seq tracks for TCF3 target genes (20 kb window for each gene, from left to right: IgH 3' enhancer, Ig κ 3' enhancer, CD79A, CD79B, ID1, ID2, ID3, LCK, PTPN6 and CCND3). **b**, Shown are representative genomic regions with enhanced binding of N551K -TCF3 (left) (BCL10, CNTNAP2, PBK and RPL30) or reduced binding N551K-TCF3 (right) (AKT1, EIF5, RASGRF2 and UBE3C). Differential binding regions were included in which $p < 10^{-10}$ and a fold change difference of WT and N551K of at least 4 was observed (see methods).

E47 wild type (black); E47 V557E (blue); E47 D561E (green); E47 N551K (red); Empty vector (gray). Antibody based (endogenous) TCF3 ChIP-seq tracks are shown at the bottom of the panel: anti-TCF3 (black), anti-Flag (gray).

Supplemental Table 1: pSNV from RNA-seq analysis in BL and DLBCL and FL from reference ¹⁴.

Supplemental Table 2: Sanger sequence verification of pSNVs in BL identified by RNA-seq.

Supplemental Table 3: Sanger sequencing analysis of Exon 5 of CCND3 (NM_001760) in 604 cases of various lymphoma subtypes.

Supplemental Table 4: Sanger sequencing and gene copy number analysis of CDKN2A (NM_000077) in 317 cases of DLBCL and BL.

Supplemental Table 5: Sanger sequencing analysis of Exon 16 of TCF3 (NM_001136139 - E47 isoform) and the coding sequence of ID3 (NM_002167) in 412 cases of various lymphoma subtypes.

Supplemental Table 6: Clinical characteristics of primary biopsies of BL and DLBCL used for RNA-seq and Sanger sequencing.

Supplemental Table 7: Primers used for Sanger sequencing.

Supplemental Table 8: shRNA sequences used in manuscript.

Supplemental Table 9: Primers used qChIP.

Supplemental Table 10: Rapamycin gene signature

Supplemental Table 11: Results from RNA interference screen in BL

Supplemental Table 12: Overlap of WT and mutant TCF3 Chip-seq binding peaks

Supplemental Table 13: TCF3 ChIP-seq peaks present in both BL41 and Namalwa Burkitt lymphoma data sets.

Supplemental Table 14: Genomic regions differentially bound by WT and N551K TCF3

Gene expression profiling data have been submitted to GEO under accession number GSE35163

RNA-seq data has been deposited in NCBI Sequence Read Archive (SRA048058).

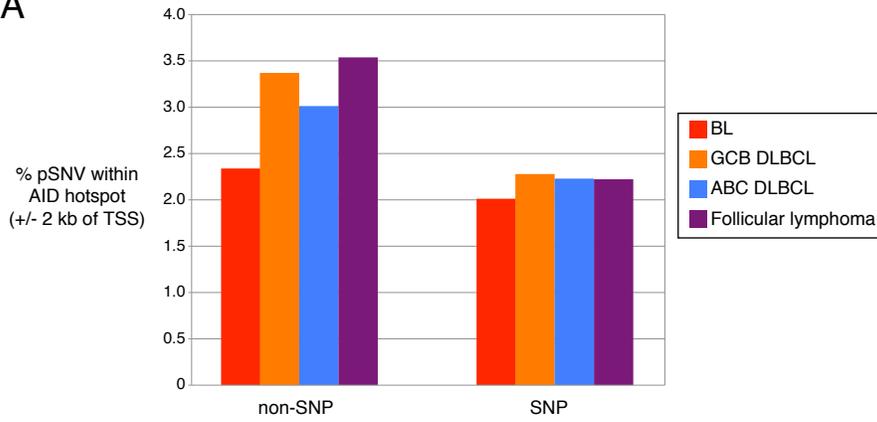
ChIP-seq data has been deposited in NCBI Sequence Read Archive (SRA052618).

