Supplemental Materials

Supplemental Methods

Mice

To generate mutant *Tfap4* alleles lacking a 1-kb enhancer region bound by c-MYC or each of the MYC-binding sites, *in vitro* transcribed RNA containing a 21-bp complementary sequence corresponding to and PAM following short guide RNA sequence complementary to MYCBS1 or MYCBS2 was generated using MEGAshortscript T7 Transcription Kit (Thermo Fisher) from DNA templates which were amplified using oligonucleotide primers, TTA ATA CGA CTC ACT ATA GGG GAC TGC TGC AGC ACC ACG TGT TTT AGA GCT AGA AAT AGC AAG or TTA ATA CGA CTC ACT ATA GGG GCT GGG GTG TCG GAG CAC GGT TTT AGA GCT AGA AAT AGC AAG, and AAA AGC ACC GAC TCG GTG CC, the pX-330 plasmid (Addgene) and Phusion PCR polymerase (Thermo Fisher). A mixture of sgRNAs and Cas9 mRNA was microinjected into C57BL/6N zygotes and founders were bred to C57BL/6NCr mice at the Transgenic, Knockout and Micro-Injection Core affiliated by the department of Pathology and Immunology and Rheumatic Diseases Research Resource-based Center of Washington University.

Retroviral infection

The retroviral plasmids encoding mouse MYC and mouse ERG were generated by cloning PCR-amplified fragment into pMIGR or pMSCV-Thy1.1. A retroviral plasmid encoding human BCR-ABL P190 was purchased from Addgene (Plasmid #27483). Viral supernatants were prepared by transiently transfecting PlatE cells[1] with a plasmid encoding either mouse MYC (T58A), mouse ERG, or human BCR-ABL P190, as indicated in the figures, together with a helper plasmid pCL-10A1 using TransIT-293 transfection reagent (Mirus Bio). Viral supernatant was harvested 48 hrs

after transfection and used to infect freshly isolated CD19^{+} BM cells at 1000 x *g*, 30°C for 1.5-2 hrs in the presence of 10 μg/ml of polybrene (Sigma).

Flow cytometry

Single cell suspension was prepared by mechanical dissociation of bones, spleen and lymph nodes in phosphate buffered saline supplemented with 2% fetal calf serum (Thermo Fisher) and 2 mM EDTA, stained with monoclonal antibodies listed below and analyzed using LSR Fortessa, LSR-II or X-20 flow cytometer (BD Biosciences), or sorted using a FACS Aria II (BD Biosciences). After cell sorting, the purity of sorted cells was confirmed to be 95% or higher.

Antibodies

The following monoclonal antibodies were purchased from Biolegend unless specified otherwise: Fluorescein isothiocyanate-conjugated anti-CD8α (53-6.7), anti-CD24 (M1/69, Biolegend), anti-IgM (II/41), anti-Ly-51 (BP-1); Alexa Fluor (AF) 488-conjugated anti-CD45.1 (A20); phycoerythrin (PE)-conjugated anti-human CD2 (RPA-2.10), anti-CD25 (PC61), anti-CD117 (2B8), anti-IgM (RMM-1); peridinin chlorophyll protein–cyanin 5.5-conjugated anti-CD4 (GK1.5), anti-CD11b (M1/70), anti-CD19 (6D5), anti-CD43 (S7, Thermo Fisher), anti-CD45R/B220 (RA3-6B2), anti-CD90.1 (OX-7); PE-indotricarbocyanine-conjugated anti-CD8α (53-6.7), anti-CD11b (M1/70), anti-CD45.1 (A20), anti-CD45R/B220 (RA3-6B2), anti-CD95 (Jo2, Thermo Fisher), anti-Ly-51 (BP-1); allophycocyanin (APC)-conjugated anti-CD93 (AA4.1), anti-CD117 (2B8), anti-Ly6A/E (E13-161.7), anti-TCRβ (H57-597, Thermo Fisher); AF647-conjugated anti-CD11b (M1/70), anti-GL-7 (GL-7); APC-indotricarbocyanine conjugated anti-CD45.2 (104), anti-CD45R/B220 (RA3-6B2), anti-CD117 (2B8), anti-Igκ (187.1, BD

Pharmingen); AF700-conjugated anti-CD11b (M1/70); Pacific Blue-conjugated anti-CD11b (M1/70), anti-CD44 (IM7), anti-CD45R/B220 (RA3-6B2), anti-IgD (11-26c.2a), anti-TER119 (TER-119); Brilliant Violet (BV) 510-conjugated anti-CD4 (RM4-5), anti-CD45R/B220 (RA3-6B2); BV605-conjugated anti-CD11b (M1/70), anti-CD45.1 (A20), anti-IgD (11-26c.2a); BV650-conjugated streptavidin; biotin-conjugated anti-IgM (AF6-78).

RNA-seq

Total RNA was extracted from sorted cells by Trizol and treated with recombinant DNase using Nucleospin RNA XS extraction kit (Macherey-Nagel). The integrity of the extracted RNA was determined using an Agilent Bioanalyzer or Tapestation. mRNA was enriched by depleting ribosomal RNA using a Ribo-ZERO kit (Illumina), and was then fragmented in buffer containing 40 mM Tris-Acetate (pH 8.2), 100 mM Potassium Acetate and 30 mM Magnesium Acetate at 94 ˚C for 150 seconds. mRNA was reverse transcribed to yield cDNA using SuperScript III RT enzyme (Thermo Fisher) and random hexamers followed by synthesis of the second strand to yield ds-cDNA. cDNA was blunt ended, had an A base added to the 3' ends, and then had Illumina sequencing adapters ligated to the ends. Ligated fragments were then amplified for 15 cycles using primers incorporating the p5 and p7 sequences and unique index tags. Fragments were sequenced on an Illumina HiSeq2500 or HiSeq3000 using single reads extending 50 bases. Sequencing reads were mapped to the mouse genome mm10 using STAR with default options. Gene counts were derived from the number of uniquely aligned reads by Subread:featureCount version 1.4.5. Sequencing performance was assessed with RSeQC version 2.3. Genes with RPKM>2 in at least one sample were initially filtered, resulting in a list of 11 637 genes further analyzed using the Phantasus web application (https://fgenome.ifmo.ru/phantasus/) for PCA. Reads were $\log_2$

transformed and differential expression analysis was conducted using a limma tool Release 3.9 (https://bioconductor.org/packages/release/bioc/html/limma.html). This output was further analyzed for gene set enrichment using the Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp) using Phantasus. Additionally, for transformed pre-B cells unsupervised clustering of expression data was performed, with indicated clusters being used for additional GSEA analysis. For Figure 4B and Supplemental Figure 4C and F, only annotated genes are shown. Human data was analyzed using limma tool Release 3.9 and GSEA was performed using the human Hallmark Molecular Signatures Database in Phantasus following processing described in the main methods.

