PNAS

Supplementary Information for

Tcf-1 promotes genomic instability and T cell transformation in response to aberrant β-catenin activation.

Stephen Arnovitz, Priya Mathur, Melissa Tracy, Azam Mohsin, Soumi Mondal, Jasmin Quandt, Kyle M. Hernandez, Khashayarsha Khazaie, Marei Dose, Akinola Olumide Emmanuel, Fotini Gounari

Fotini Gounari
Email: fgounari@uchicago.edu

This PDF file includes:

Supplementary text
Figures S1 to S5
Tables S1 to S2

Supplementary Information Text

Materials and Methods

Mice

Cd4-Cre/Ctnnb1^{ex3fl} (CAT), *Cd4-Cre/ Ctnnb1*^{ex3fl}/*Tcf7*$^{fl/fl}$ (CAT-*Tcf7*$^{\Delta}$), *CD4-Cre/Ctnnb* ^{ex3fl}/*BclXL*$^{fl/fl}$(CAT-*BclXL*$^{\Delta}$), *Cd4-Cre/Tcf7*$^{fl/fl}$ (*Tcf7*$^{\Delta}$), *CD4-Cre/BclXL*$^{fl/fl}$(*BclXL*$^{\Delta}$) or littermate control *Cd4-Cre (*Cre*)* were used in all described experiments. *Ctnnb*^{ex3fl} allele was previously reported (1), and mice carrying *CD4-Cre*, *Tcf7*^{fl} or *BclXl*^{fl} alleles and Rag$^{-/-}$ mice were procured from Jackson lab. Mice were maintained on a C57BL/6 background and housed at the University of Chicago animal facility in accordance with protocol #71880, approved by the University of Chicago Institutional Animal Care and Use Committee.

Isolation of thymocytes.

Thymi were resected from 6–8-week-old mice in ice-cold FACS buffer (2%FBS in PBS) and dissociated to single cell suspension using 0.70 μM strainers. Red blood cells were lysed with ACK buffer (150 mM NH4Cl, 10 mM KHCO3 and 0.1 mM EDTA (pH 7.4)) for 2 min on ice. Reaction was halted by flooding with ice cold 1x PBS. Cells were immediately spun down and

resuspended in fresh FACS buffer (2%FCS in PBS). Lymphoma samples were collected as above from 2–3-month-old *Cd4-Cre/Ctnnb1*^{ex3fl} mice with enlarged and transformed thymi.

Flow cytometry and fluorescence-activated cell sorting of murine lymphocytes

Thymocytes were surface stained in flow cytometry buffer (2% FBS in PBS) for 30 min on ice. Samples were washed with flow cytometry buffer, and data were acquired on an LSRII flow cytometer (Becton Dickinson). The data were analyzed with FlowJo software (Becton Dickinson). Surface antibodies were to CD4 (clone GK1.5, BD Biosciences), CD8 (clone 53-6.7, eBiosciences), Tcrβ (clone H57-597, eBiosciences), and Ccr7(clone 4B12, eBiosciences). Cells were stained for viability using LIVE/DEAD Aqua fluorescent reactive dye (Molecular Probes–Life Technologies, L34963). For γH2ax experiments, cells were then fixed and permeabilized using the Foxp3/Transcription Factor Fixation kit (eBioscience 00-5521) according to manufacturer recommendations. Cells were stained intracellularly for γH2ax (Anti-H2AX (pS139), BD Biosciences, BDB562377) for 30 minutes on ice in permeabilization buffer, washed 2x, and acquired as above.

Translocation Breakpoint Detection

Sequencing libraries from 4 independent CAT leukemia samples were created with the Nextera Mate Pair Preparation Kit (FC-132-1001), which allows for large insert-size (up to 12kb) and is ideal for the detection of structural variation. Each biological sample was replicated across two lanes and sequenced using HiSeq2500. Reads were then preprocessed and special adaptor clipping procedures were necessary before any alignment procedure NextClip (2). Aligned MP reads were post-processed with Samtools v0.1.18 (3) and Picard v1.70 (http://picard.sourceforge.net/). BreakDancer v1.3.6 (4) was used to detect inter-chromosomal translocations. Finally, the distribution of breakpoints across chromosomes was analyzed using the 5kb binned data. Primers were then designed for consensus translocations involving *Myc-Pvt1* loci and precise clonal breakpoint sequences were determined with Sanger sequencing (Table S1).

