Supplementary Material for:

An Oncogenic Super-Enhancer Formed through Somatic Mutation of a Noncoding Intergenic Element

Marc R. Mansour, Brian J Abraham, Lars Anders, Alla Berezovskaya, Alejandro Gutierrez, Julia Etchin, Lee Lawton, Stephen E. Sallan, Lewis B. Silverman, Mignon L. Loh, Stephen P. Hunger, Takaomi Sanda, Richard A. Young, and A. Thomas Look

Corresponding authors. E-mail: thomas_look@dfci.harvard.edu and young@wi.mit.edu

This pdf file includes:

Material and Methods

Cell lines

The identity of T-ALL cell lines was analyzed by analysis of short tandem repeats using the PowerPlex 1.2 system (Promega) in January 2013, and the Jurkat and MOLT-3 cell lines were retested in January 2014. T-ALL cells were maintained in RPMI-1640 medium supplemented with 10% FBS, L-glutamine, and penicillin/streptomycin (Invitrogen). HEK-293T cells were recently obtained from American Type Culture Collection, and maintained in Dulbecco's modified Eagle's medium supplemented with 10% FBS, L-glutamine, and penicillin/streptomycin.

T-ALL Patient samples

A total of 146 pediatric T-ALL diagnostic specimens were collected with informed consent and IRB approval from children treated on Children's Oncology Group (COG) P9404, Dana-Farber Cancer Institute (DFCI) 00-01, and St. Jude Children's Research Hospital (SJCRH) Total Therapy XI–XIII clinical trials (*25-28*).

Sequencing of the STIL-TAL1 enhancer region

The 766-bp region of the STIL-TAL1 enhancer was amplified by 35 cycles of PCR using primers 5'-TGAACGGTGACTTTCCAAATC and 5'-CTGTACACTAAAATGAGTACATCT with an annealing temperature of 59°C using Phusion reagents (Finnzymes). PCR products were sequenced in both forward and reverse orientation, and positive samples were confirmed by repeat PCR and sequencing.

Quantitative Reverse Transcriptase PCR (qRT-PCR)

Total RNA was harvested using the RNeasy (Qiagen) kit and reverse transcribed with Superscript III (Invitrogen). Quantitative PCR analysis was conducted with the ViiATM 7 system (Life Technologies) using SYBR Green PCR Master Mix (Roche) and specific primers sets for each gene. Primer sequences for qRT-PCR were TAL1-F 5'-AGGGCCTGGTTGAAGAAGAT; TAL1-R 5'-AAGTAAGGGCGACTGGGTTT; GAPDH-F 5'-TGCACCACCAACTGCTTAGC; GAPDH-R 5'GGCATGGACTGTGGTCATGAG; MYB-F 5'-TGTTGCATGGATCCTGTGTT; MYB-R 5'-AGTTCAGTGCTGGCCATCTT. *TAL1* expression was classified as absent if no signal was detected after 40 cycles of PCR. Primers directed at the 3'UTR of TAL1 (to detect only endogenous TAL1, and not the MSCV-TAL1 lacking the 3'-UTR sequence) were 5'-CTAGTGGCTTGTCCTCACC and 5'-TCACAGGATCTTCCAAACGT.

Analysis of SNPs in 3'-UTR of TAL1

All RNA samples were DNase-treated (Qiagen) prior to Superscript III RT-PCR (Invitrogen). The three most frequent SNPs of the TAL1 3'-UTR reported on dbSNP (rs977747, rs22070929 and rs7664) were analyzed by PCR of genomic DNA and paired cDNA samples, and sequencing with the following primer pairs (with primer names reflecting the distance downstream of the TAL1 stop codon): 568F 5'-CCTTCTTCAGGGCCTGGTTG and 852R 5'-CAGCACACTGGCATTCACTC; 1660F

5’-CTAGTGGCTTGTCCTCACC and1867R 5’-TCACAGGATCTTCCAAACGT; 3088F5’-TCTGTAGTCAGCCGACAACT and 3401R 5’-CTTCCCGATACATCCTCACA.

Luciferase reporter experiments

A 395-bp region of the STIL-TAL1 enhancer site was amplified from TAL1 enhancer mutation positive patients by 35 cycles of PCR using Phusion reagents (Finnzymes) and primers 5’-CTGTGCACAGCTGGAGCTCT and 5’-TATACTCGAGTCATGCTGCTCAGGGCCA. PCR products were digested with SacI and XhoI (New England Biolabs), and cloned into the respective sites of the PGL3-promoter vector (E176A, Promega), encoding a minimal SV40 promoter upstream of Firefly luciferase (PGL3-luc). For reporter assays, $1x10^6$ Jurkat cells resuspended in 100μl of Nucleofector Solution (Mirus) with the addition of 1μg of PGL3-luc and 500ng of renilla plasmid (pTK). Cells were electroporated on program D-23 (Amaxa) and resuspended in 500μl of RPMI/10%FCS and incubated at 37°C/5%CO_2 for 48 hrs. Luciferase activity was measured using the Dual-Glo Luciferase Assay system (Promega) as per the manufacturer’s guidelines. Experiments were performed in triplicate. To control for cell number and transfection efficiency, Firefly luciferase activity was normalized to renilla luciferase. Measurements were expressed as a ratio relative to activity of the wild-type STIL-TAL enhancer construct. For reporter assays incorporating MYB knockdown, 50nM MYB siRNA#1 or #2, or 50nM control siRNA were nucleofected into Jurkat cells as described above and luciferase activity was measured at 24 hrs.

MYB siRNA experiments

Control siRNA was purchased from Dharmacon, and MYB siRNAs sequences were: MYB siRNA#1 5’-UAUAGUGUCUCUGAAUGGCUGCGGCUU; MYB siRNA#2 5’-UAUCAGUUCGUCCAGGCAGUU, and have been previously reported (*29, 30*). siRNAs were used at 50nM in Jurkat cells and 500nM in MOLT-3 cells.

Western blotting and Co-Immunoprecipitation experiments

Whole cell lysates were prepared in RIPA buffer. Immunoblotting was carried out with the following antibodies as previously described (*9, 10*): mouse monoclonal anti-TAL1 antibody (clone BTL73; Millipore) 1 in 1000, rabbit anti-MYB (ab45150, Abcam) 1 in 2000, rabbit anti-RUNX1 (D33G6, Cell Signaling) 1 in 2000, rabbit anti-GATA3 (D13C9, Cell Signaling) 1 in 1000, and mouse monoclonal anti-α-tubulin antibody diluted 1 in 10,000 (clone B-5-1-2; Sigma-Aldrich). Secondary horseradish peroxidase (HRP)-linked IgG antibodies to mouse or rabbit diluted 1 in 10,000 were from Cell Signaling Technology. For co-immunoprecipitation experiments, Jurkat cells were lysed in protein lysis buffer (25mM HEPES, 5mM MgCl2, 25mM KCl, 0.05mM EDTA, 10% glycerol and 0.1% NP40). Following clearance, nuclear lysates were prepared in lysis buffer (10mM HEPES, 3mM $MgCl_2$, 100mM KCl, 0.1mM EDTA, 10% Glycerol) by sequential centrifugation. 2000 μg of protein was immunoprecipitated by Dynabeads, with either 5 μg of TAL1 clone BTL73 (Millipore, Billerica, MA), MYB clone EP768Y (Abcam, Cambridge, MA) or anti-mouse or rabbit IgG (Santa Cruz Biotechnology, Dallas, TX, USA), as appropriate. Immunoprecipitation was carried out per the manufacturers instructions (Life Technologies, Waltham, MA). Protein was resolved by SDS-PAGE by standard methodologies and visualized with Dura ECL reagent (Thermo Fisher Scientific, Waltham, MA).

