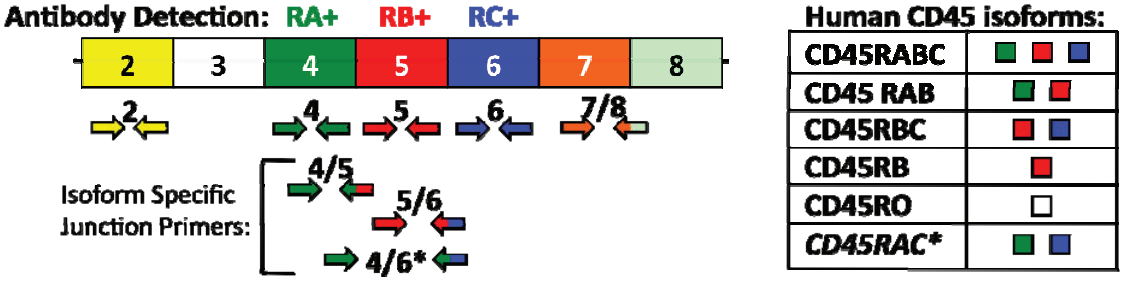
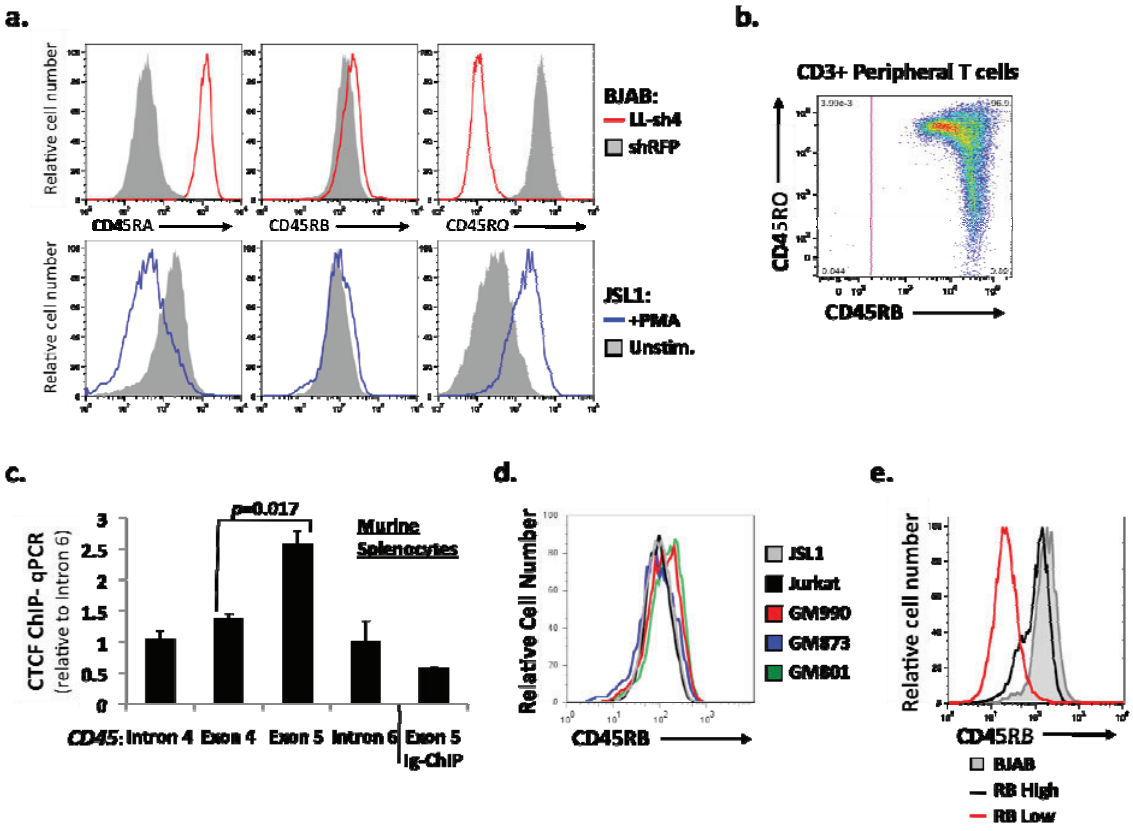
doi:10.1038/nature10442