Supplemental References

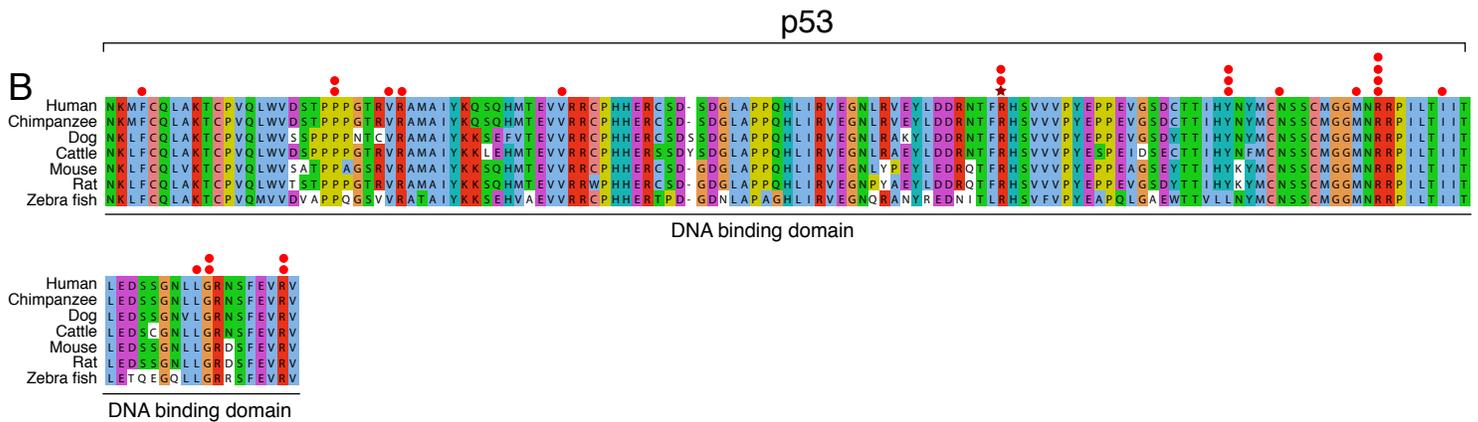
- 1 Dave, S.S. *et al.*, Molecular diagnosis of Burkitt's lymphoma. *N Engl J Med* **354**, 2431-2442 (2006).
- 2 Dunleavy, K. *et al.*, The role of tumor histogenesis, FDG-PET, and short-course EPOCH with dose-dense rituximab (SC-EPOCH-RR) in HIV-associated diffuse large B-cell lymphoma. *Blood* **115**, 3017-3024 (2010).
- 3 Kelly, G., Bell, A., & Rickinson, A., Epstein-Barr virus-associated Burkitt lymphomagenesis selects for downregulation of the nuclear antigen EBNA2. *Nat Med* **8**, 1098-1104 (2002).
- 4 Ngo, V.N. *et al.*, A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106-110 (2006).
- 5 Li, H. & Durbin, R., Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
- 6 Larsson, T.P., Murray, C.G., Hill, T., Fredriksson, R., & Schioth, H.B., Comparison of the current RefSeq, Ensembl and EST databases for counting genes and gene discovery. *FEBS Lett* **579**, 690-698 (2005).
- 7 Langmead, B., Trapnell, C., Pop, M., & Salzberg, S.L., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
- 8 Ng, P.C. & Henikoff, S., SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812-3814 (2003).
- 9 Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., & Wold, B., Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628 (2008).
- 10 Lenz, G. *et al.*, Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A* **105**, 13520-13525 (2008).
- 11 Ngo, V.N. *et al.*, Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470**, 115-119 (2011).
- 12 Wright, G.W. & Simon, R.M., A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* **19**, 2448-2455 (2003).
- 13 Machanick, P. & Bailey, T.L., MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696-1697.
- 14 Morin, R.D. *et al.*, Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* (2011).