Genome-wide occupancy analysis
ChIP coupled with massively parallel DNA sequencing (ChIP-seq) was performed as previously described (*31, 32*). The following antibodies were used for ChIP: anti-TAL1 (Santa Cruz, sc-12984), anti-H3K27ac (Abam, ab4729) and anti-cMYB (Abcam, ab45150 and Millipore, 05-175). For each ChIP, 10 μg of antibody was added to 3 ml of sonicated nuclear extract. Illumina sequencing, library construction and ChIP-seq analysis methods were previously described (*31*). ChIP-seq datasets are available under superseries GSE59657, and relevant accession numbers are shown in table S3.

ChIP-Seq processing
Reads were aligned to build hg19 of the human genome using bowtie with parameters –k 2 –m 2 –e 70 –l 36 –best (*33*). For visualization in the UCSC genome browser (*34*), WIG files were created from aligned ChIP-Seq read positions using MACS with parameters –w –S –space=50 –nomodel –shiftsize=200 to artificially extend reads to be 200bp and to calculate their density in 50bp bins (*35*). Read counts in 50bp bins were then normalized to the millions of mapped reads, giving reads per million (rpm) values.

Enriched regions
Regions enriched in ChIP-Seq signal were identified using MACS with corresponding control and parameters –keep-dup=auto and –p 1e-9. Heatmaps of ChIP-Seq signal in MYB-binding sites (Figure 3A) were made by counting reads in bins dividing each site using bamToGFF_turbo (https://github.com/bradnerComputation). Reads were artificially extended to be 200bp, MYB-binding sites were +/- 5000bp of the center of the MACS-defined peak, and read densities in 100 equally sized bins per MYB-binding site were calculated. The display heatmap was created using heatmap.2 in R.
Super-enhancers in Jurkat were identified using ROSE (https://bitbucket.org/young_computation/rose), as described in Hnisz et al. (*15*). Briefly, peaks of H3K27ac were identified using MACS as described above. These peaks were stitched computationally if they were within 12500bp of each other, though peaks fully contained within +/- 2000bp from a RefSeq promoter were excluded from stitching. These stitched enhancers were ranked by their H3K27ac signal (length * density) with input signal subtracted. Regions co-bound by CBP and MYB were determined by intersecting enriched regions called separately for each. Super-enhancers co-bound by CBP and MYB were defined as those overlapping one of these co-enriched regions.

Motif analyses
Wild-type and mutant enhancer sequences were analyzed using UniPROBE (*19*). Locations of DNA sequence motifs preferred by important T-ALL regulators were identified in the TAL1 super-enhancer using FIMO with motif libraries from Transfac and HOCOMOCO (*36-38*).

Allele-specific re-mapping
Reads aligning to specific allele sequences were determined using bowtie. Artificial bowtie indices were created from the sequences below. Reads aligning to each of these sequences

using bowtie with parameters -S -k 1 -m 1 --best were retained. These retained reads were remapped to the same loci with parameters -S -k 1 -m 1 -n 0 --best. Reads mapping to the reference locus at bases 210-210, the MOLT3 locus at bases 210-212, or the Jurkat locus at bases 210-222 were counted using samtools view and extracting reads with NM:i:0.

Reference locus:
ATGGAGAGACAGGGCCGGGGTAGGAAAGAGGAGCAGACTTAGAGACAGAGAGAATGCACATGCGCTTAAAATAG
AGATGGCATGATGAGAAGCAGAGTGAAAGAGATGATAAGAGATAAAAGGGAGAGAGAGATCCTGTGGTTGCCCC
ACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAAGACGTAACCCTACTTCCTGGCAGATGTCTCTGTCA
CTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGCCCTGTGTGCCTGGCCCTG
AGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAGGGTATGCTAATCAGGAGAGCAAGTGGTGAGAGGAGGAGA
TATAAATGCAGCTGAGGACAGTATTGATCAG

MOLT3 locus:
ATGGAGAGACAGGGCCGGGGTAGGAAAGAGGAGCAGACTTAGAGACAGAGAGAATGCACATGCGCTTAAAATAG
AGATGGCATGATGAGAAGCAGAGTGAAAGAGATGATAAGAGATAAAAGGGAGAGAGAGATCCTGTGGTTGCCCC
ACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAAGACGGTTAACCCTACTTCCTGGCAGATGTCTCTGT
CACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGCCCTGTGTGCCTGGCCC
TGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAGGGTATGCTAATCAGGAGAGCAAGTGGTGAGAGGAGGA
GATATAAATGCAGCTGAGGACAGTATTGATCAG

Jurkat locus:
ATGGAGAGACAGGGCCGGGGTAGGAAAGAGGAGCAGACTTAGAGACAGAGAGAATGCACATGCGCTTAAAATAG
AGATGGCATGATGAGAAGCAGAGTGAAAGAGATGATAAGAGATAAAAGGGAGAGAGAGATCCTGTGGTTGCCCC
ACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAAGACGGTTAGGAAACGGTAACCCTACTTCCTGGCAG
ATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGCCCTGTG
TGCCTGGCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAGGGTATGCTAATCAGGAGAGCAAGTGGT
GAGAGGAGGAGATATAAATGCAGCTGAGGACAGTATTGATCAG

CRISPR/Cas9 deletion of the TAL1 enhancer in Jurkat cells
CRISPR/Cas9 target sites were identified using the CRISPR design tool available at crispr.mit.edu, and targets were chosen with the best specificity, based on analyses performed by Hsu et al. (*39*). To create genomic deletions across the MuTE site of approximately 177-193-bp, two guides RNA were designed to the target sequences shown in Fig. S6, using techniques as previously described by Canver *et al*. (*40*). Guide RNAs were cloned into the BbsI sites of the PX330 vector, which encodes both the guide RNA and mammalian Cas9 enzyme (*41, 42*). Sequence of the oligonucleotides to create double-stranded guide RNAs including the BbsI overhangs are shown below. Oligonucleotides were annealed by heating to 95°C for 5 minutes in NEB buffer 2 (New England Biolabs) and slow cooling to room temperature over 1 hour.