Identification of RSS and cryptic RSS sites

Genomic sequences from translocation breakpoints +/- 200 bp were investigated for RSS and cryptic RSS (cRSS) sites using the online reference database and predication tool RSSsite (5). For *Myc-Pvt1* loci, the DnaGrab tool was used to score potential RSSs in this region. The top scoring potential cRSS site containing a 23-bp spacer was selected from each of 4 CAT lymphomas, representing the most likely substrate for fusion with 12-bp spacer RSS sites in *Tcra* loci according to the 12/23 rule.

Rag Recombination Assay

As previously described, a plasmid-based Rag-recombination assay (6) was adapted to evaluate potential RSS sites for Rag-cleavage. Briefly, a modified pMX-INV vector harbors an inverted eGFP sequence flanked by canonical RSS sites such that successful Rag-recombination flips the cassette in frame, leading to eGFP expression (pMX-INV-GFP) (gift Sleckman). An IRES-Thy1.2 marker is 3' to the eGFP cassette, to mark successful viral transduction. The top scoring potential cryptic RSS site from *Myc-Pvt1* break sites were synthesized by IDT, annealed, and cloned to replace the upstream RSS site (Figure S1a). Viral supernatants were produced by transfection into Plat-E cells using FuGENE reagents according to manufacturer recommendations (Promega E2311). Viral supernatants were collected, filtered, and used to transduce v-Abl transformed pro-B cells (gift Sleckman). These highly proliferative cells are forced into G1 by treatment with 3uM STI571 (Imatinib, Selleck Chem, S2475) for 3 days, inducing Rag expression and recombination (Figure S1b). After Imatinib treatment, cells are collected, stained for Thy1.2 on ice for 30 minutes in flow cytometry staining buffer, (anti-CD90.2/Th1.2-PE, Biolegend #553005) and evaluated for eGFP expression by flow cytometry using an LSRII (Becton Dickinson).

Nascent DNA Fiber Assays

For gentler selection conditions prior to culture, thymocytes were enriched for DPs using a CD4+ T Cell Isolation kit (Miltenyi, 130-104-454) according to manufacturer protocols. Cells were subsequently cultured in thymocyte media (IMDM, 10%FBS, 1%Pen/Strep, 50mM β-mercatopethanol) for 1 hour at 5% CO_2 and 37°C to acclimate cells to culture conditions. For untreated fibers, cells were pulse labeled first with 25 μM IdU (MP Biochemicals, 0210025701) for 25 minutes, washed once in 1xPBS, and subsequently resuspended in media with 250 μM CldU (Sigma, C6892) for an additional 25 minutes. For stressed replication assays, cells were co-cultured in 2mM Hydroxyurea (Sigma, H8627) during the second pulse. After labeling, cells were immediately collected on ice and diluted to $7.5x10^5$ cells/mL. Cells were then spotted onto the top of glass slides, briefly dried, and lysed dropwise (100mM Tris pH 7.5, 0.5% SDS, 50mM EDTA). After 2-3 minutes, slides were tilted at a 20-40° angle allowing droplet to roll down the slide length. Slides were then air dried for 45 minutes, fixed in a 3:1 methanol/ acetic acid solution for 10 minutes, and denatured in 2.5M HCl for 80 minutes. Fibers were stained at 4°C overnight in a humidified chamber with 1%BSA containing primary antibodies rat-α-BrdU (1:200, BD Biosciences, B44) and mouse-α-BrdU (1:25, Abcam, BU1/75(ICR1)), which label IdU and CldU, respectively. Slides were then washed in PBS, fixed for 10 minutes (3% PFA, 3.4% sucrose in PBS), and stained with α-rat Alexa Fluor 488 (1:500, Invitrogen, A-11006) and α-mouse Alexa Fluor 594 (1:400, Invitrogen, A-11005) for 1.5 hours at room temperature. Slides were then washed and mounted with ProLong Anti-fade mounting media (Invitrogen, P36930) and #1.5 cover glass. Images were acquired with an Olympus IX81 inverted microscope with a 100X, NA 1.45 objective at the Integrated Light Microscopy Core at the University of Chicago. Images were visualized and individual DNA fibers were hand selected using ImageJ. Individual fiber track lengths and CldU:IdU ratios were measured using custom Matlab scripts to avoid user bias.