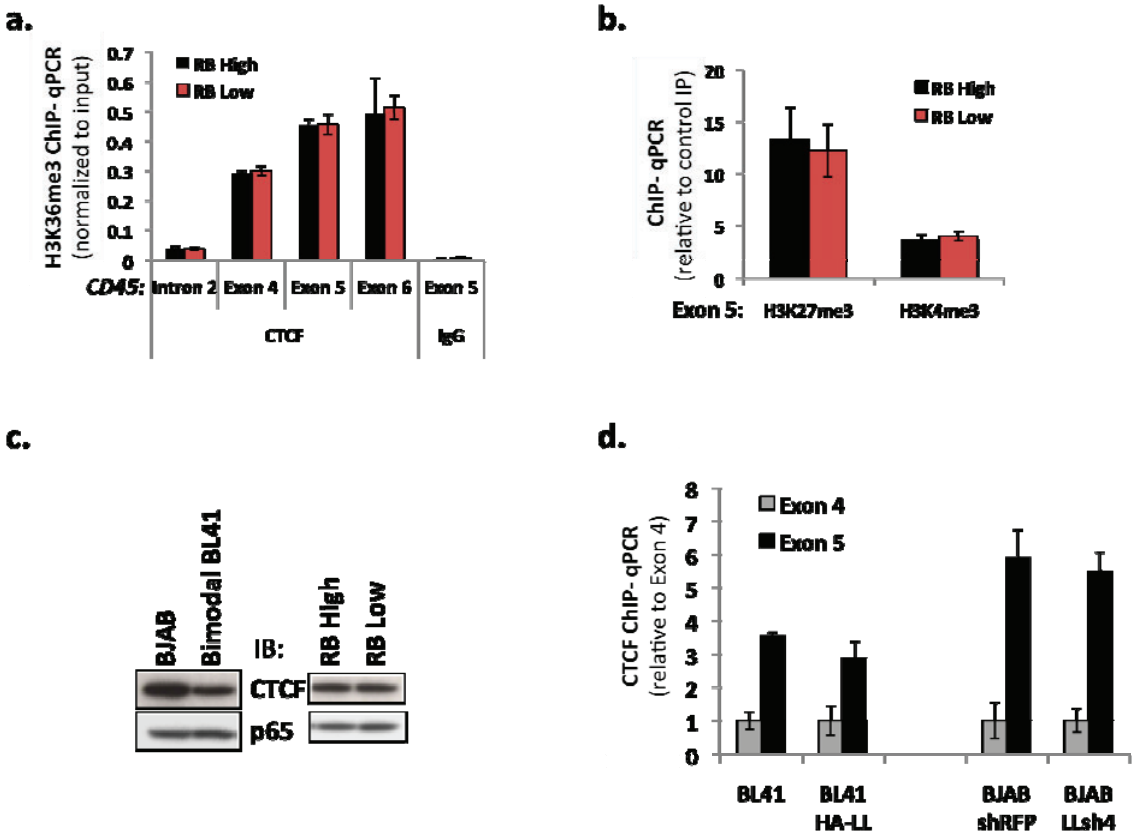
Supplementary Figures and Legends



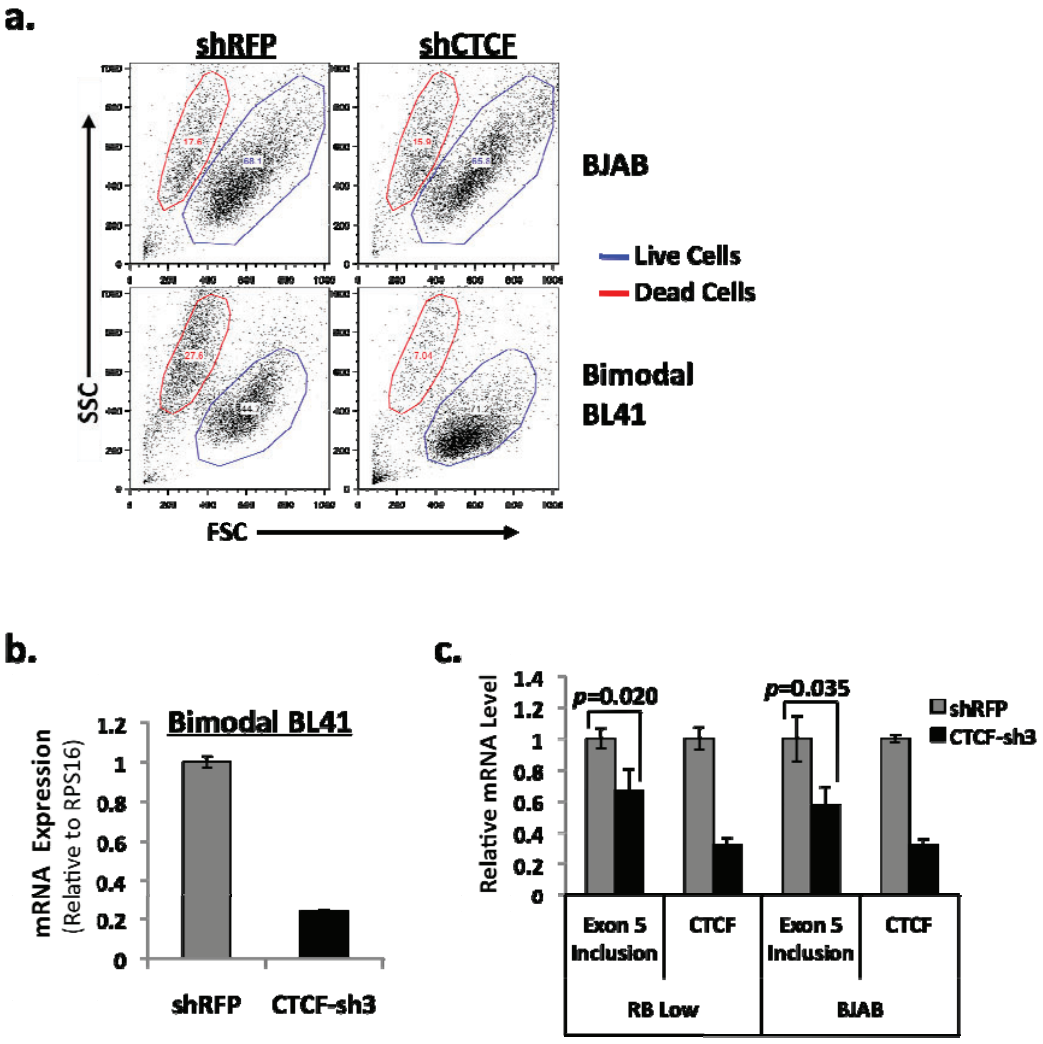
Supplementary Figure 1. Schematic of CD45 isoforms found in human cells and methods of detection. Exons 4-6 of *CD45* transcripts are variably lost during lymphocyte development to produce the top five isoforms shown in the table. Isoforms are named according to which of the variable exons are included, with the exception of CD45RO, in which all three variable exons are excluded. Isoform-specific antibodies to cell-surface CD45 are directed against the indicated variable exon, and cannot faithfully distinguish the various isoforms including the target exon. To complement flow cytometric analysis of cell-surface CD45, quantitative RT-PCR was utilized to quantify relative changes in exon inclusion. In addition to primer sets directed against the exon as a whole, exon junction spanning primer sets were used to specifically measure spliced products, as opposed to unspliced precursor, as indicated above. * The CD45RAC isoform is not thought to be produced in human cells, but emerges following aberrant loss of CTCF binding, as demonstrated in Figures 1 and 2.