Guide#1-up 5'-CACCGGAATGGGGTGGGGCAACCAC
with guide#1 down-5'-AAACGTGGTTGCCCCACCCCATTCC
and
guide#2-up 5'- CACCGATGAGTTAGACTGTAACGGA
with guide#2-down 5'-AAACTCCGTTACAGTCTAACTCATC.

To specifically target the 12-bp MuTE insertion found in Jurkat cells, a guide RNA#3 was designed against the target sequence 5'-CACAGAAAGACGGTTAGGAAA, using oligonucleotides:
guide#3-up 5'- CACCGCACAGAAAGACGGTTAGGAAA
with guide#3-down 5'-AAACTTTCCTAACCGTCTTTCTGTGC.

In order to maintain the survival of Jurkat cells after CRISPR/Cas9 deletion of the MuTE, Jurkat cells were engineered to express TAL1 from the MSCV-TAL1-neomycin vector, to create Jurkat-MSCV-TAL1-neomycin cells (Jurkat-MTN). Retroviral supernatants, transductions and cell selection with G418 were performed as previously described (*9, 10*). $1x10^6$ total Jurkat-MTN cells were resuspended in 100µl of Nucleofector Solution (Mirus) with the addition of 2µg total PX330 vector (encoding 1µg each of guides#1 and guide#2, or 2µg of guide#3 only). Cells were electroporated on program D-23 (Amaxa) and resuspended in 500µl of RPMI/10%FCS and incubated at 37°C/5%CO_2 for 48 hrs. Single cell clones were established by limiting dilution in 96-well plates; the detailed protocol is available at: http://catalog2.corning.com/LifeSciences/en-BR/TDL/techInfo_abstract.aspx?productid=3124
DNA was extracted from cells after 2-3 weeks of incubation using QuickExtract solution (QE09050, Epicentre, Illumina, Wisconsin, USA), and deletions and mutations were analyzed by PCR and sequencing as described above.

Figure S1. Mutation of the *TAL1* enhancer (MuTE) is associated with monoallelic expression of TAL1. Sequencing chromatograms from the 3' UTR of TAL1 from genomic DNA (gDNA; top) or complementary DNA (cDNA; bottom) made from DNAse treated RNA. Single nucleotide polymorphisms (SNPs) identification numbers are from dbSNP and highlighted with black arrows.

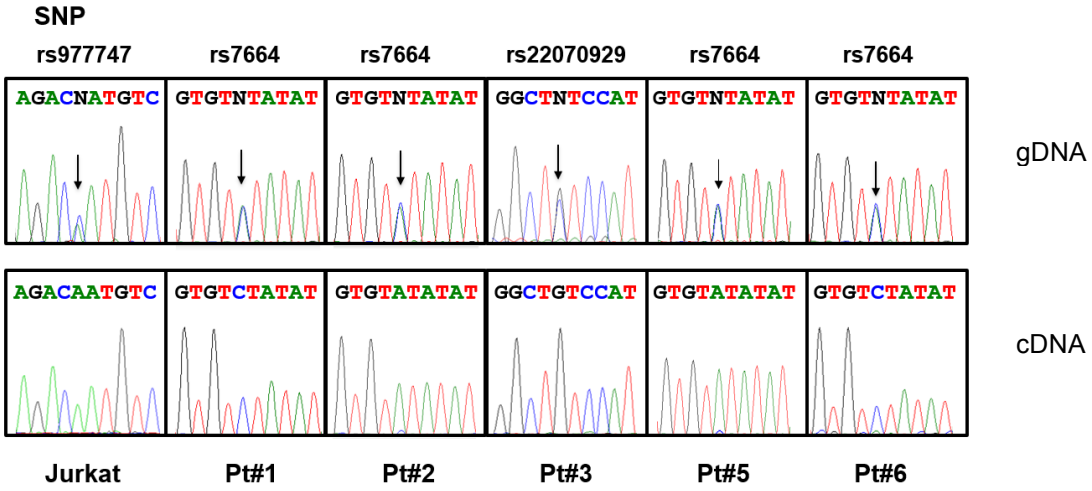


Figure S2. TAL1 Enhancer Mutations activate in T-ALL cells, but not in HEK-293T cells. A 400 bp fragment of the -7.5 kb *TAL1* enhancer containing either the wild-type sequence or each of the mutant alleles was cloned upstream of luciferase and a minimal promoter. Constructs were transfected into HEK-293T or Jurkat cells. Firefly luciferase activity was measured at 48 hrs, normalized to renilla luciferase to control for cell number and transfection efficiency, and expressed as a ratio relative to activity of the wild-type STIL-TAL enhancer construct. Error bars are ±SEM from two independent experiments performed in triplicate.

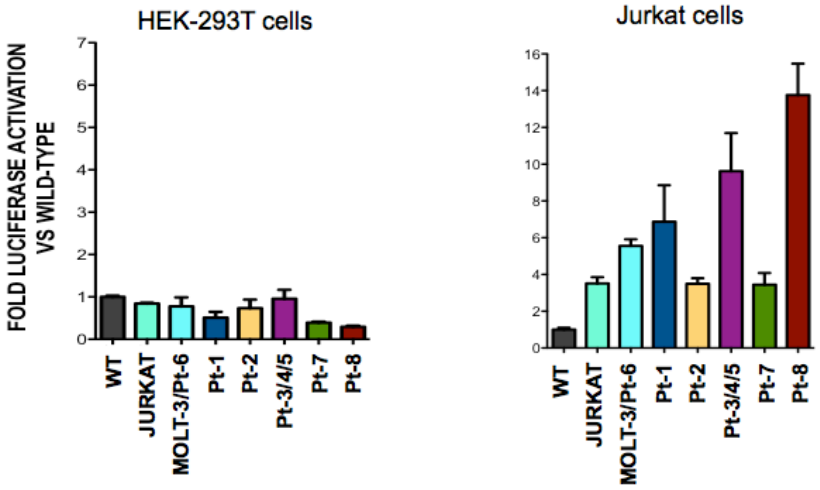


Figure S3. TAL1 does not bind at the -7.5 kb site in CD34+ cells or T-ALL cells harboring the TAL1^{d}. ChIP-seq tracks for TAL1 at the *STIL-TAL1* locus performed in MuTE-positive Jurkat cells compared to two normal CD34+ HSC samples, TAL1^{d}-positive cells from a T-ALL patient derived xenograft (PDX), and TAL1^{d}-positive RPMI-8402 and CCRF-CEM T-ALL cell lines.