Chromatin immunoprecipitation and sequencing

$1x10^7$ Live $CD4^+CD8^+$ DP thymocytes were sorted from replicate *Cd4-Cre*, *Cd4-Cre/Ctnnb1*^{ex3fl}, *Cd4-Cre/ Ctnnb1*^{ex3fl}/*Tcf7*$^{fl/fl}$ and prepared for chromatin immunoprecipitation and library generation as previously described [7]. Antibodies were to H3K27Ac (Abcam, #4729), H3K27me3 (Millipore #07-449) and H3K9K14Ac (Millipore, #06-599). Additional ChIP data for Tcf-1 and H3K4me3 and some CD4-Cre profiles are publicly available from previous reports (GSE46662, GSE32311, SRP142342). Sequencing was performed at the University of Chicago Genomics Facility using a HiSeq 4000 sequencer.

Isolation and preparation of samples for RNA-seq

$1x10^7$ Live $CD4^+CD8^+$ DP thymocytes were sorted from three *Cd4-Cre, Cd4-Cre/Ctnnb1*^{ex3fl}, *Cd4-Cre/ Ctnnb1*^{ex3fl}/*Tcf7*$^{fl/fl}$ and total RNA was extracted using TRIzol (15596026, Invitrogen) according to the protocol described by the Immunological Genome Project (https://www.immgen.org/). Libraries were generated and sequenced by the University of Chicago Genomics Facility.

Genome mapping and data analysis

Sequenced ChIP datasets were mapped with the Galaxy (https://usegalaxy.org/) suite of tools. Data were groomed and aligned to the mouse mm9 genome with Bowtie, allowing up to one mismatch and retaining only uniquely mapped reads, and unmapped reads were filtered. Peak calling and further analyses were done as described previously (7). Motif enrichment analysis was performed with HOMER motif discovery algorithm. Tcf-1 and histone ChIP profiles are represented as heatmaps and enrichment histograms using ngs.plot software (8). Sequenced RNA datasets were aligned to the mouse mm9 genome similarly to ChIP–seq datasets. Batch correction was performed with the RUVs function in RUVSeq (9). Kmeans clustering analysis of normalized RNA-seq reads was performed with ExpressCluster1.3 (www.GenePattern.org) (10) using Euclidean distance and filtering for minimum expression of 1 FPMK and minimum fold change of 1.2. Differential gene expression analysis was performed with EdgeR (11). Only genes

that exhibited a counts per million (CPM) greater than 0.5 in at least 3 samples were kept, and TMM normalization was applied using the calcNormFactors (12) function in EdgeR. Genes with transcript abundance differences below $P < 0.05$ were considered to be significantly differentially expressed. Heat maps of normalized reads for gene subsets and further analyses were as described in (7). Pathway enrichment analysis was performed using Metascape. Gene set enrichment analysis (GSEA) was run against all gene sets within KEGG, HALLMARK, Reactome and GO databases (Broad, MsigDB).

Spearman correlation density plots

Density plots were created with the stat_bin2d function in the ggplot2 package in R, with 30 bins in each dimension. For visualization purposes, the axis ranges of some density plots were limited to highlight the high-probability regions of the plot. Spearman correlation coefficients and P values were computed in R with the cor and cor.test functions. Genes were then filtered for the greatest relevance to the rescue condition by selecting those with the tightest fit to the spearmen correlation (i.e. distance ≤ 1 from the fit line) to perform pathway analysis on the ”restored” genes relative to CAT.

Olaparib treatment in Mice

Suspensions of CAT lymphomas were isolated and $2x10^6$ cells were transferred into sub lethally irradiated (450 rads) $Rag^{-/-}$ mice by tail vein injection. Three days post-transplant, mice were treated 5x/week for 3 weeks with intraperitoneal injections of 50mg/kg Olaparib (AstraZeneca, Lynparza). The progression of CAT lymphomas was enumerated by the fraction of $CD4^+CD8^+$ cells identified in weekly peripheral blood sampling (100ul) and flow cytometry after staining for CD4 and CD8, as described above.