Genomic PCR and variant validation

Genomic DNA was purified from tumor cells and a tail-tip using QIAmp Fast DNA Tissue Kit (Qiagen). Genomic sequences of wild-type and targeted Tfap4 alleles as well as Runx1, as a control positioned on the same chromosome, were amplified using Taq polymerase (MidSci) and primers listed below. To confirm mutations identified by whole exome sequencing, cDNA from sorted pro/pre-B tumor cells was used to PCR amplify Kras using Phusion high-fidelity PCR master-mix (Thermo Fisher) and the fragments were sequenced following TA-cloning into pCR2.1-TOPO (Thermo Fisher). Five to seven clones from each sample were sequenced using the T7 primer. Obtained sequencing data were manually inspected for Kras point mutations. The following primers were used:

Tfap4 (WT), CAG ACC CAG GGT ATG TGC TC and GCA GCT CAC TCT GTG TAA AGA ACT G;

Tfap4 (targeted), CAG ACC CAG GGT ATG TGC TC and GTA ACA AGA TCT GAT GCC CTC TTC;

Runx1, GCG TTC CAA GTC AGT TGT AAG CC and CTG CAT TTG TCC CTT GGT TGA CG;

Kras, CGC GGA GAG AGG CCT GCT GAA A and GTG TGC CTT AAG AAA GAG TAC A.

Real-time quantitative RT-PCR

RNA was extracted with Trizol (Thermo Fisher) and was reverse-transcribed with qScript Supermix (Quanta Bio). Luminaris SYBR green qPCR mix (Thermo Fisher) and a LightCycler 480 (Roche) were used for real-time quantitative RT-PCR. Quantities of transcripts were normalized to that of *Hprt1* RNA unless specified otherwise. For quantification of gene expression normalized to that of 'spiked-in' RNA, 5 µl of ERCC (External RNA Controls Consortium) RNA Spike-In Control Mixes (Thermo Fisher) at a dilution of 1:1000 was added to cell lysates in Trizol per 0.5×10^6 cells prior to RNA extraction. The following oligonucleotide primers (Sigma) were used:

Tfap4, GGA GAA GCT AGA GCG GGA AC and TTT TGC CGG GAT GTA GAG AC;

Myc, AGT GCT GCA TGA GGA GAC AC and GGT TTG CCT CTT CTC CAC AG;

Erg, GGA GTG CAA CCC TAG TCA GG and GTA GCT GCC GTA GCT CAT CC;

Rag1, CAA CCA AGC TGC AGA CAT TC and AAT TTC ATC GGG TGC AGA AC;

Vpreb1, ATG CTG CTG GCC TAT CTC AC and GAT GCT AAT GGT GGC TGA TG;

Hprt1, AGG TTG CAA GCT TGC TGG T and TGA AGT ACT CAT TAT AGT CAA GGG CA;

ERCC-00108, GCT ATC AGC TTG CGC CTA TTA T and GTT GAG TCC ACG GGA TAG AGT C.

Immunoblot analysis

Whole cell lysates were prepared in Laemmli buffer containing 1% SDS and 2% 2-mercaptoethanol. Lysates from 5 x 10^5 $B220^+$ IgM^- cells or 1 x 10^5 activated T cells were separated on 10% SDS-PAGE and transferred to nitrocellulose membranes (GE Healthcare). The blots were incubated with primary antibodies (identified below), followed by detection with horseradish peroxidase–conjugated antibody to rabbit immunoglobulin light chain (211-032-171; Jackson ImmunoResearch) and a Luminata HRP substrate (Millipore). Anti-AP4 was described previously [2] and anti-c-MYC was purchased from Santa Cruz Biotechnology (sc-764). Anti-Histone H3 (ab1791; Abcam) or anti-HDAC1 (ab7028; Abcam) were used as loading controls. For inhibition of translation in an assay for AP4 protein stability assay, 10 μM cycloheximide (Sigma) was added to the activated T cell culture as described.[2]

***In vitro* stimulation of T cells**

Naive $CD8^+$ T cells were purified using Dynabeads CD8 positive isolation kit (Thermo Fisher) and cultured in RPMI medium supplemented with 10% FBS (Thermo Fisher) in the presence of soluble anti-CD3 (145-2C11; Biolegend) and anti-CD28 (37.51; Bio X Cell) at concentrations of 0.1 μg/ml and 1 μg/ml, respectively, in multiwell tissue culture plates coated with goat anti-hamster IgG (55397; MP Biomedicals). For retroviral transduction of activated T cells, T cells were activated overnight and then 'spin-inoculated' as described.[2] Following retroviral infection T cells were cultured in original media with anti-CD3 and anti-CD28 stimulation for additional 24 hrs. Subsequently, infected T cells were cultured in the presence of 100 U/mL of recombinant IL-2 (Thermo Fisher) for 24 hrs and were transferred to culture with 1 U/ml of IL-2 for analysis of CD25 expression 12-16 hrs later.

Chromatin Immunoprecipitation and sequencing (ChIP-seq)

ChIP-seq was performed essentially as described previously [2,3]. Briefly, $CD19^+$ cells from 3-4-week-old Eμ-*Myc* *Tfap4*$^{+/+}$ and Eμ-*Myc* *Tfap4*$^{+/-}$ mice were purified from BM using anti-CD19 microbeads and LS columns (Miltenyi), and fixed with 1% paraformaldehyde for 10 min at room temperature. Sonicated chromatin was immunoprecipitated using 5 µg of anti-AP4, 1 µg of anti-H3K27ac (ab4729, Abcam) and Dynabeads protein-G magnetic beads (Thermo Fisher). After reverse crosslinking, precipitated DNA was purified using a GenElute PCR Clean Up Kit (Sigma) and quantitated using a Qubit DNA quantitation kit (Thermo Fisher). DNA size range was assayed on Agilent Bioanalyzer High Sensitivity DNA chips. The purified DNA was blunt ended, had addition of "A" base to 3' end, and had sequencing adapters ligated to the ends. The fragments were size-selected to 200-600 base pairs, and underwent amplification for 15 cycles with primers incorporating p5 and p7 sequences and a unique index tag for multiplexing. The resulting libraries were sequenced using the Illumina HiSeq3000 as single reads extending 50 bases. Sequenced reads were mapped to the mouse genome mm9 using bowtie-2 with default parameters and the Homer software package[4] was used for peak calling, motif analysis and visualization on the UCSC genome browser. Homer was used for peak 'calling' with a –style factor option and default False Discovery Rate of 0.001. Previously published ChIP-seq and ATAC-seq data were obtained from the NCBI-GEO database with the following accession IDs. GEO: Eμ-*Myc* anti-MYC ChIP-seq: GSE51011; ATAC-seq GSE100738; activated T cell ChIP-seq GSE58081; activated B cell ChIP-seq GSE80669, T cell ATAC-seq GSE87646.