Supplemental Figure 1

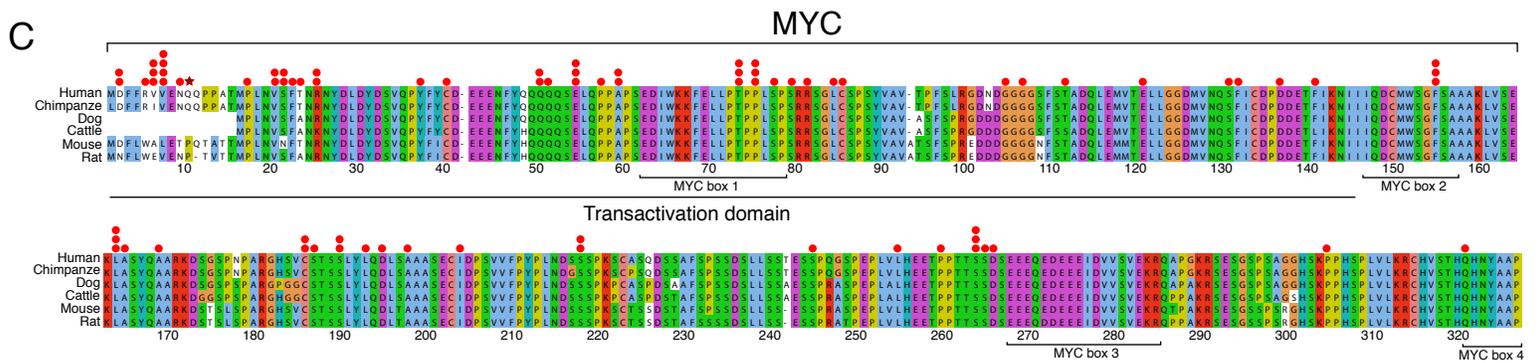
A



B

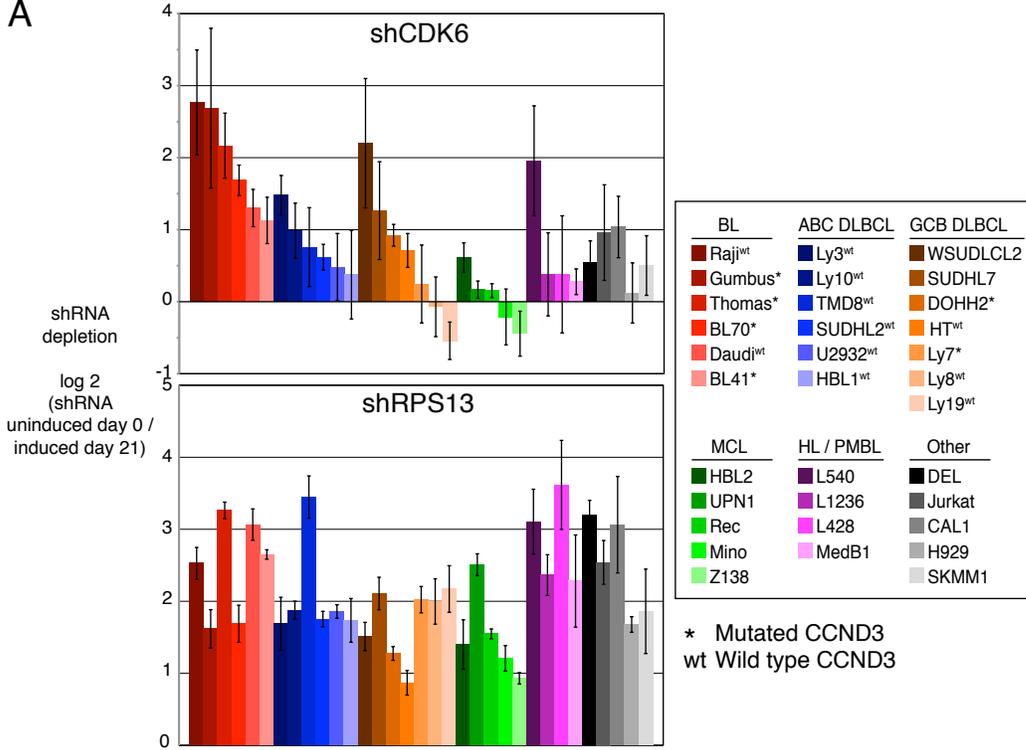


