Supplemental Information

SI Materials and Methods

Cell Culture. P2C2 (SCID.adh2C2) cells were grown in RPMI medium supplemented with 10% fetal bovine serum, penicillin-streptomycin-glutamine, nonessential amino acids, sodium pyruvate, and 2-mercaptoethanol. RAW 264.7 and NIH 3T3 cells were grown in Dulbecco's modified Eagle's medium with 10% fetal bovine serum, penicillin-streptomycin-glutamine. 32D cells were grown in RPMI supplemented with 10% fetal bovine serum, penicillin-streptomycin-glutamine, nonessential amino acids, sodium pyruvate, 2-mercaptoethanol and Interleukin-3 (WEHI3B supernatant).

Bisulfite-DNA Sequencing. Ten μg of DNA was denatured in 22 μl of 0.3 M NaOH at 37ºC for 15 minutes. Freshly prepared 2.2 M sodium metabisulfite, containing 0.5 M hydroquinone (pH 5.0), was added to the denatured DNA and incubated at 50 ºC for 8 hours. Bisulfite-modified DNA was recovered by a mini-Sephadex G-50 gel column and desulfonated with 0.3 M NaOH at 37 ºC for 20 minutes. The solution was neutralized with 30 μl of 10 M ammonium acetate, precipitated with ethanol and dissolved in 25 μl of nuclease-free water. The PCR primers for 5'-flanking region of Bcl11b are: sense 5'- AGT TGT GAG ATT AGA TAT TAA TAG TTT and antisense 5'- AAA ATC AAC TAA CCT TCT ATC TCT TAA; sense 5'- GGA GGT TAG GTT ATT TAG TAG ATT TA and anrisense 5'- CCC CAC ACC CTA ATC TCC TTA AAA AAC. The PCR primers for exon 4 of Bcl11b are: sense 5'- AAG AAG AGG AGT TGT TGT TGG AGA A and antisense 5'- AAC TCC AAA TCT TTC TCC ACC TTA AT; sense 5'- ATT AAG GTG GAG AAA GAT TTG GAG TT and antisense 5'- CTT TTT CAT ATA TTT CTC CAA AAT ACT. The PCR product was purified with QIAquick Gel Extraction KIT (Qiagen), ligated into TA cloning pGEM T-Easy vector (Promega) and transformed into chemically competent

E.coli under standard protocol. Plasmid DNA was isolated with QIAprep Spin Miniprep Kit and analysis by DNA sequencing (Laragen, Inc).

Cloning of Reporters and Luciferase Assays. Four fragments that cover different regions of Bcl11b promoter area were cloned (Table S1 for primer information, Fig. 1B and Fig. S1A for positions of the fragments.). The fragments were cloned into pGL3-basic vector (Promega). Three conserved elements upstream of the DMR were then cloned into the Bcl11b promoter-pGL3 vectors to test whether the conserved elements had the ability to control the promoter activity to recapitulate the expression pattern of Bcl11b. Sequences from the 3'UTR of Bcl11b were cloned into the pGL3–control vector (Promega) between the luciferase cDNA and poly(A) at the *Xba*I restriction site (Fig. S4B). Major Peak (chr12:108,396,825-108,398,672) and deletion derivatives of the major peak were first used the PCR primers listed in Table S1 to cloned into pGEM-T Easy (Promega) TA cloning vector, and then were sub-cloned into the PR-pGL3 vectors downstream of the SV40 late poly(A) signal at *BamH*I/*Sal*I sites. The Major Peak fragment was also cloned into pGL3-promoter vector (Promega) downstream of the late poly(A) signal at *BamH*I/*Sal*I sites. Positions of the fragments cloned in the experiments were list in Table S2. For transient transfections, 2 μg of reporter constructs were mixed with 1 μg of pTracer-Renilla Luciferase and electroporated into P2C2 and Raw264.7 cells. Luciferase activities were measured 24 hours after transfections. For stable transfection, the constructs were linearized and electroporated to P2C2 and Raw264.7 cells. Two days after transfection, blasticidine was added at concentrations at 4 ng/ml for Raw264.7 cells and 8 ng/ml for P2C2 cells. In the first stable selection experiments, we noticed that gradual increase of blasticidine to the final concentrations yielded more reliable results. Therefore, in the later experiments, concentrations of blasticidine in P2C2 were 2 ng/ml on day 3 after electroporation, 4 ng/ml on

day 5, 6 ng/ml on day 7, 8 ng/ml on day 9. The concentrations of blasticidine in Raw264.7 cells were 1 ng/ml on day 3 after electroporation, 2 ng/ml on day 5, 3 ng/ml on day 7, and 4 ng/ml on day 9. Luciferase activities were measured by Dual-Luciferase Assay System (Promega) in 6-12 independently drug-selected subcultures per sample after two weeks of blasticidine selection in parallel.