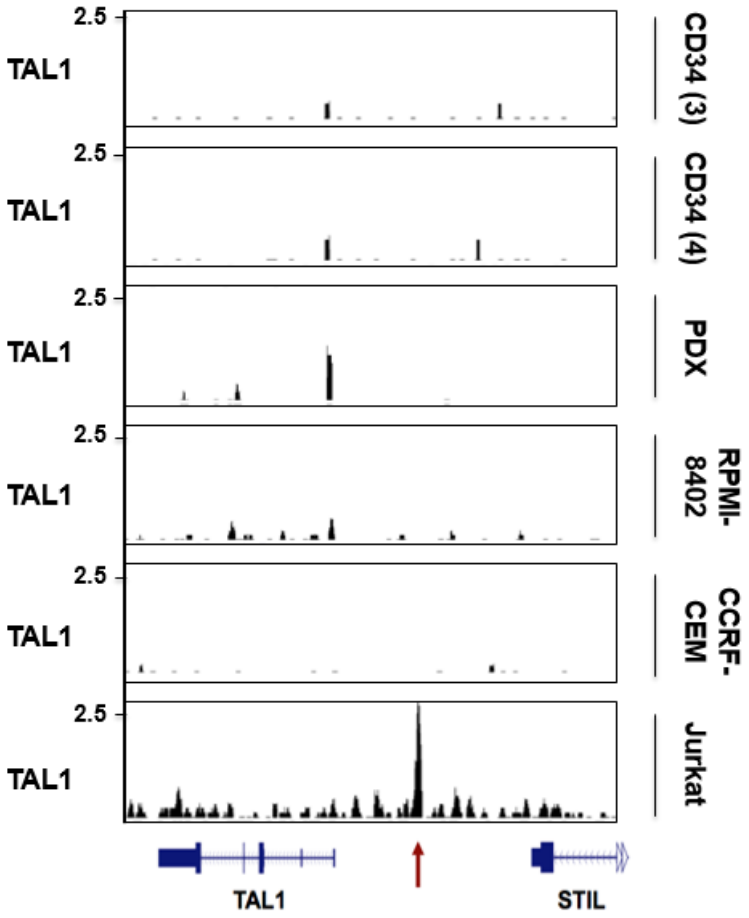


Figure S4. Knockdown of MYB depletes TAL1 expression in MuTE positive T-ALL cells. Jurkat and MOLT-3 cells were nucleofected with MYB or control siRNAs and cells harvested for RNA at 24 hrs. *MYB* and *TAL1* mRNA were measured by q-RT-PCR.

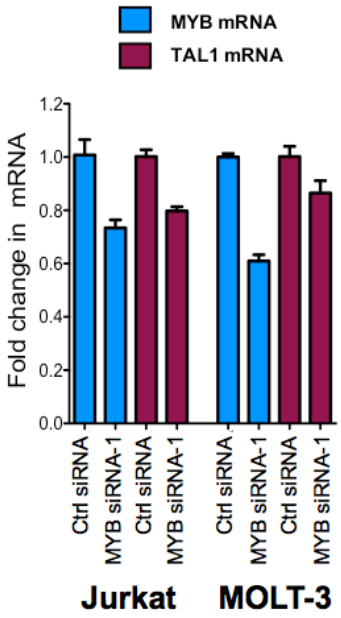
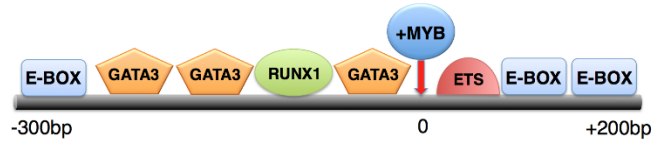


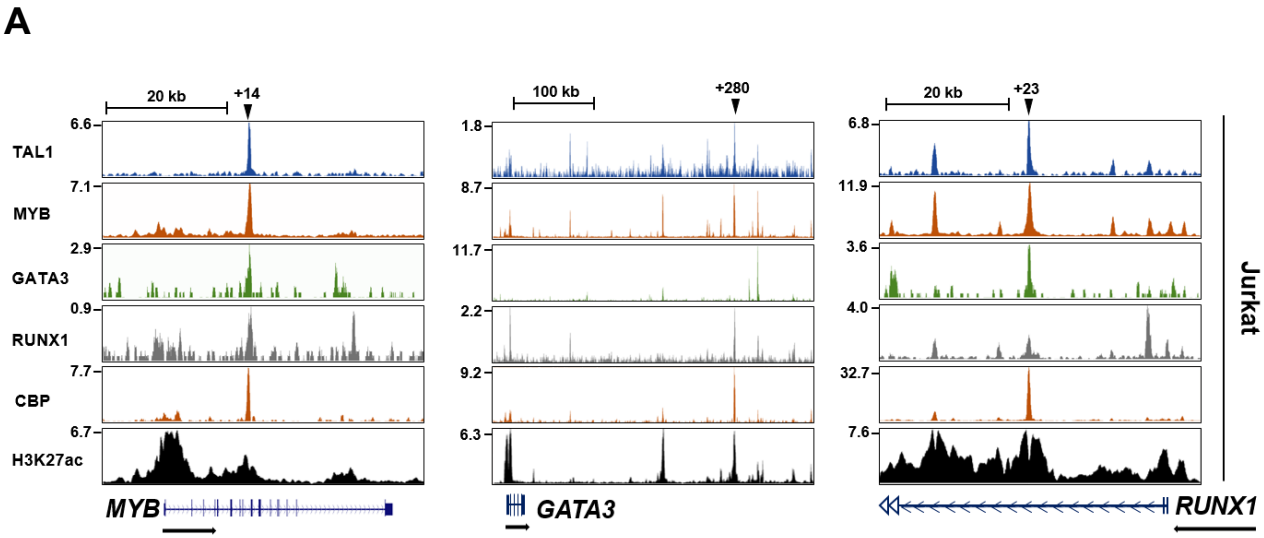
Figure S5. Schematic and motif analysis of a region 7.5 kb from the transcriptional start site of *TAL1* identifying binding sites for members of the TAL1 complex. Locations of DNA sequence motifs preferred by important T-ALL regulators were identified in the TAL1 super-enhancer using UniPROBE and FIMO, with motif libraries from Transfac and HOCOMOCO (*19, 36-38*). The mutation hotspot region is shown underlined in bold. Note no MYB motifs are identified in the wild-type allele.