C

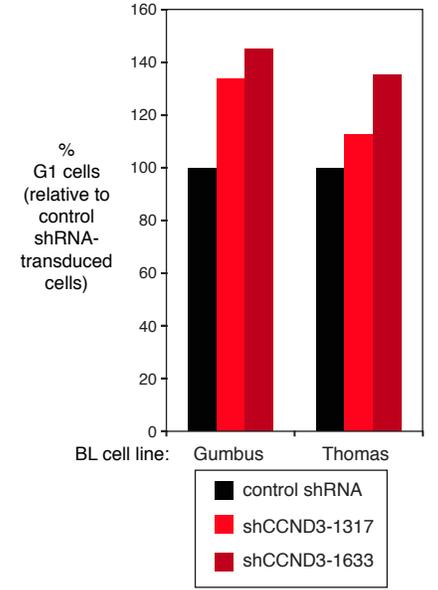


Supplemental Figure 2

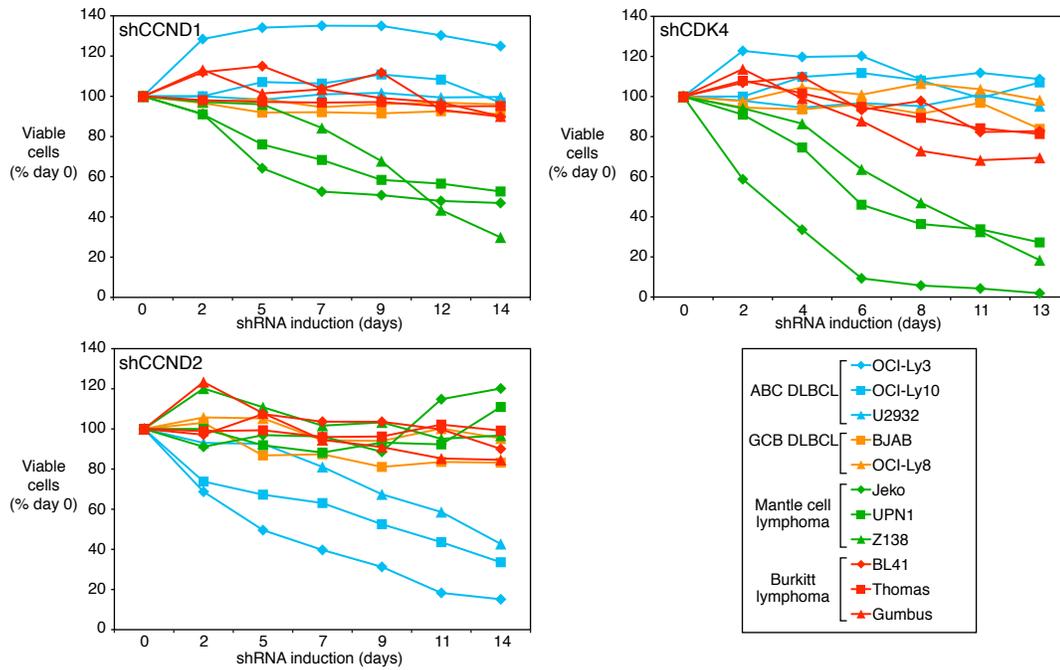