Chromatin Conformation Capture Assay. Ten million cells were centrifuged, and resuspended in 500 ul of 10% FBS/PBS. The cells were filtered through 40 μm cell strainer to make single cell suspension. Then, 9.5 ml of 2% formaldehyde/10% FCS/PBS was added and the sample was incubated for 10 min at room temperature. The reaction tubes were transferred to ice and added 1.425 ml of 1 M ice-cold glycine. The tubes were spun for 8 min at 225g at 4 ºC. The supernatant was removed and the pellet was lysed and incubated on ice for 10 minutes. Samples were centrifuged for 5 minutes at 4 ºC. Nuclear pellets were resuspended in 0.5 ml of 1.2× *Bgl*II buffer, 7.5 μl of 20% SDS at 37 ºC for 1 hour while shaking at 900 rpm. Then, 50 μl of 20% Triton X-100 was added, and the samples were incubated for 1 hour at 37 ºC while shaking at 900 rpm. *Bgl*II of 400 U was added, and the samples were incubated at 37 ºC for 16 hours while shaking at 900 rpm. The samples were added 40 μl of 20% SDS, and incubated for 20–25 min at 65 ºC while shaking at 900 rpm. Then, 6.125 ml of 1.15× ligation buffer and 375 μl of 20% Triton X-100 were added, and the samples were incubated for 1 hour at 37 ºC while shaking gently. To each sample was added 5 μl ligase (100 U total) and incubated for 4 h at 16 ºC followed by 30 min at room temperature. The samples were then treated with 30 μl of 10 mg/ml proteinase K, and incubated at 65 ºC overnight to de-crosslink. The next day, DNA was extracted by phenol-chloroform, and precipitated by ethanol at -20 ºC overnight. The DNA was washed twice with 80% ethanol and dissolved in 100 μl of buffer TE for qPCR analysis later.

The BAC amplification efficiency controls were made from equal molar mixtures of BACs covering the relevant loci being assessed for interaction (*Bcl11b* and Major Peak, Major Peak and *Vrk1*, *Bcl11b* and *Sfpi1*, and Major Peak and *Cd3gde* cluster). The BAC mixtures were digested by *Bgl*II for 1 hour at 37 ºC, and purified by phenol-chloroform extraction followed by ethanol precipitation. The DNA pellet was dissolved in Tris-HCl pH7.4 and ligated by T4 DNA ligase. After ligation, DNA was purified and dissolved in buffer TE.

The 25-kb upstream region and *Bcl11b* locus were digested by *Bgl*II and 18 fragments were generated. Nine fragments were generated from digestion of the 40-kb region around Major Peak. Eleven fragments were generated from digestion of *Vrk1* locus, eight from digestion of *Sfpi1*, and ten from digestion of *Cd3gde* loci. qPCR primers to measure the 3C interactions were designed to prime across fragment boundaries, and the efficiency of each primer set was tested on the mixed, digested, and randomly ligated BAC controls. The efficiency was later used to normalize the relative interaction frequency. The specificity of the primer sets was determined by using them to perform qPCR on a DNA sample that was digested by *Bgl*II without the later ligation reaction (un-ligated control). Additional negative controls were tests of each primer pair against genomic DNA that had been purified without 3C processing from B6 thymocytes. Primers used for results shown in this paper (listed in Table S1) were a subset selected for low or undetectable background reactivity with unprocessed genomic DNA; the exception was the R5-F5 combination which yielded an anomalous amplification product in combination, despite good specificity for each primer in other combinations.

For PCR reactions, equal loading of DNA was normalized by matching GAPDH levels. qPCR were performed using SybrgreenER (Invitrogen) and a 7900HT real-time PCR machine (ABI). Values shown are calculated from the ΔC_T between the product of a given primer pair on

the 3C sample and on mixed, cleaved, and religated BAC control DNAs in parallel reactions. BAC efficiency standards were tested at the same concentration for all samples using relevant primer pairs.