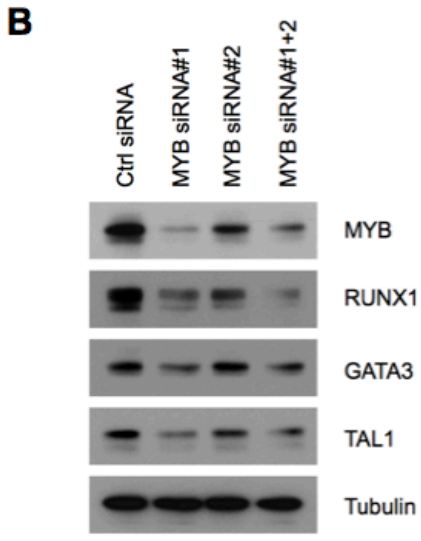


CTGTGCACAGCTGGAGCTCTTATTGACAGCCTGAACAGAGACCAAAAAAGAGGAAGGAGGCAATGATGAGAGGCAGATTAAA
GACAGAGATGGAGAGACAGGGCCGGGGTAGGAAAGAGGAGCAGACTTAGAGACAGAGAGAATGCACATGCGCTTAAAATAGA
GATGGCATGATGAGAAGCAGAGTGAAAGAGATGATAAGAGATAAAAGGGAGAGAGAGATCCTGTGGTTGCCCCACCCCATTC
CTATTACAGATAAACTGAGGGTCACAGAAA**GACGTA**ACCCTACTTCCTGGCAGATGTCTCTGTCACTTTTTCAAATACCATT
AATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGCCCTGTGTGCCTGGCCCTGAGCAGCATGAGTTAGACTGTAACGG
AGGGCTCCAGGGTATGCTAATCAGGAGAGCAAGTGGTGAGAGGAGGAGATATAAATGCAGCTGAGGACA

Highlighted are:
E-BOX motif CAG[N]TG (*9, 43*)
RUNX1 motif TGTGGNNN (HOCOMOCOv9.RUNX1_f1)
GATA3 motifs AGATAA or TGATAA (TRANSFAC_V$GATA3_02.pwm)
ETS motif AGGAA (TRANSFAC_V$CETS168_Q6.pwm)

Figure S6. MYB is a component of a positive interconnected autoregulatory loop through which the TAL1 complex regulates its own enhancers. (A) ChIP-seq tracks for TAL1, MYB (05175), GATA3, RUNX1, CBP and H3K27ac showing binding of the TAL1 complex, including MYB, at the regulatory sites of the *MYB, GATA3* and *RUNX1* genes. **(B)** Western blots for members of the core TAL1 complex members performed 48 hrs after MYB knock-down with siRNA#1 (100nM), the less potent MYB siRNA#2 (100nM), or both MYB siRNAs#1 and #2 (50nM each) (C) Positive interconnected autoregulatory loop depicting a model through which the MYB, TAL1, GATA3 and RUNX1 proteins positively regulate their own enhancers.





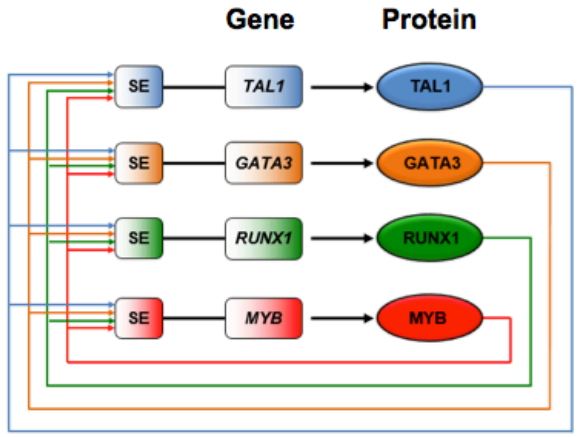


Figure S7. Sequences of CRISPR/Cas induced deletions of the wild-type and mutant TAL1 enhancer site in Jurkat cells A) Wild-type and B) mutant allele sequences are shown, together with CRISPR guide target sites in blue font, with PAM sequences highlighted in pink. The mutation hotspot is shown in bold underlined, and mutant inserted sequences in red. C), D), E), F) and G) are Jurkat clones grown from single cells after CRISPR/Cas targeting. The shaded yellow areas represent deleted sequences.

A) Wild-type allele

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GACGTA**ACCCTACTTCCTGGC
AGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGCCCTGTGTGCCTG
GCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

B) Mutant allele

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GAC**GGTTAGGAAACG**GTA**ACC
CTACTTCCTGGCAGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGC
CCTGTGTGCCTGGCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

C) Clone 1 (55) – biallelic deletion

Allele 1: deletion of either wild-type or mutant (undetermined)

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GACGTA**ACCCTACTTCCTGGC
AGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGCCCTGTGTGCCTG
GCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

Allele 2: deletion of either wild-type or mutant (undetermined)

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GAC**GGTTAGGAAACG**GTA**ACC
CTACTTCCTGGCAGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGC
CCTGTGTGCCTGGCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

D) Clone 2 (60) – deletion of wild-type allele

178bp deletion of wild-type allele

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GACGTA**ACCCTACTTCCTGGC
AGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGCCCTGTGTGCCTG
GCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

E) Clone 3 (83-01) - deletion of mutant allele

190bp deletion of mutant allele

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GAC**GGTTAGGAAACG**GTA**ACC
CTACTTCCTGGCAGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGC
CCTGTGTGCCTGGCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

F) Clone 4 (83-08) - deletion of mutant allele

190bp deletion of mutant allele

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GAC**GGTTAGGAAACG**GTA**ACC
CTACTTCCTGGCAGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGC
CCTGTGTGCCTGGCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

G) Clone 5 (83-16) - deletion of mutant allele

193bp deletion of mutant allele

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GAC**GGTTAGGAAACG**GTA**ACC
CTACTTCCTGGCAGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGC
CCTGTGTGCCTGGCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

25bp deletion at first CRISPR site of wild-type allele

AGAGAGATCCTGTGGTTGCCCCACCCCATTCCTATTACAGATAAACTGAGGGTCACAGAAA**GACGTA**ACCCTACTTCCTGGC
AGATGTCTCTGTCACTTTTTCAAATACCATTAATTCAGTTCCTCCCTTGTCTACAGATCCTTTTTTGTGCCCTGTGTGCCTG
GCCCTGAGCAGCATGAGTTAGACTGTAACGGAGGGCTCCAG

Figure S8. ChIP-seq tracks for MYB performed for CRISPR/Cas9 Jurkat clones harboring deletions of the MuTE site. Zoomed out view of the STIL-TAL1 locus showing MYB enrichment is broadly preserved apart from a lack of MYB binding at the MuTE site in clone#3 that harbors a deletion of the mutant allele.