A



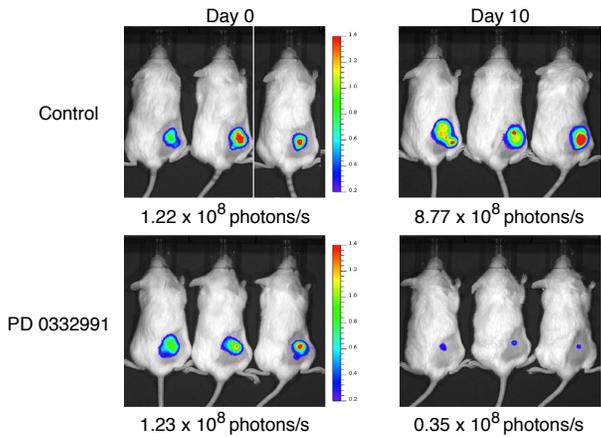
B



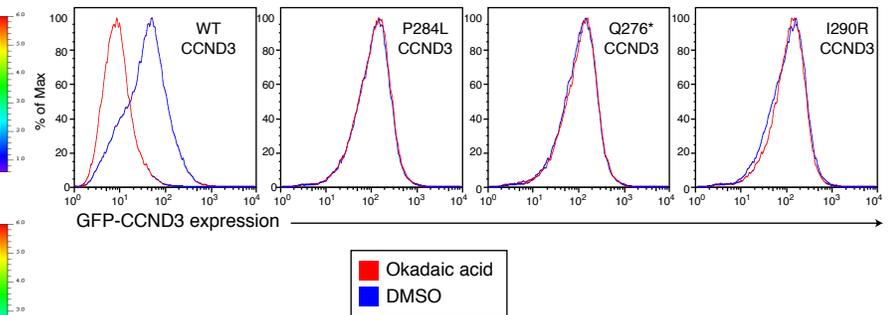
C