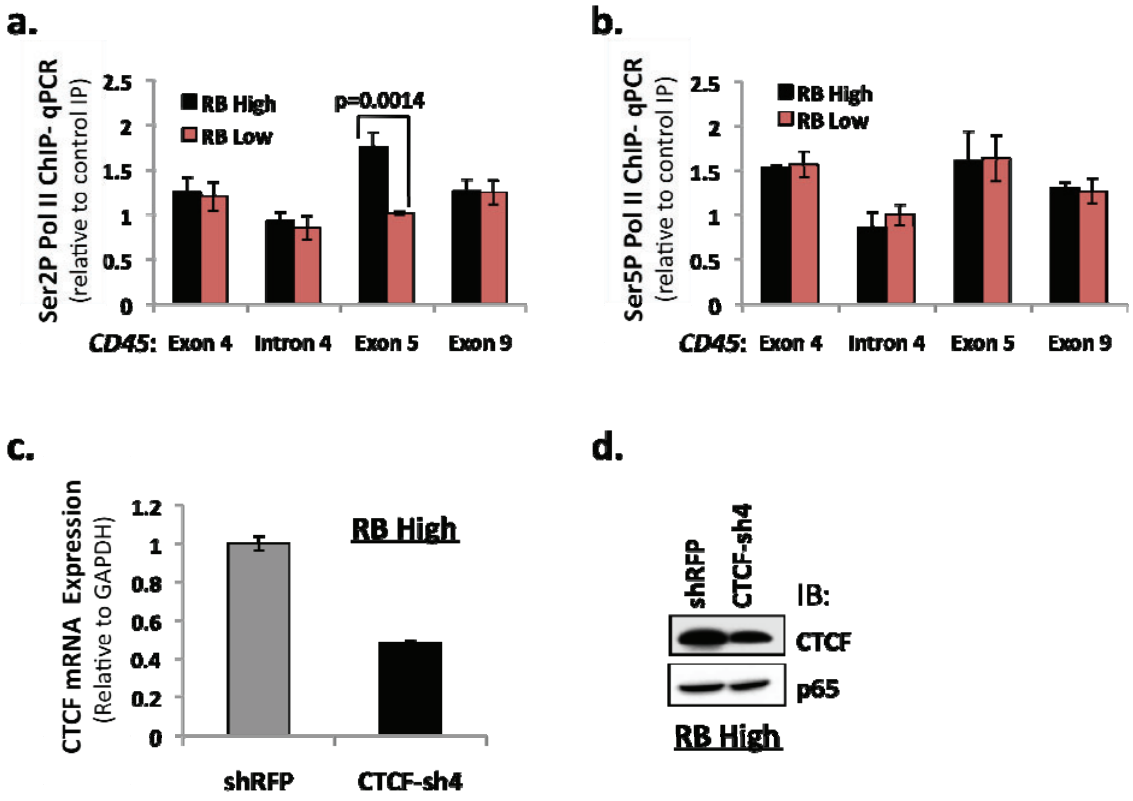
Supplementary Figure 2. (a) Cell-surface staining of CD45 isoforms in BJAB B and JSL1 T cells. shRNA-mediated depletion of hnRNPLL (LL-sh4) from BJAB cells reduces CD45RO, increases CD45RA and has no effect on CD45RB, as compared to control-transduced cells (shRFP). PMA-induced hnRNPLL expression in JSL1 cells increases CD45RO, decreases CD45RA and does not affect CD45RB expression. **(b)** CD3+ T cells purified from human peripheral blood express high cell-surface CD45RB and variable CD45RO. Naïve T cells are $CD45RB^{high}/CD45RO^{low}$ and transition to $CD45RB^{low}/CD45RO^{high}$ following activation. **(c)** CTCF ChIP in murine splenocytes (distinct mouse from Fig. 1a) and qPCR to detect *CD45* exons shows CTCF binding at exon 5, and not at surrounding regions. Values represent means +/- SD from two independent immunoprecipitation experiments. *p*= two-tailed Student's *t* test comparing exon 5 values to exon 4 values. **(d)** Cell-surface CD45RB expression is uniformly high in tested T (JSL1, Jurkat) and B (GM) cell lines, as indicated. **(e)** Cell-surface staining for CD45RB in BJAB cells and stably maintained RB^{high} and RB^{low} BL41 cells.