Genomics of Drug Sensitivity in Cancer (GDSC) database

Normalized IC50 data was extracted for all compounds and all cancer cell lines from the Genomics of Drug Sensitivity in Cancer database (13). Cell line metadata was also extracted and used to define T-cell and leukemia/ lymphoma cell lines, which was integrated to IC50 data using custom R-scripts. Metadata for inhibitors was used to classify pathway targets, and data was reduced to drugs of interest (DOIs) targeting WNT, PARP, or AKT pathways. Means of normalized IC50 data for DOIs in T-ALL or leukemia and lymphoma cell lines were compared to all other cancer lines using two-sided, unpaired t-tests. Linear correlations of IC50 data for Olaparib versus other WNT or PARP inhibitors was performed in R using the lm() function and data is depicted using the ggplot2 package.

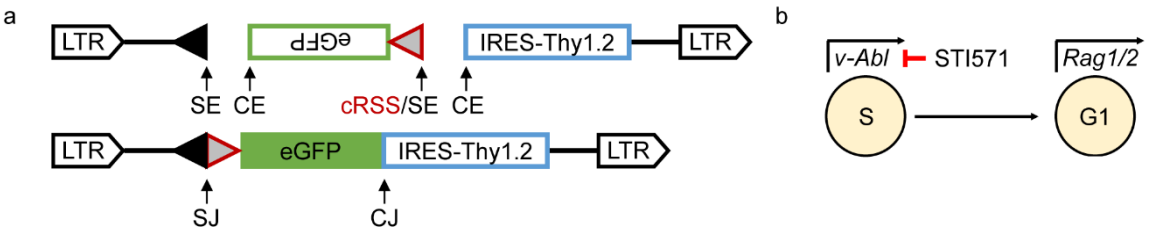

Fig. S1. Schematic of Rag-reporter system (a) Schematic of pMX-INV-GFP construct. An eGFP cassette is flanked by directional RSS or cRSS sites (triangles) where Rag recombinases act to generate hair-pins at coding ends (CE) and blunt signal ends (SE). Rag mediates inversion of the eGFP cassette to generate a coding join (CJ) and signal join (SJ) that puts eGFP in frame to be expressed. A flanking IRES-Thy1.2 reports on successful transduction of the construct (adapted from Bredemeyer et al.) (39) (b) Schematic of highly proliferative (v)-Abl transformed pre-B-cells (gift, Sleckman) in S-phase (S) that are driven into G1 by treatment with Abelson kinase inhibitor ST1571 (Imatinib) to initiate Rag expression. or paste legend here.

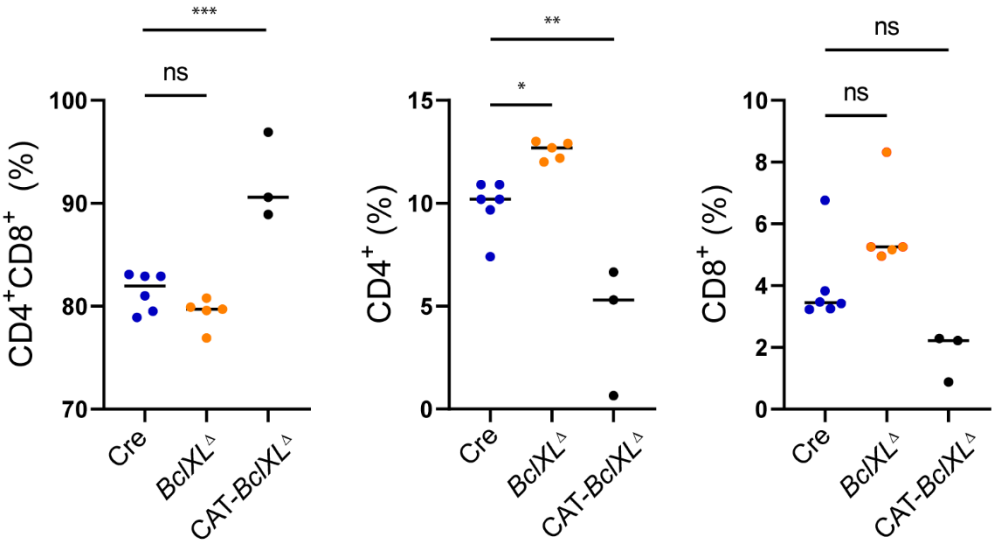

Fig. S2. Co-deletion of BclXL does not rescue CAT lymphoma phenotypeype. Flow cytometric histograms with the percentage of live thymocytes in the indicated late developmental stages in Cre (n=6), *BclXL*$_{\Delta}$ (n=5), and CAT-*BclXL*$^{\Delta}$ (n=3) mice); data are represented as the mean ± s.e.m., and statistical testing is depicted as two-sided, unpaired t-tests; *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001.here.