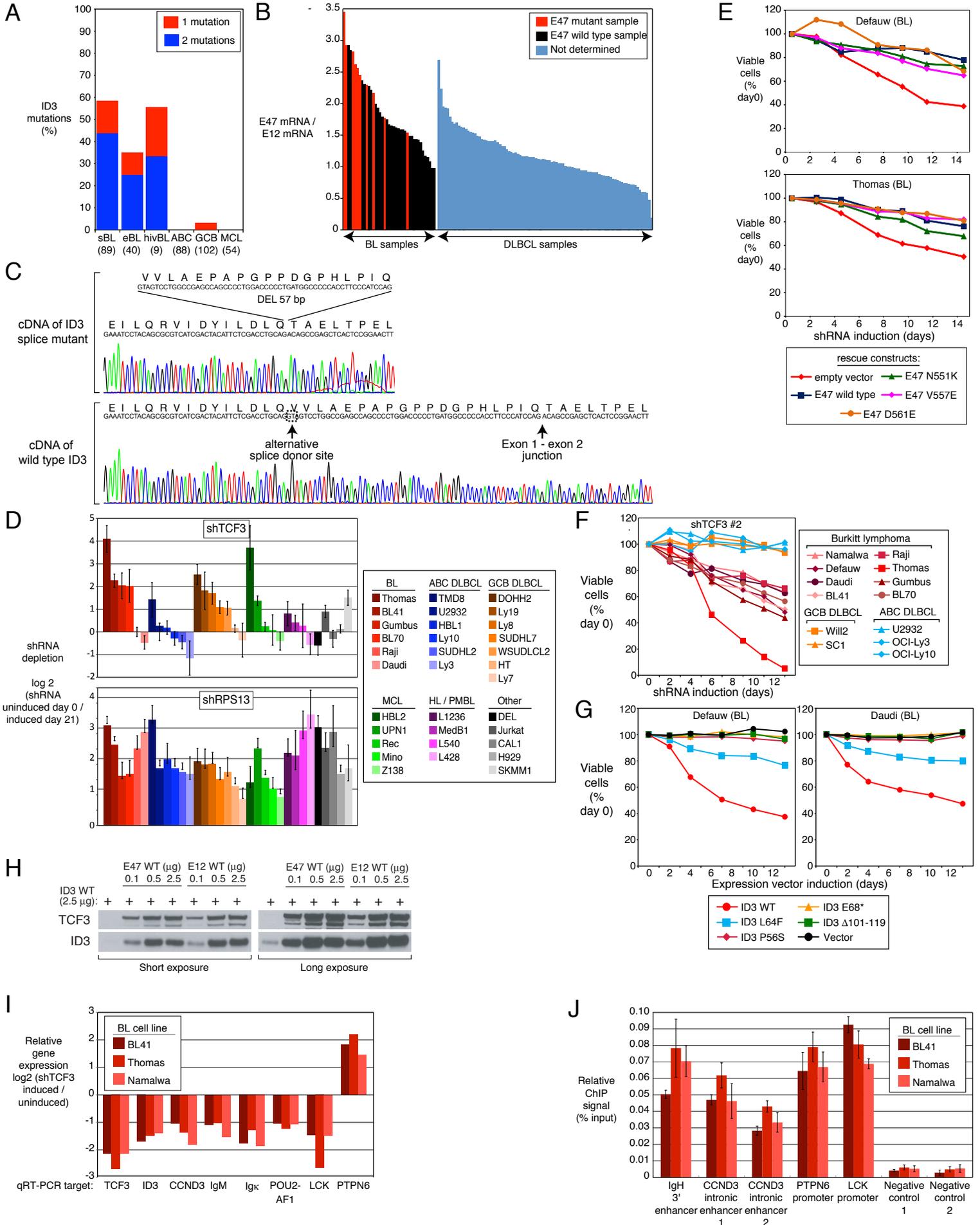
D



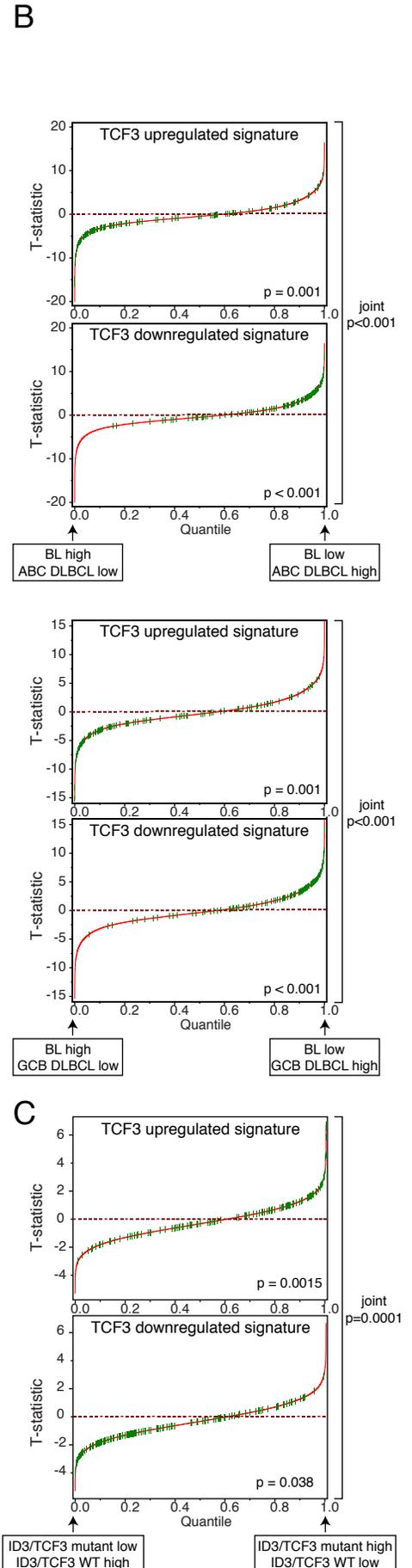
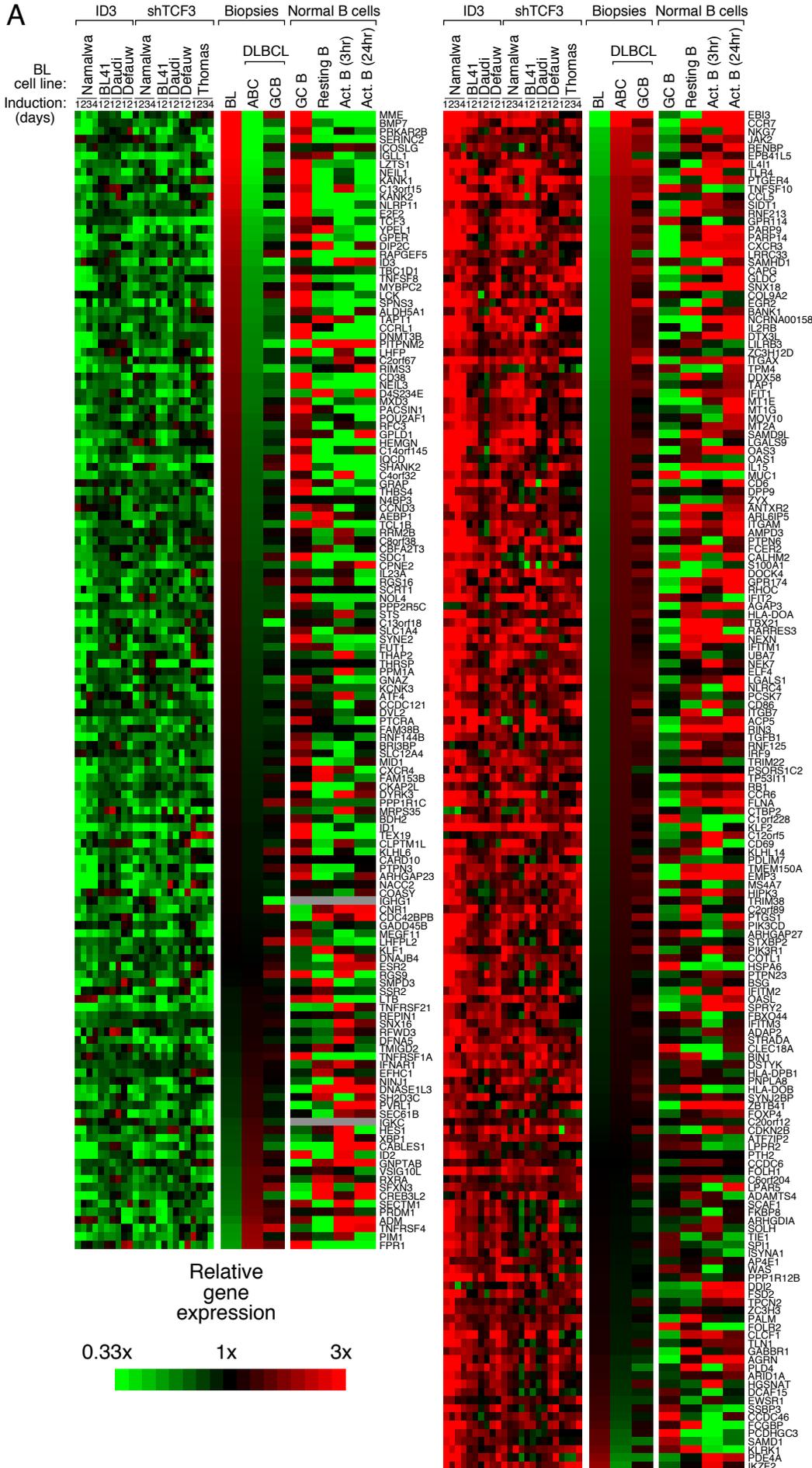
E