Whole Exome Sequencing and somatic variant identification

Library preparation was performed with 1 µg genomic DNA from sorted tumor cells and a tail-tip of each corresponding tumor bearing mouse. The integrity of genomic DNA was determined using Agilent Tape station. Genomic DNA was sonicated to an average size of 175bp. The fragments were blunt ended, had addition of "A" base to 3' end, and had Illumina's sequencing adapters ligated to the ends. The ligated fragments underwent amplification for 8 cycles. Fragments were hybridized to biotinylated RNA oligos (Agilent Sure Select Mouse Exome) specific to regions of interest, and selected from remaining fragments using streptavidin beads. Enriched fragments were amplified for 11 cycles with primers that incorporate a unique indexing sequence tag. The resulting libraries were sequenced using the Illumina HiSeq-3000 as paired end reads extending 150 bases from both ends of the fragments. to achieve at least 25-fold coverage in >90% of exome for tumor samples and at least 10-fold coverage in >95% of exomes for tail samples.

Whole exome sequencing data were aligned to the mouse reference sequence mm9 using BWA [5] version 0.7.10. BWA-MEM algorithm was used with the parameter, -t 8. BAM files were de-duplicated by using picard version 1.113 (https://broadinstitute.github.io/picard/). Somatic variants in tumor samples were called against matched normal tail tip samples. In detail, single nucleotide variants (SNVs) were detected using the union of three callers: 1) Samtools [6] version r982 (params: mpileup -BuDS) intersected with Somatic Sniper [7] version 1.0.4 (params: -F vcf -G -L -q 1 -Q 15) and processed through false-positive filter v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15 --min-mapping-quality 40 --min-somatic-score 40) 2) VarScan [8] version 2.3.6 (params: --nobaq --version r982:) filtered by varscan-high-confidence filter v1 and processed through false-positive filter v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15), and 3) Strelka [9] version 1.0.11 (params: isSkipDepthFilters

= 1). INDELs were detected using the union of 4 callers: 1) GATK somatic-indel [10] version 5336, 2) Pindel [11] version 0.5 filtered with pindel somatic-calls v1 and vaf-filter v1 (params: --variant-freq-cutoff=0.08), 3) VarScan version 2.3.6 (params: --nobaq --version r982:) filtered by varscan-high-confidence-indel v1, and 4) Strelka version 1.0.11 (params: isSkipDepthFilters = 1). Furthermore, low confidence SNVs were filtered out given known tumor and normal contamination rates, by using "gmt validation identify-outliers" tools (params: --llr-cutoff 3 –tumor-purity 0.95) (Genome Modeling System [12]: (https://github.com/genome/gms). Finally, at least 20x coverage was required at each variant site, then variant calls by read alignment in paralogous sequence regions or unstable contigs were filtered out using the mapping quality score.

EdU labeling

Three week-old mice received 1.5 mg EdU (5-ethynyl-2´-deoxyuridine) by intraperitoneal injection 1 hr before euthanasia. EdU incorporation was determined by a Click-iT Plus EdU Alexa Fluor 647 Flow Cytometry Assay Kit according to the manufacturer's instructions (Thermo Fisher).

Annexin V staining

Developing B cells from the bone marrow of 3-week-old mice were stained for Annexin V-APC (Biolegend) and Live/Dead Fixable Aqua (Thermo Fisher) according to the manufacturer's instructions.

***Ex vivo* B cell survival assay**

CD19+ pro/pre-B cells of 3–4-week-old Eμ-*Myc* mice were cultured 10% FBS-containing medium without or with 10 ng/μl of recombinant IL-7 (Peprotech). Cell viability was determined by Live/Dead Fixable Aqua staining at 24, 48 and 72 hrs post isolation from the bone marrow.