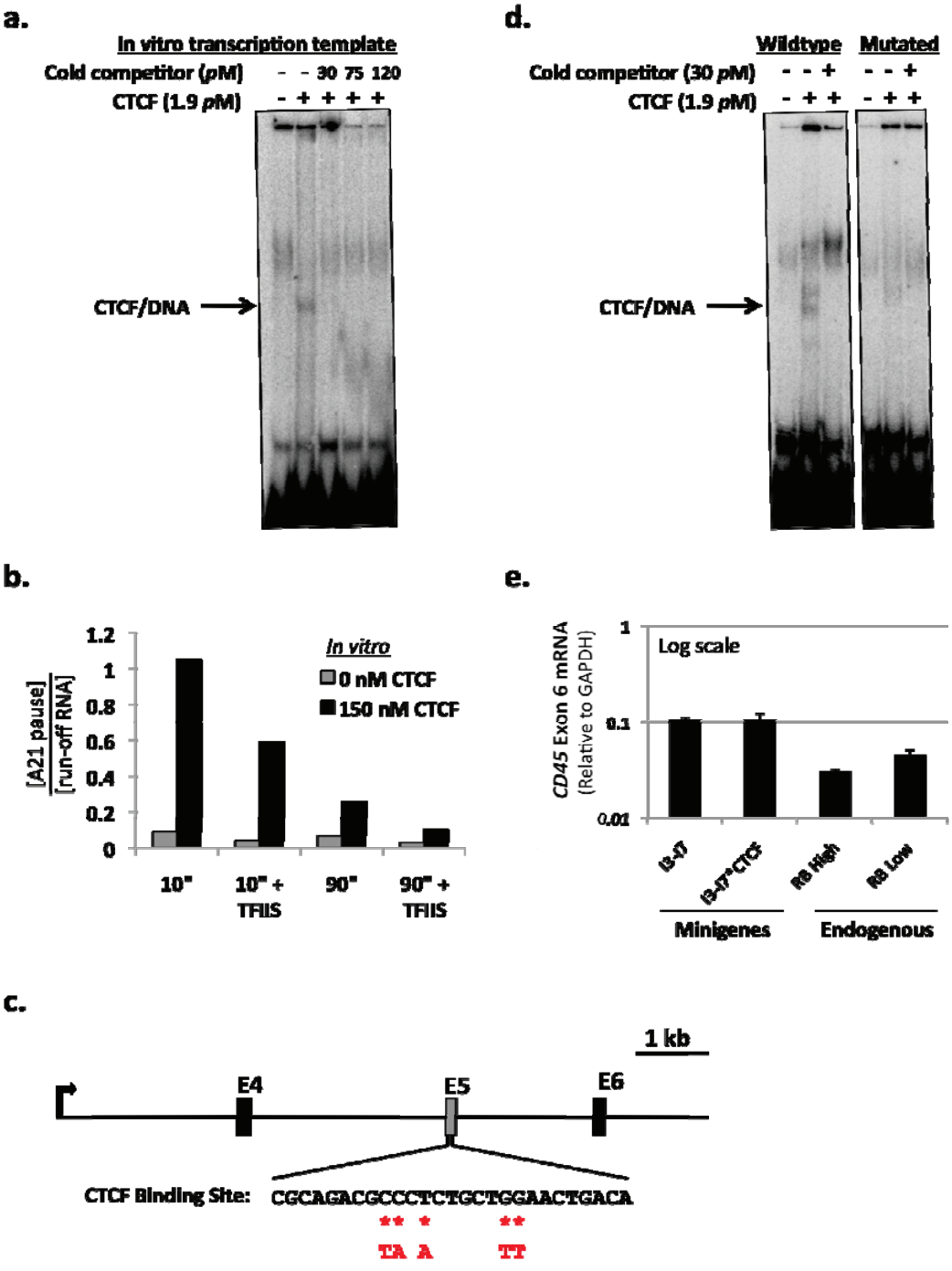
Supplementary Figure 3. ChIP with antibody specific for **(a)** H3K36me3 and **(b)** H3K4me3 and H3K27me3 in RB^{high} and RB^{low} BL41 cells and qPCR to detect *CD45* exons. H3K36me3, H3K27me3, and H3K4me3 levels are comparable at exon 5 in RB^{high} and RB^{low} cells. **(c)** Western Blot for CTCF expression in BJAB, bimodal BL41, RB^{high} and RB^{low} cells. Anti-p65 immunoblot serves as a loading control. **(d)** CTCF-ChIP in parental BL41 (RB^{high}) and BJAB cells with stable expression (BL41, HA-LL) or depletion (BJAB, LL-sh4) of hnRNPLL versus control-transduced cells. Quantitative PCR of *CD45* exons indicates that hnRNPLL expression does not impact on CTCF binding at exon 5. For all graphs, ChIP values represent means +/- SD from three independent immunoprecipitation experiments.



Supplementary Figure 4. (a) Forward scatter and side scatter plot of BJAB and bimodal BL41 cells transduced with shRNA directed against CTCF and RFP. Although infection resulted in some cell-death in both BJAB and BL41 cells, CTCF depletion did not increase cell-death as compared to control-transduced cells. **(b)** Quantitative RT-PCR analysis of *CTCF* levels in shCTCF and shRFP-transduced bimodal BL41 cells. These samples were used for the RNA-seq analysis presented in Figure 5 and Supplementary Figs. 8 and 9. **(c)** qRT-PCR data from three independently performed CTCF-depletions in RB^{low} (left) and BJAB cells (right) to detect exon 5 inclusion as a ratio of 5/6 to 4/6 junctions, or CTCF knockdown efficiency relative to several housekeeping genes (*GAPDH*, *aldolase*, *L32* and *beta-actin*).