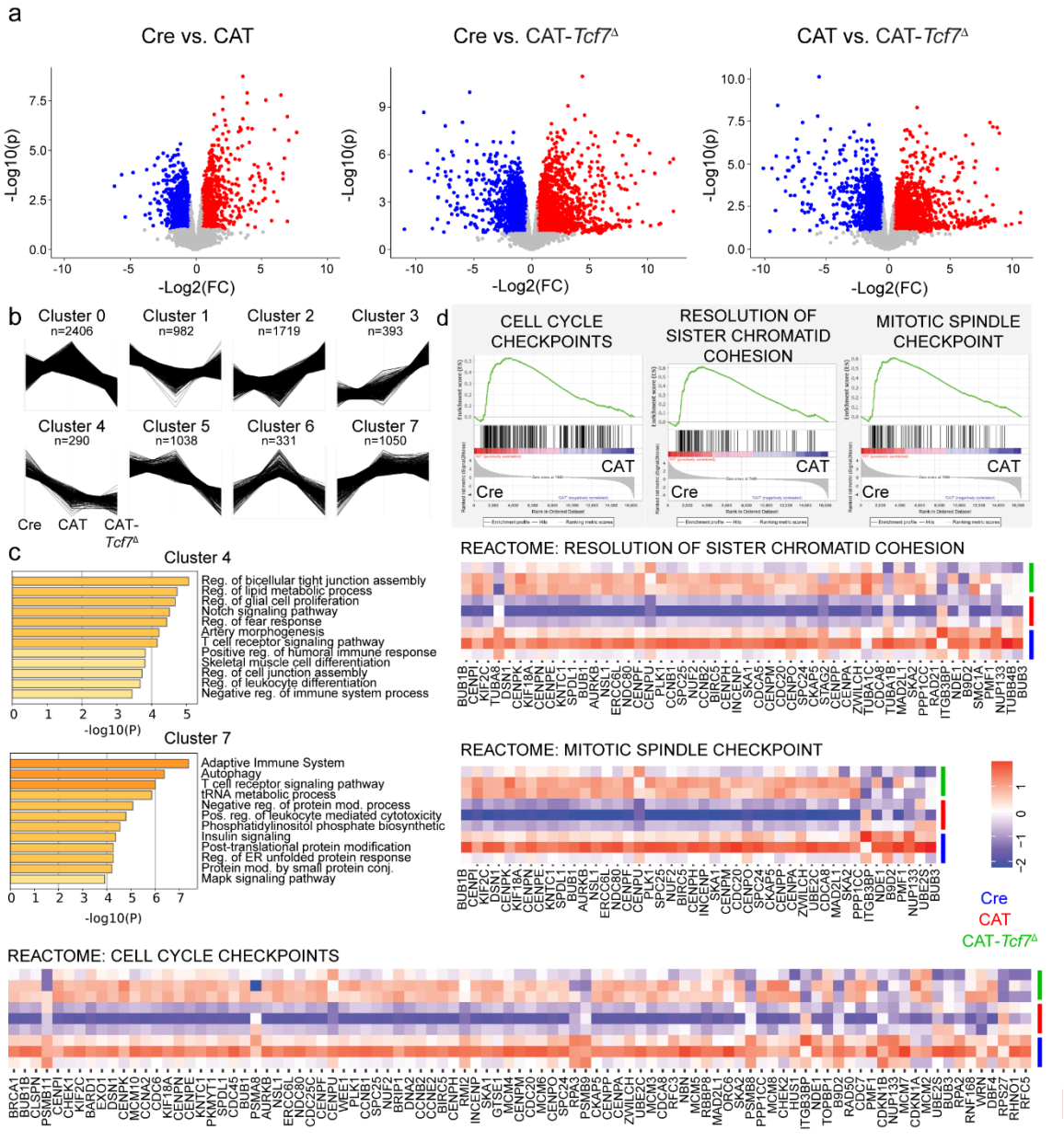

Fig. S3. Transcriptional profiling of Tcf-1 regulated genes in CAT. (a) Volcano plots of differentially expressed genes (DEGs) in DP thymocytes between the indicated conditions. (b) Full profiles from unsupervised clustering of triplicate RNA-seq profiles (FPKMs) for Cre (n=3), CAT (n=3), and CAT-*Tcf7*$^{\Delta}$ (n=3) DP (CD4^{+}CD8^{+}) thymocytes. (c) Pathway analysis (Metascape) for gene clusters associated with altered expression in CAT thymocytes that remain dysregulated in CAT-*Tcf7*$^{\Delta}$ (Cluster 4 and Cluster 7). (d) GSEA profiles for the indicated pathways highlighted in Fig. 2e. and heatmaps of core enrichment genes of the associated pathways.