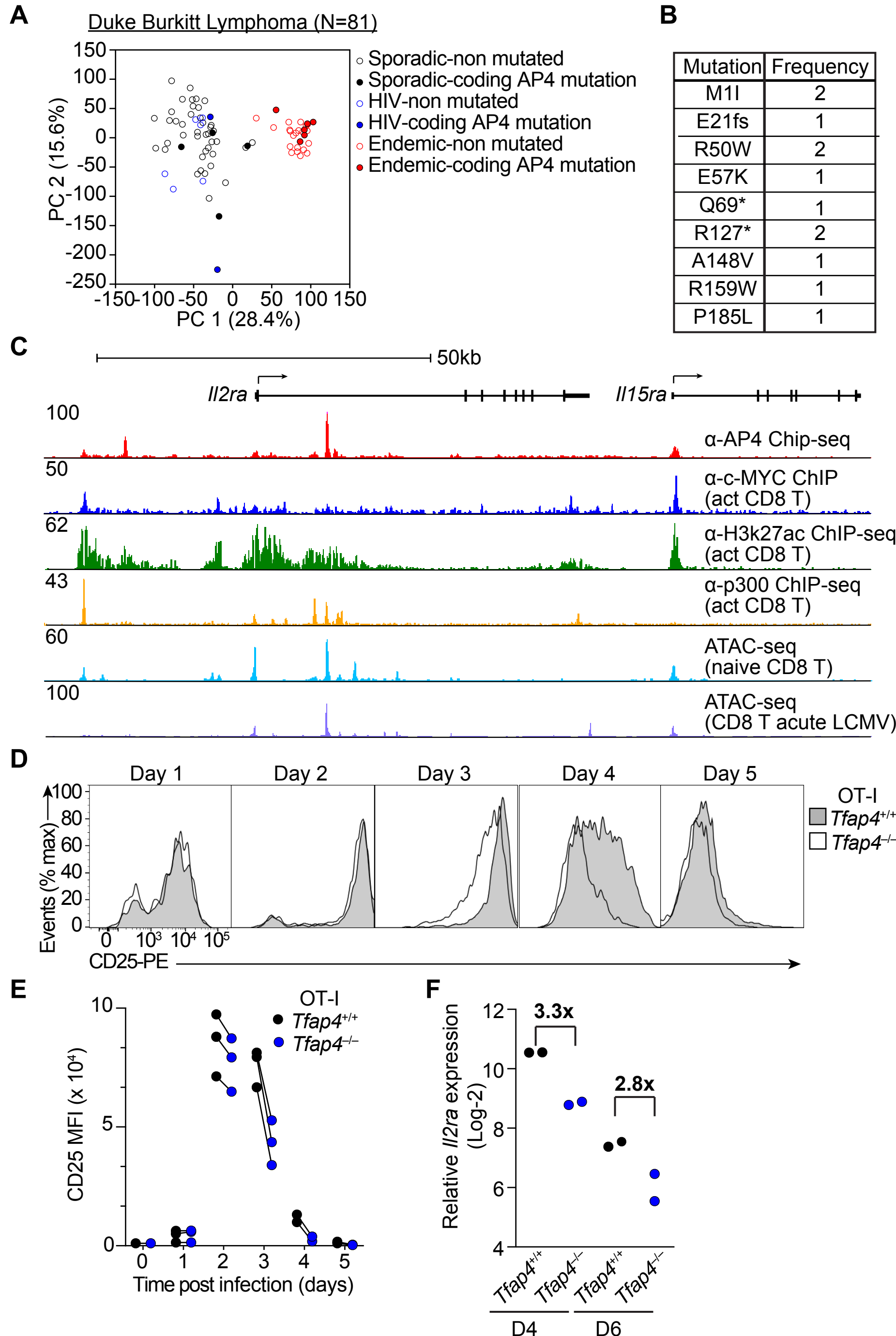
References

1. Morita S, Kojima T, Kitamura T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene therapy*. 2000;7(12):1063-1066.
2. Chou C, Pinto AK, Curtis JD, et al. c-Myc-induced transcription factor AP4 is required for host protection mediated by CD8+ T cells. *Nat Immunol*. 2014;15(9):884-893.
3. Chou C, Verbaro DJ, Tonc E, et al. The Transcription Factor AP4 Mediates Resolution of Chronic Viral Infection through Amplification of Germinal Center B Cell Responses. *Immunity*. 2016;45(3):570-582.
4. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-589.
5. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
6. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
7. Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*. 2012;28(3):311-317.
8. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568-576.
9. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-1817.

10. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.
11. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865-2871.
12. Griffith M, Griffith OL, Smith SM, et al. Genome Modeling System: A Knowledge Management Platform for Genomics. *PLoS Comput Biol*. 2015;11(7):e1004274.

Supplemental figures

Figure S1.

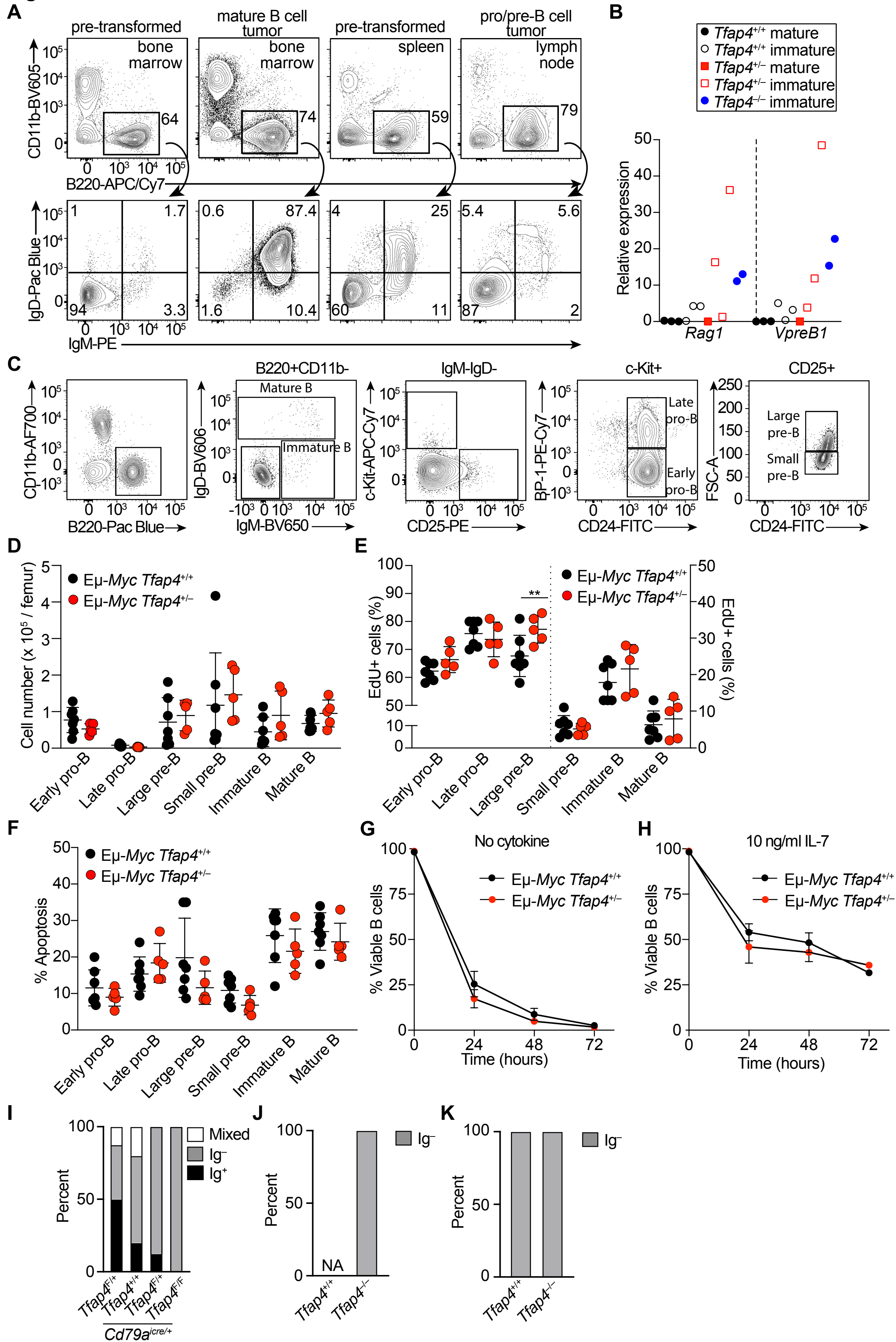


Supplemental Figure 1. The transcription factor TFAP4 is mutated in lymphoid cancers and directly upregulates the expression of *Il2ra,* encoding the CD25 subunit of interleukin-2 receptor. Related to Figure 1.

(**A**) PCA of RNA-seq data from patient samples across sporadic, HIV positive, and endemic subtypes of Burkitt lymphoma. (**B**) List of coding TFAP4 mutations in Burkitt lymphoma samples from (A). (**C**) ChIP-sequencing data showing binding of AP4, MYC, and p300, deposition of

H3K27ac, and accessible chromatin states in a genomic region containing the *Il2ra* locus in activated CD8 T cells. AP4 binding colocalizes with chromatin accessibility within the *Il2ra* locus and p300 binding, which is predictive of enhancer activity. Y-axes show tag counts in 10^7 mapped reads. (**D**) Expression of CD25 in adoptively transferred *Tfap4*$^{-/-}$ OT-I cells compared to co-transferred *Tfap4*$^{+/+}$ OT-I cells in congenic recipient mice at different time points after *Listeria monocytogenes* expressing ovalbumin (Lm-Ova) infection. (**E**) Mean fluorescence intensity (MFI) of CD25 on *Tfap4*$^{-/-}$ OT-I cells and *Tfap4*$^{+/+}$ OT-I T cells after Lm-Ova infection. Data are representative of three independent experiments. (**F**) Microarray data showing *Il2ra* mRNA expression in *Tfap4*$^{-/-}$ OT-I cells compared to *Tfap4*$^{+/+}$ OT-I cells on days four and six post Lm-Ova infection as in (E).