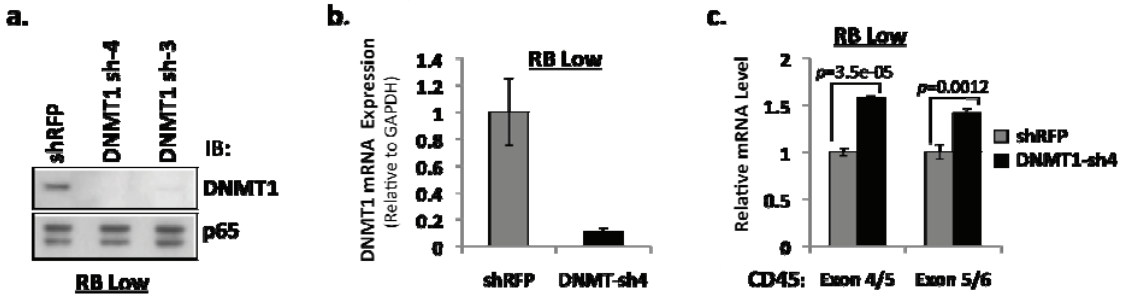


Supplementary Figure 5. ChIP with antibody specific for **(a)** pol II CTD phosphorylated on serine 2 (Ser2P) or **(b)** on serine 5 (Ser5P) in RB^{high} and RB^{low} cells and quantitative PCR detection of *CD45* DNA relative to mouse Ig control ChIP. ChIP values represent means +/- SD from three independent experiments. *p*= two-tailed Student's *t* test comparing RB^{high} to RB^{low} values. **(c)** Quantitative RT-PCR in CTCF depleted RB^{high} cells relative to shRFP-transduced cells from Fig. 3b, c. **(d)** Western blot for CTCF expression in cells from (c), and p65 immunoblotting as a loading control**.** CTCF depletion leads to significant cell death in RB^{high} cells. To achieve sufficient cell numbers for ChIP experiments, transductions were performed in individual wells of a 96 well plate and pooled prior to further analysis.

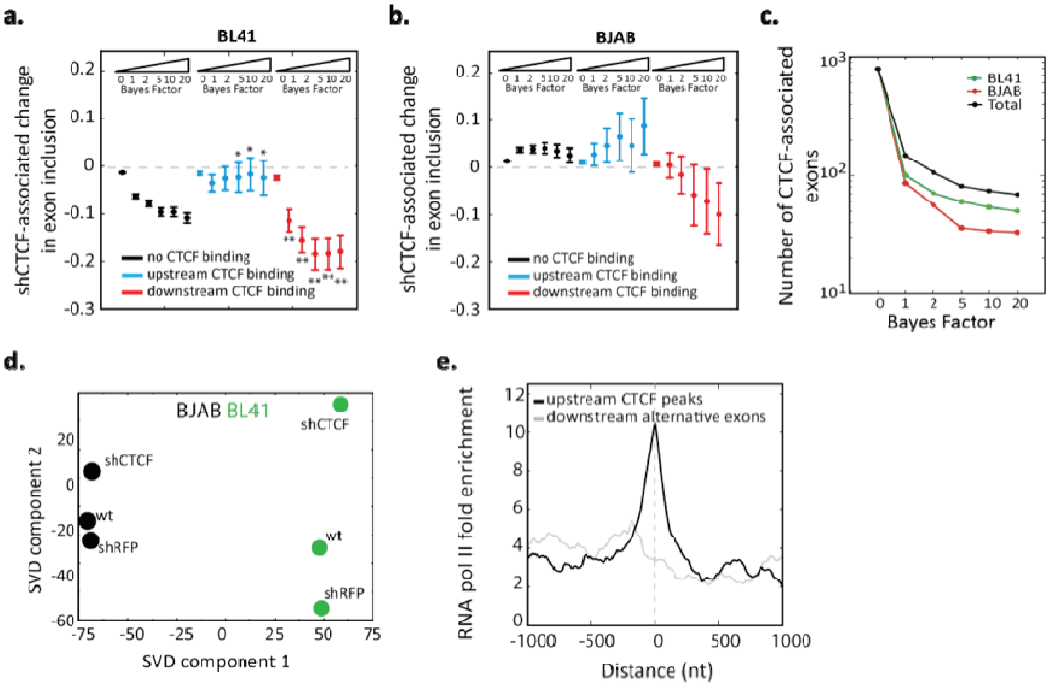


Supplementary Figure 6. (a) EMSA confirms CTCF binding to the DNA oligonucleotide used in the *in vitro* transcription assays (Fig. 3d). The arrow indicates the location of the DNA/protein complex. Incubation in the presence of unlabeled oligo effectively competed away formation of the labeled complex, indicating specific binding. **(b)** Quantification of pol II paused at A21 versus

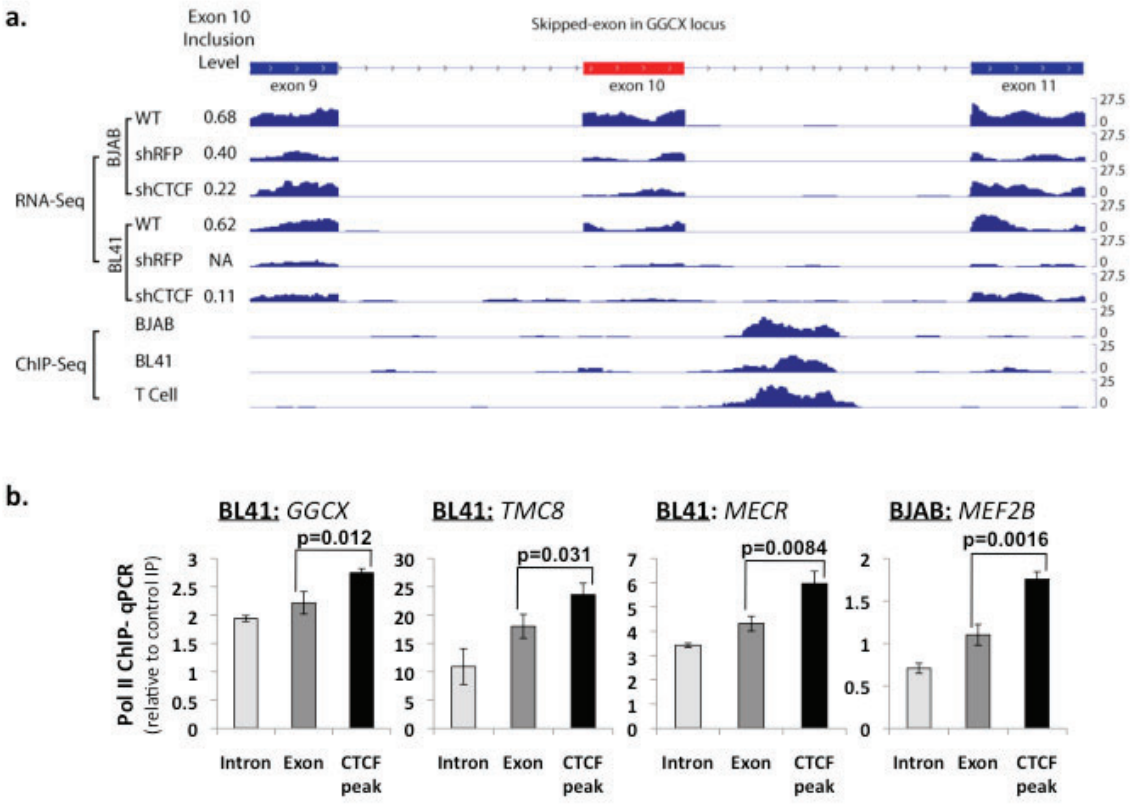
run-off accumulation from Fig. 3d. **(c)** Description of nucleotides mutated to disrupt CTCF binding to exon 5 of the *CD45* minigene (Fig. 3f-j). A total of 5 substitutions were made within the binding core. **(d)** EMSA with a 25 base pair oligo encompassing the *CD45* exon 5 CTCF binding site (wildtype), as well as an oligo in which 5 nucleotides were mutated as shown in (c). In either case, incubation in the presence of unlabeled oligo competes away formation of the labeled complex, indicating specific binding. **(e)** Quantitative RT-PCR to detect expression of the *CD45* minigene in HEK293 cells as a function of exon 6 expression relative to *GAPDH*. CD45 is highly transcribed in immune cells and levels derived from the *CD45* minigene constructs were within the physiological range, as shown for BL41 RB^{high} and RB^{low} cells.