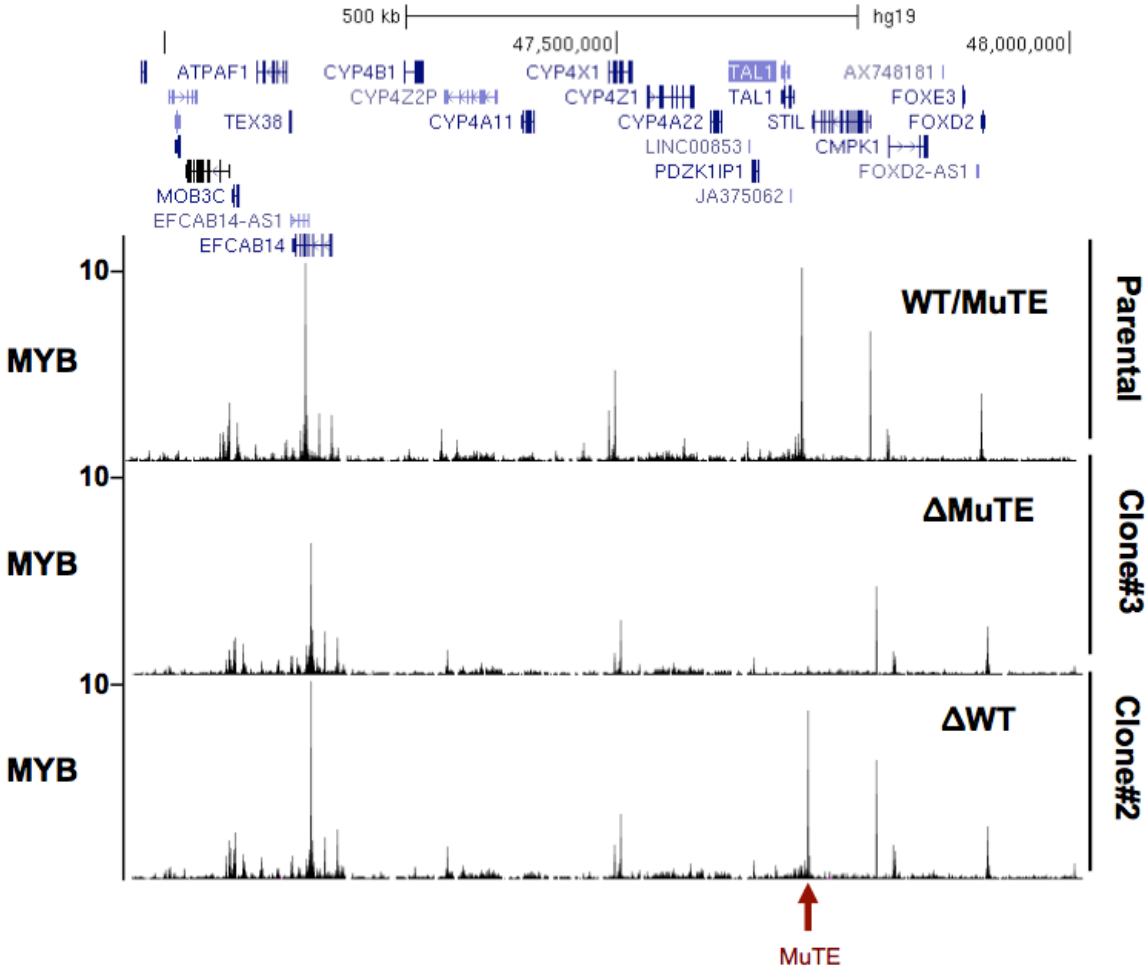
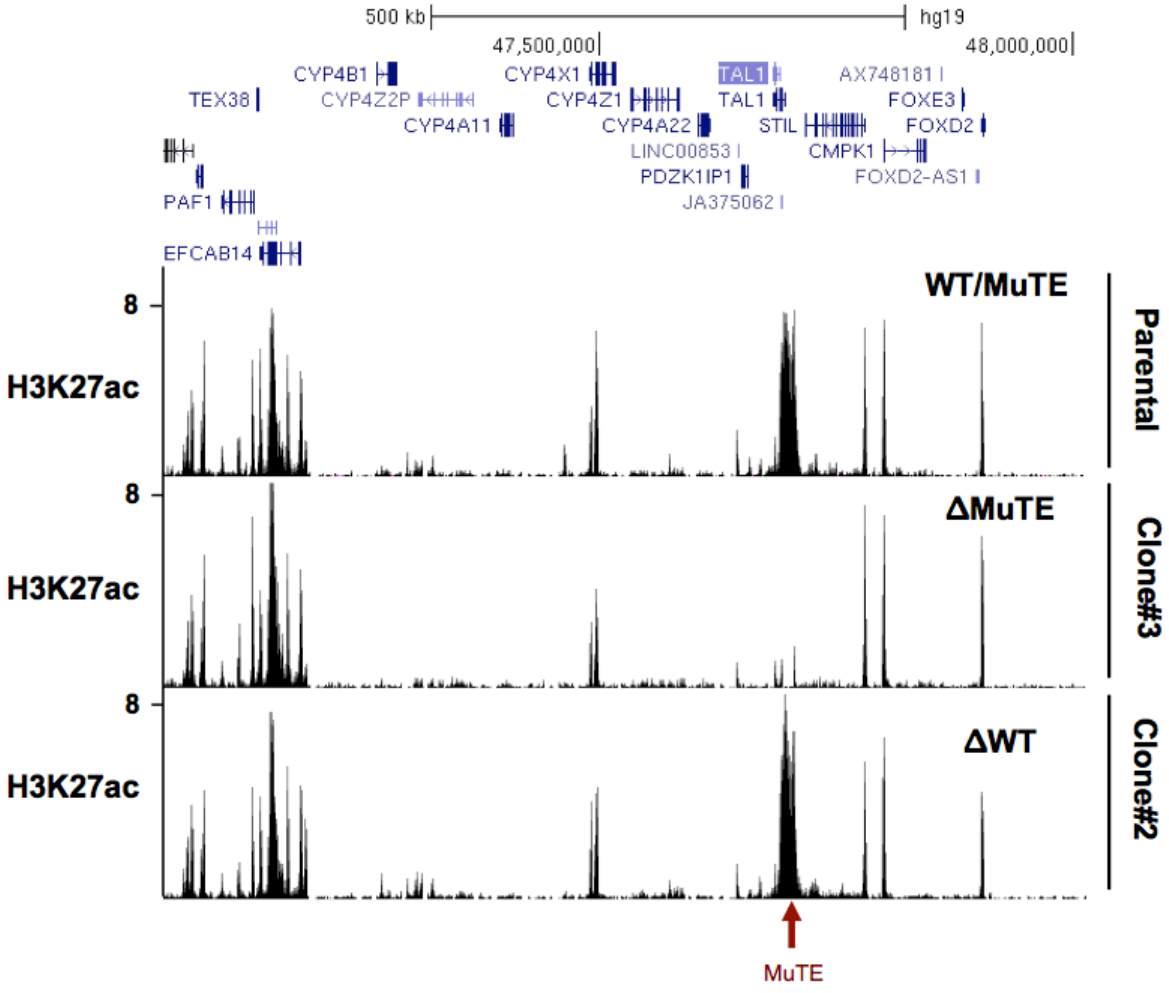


Figure S9. ChIP-seq tracks for H3K27ac performed for CRISPR/Cas9 Jurkat clones harboring deletions of the MuTE site. Zoomed out view of the STIL-TAL1 locus showing disruption of the super-enhancer in clone#3 that lacks the MuTE site, while enhancers elsewhere are broadly preserved in that clone.



Supplemental Tables

Table S1. Mutation status of the TAL1 enhancer in 20 human T-ALL cell lines

Table S2. Transcription factor binding sites of TAL1 enhancer mutant alleles as analyzed by UniPROBE.