Modification of Bcl11b-BAC and BAC Transfections. A BAC fluorescent reporter for Bcl11b expression was generated by lambda Red-mediated recombineering of an mCitrine YFP into a BAC containing the genomic *Bcl11b* locus (RP24-282D6, BACPAC Resource Center, Oakland, CA), using the *E. Coli* recombineering strain SW102. Homologous sequences flanking exon 1of Bcl11b were attached to the ends of an mCitrine YFP cassette containing a kanamycin/neomycin resistance gene using fusion PCR. This mCitrine-YFP cassette was then knocked into Bcl11b exon 1 of the BAC using recombineering, and recombinants were selected using kanamycin (Sigma-Aldrich). The BAC-Bcl11b-YFP reporter was used directly for transfection, or further modified to incorporate the major peak (MP) *cis*-regulatory region.

The MP sequence was inserted into a region downstream of the Bcl11b gene in the Bcl11b-YFP reporter BAC, using a two-step recombineering procedure involving GalK selection. Homologous sequences flanking an insertion site (between chr12:109,138,812 and chr12:109,138,813) 9.8 kb downstream of the 3' UTR of the Bcl11b gene were attached to either a GalK cassette, or a DNA sequence containing the MP regulatory region using fusion PCR. The GalK cassette was knocked into the Bcl11b-YFP reporter BAC using recombineering, and GalK-containing recombinants were selected for on minimal media containing galactose. The MP region was knocked into the Bcl11b-YFP reporter BAC containing the GalK cassette, and recombinants in which the GalK-cassette was replaced by the MP region were selected for using 2-deoxygalactose (Sigma-Aldrich). The BAC-Bcl11b-YFP-MP reporter containing the inserted MP region was then used for transfection into tissue culture cells.

For stable transfection of BAC reporters to P2C2 and Raw264.7 cells, BAC-Bcl11b-YFP and BAC-Bcl11b-YFP-MP were digested by PI-SceI (New England Biolabs) for 30 minutes and then extracted by phenol/chloroform. Then, 2 μg of BAC was mixed with 0.01 μg of pTracer and transfected into P2C2 and Raw264.7 cells by electroporation. Two days after transfection, blasticidine was added at concentrations of 4 ng/ml for Raw264.7 cells and 8 ng/ml for P2C2 cells. After two weeks of blasticidine selection, the expression of YFP was measured by flow cytometry.

Datasources for transcription factor chromatin immunoprecipitation results. Results for PU.1 and GATA-3 binding were from J. A. Zhang, A. Mortazavi, et al (ref. 9, accession # GSE31235). Sites of occupancy by the Notch-activated transcription factor RBPJ (CSL) were taken from ChIP enrichments analyzed by realtime PCR, reported by P. Li, S. Burke, et al. (ref. 3). The binding sites themselves are bracketed between the locations of the primer pairs used to amplify the three binding regions, which are shown in Fig. 1A. Data for Runx1 binding (Fig. 1A and Fig. 6) were from M. Yu, T. Mazor, et al., *Molecular Cell* **45**, 330–343, 2012 (Ref. 25, accession # GSE33653). Data for Ikaros binding (Fig. 6) were from J. Zhang, A. F. Jackson, et al., *Nat. Immunol.* **13**, 86-94, 2012 (Ref. 37, accession # GSE32311). Data for TCF-1 binding (Figs. 1A and 6) were from ref. 24 (accession #GSE46662). Histone modification (Fig. 2), PU.1, and GATA-3 data (Fig. 6) were processed into bigWig files as described in ref. 9. Runx1, Ikaros, and TCF-1 data were first processed by MACS to select signals more significant than a 1e-05 or 1e-13 probability threshold, as indicated, and then displayed as Bedgraphs. PU.1 data from DN1 cells processed in the same way (1e-13 threshold) are also shown (Fig. 1A).