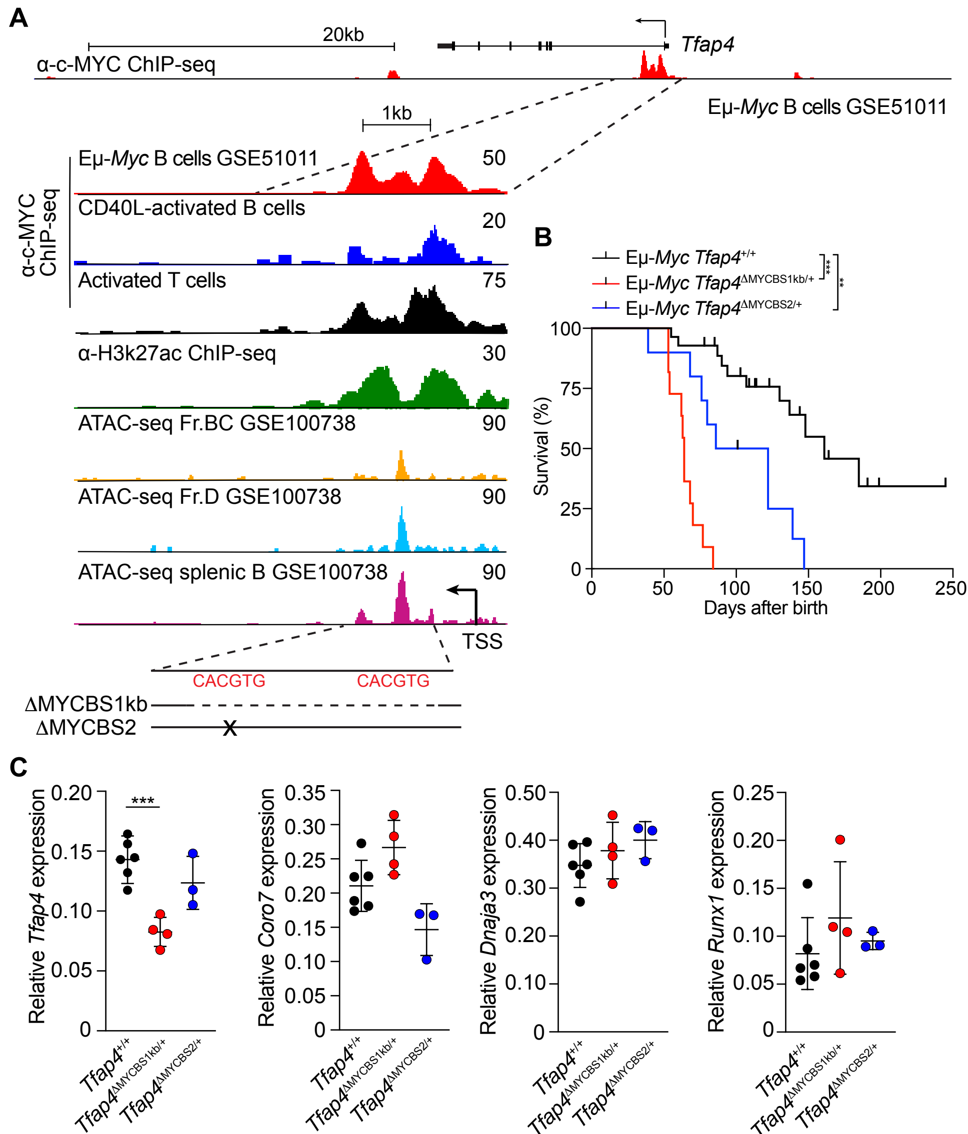
Figure S2.



Supplemental Figure 2. AP4 is a cell-intrinsic suppressor in c-MYC-induced tumorigenesis of developing B cells. Related to Figure 2.

(**A**) A representative gating strategy for pre-transformed pro/pre-B cells in the BM and spleen of Eμ-*Myc* mice, or mature B and pro/pre-B cell tumor cells, based on surface expression of B220, IgM, and IgD. Transformed mature B cells can be found in the BM (second column), and, similarly, transformed pro/pre-B cells can be found at a high frequency with an almost complete lack of surface Ig^{+} cells in peripheral lymphoid organs (last column). (**B**) Representative expression of *Rag1* and *VpreB1* mRNA, measured by qPCR, in mature B and pro/pre-B (immature) tumor cells. (**C**) A representative gating strategy for subpopulations of developing B cells in the bone marrow defined as follows: early pro-B ($B220^{+}$ $IgM^{–}$ $IgD^{–}$ c-Kit^{+} $CD25^{–}$ $CD24^{+}$ BP-$1^{–}$), late pro-B ($B220^{+}$ $IgM^{–}$ $IgD^{–}$ c-Kit^{+} $CD25^{–}$ $CD24^{+}$ BP-1^{+}), large pre-B ($B220^{+}$ $IgM^{–}$ $IgD^{–}$ c-$Kit^{–}$ $CD25^{+}$ $CD24^{+}$ large), small pre-B ($B220^{+}$ $IgM^{–}$ $IgD^{–}$ c-$Kit^{–}$ $CD25^{+}$ $CD24^{+}$ small), immature B ($B220^{+}$ IgM^{+} $IgD^{–}$), and mature B ($B220^{+}$ IgM^{+} IgD^{+}). (**D**) Number of developing B cells in populations defined in (C) from 3-week old Eμ-*Myc* $Tfap4^{+/+}$ (n=7) and Eμ-*Myc* $Tfap4^{+/–}$ (n=5) mice. EdU (**E**) and Annexin V (**F**) staining of developing B cells from bone marrow of 3-week old mice in (D). (**G-H**) An *in vitro* survival assay of BM $CD19^{+}$ pro/pre-B cells of 3-4 week old Eμ-*Myc* mice with indicated *Tfap4* genotypes in an FBS-containing medium without (G) or with 10ng/μl of recombinant IL-7 (H). Data shown are combined from two independent experiments. (**I-K**) Frequencies of surface $Ig^{–}$ (immature) or Ig^{+} (mature) B cells from tumor models in Figure 2 C-E, respectively. (I) Eμ-*Myc* $Tfap4^{F/+}$, n=8; Eμ-*Myc* $Tfap4^{+/+}$ $Cd79a^{icre}$, n=5; Eμ-*Myc* $Tfap4^{F/+}$ $Cd79a^{icre}$ n=8; and Eμ-*Myc* $Tfap4^{F/+}$ $Cd79a^{icre}$ mice, n=1. (J) $Tfap4^{–/–}$ + RV-MYC, n=8; no $Tfap4^{+/+}$ donor derived tumors developed in this model. (K) $Tfap4^{+/+}$ + RV-P190 *bcr-abl*, n=9; $Tfap4^{–/–}$ + RV-P190 *bcr-abl*, n=9. $^{**}P<0.01$, $^{****}P<0.0001$ by one-way ANOVA followed by Tukey's multiple comparisons test (D-E).