Table S3. Table of ChIP-seq datasets with GSE accession number.

Table S1. Mutation status of the TAL1 enhancer in 20 human T-ALL cell lines

T-ALL cell line	TAL1 expression status	TAL1 lesion	Sequencing of TAL1 enhancer
DU.528	TAL1 POS	t(1;14)	WT
HSB2	TAL1 POS		WT
JURKAT	TAL1 POS		12-bp insertion
CCRF-CEM	TAL1 POS	TAL1d	WT
MOLT-3	TAL1 POS		2-bp insertion
MOLT-16	TAL1 POS		WT
PF-382	TAL1 POS		WT
RPMI-8402	TAL1 POS	TAL1d	WT
SKW-3/Ke-37	TAL1 POS		WT
SUP-T13	TAL1 POS	TAL1d	WT
ALL-SIL	TAL1 NEG		WT
CUTLL1	TAL1 NEG		WT
DND-41	TAL1 NEG		WT
HPB-ALL	TAL1 NEG		WT
KOPT-K1	TAL1 NEG		WT
LOUCY	TAL1 NEG		WT
P12-ICHIKAWA	TAL1 NEG		WT
SUP-T1	TAL1 NEG		WT
SUP-T11	TAL1 NEG		WT
TALL1	TAL1 NEG		WT

t(1;14) in DU.528 cell line involves translocation of *TAL1* to TCR delta (*44*).

Table S2. Transcription factor binding sites of *TAL1* enhancer mutant alleles as analyzed by UniPROBE. Mutant inserted sequences are show in red. Wild-type and mutant sequences were analyzed at: http://the_brain.bwh.harvard.edu/uniprobe/index.php

Sequence	Gene Match	K-mer	Reverse Complement	Position	Enrichment Score
Wild-type sequence	Tbf1	GTAACCCT	AGGGTTAC	8	0.489511
AGAAAGACGTAACCCT	Gmeb1	AGACGTAA	TTACGTCT	4	0.456442
	Gmeb1	GACGTAAC	GTTACGTC	5	0.462877
JURKAT	Tbf1	GTAACCCT	AGGGTTAC	20	0.489511
AGAAAGACGGTTAGGAAACGGTAACCCT	Mybl1	AGACGGTT	AACCGTCT	4	0.470742
	Myb	AGACGGTT	AACCGTCT	4	0.477693
	Mybl1	GACGGTTA	TAACCGTC	5	0.48833
	Myb	GACGGTTA	TAACCGTC	5	0.485992
	Mybl1	ACGGTTAG	CTAACCGT	6	0.465502
	Myb	ACGGTTAG	CTAACCGT	6	0.468785
	Irf3	TAGGAAAC	GTTTCCTA	11	0.462063
	Myb	AAACGGTA	TACCGTTT	15	0.451796
	Mybl1	AACGGTAA	TTACCGTT	16	0.470917
	Myb	AACGGTAA	TTACCGTT	16	0.471237
	Rfx4	CGGTAACC	GGTTACCG	18	0.453085
	Rfxdc2	CGGTAACC	GGTTACCG	18	0.467973
	Etv6	AGGAAACG	CGTTTCCT	12	0.45145
	Ehf	AGGAAACG	CGTTTCCT	12	0.4561
	Etv6	GGAAACGG	CCGTTTCC	13	0.45151
MOLT-3 and Patient#6	Tcf2	CGGTTAAC	GTTAACCG	7	0.47813
AGAAAGACGGTTAACCCT	Hmbox1	GGTTAACC	GGTTAACC	8	0.46841
	Tcf2	GGTTAACC	GGTTAACC	8	0.48176
	Tcf2	GTTAACCC	GGGTTAAC	9	0.46975
	Tbf1	TTAACCCT	AGGGTTAA	10	0.485447
	Mybl1	AGACGGTT	AACCGTCT	4	0.470742
	Myb	AGACGGTT	AACCGTCT	4	0.477693
	Mybl1	GACGGTTA	TAACCGTC	5	0.48833
	Myb	GACGGTTA	TAACCGTC	5	0.485992
	Mybl1	ACGGTTAA	TTAACCGT	6	0.473091
	Myb	ACGGTTAA	TTAACCGT	6	0.464134
	Tcf1	CGGTTAAC	GTTAACCG	7	0.47691
	Tcf1	GGTTAACC	GGTTAACC	8	0.476288

	Tcf1	GTTAACCC	GGGTTAAC	9	0.456727
Patient#1	Barhl2	ACCGTTTA	TAAACGGT	6	0.492
AGAAAGACCGTTTAACCCT	Hmx2	ACCGTTTA	TAAACGGT	6	0.48109
	Barhl1	ACCGTTTA	TAAACGGT	6	0.49542
	Barhl2	CCGTTTAA	TTAAACGG	7	0.48433
	Hmx2	CCGTTTAA	TTAAACGG	7	0.48387
	Barhl1	CCGTTTAA	TTAAACGG	7	0.48604
	Barhl2	CGTTTAAC	GTTAAACG	8	0.47988
	Hmx2	CGTTTAAC	GTTAAACG	8	0.46721
	Barhl1	CGTTTAAC	GTTAAACG	8	0.48348
	Tbf1	TTAACCCT	AGGGTTAA	11	0.485447
	Mybl1	AGACCGTT	AACGGTCT	4	0.461165
	Myb	AGACCGTT	AACGGTCT	4	0.460314
	Mybl1	GACCGTTT	AAACGGTC	5	0.469869
	Myb	GACCGTTT	AAACGGTC	5	0.471101
	Mybl1	ACCGTTTA	TAAACGGT	6	0.461873
	Myb	ACCGTTTA	TAAACGGT	6	0.462745
	Rxra	TTTAACCC	GGGTTAAA	10	0.453132
Patient#2		GTTAACAG	CTGTTAAC	9	0.46743
AAAGACGCCGTTAACAGACGGTAAACTACTT	Tcf2	GTTAACAG	CTGTTAAC	9	0.45382
	Nkx2-6	AACTACTT	AAGTAGTT	23	0.45054
	Stp2	GACGCCGT	ACGGCGTC	3	0.470979
	Mybl1	GCCGTTAA	TTAACGGC	6	0.472739
	Myb	GCCGTTAA	TTAACGGC	6	0.467223
	Mybl1	CCGTTAAC	GTTAACGG	7	0.469093
	Myb	CCGTTAAC	GTTAACGG	7	0.465776
	Tcf1	GTTAACAG	CTGTTAAC	9	0.465737
	Smad3	CAGACGGT	ACCGTCTG	14	0.45145
	Nkx3-1	AACTACTT	AAGTAGTT	23	0.450517
Patient #3, #4 and #5	Tcf2	CGTTAACC	GGTTAACG	7	0.46472
GAAAGACCGTTAACCCT	Tcf2	GTTAACCC	GGGTTAAC	8	0.46975
	Tbf1	TTAACCCT	AGGGTTAA	9	0.485447
	Mybl1	AGACCGTT	AACGGTCT	3	0.461165
	Myb	AGACCGTT	AACGGTCT	3	0.460314
	Mybl1	GACCGTTA	TAACGGTC	4	0.487305
	Myb	GACCGTTA	TAACGGTC	4	0.482084