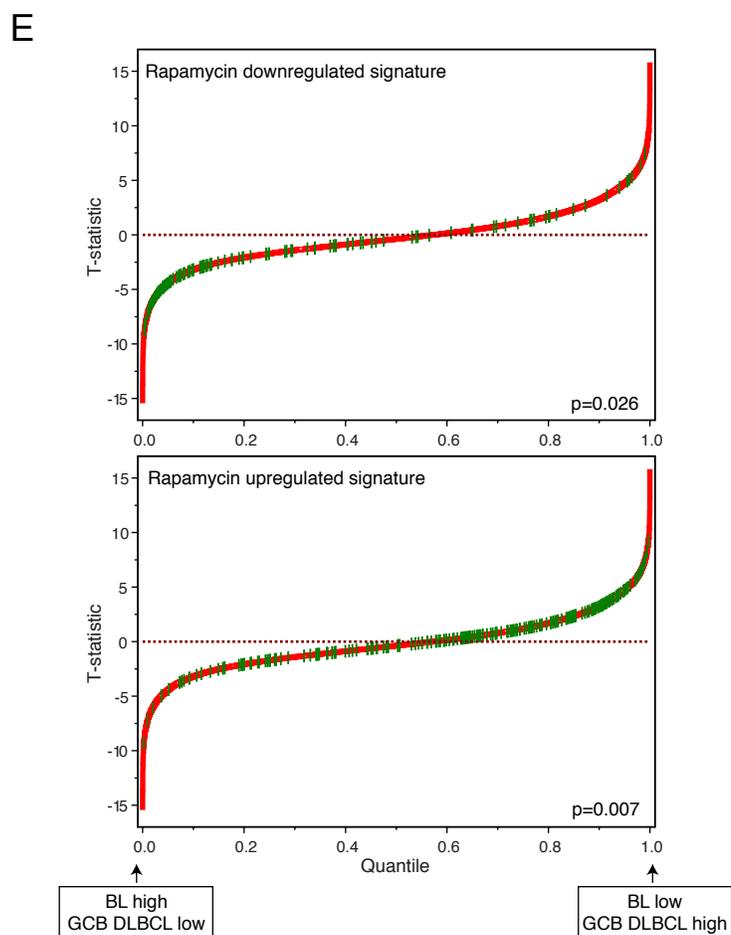
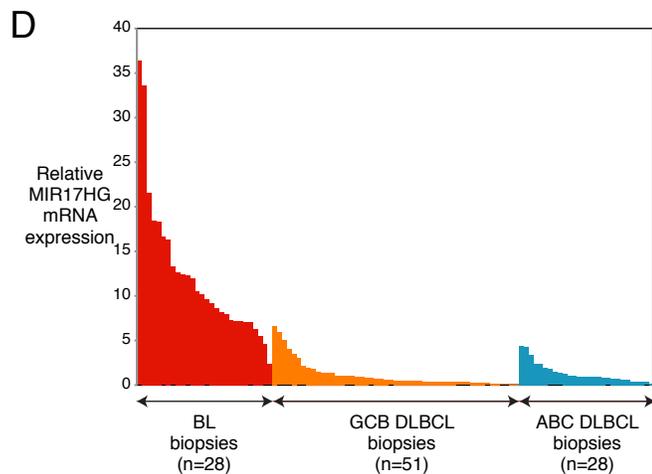
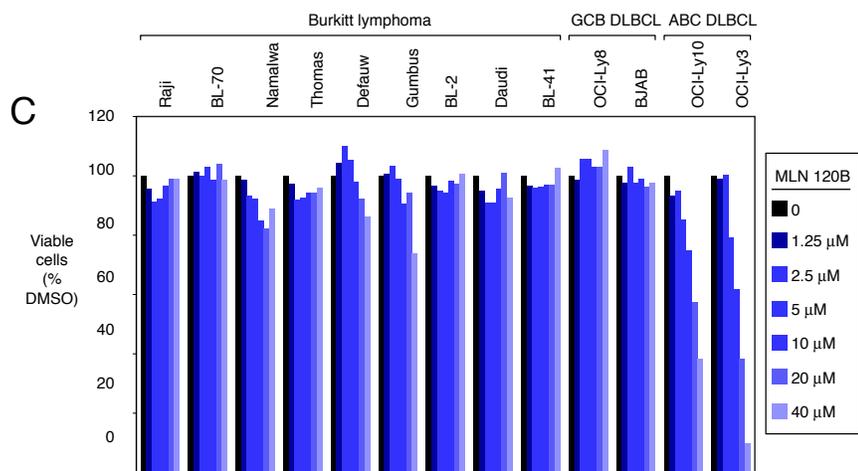
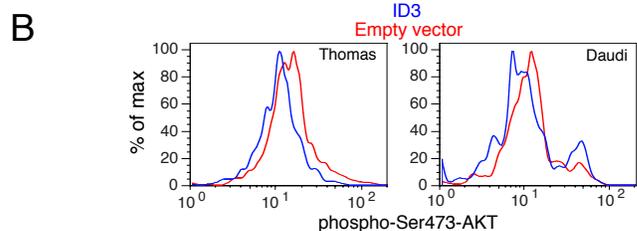
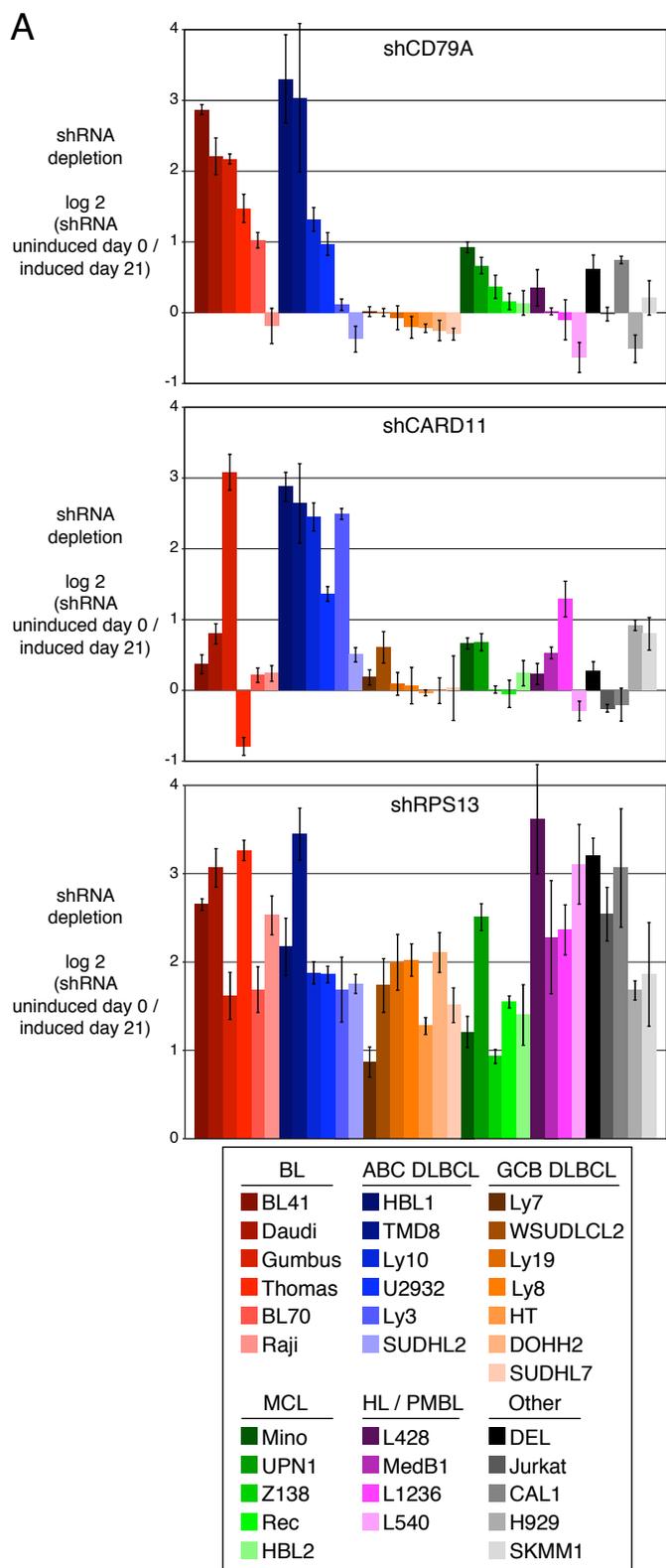
Supplemental Figure 3