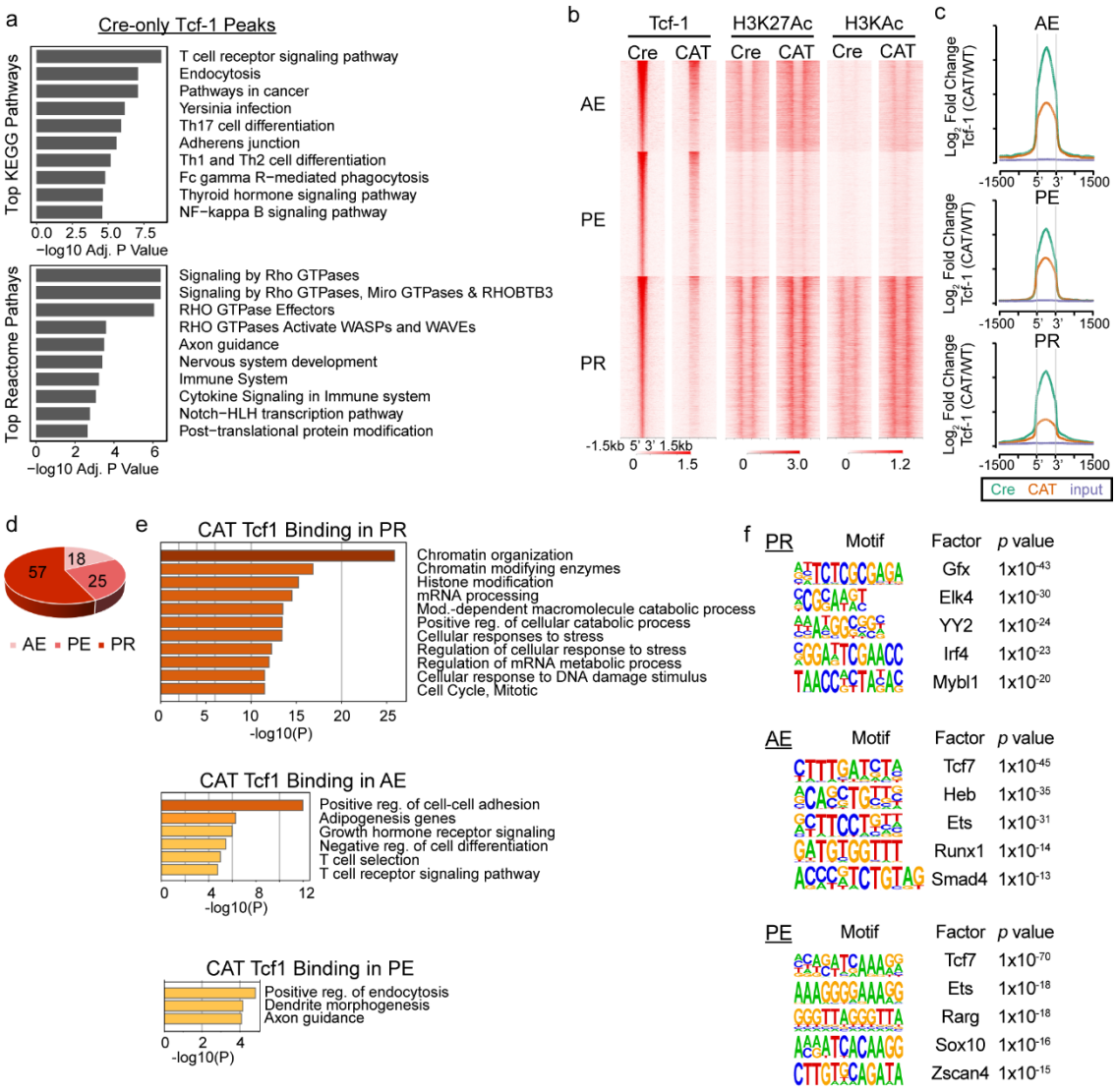

Fig. S4. Epigenetic analysis of Tcf-1 in DP thymocytes. (a) Pathway enrichment analysis (g:Profiler, biit.cs.ut.ee/gprofiler) for genes associated with Tcf-1 binding in Cre that are lost in CAT mice (n=8,195 peaks); top significantly enriched KEGG and Reactome pathways are shown. (b) Heat maps of ChIP-seq peaks for Tcf-1 and indicated histone marks centered on Tcf-1 binding (+/- 1.5kb) at promoters (PR), poised enhancers (PE), or active enhancers (AE), as previously defined by histone profiles in Cre DP thymocytes (33) and (c) corresponding enrichment histograms for Tcf-1 binding. (d) Distribution of CAT unique Tcf-1 binding sites to PR, PE, or AE as in b and c. (e). Pathway analysis of genes with CAT unique Tcf-1 sites in the indicated regulatory regions. f. The most significantly enriched transcription-factor-binding motifs (HOMER) of CAT unique Tcf-1 sites at the indicated regulatory regions.