	Mybl1	ACCGTTAA	TTAACGGT	5	0.487881
	Myb	ACCGTTAA	TTAACGGT	5	0.485773
	Mybl1	CCGTTAAC	GTTAACGG	6	0.469093
	Myb	CCGTTAAC	GTTAACGG	6	0.465776
	Tcf1	CGTTAACC	GGTTAACG	7	0.457345
	Tcf1	GTTAACCC	GGGTTAAC	8	0.456727
Patient#7	Hmbox1	TTACCAGT	ACTGGTAA	10	0.45989
GAAAGACGGTTACCAGTTTGAAACCCT	Nrg1	GAAACCCT	AGGGTTTC	20	0.465692
	Tbf1	GAAACCCT	AGGGTTTC	20	0.489162
	Mybl1	AGACGGTT	AACCGTCT	4	0.470742
	Myb	AGACGGTT	AACCGTCT	4	0.477693
	Mybl1	GACGGTTA	TAACCGTC	5	0.48833
	Myb	GACGGTTA	TAACCGTC	5	0.485992
	Mybl1	ACGGTTAC	GTAACCGT	6	0.46042
	Myb	ACGGTTAC	GTAACCGT	6	0.453073
	Rfxdc2	CGGTTACC	GGTAACCG	7	0.464773
	Rfx4	GGTTACCA	TGGTAACC	8	0.481666
	Rfxdc2	GGTTACCA	TGGTAACC	8	0.482793
	Rfx3	GGTTACCA	TGGTAACC	8	0.45019
	Rfx4	GTTACCAG	CTGGTAAC	9	0.463289
	Rfxdc2	GTTACCAG	CTGGTAAC	9	0.452098
Patient#8	Tbf1	TTAACCCT	AGGGTTAA	10	0.485447
GAAAGACGGTTTAACCCT	Tbf1	TAACCCTA	TAGGGTTA	11	0.496212
	Tbf1	AACCCTAC	GTAGGGTT	12	0.493967
	Tbf1	ACCCTACT	AGTAGGGT	13	0.467558
	Mybl1	AGACGGTT	AACCGTCT	3	0.470742
	Myb	AGACGGTT	AACCGTCT	3	0.477693
	Mybl1	GACGGTTT	AAACCGTC	4	0.468374
	Myb	GACGGTTT	AAACCGTC	4	0.477396
	Mybl1	ACGGTTTA	TAAACCGT	5	0.458226
	Myb	ACGGTTTA	TAAACCGT	5	0.470212
	Rxra	TTTAACCC	GGGTTAAA	9	0.453132

Table S3. ChIP-seq data accession numbers for relevant datasets

Sample	ChIP	GEO	REFERENCE
CD34 (1)	H3K27ac	GSM772885	Roadmap
CD34 (2)	H3K27ac	GSM772894	Roadmap
CD34 (3)	TAL1	GSM1442001	
CD34 (4)	TAL1	GSM1442002	
RPMI-8402	H3K27ac	GSM1442003	
DND-41	H3K27ac	GSM1003462	ENCODE/Broad
Fetal thymus	H3K27ac	GSM1036258	Roadmap
Jurkat	H3K27ac	GSM1296384	Kwiatkowski et al. (*17*)
Jurkat	GATA3	GSM722168	Sanda et al. (*9*)
Jurkat	HEB	GSM722166	Sanda et al. (*9*)
Jurkat	RUNX1	GSM722170 GSM722169 GSM722171	Sanda et al. (*9*)
Jurkat	TAL1	GSM722165 GSM722164	Sanda et al. (*9*)
Jurkat	CBP	GSM449527	Sanda et al. (*9*)
Jurkat	RNA POL II	GSM1224784	Kwiatkowski et al. (*17*)
Jurkat	MED1	GSM1442004	
Jurkat	MYB (45150)	GSM1442006	
Jurkat	MYB (05175)	GSM1442005	
T-ALL PDX	TAL1	GSM838000	Sanda et al. (*9*)
CCRF-CEM	TAL1	GSM837996	Sanda et al. (*9*)
MOLT-3	TAL1	GSM1519646	
MOLT-3	MYB	GSM1519643	
MOLT-3	H3K27ac	GSM1519644	

Roadmap data is from NIH Roadmap Epigenomics Mapping Consortium
http://www.roadmapepigenomics.org/

REFERENCES

25. A. Gutierrez *et al.*, Inactivation of LEF1 in T-cell acute lymphoblastic leukemia. *Blood* **115**, 2845 (Apr 8, 2010).
26. A. Gutierrez *et al.*, Absence of biallelic TCRgamma deletion predicts early treatment failure in pediatric T-cell acute lymphoblastic leukemia. *J Clin Oncol* **28**, 3816 (Aug 20, 2010).
27. A. Gutierrez *et al.*, High frequency of PTEN, PI3K, and AKT abnormalities in T-cell acute lymphoblastic leukemia. *Blood* **114**, 647 (Jul 16, 2009).
28. A. A. Ferrando *et al.*, Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75 (Feb, 2002).
29. I. Lahortiga *et al.*, Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. *Nat Genet* **39**, 593 (May, 2007).
30. S. Jin *et al.*, c-Myb binds MLL through menin in human leukemia cells and is an important driver of MLL-associated leukemogenesis. *J Clin Invest* **120**, 593 (Feb, 2010).
31. A. Marson *et al.*, Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521 (Aug 8, 2008).
32. T. I. Lee, S. E. Johnstone, R. A. Young, Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nature protocols* **1**, 729 (2006).
33. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
34. W. J. Kent *et al.*, The human genome browser at UCSC. *Genome Res* **12**, 996 (Jun, 2002).
35. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
36. V. Matys *et al.*, TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34**, D108 (Jan 1, 2006).
37. I. V. Kulakovskiy *et al.*, HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res* **41**, D195 (Jan, 2013).
38. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017 (Apr 1, 2011).
39. P. D. Hsu *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827 (Sep, 2013).
40. M. C. Canver *et al.*, Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of biological chemistry* **289**, 21312 (Aug 1, 2014).
41. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819 (Feb 15, 2013).
42. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281 (Nov, 2013).

43. C. G. Palii *et al.*, Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *The EMBO journal* **30**, 494 (Feb 2, 2011).
44. O. Bernard *et al.*, Two distinct mechanisms for the SCL gene activation in the t(1;14) translocation of T-cell leukemias. *Genes, chromosomes & cancer* **1**, 194 (Jan, 1990).