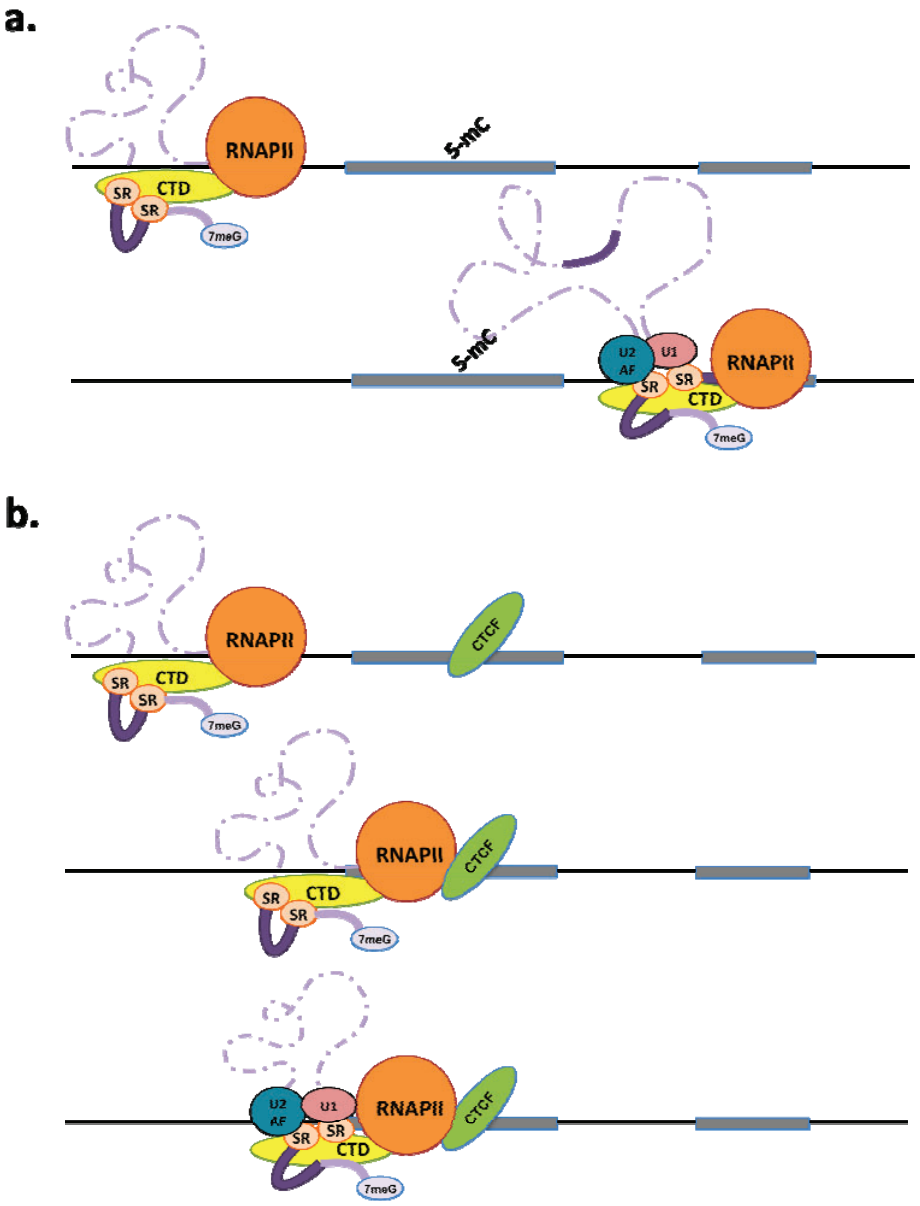
Supplementary Figure 7. (a) Western blot analysis for DNMT1 in RB^{low} cells transduced with shRNA directed against DNMT1 or control shRNA against RFP (from Fig. 4e-h). p65 immunoblotting served as a loading control. **(b)** Quantitative RT-PCR detection of *DNMT1* transcript levels relative to *GAPDH* in RB^{low} cells from (a). **(c)** Quantitative RT-PCR in cells from (a) to detect Exon 4/5 and Exon 5/6 junctions relative to Exon 6. In order to achieve sufficient cell numbers for ChIP following DNMT1 depletion, lentiviral transductions were performed in individual wells of a 96 well plate and pooled prior to assay. Values represent mean +/- SD from three independent RNA preparations from the pooled, bulk, polyclonal culture.