Table S1. PCR Primers

3C Primers

Bcl11b Locus	
F1	AGCTGTGGTGCTGTTACGTG
F2	ACAATGCTTCGTCGTAGGTG
F3	GCCAGCAGAGCAGGTGTTAT
F5	GAAGAGGGAGGGGAGGTAGTT
F6	TAGGGATGGGGGTGATCATA
F7	CATTGGCCTCTTCCCTGAA
F8	GGGGGATCGCAGATAGAAGT
F9	GGTCAGGTACAGTCTGTTAG
F10	ACTGTGAGGTGTTCCAGTCT
F13	GGCCTAGGCTTGTCCTCTTC
F14	AGAGGCTGGTACAGCTGTT
F15	AGGGTCTGGGCTTTATGGTT
F16	GGTGAGTTTGTCCATCAGCA
F17	GCTCCCTGTATTTGCTCAGG
F18	GCCCCAGCCATTCTAAGTAA
Gm16084 Locus	
R3	ATCACCCCACCTCTCATCAA
R4	TGTCCTTGAACCCTTGAACC
R5	CTCCTGCCCTGAGTTCCTAA
Cd3 Cluster	
F1	ATCCGAGGCATACATTTGGA
F2	GGCACCTTTACCTGTTGAGC
F3	GGTTTAGGAGCCCCTGTTTC
F4	GCCTGTTGCTGCATCTCC
F5	CATCACTGTTGGCTTCTCCA
F6	AACCCCACAAAGCTGAAGTG
F7	CTGGAGGTACTTGCCTTTG
F8	GACACGGTAGTTCATGCCTTT
Vrk1 Locus	
F4	TCCCAGCATTGTTGAGACAG
F5	GTGGCCCACAGCACTTTAAT
Sfpi1 Locus	
F2	CCCGCTCTTAACCTTCACAA
F3	GGGAGGTTGATAACGGAGGT
F6	AGGGTTCTGGGTTGAAGATG

Cloning Primers

	Forward	Reverse
PR1	CCAGAAAGAGCCGAACCAAGGCA	TGCCCCGGCATCTATTCTGGCAT
PR2	CCAGAAAGAGCCGAACCAAGGCA	TCTTTATTTTGCTCTTTCTTC
PR3	TTGGACTCAGTGTACCTCAGTTA	TGCCCCGGCATCTATTCTGGCAT
PR4	TTGGACTCAGTGTACCTCAGTTA	ATAAAACTGCTGTTGCTCGCCGCT
CS1	TGGGATTGAGACAACTCAATAAA	GAGGCTATGGTGGGTGACTC
CS2	AGATTTCAATTGGAGTGTGAGTT	CCTATCTGCTTAACAGGAACAAT
CS3	TGTGGAGTTAGGTCCACCTCA	TGCAGCCTGGAAGACTTAGAAA
3UTR	CTAATGATGTCAAAATCGAGCAG	AAATCAGAAGGGAAAACAGAGAC
Major Peak	GATCCCCCTCCAGAGCCGGCT	ACTATTCCAAACAAGAACTGGGA
MPF1	GATCCCCCTCCAGAGCCGGCT	GACTGCCATCTGGGTCCT
MPF2	GATCCCCCTCCAGAGCCGGCT	GACACCTCTGTACTTATAG
MPF3	CTATAAGTACAGAGGTGTC	ACTATTCCAAACAAGAACTGGGA
MPF4	AGGACCCAGATGGCAGTC	ACTATTCCAAACAAGAACTGGGA
MPF5	GATCCCCCTCCAGAGCCGGCT	GTCGACAGCGTTTGACAAGGGCTT
MPFusion	GATCCCCCTCCAGAGCCGGCT	ATTGGCTCGCTGCCCAGACAGATAGAGTTTCA
	CTGTCTGGGCAGCGAGCCAATTAGCAGTTTC	ACTATTCCAAACAAGAACTGGGA
dmTCF	GATCCCCCTCCAGAGCCGGCT	GGCCAGGATCCTAGCAGTGGTCA
	CTAGGATCCTGGCCTGTCGCAAGA	ACTATTCCAAACAAGAACTGGGA
	GATCCCCCTCCAGAGCCGGCT	TGTCAGAACCACGGAACTGTGACAC
	GTGGTTCTGACATGTCATCGGGTG	ACTATTCCAAACAAGAACTGGGA

Table S2. Positions of the Sequences Cloned

Fragments	mm9 Coordinates
DMR	chr12:109,241,957-109,242,544
PR3	chr12:109,241,354-109,242,688
Major Peak	chr12:108,396,825-108,398,672
MPF1	chr12:108,398,141-108,398,672
MPF2	chr12:108,397,531-108,398,672
MPF3	chr12:108,396,825-108,397,549
MPF4	chr12:108,396,825-108,398,158
MPF5	chr12:108,397,242-108,398,672
dmTCF	chr12:108,397,789-108,398,672 + chr12:108,397,518-108,397,774 + chr12:108,396,825-108,397,470
MPFusion	chr12:108,398,397-108,398,672 + chr12:108,396,825-108,397,868