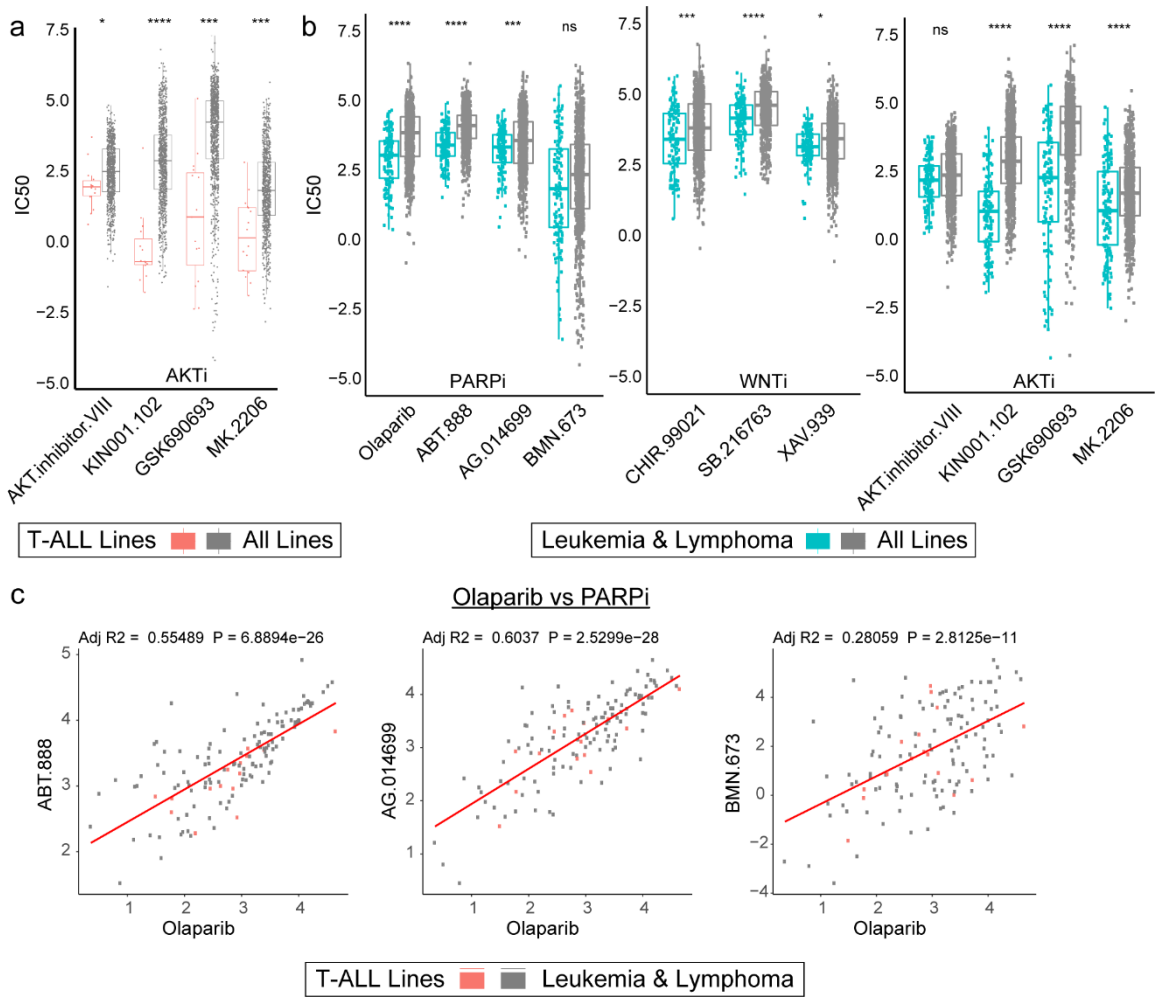

Fig. S5. Drug sensitivities in T-ALL, leukemia, and lymphoma cell lines. Normalized IC50 data from the Genomics of Drug Sensitivity in Cancer (GDSC) database comparing (a) T-ALL cell line (n=16) sensitivity to all cancer cell lines (n=974) for inhibitors targeting AKT (AKTi). (b) IC50 data comparing leukemia and lymphoma cell lines (n=150) to all other cancer cell lines for sensitivity to inhibitors targeting PARP, WNT or AKT. Data represented as the mean ± s.e.m. and statistical testing depicted as two-sided, unpaired t-tests; *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001. (c) Correlations of Olaparib sensitivity (IC50) to other PARPi compounds in all leukemia and lymphoma cell lines (n=150) with T-ALL lines highlighted (n=16, red). Linear regressions analysis performed in R assuming Gaussian distribution and correlations represented as adjusted R-squared values (adj R2).

Table S1. Primers for Sanger sequencing of translocation breakpoints.

Translocation	breakpoint Chr15 (*Myc-Pvt1*)	breakpoint Chr14 (*Tcra*)	PCR primer Chr15 (5'→3')	PCR primer Chr14 (5'→3')
1	62025135	1606036	GAATTCTGAACCTGCAGAAGGAGC	GATTGGGAGTCACAGCAACAGTTG
2	61993565	1606036	ATAAGGACCCATGCTTGACACCAC	GTTGAAGTGTGAGCATGGCAGAAG
3	62295692	1595874	GCAGCTCTCAGAGTTTCAAAGCTG	CTTCCAGGCACTTGGAAATGTTGG
4	62214890	1567387	AACCAGGCTATTGTCTTGCAGGTG	CCTAGTGATCCAGTCGTGTTGAGT

Primers designed to detect precise translocation breakpoints in 4 CAT lymphomas based on discordant reads from mate-pair sequencing. (Genomic position of breakpoints correspond to mm10)

Table S2. Gene set for PTEN targets in human T-ALL.

PTEN_targets_T-ALL
ALDH1B1
ARL4C
ART2A-PS
BACH2
CACNA1E
CARD6
CCR7
CD226
CD53
CD7