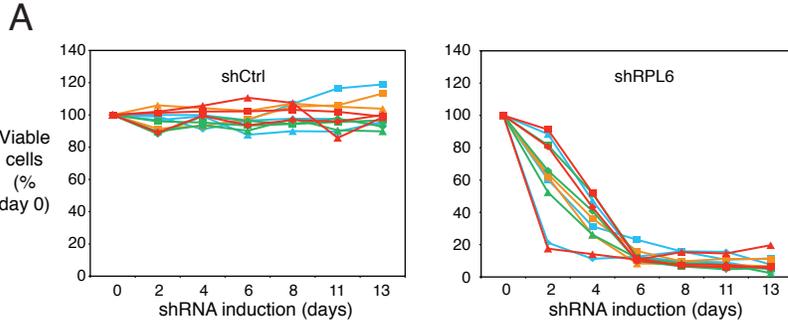
Supplemental Figure 4



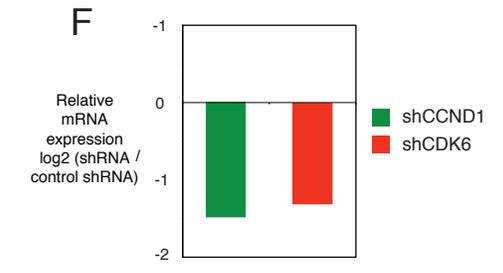
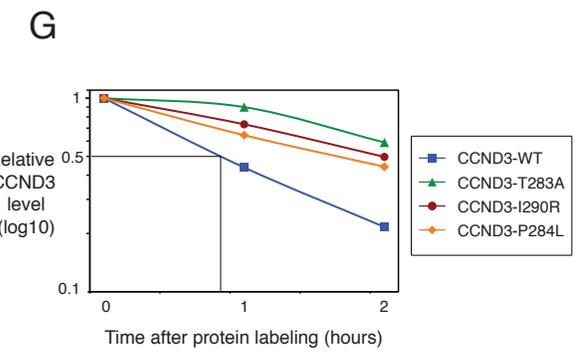
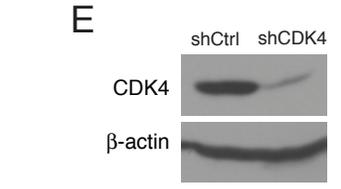
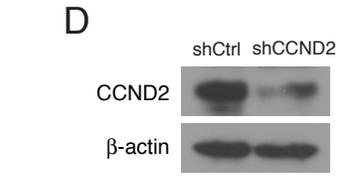
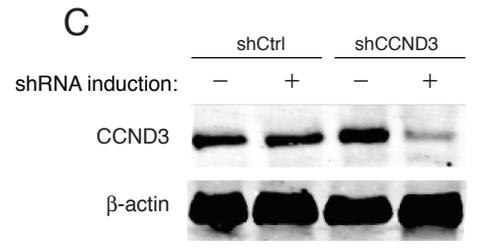
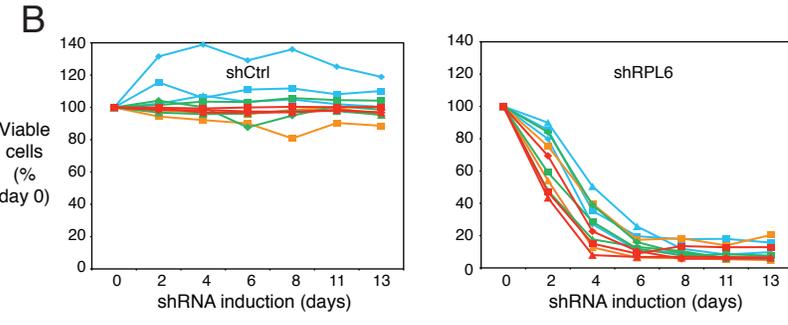
Supplemental Figure 5



Supplemental Figure 6



- BL41 } BL
- Thomas } BL
- Gumbus } BL
- Jeko } MCL
- UPN1 } MCL
- Z138 } MCL
- BJAB } DLBCL
- OCI-Ly8 } DLBCL
- OCI-Ly3 } ABC
- OCI-Ly10 } DLBCL
- U2932 } DLBCL



Supplemental Figure 7