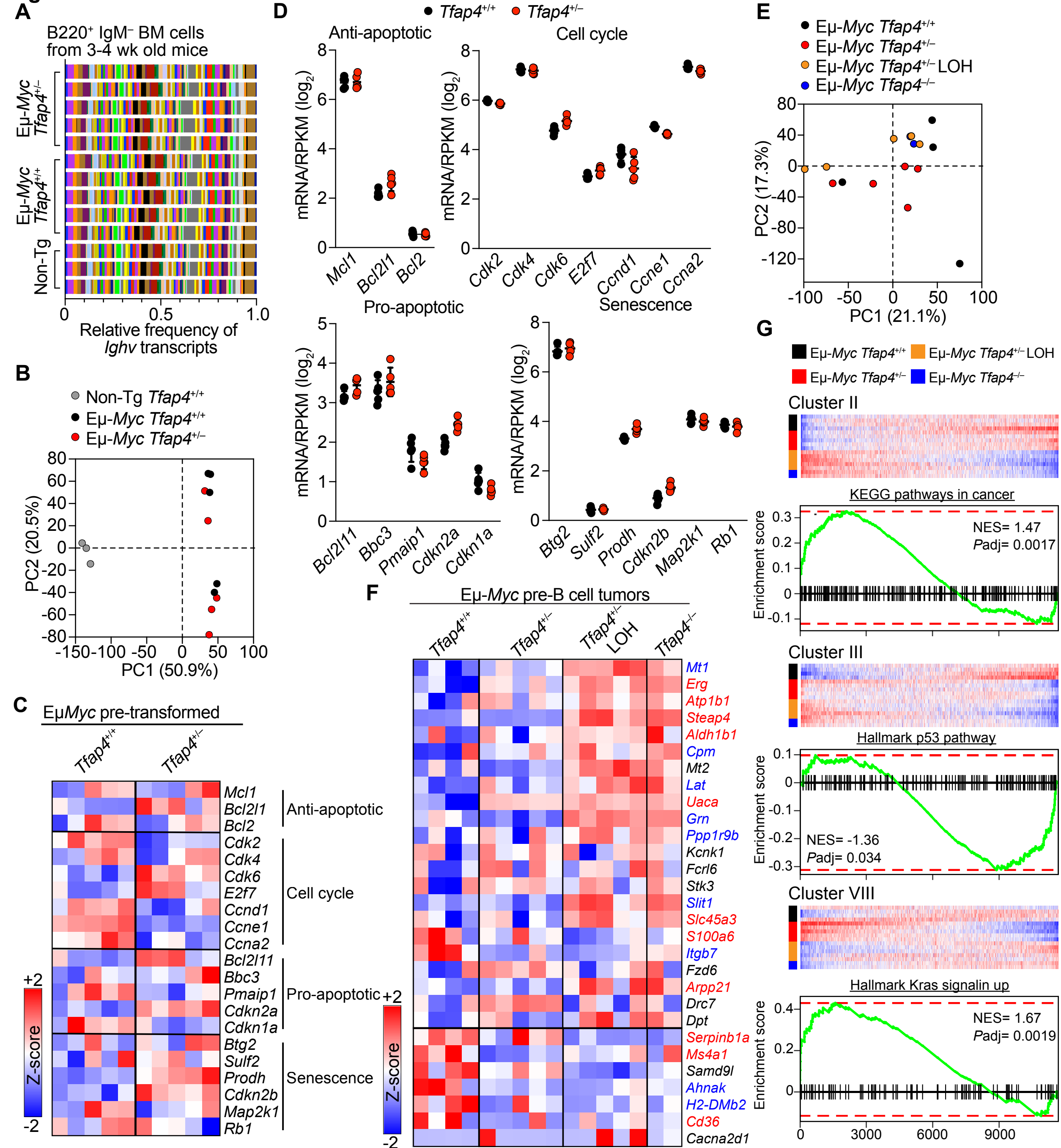
Figure S3.



Supplemental Figure 3. Direct binding of c-MYC in the *Tfap4* locus is essential for tumor suppression in Eμ-*Myc* mice. Related to Figure 2.

(**A**) Binding of c-MYC to the *Tfap4* locus in Eμ-*Myc* B cells, CD40L-activated B cells, and activated T cells. Histograms of sequenced tags pulled down with anti-c-MYC or anti-H3K27ac antibody are shown, as well as ATAC-seq histograms for Hardy's fractions B, C, D, and mature splenic B cells from the Immgen database. Y-axes show tag counts in 10^7 mapped reads. (**B**) Survival of mice with one allele deletion of the 1kb region containing the two c-MYC consensus binding sites in the *Tfap4* locus (ΔMYCBS1kb), and those harboring one allele with a 32-nucleotide deletion encompassing the 3' MYC binding site (ΔMYCBS2). Eμ-*Myc Tfap4*$^{+/+}$, n=28, median survival: 161 days; Eμ-*Myc Tfap4*$^{\Delta MYCBS1kb/+}$, n=11, median survival: 64 days; Eμ-*Myc Tfap4*$^{\Delta MYCBS2/+}$, n=10, median survival: 122 days. (**C**) Expression of *Tfap4*, *Coro7*, *Dnaja3*, and *Runx1* in B220^{+} IgM^{-} pro/pre-B cells from 3-4 week old Eμ-*Myc* mice with indicated genotypes, prior to tumor development. ***P*<0.01, ****P*<0.001 by log-rank test, adjusted for multiple comparisons.

Figure S4.