CMAH
DTNA
FAM101B
GIMAP3
GIMAP4
GIMAP7
GM11346
GM14085
HSD17B11
IL2RB
IL6RA
INPP4B
ITGA4
ITGB3
KLF2
S1PR1
SCEL
SELL

PTEN gene signature derived from RNA expression profiling after reintroduction of PTEN in PTEN-deficient human T-ALL cells. Established by the lab of Dr. Adolfo Ferrando (14).

SI References

1. N. Harada *et al.*, Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. *EMBO J* **18**, 5931-5942 (1999).
2. R. M. Leggett, B. J. Clavijo, L. Clissold, M. D. Clark, M. Caccamo, NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* **30**, 566-568 (2014).

3. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
4. K. Chen *et al.*, BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* **6**, 677-681 (2009).
5. I. Merelli *et al.*, RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. *Nucleic Acids Res* **38**, W262-267 (2010).
6. A. L. Bredemeyer *et al.*, ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466-470 (2006).
7. A. O. Emmanuel *et al.*, TCF-1 and HEB cooperate to establish the epigenetic and transcription profiles of CD4(+)CD8(+) thymocytes. *Nat Immunol* **19**, 1366-1378 (2018).
8. L. Shen, N. Shao, X. Liu, E. Nestler, ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
9. D. Risso, J. Ngai, T. P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* **32**, 896-902 (2014).
10. M. Reich *et al.*, GenePattern 2.0. *Nat Genet* **38**, 500-501 (2006).
11. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
12. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
13. F. Iorio *et al.*, A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166**, 740-754 (2016).
14. T. Palomero *et al.*, Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. *Nat Med* **13**, 1203-1210 (2007).