Supplementary Figure 8. Global identification of CTCF-dependent exons. **(a)** Difference in mean exon inclusion level between shCTCF transduced bimodal BL41 cells (from Fig. 2a) and untransduced control cells for exons with CTCF peak in upstream (blue) or downstream (red) regions but not in the exon body and for exons with no CTCF binding (black). The mean and standard error of the mean for each class of exons are plotted against increasing Bayes Factor thresholds. Statistically significant differences in exon inclusion levels between CTCF bound and unbound exons at different thresholds are indicated *(p<0.05), ** (p<0.01), ***(p<0.001), wilcoxon rank sum test. The bias towards reduced exon inclusion for unbound exons in the comparison to untransduced cells, likely reflects alternative splicing in BL41 cells in response to lentiviral transduction. BL41 cells are particularly sensitive to transduction as compared to BJAB cells (Supplementary Fig. 4a). This bias was eliminated in the appropriate controlled comparison between shCTCF and shRFP transduced BL41 cells (Fig. 5b). **(b)** Same as (a) for BJAB shCTCF compared with shRFP control cells (from Fig. 2a). **(c)** Number of exons with downstream CTCF binding plotted against increasing Bayes Factor for BL41 (Figure 5b), BJAB (Figure 5c) and the union of exons. **(d)** Singular value decomposition analysis of gene expression levels in BJAB and BL41 for shCTCF, shRFP and wildtype cells. Samples are plotted against the two most informative components; where the first one separates BJAB cells from BL41 cells and the second component distinguishes shCTCF samples from control cells. **(e)** Same as in Figure 5d for CTCF peaks upstream of alternative exons and the corresponding downstream alternative exons.



Supplementary Figure 9. (a) RNA-seq and CTCF-ChIP-seq in BJAB and BL41 cells at *GGCX* exon 10. CTCF ChIP-seq shows the location of a peak downstream of exon 10. Analysis of RNA-seq data shows a reduction in the relative inclusion of exon 10 in response to CTCF depletion. BL41 shRFP exon 10 inclusion levels could not be calculated due to insufficient total reads. **(b)** Quantitative PCR of pol II ChIP DNA from BJAB and BL41 cells at the location of the CTCF ChIP-seq peak (Supplementary Table 4), alternative upstream exon and control upstream intron. ChIP values represent means +/- SD from three independent IP experiments. *p*= two-tailed Student's *t* test comparing peak values to upstream exon values.



Supplementary Figure 10. Model for CTCF-mediated regulation of CD45 exon 5 alternative splicing. **(a)** The presence of methylated DNA at *CD45* exon 5 inhibits CTCF binding. In the absence of CTCF-mediated pol II pausing, exon 5 is excluded from spliced CD45 transcripts. **(b)** CTCF binds to and promotes local pol II pausing at *CD45* exon 5 when the target DNA is not methylated on cytosine. Pausing provides favorable conditions for spliceosome assembly prior to reactivation of elongation.

Supplementary Tables

Supplementary Table 1. Primer sequences utilized for qPCR and mutagenesis.

Primer Name	Sequence (5'-3')
CD45 In2 Fw	GCT ATG GTG CCA CCT ACT GAA A
CD45 In2 Rev	ACA GTT ACT ACA TTC TAC ACT TTG ACT
CD45 ex4 Fw	GCA AAG ATG CCC AGT GTT CCA CTT
CD45 ex4 Rev	TTC TCT TTC AAA GGT GCT TGC GGG
CD45 In4 Fw	AAC AGA TTC AGT TAC TTC ACA GTT TG
CD45 In4 Rev	CCA CTG GAG ATG AAT AAT TTA CTA AG
CD45 ex5 Fw	TCA TCA GTA CAG ACG CCT CAC CTT
CD45 ex5 Rev	CTG AAT GTC TGC GTG TCA GTT CCA
CD45 In5 Fw	AGC TCT TGA ATA GAG GGC AGC TCA
CD45 In5 Rev	ACA CCT TTG CCC AAC ATC ACA TCC
CD45 ex6 Fw	AGC ACC TTT CCT ACA GAC CCA GTT
CD45 ex6 Rev	TGT TCG CTG TGA TGG TGG TGT T
CD45 In6 Fw	TGG AAA GCC GTG AAT GAA AGG G
CD45 In6 Rev	CTT TGC TCT AGG AGT CCT CGT TTG
CD45 Ex9 Fw	GCACAAACAATGAGGTGCATAACC
CD45 Ex9 Rev	ATGTCTTATCAGGAGCAGTACATGA
CD45-Ex4 FW	CCC GCA AGC ACC TTT GAA AGA GAA
CD45-Ex4/5 Rev	TGT ACT GAT GAA ACA CCT GTG GTA
CD45-Ex5 FW	TTC ATC AGT ACA GAC GCC TCA CCT
CD45-Ex5/6 Rev	TCT CCT GGG ACA TCT GAG ATA GCA
CD45- Ex4 FW	GCA AAG ATG CCC AGT GTT CCA CTT
CD45- Ex4/6 Rev	CTC TCC TGG GAC ATC TGT GGT ATT
CTCF ex2 Fw	ACG GAC CTG AAG CCA AAG AAC AAG
CTCF ex2 Rev	GTG GCT GCA AAG TCA GTT TGA GGA
RPS16 SET5 Fw	AAA CGC GGC AAT GGT CTC ATC AAG
RPS16 SET5 Rev	TGG AGA TGG ACT GAC GGA TAG CAT
GAPDH Fw	TCG ACA GTC AGC CGC ATC TTC TTT
GAPDH Rev	ACC AAA TCC GTT GAC TCC GAC CTT
DNMT1 Ex 4 Fw	TGC TTA CAA CCG GGA AGT GAA TGG
DNMT1 Ex 4 Rev	TTG GCA TCT GCC ATT CCC ACT CTA
GGCX CTCF peak Fw	TTG TGG GCA TAG CTG GTT GCT TC
GGCX CTCF peak Rev	GGG TTA GTA GCT ACC TGA GAA GG
GGCX Ex 10 Fw	AAG CAA TAT GCC ACT TGC CTG AGC
GGCX Ex 10 Rev	GCG GTC ATT GAT GGA GAC CCA AAT
GGCX In 8 Fw	AGGCTGGACTTAGAAAGGAACGGA
GGCX In 8 Rev	AGTTGTTATAGCCCTGGGAAGGCA
TMC 8 Ex 10 Fw	ATGTTCTCCGTCTCCCTGGGTCA
TMC 8 Ex 10 Rev	AGTCACAAACGTTGTAGCCGTAGG
TMC 8 CTCF peak Fw	CACCTTTGAGAGAAGCTGTGAGCA
TMC 8 CTCF peak Rev	TTGAGGATGGGATTAGAGCAGGAAGC
TMC 8 In 9 Fw	AAGCTGAGGTGGGAGGATCACTTA
TMC 8 In 9 Rev	AGATAGGGTCTCACTCCATTGCCT
MECR Ex 5 Fw	TCATCCAGAATGCATCCAACAGCG
MECR Ex 5 Rev	ACCACATTGATGGTTCTTAGGCCC