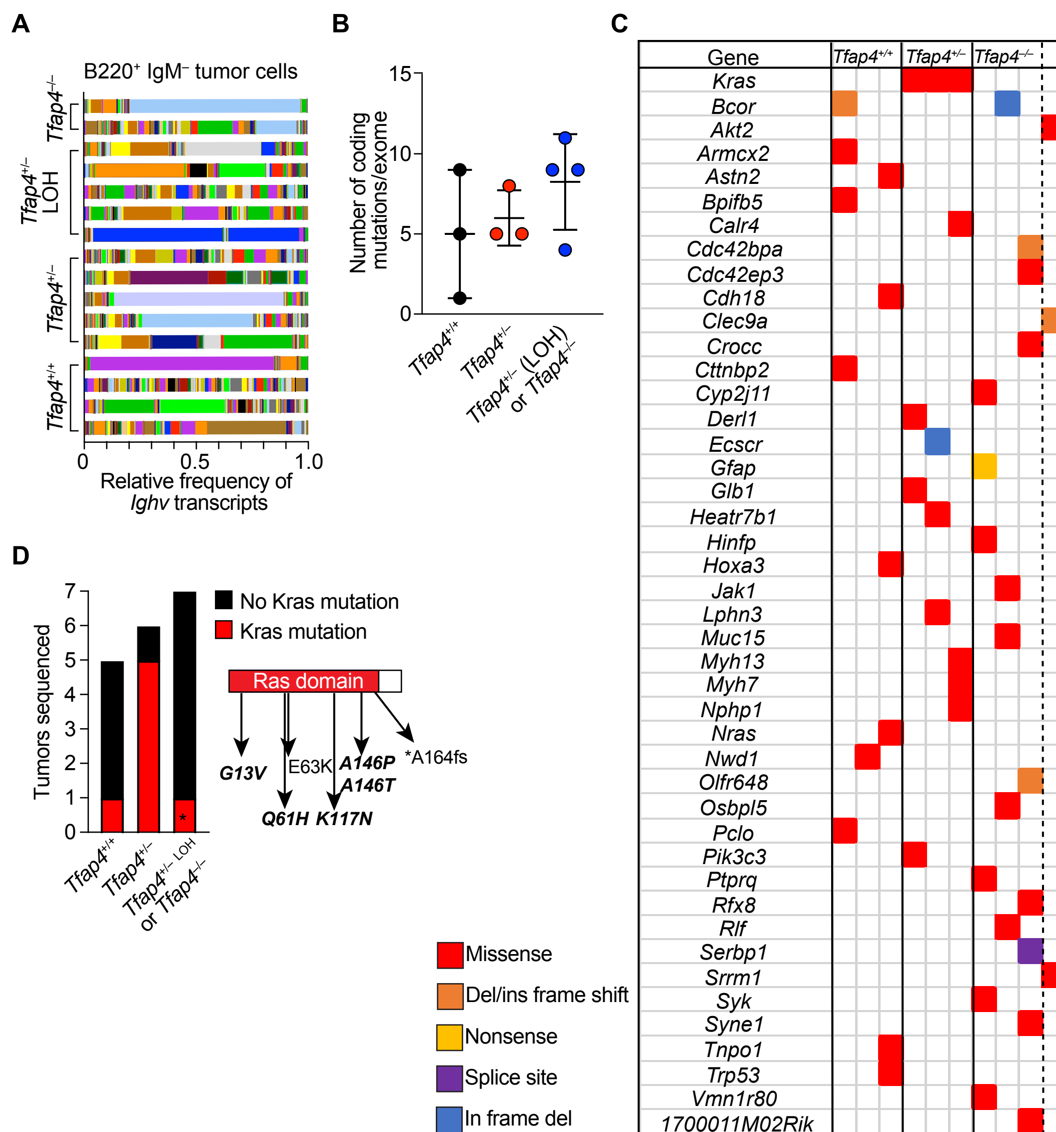


Supplemental Figure 4. Gene expression changes in Eμ-*Myc Tfap4*$^{+/+}$, Eμ-*Myc Tfap4*$^{+/–}$ and Eμ-*Myc Tfap4*$^{–/–}$ pre-transformed and transformed pro/pre-B cells. Related to Figure 4.

(**A**) Usage of variable gene segments of *Igh* in B220^{+} IgM$^{–}$ BM cells sorted from 3-4-week-old Eμ-*Myc*$^{–}$ *Tfap4*$^{+/+}$ (non-Tg) (n=3), Eμ-*Myc Tfap4*$^{+/+}$ (n=5) and Eμ-*Myc Tfap4*$^{+/–}$ (n=5) mice. A fraction of individual *Ighv* reads over total read counts for all *Ighv* within each sample are shown where each color represents expressed mRNA for distinct *Ighv*. (**B**) Principal component analysis (PCA) of RNA-seq data in (A), following filtering of genes expressed at RPKM>2 in any given sample. (**C-D**) A z-score heatmap (C) and relative transcript frequencies (D) for anti-apoptosis-,

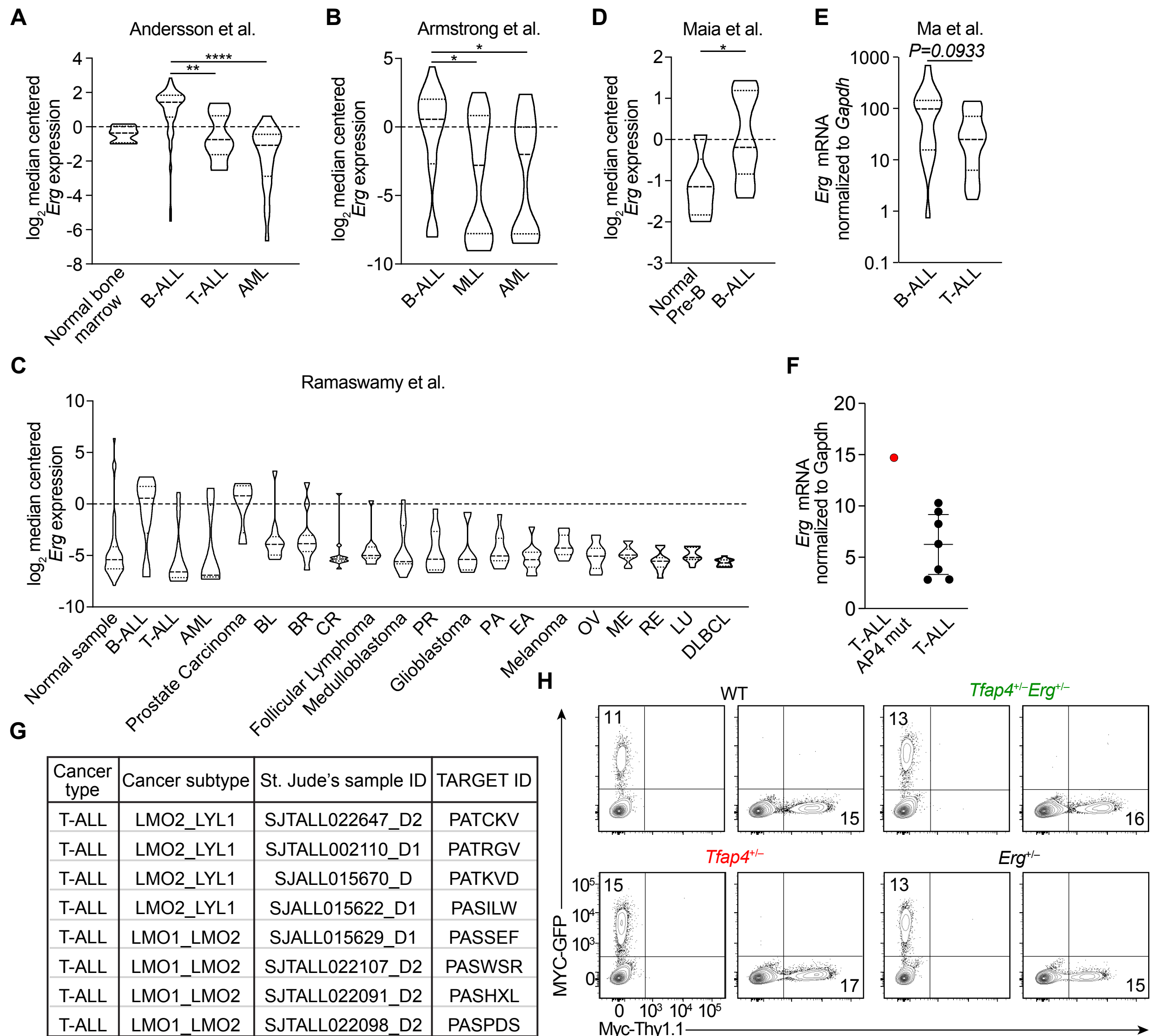
cell cycle-, pro-apoptosis- and senescence-associated genes in Eμ-*Myc Tfap4*$^{+/+}$ and Eμ-*Myc Tfap4*$^{+/–}$ BM B220^{+} IgM$^{–}$ pre-transformed cells (Eμ-*Myc Tfap4*$^{+/+}$, n=5; Eμ-*Myc Tfap4*$^{+/–}$, n=5 mice). (**E**) PCA of RNA-seq data of Eμ-*Myc Tfap4*$^{+/+}$ (n=4), Eμ-*Myc Tfap4*$^{+/–}$ (n=5), Eμ-*Myc Tfap4*$^{+/–}$ with LOH (n=5), and Eμ-*Myc Tfap4*$^{–/–}$ (n=2) pro/pre-B tumor cells. (**F**) A z-score heatmap showing expression of differentially expressed genes defined in Figure 4B by tumor pro/pre-B cells from mice with indicated *Tfap4* genotypes. (**G**) GSEA of genes in Figure 4D using selected signatures enriched in indicated clusters.

Figure S5.



Supplemental Figure 5. *Tfap4*$^{+/–}$ Eμ-*Myc* pro/pre-B tumor cells harbor comparable overall mutation burden but frequent oncogenic *Kras* mutations. Related to Figure 4.
(**A**) Usage of variable gene segments of *Igh* in B220^{+} IgM$^{–}$ tumor B cells sorted from Eμ-*Myc Tfap4*$^{+/+}$ (n=5), Eμ-*Myc Tfap4*$^{+/–}$ (n=5), Eμ-*Myc Tfap4*$^{+/–}$ LOH (n=5), and Eμ-*Myc Tfap4*$^{–/–}$ (n=2) mice. A fraction of individual *Ighv* reads over total read counts for all *Ighv* within each sample are shown where each color represents expressed mRNA for distinct *Ighv*. (**B**) Frequencies of coding mutations of pro/pre-B tumor cells from three Eμ-*Myc Tfap4*$^{+/+}$, three Eμ-*Myc Tfap4*$^{+/–}$ and combined three Eμ-*Myc Tfap4*$^{+/–}$ mice with LOH and one Eμ-*Myc Tfap4*$^{–/–}$ mouse determined by whole exome sequencing. (**C**) List of coding mutations that were detected at the variant allele frequencies ≥10%. (**D**) Frequencies of *Kras* mutations in pro/pre-B tumor cells of indicated genotypes. Mutations found in Eμ-*Myc Tfap4*$^{+/–}$ tumor cells are italicized. * indicates a mutation detected in one out of six sequenced clones.

Figure S6.



Cancer type	Cancer subtype	St. Jude's sample ID	TARGET ID
T-ALL	LMO2_LYL1	SJTALL022647_D2	PATCKV
T-ALL	LMO2_LYL1	SJTALL002110_D1	PATRGV
T-ALL	LMO2_LYL1	SJALL015670_D	PATKVD
T-ALL	LMO2_LYL1	SJALL015622_D1	PASILW
T-ALL	LMO1_LMO2	SJALL015629_D1	PASSEF
T-ALL	LMO1_LMO2	SJTALL022107_D2	PASWSR
T-ALL	LMO1_LMO2	SJTALL022091_D2	PASHXL
T-ALL	LMO1_LMO2	SJTALL022098_D2	PASPDS

Supplemental Figure 6. *ERG* is upregulated in B-ALL compared to other human cancers. Related to Figures 5 and 6.

(**A-D**) Violin plots of $\log_2$ median-centered *ERG* expression determined by microarray-based gene expression profiling of various cancer types in indicated studies. Bladder urothelial carcinoma (BL), breast adenocarcinoma (BR), colorectal adenocarcinoma (CR), prostate adenocarcinoma (PR), pancreatic adenocarcinoma (PA), endometrial adenocarcinoma (EA), ovarian adenocarcinoma (OV), pleural mesothelioma (ME), renal cell carcinoma (RE), lung adenocarcinoma (LU), diffuse large B-cell lymphoma (DLBCL). Data presented here were obtained and analyzed through the Oncomine database (www.oncomine.org, Thermo Fisher Scientific, Ann Arbor, MI). (**E**) A violin plot of *ERG* expression in pediatric B-ALL and T-ALL, normalized to *GAPDH* from RNA-seq data from the TARGET database. (**F**) *ERG* expression in a T-ALL sample with an E57K AP4 mutation and an LMO2-LYL1 translocation (red) and control T-ALL samples of LMO2-LYL1 and LMO1_LMO2 subtypes without a somatic AP4 mutation (black). Data are shown with a mean and a 95% confidence interval. (**G**) List of samples in (F) with the subtype and sample identifiers. The first sample has E57K AP4 mutation. RNA-seq data

were obtained from the TARGET website and is also available in dbGaP under accession number phs000218. (**H**) Representative flow cytometry plots of $CD19^{+}$-enriched BM cells from WT, *Tfap4*$^{+/-}$, *Tfap4*$^{+/-}$ *Erg*$^{+/-}$ and *Erg*$^{+/-}$ mice ~24hrs post infection with MYC-GFP and MYC-Thy1.1 viruses.