MECR CTCF peak Fw	AGACCATGTCTTGGGTCTTTCCCA
MECR CTCF peak Rev	TGACTCATCTGTTTCAGCCGCCAA
MECR In 4 Fw	CAAAGCTAAACCTGGGCTTTGCCA
MECR In 4 Rev	GGTGCACTGCCTTGAAATGGGATT
MEF2B CTCF peak Fw	TTCCGGTACGGATGGGCAGCTATTT
MEF2B CTCF peak Rev	GCAAAGGCTTGGCAGGAGGA
MEF2B Ex 12 Fw	AAT TCC TGC CCA CCA GAA TAT GGC
MEF2B Ex 12 Rev	TCG TCT CAG GGC ACT TAC CTG G
MEF2B In 10 Fw	AGGGTGTGGCAGGGCTGATGA
MEF2B In 10 Rev	TTCTCTGTGCGGAGAGAGGGTGAA
Mouse CD45 Ex 4 Fw	AGAGAATGCCCTTCTTCTGCCTCA
Mouse CD45 Ex 4 Rev	TGTGGTCTCTGAAGAGCCATTTCC
Mouse CD45 In 4 Fw	GCC TGT GGC ATT CCT TCC AAC TTT
Mouse CD45 In 4 Rev	GTG CAC AAT CTG GTG TTG GCA GAA
Mouse CD45 Ex 5 Fw	TGT GTT ATC CAC GCT GCT GCC TCA
Mouse CD45 Ex 5 Rev	TGG CTG CTG AAT GTC TGA GTG TCA
Mouse CD45 In 6 Fw	TGA GGC TCT TCT CTG ACT CCA GAT
Mouse CD45 In 6 Rev	TCC TAG AGC TAC AAG CTG TTA ATG T
Mouse GAPDH Fw	ATGGGAAGCTTGTCATCAACGGGA
Mouse GAPDH Rev	AAGACACCAGTAGACTCCACGACA
FW E5 Mut CTCF	/5Phos/GCT TTA ACT GAC ACG CAG ACA TTC AGC
Rev E5 Mut CTCF	/5Phos/TGA GTA CGT CTG CGA GTC TGC GTG

Supplementary Table 2. ChIP- Seq data statistics

Sample	**Number of uniquely mapped reads**	**Number of identified peaks**	**% promoter**	**% polyA sites**	**% intragenic**	**% intergenic**
BJAB_CTCF	17,879,919	38,565	6.1	4.2	44.9	44.8
BJAB_Rabbit Ig	15,413,930	NA				
BL41_CTCF	24,090,392	36,795	6.9	4.5	43.9	44.5
BL41_Rabbit Ig	25,922,250	NA				
re-analyses of published data from Barski et al. 2007						
CD4+ T-cells	5,922,166	33,347	18.2	5.5	41.5	34.7

Supplementary Table 3. RNA-Seq data statistics

Sample	**Library-type**	**Number of reads***	**% uniquely mapped**	**% splice-junctions**	**Exonic %**	**Intronic %**	**Intergenic %**

BJAB wild type	Paired-end RNA-seq, 50bp	49,990,036	91.8	11.61	74.56	6.01	16.15
BJAB shRFP	Paired-end RNA-seq, 50bp	35,304,049	92.5	11.89	76.42	6.74	15.19
BJAB shCTCF	Paired-end RNA-seq, 50bp	53,889,834	92.0	12.02	76.05	5.51	15.81
BL41 wild type	Paired-end RNA-seq, 50bp	49,911,365	91.4	11.36	74.16	5.71	18.19
BL41 shRFP	Paired-end RNA-seq, 50bp	41,255,000	90.4	9.97	69.07	6.99	19.44
BL41 shCTCF	Paired-end RNA-seq, 50bp	49,723,317	90.6	10.27	68.3	7.32	18.79