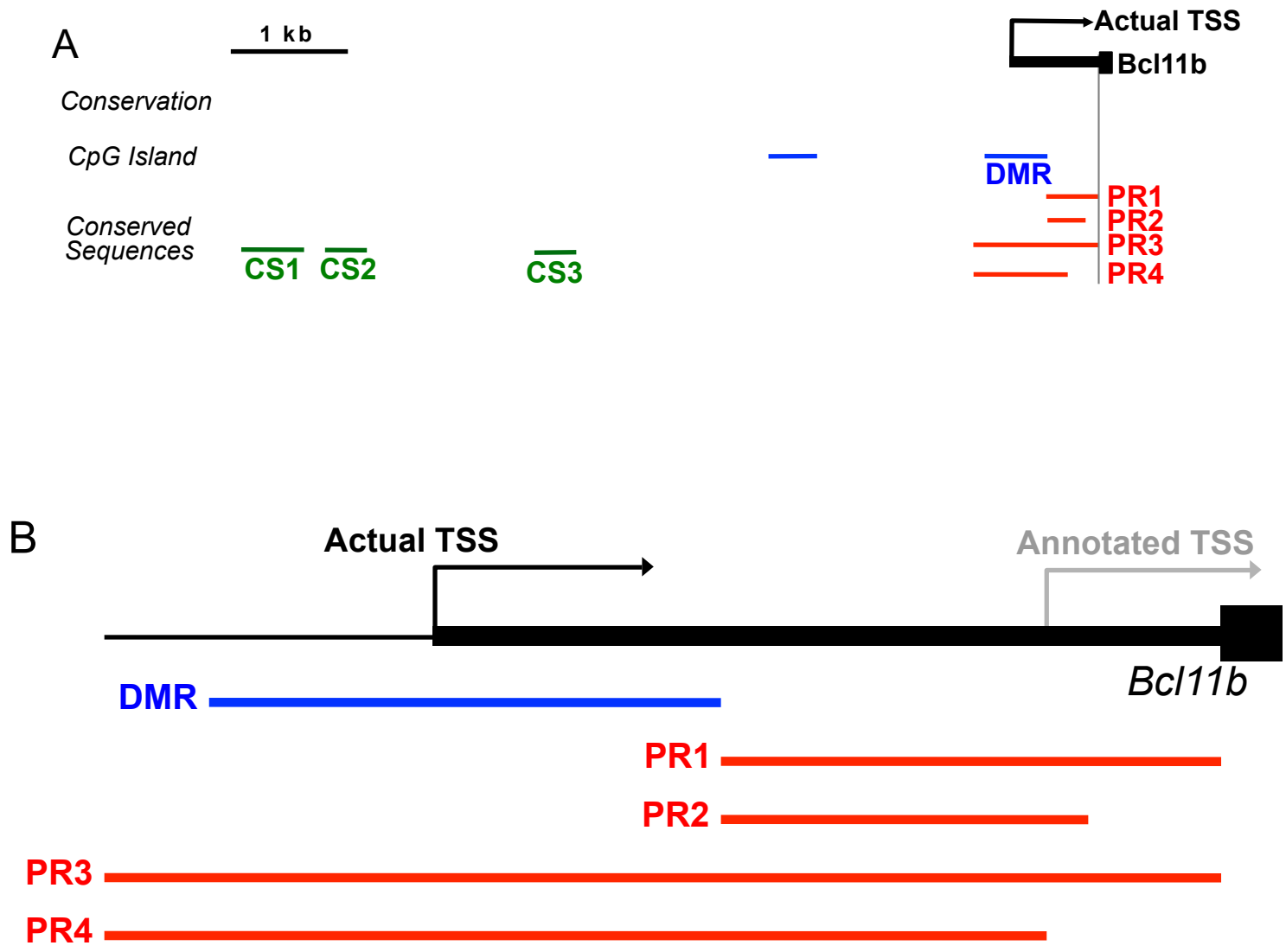


Figure S1. Positions of *Bcl11b* transcription start sites, CpG islands, promoter frangments and upstream conserved sequences cloned in the luciferase assays. DMR, differentially methylated region identified in Figure 1B. PR1-PR4, four sequences that cover different areas of *Bcl11b* promoter region. CS1-CS3, three highly conserved sequences identified by UCSC Genome Browser 30-Way Multiz Alignment and Conservation. These elements were cloned into pGL3-Basic vector to map *cis*-regulatory sequences of *Bcl11b*. The actual transcription start site identified by our RNA-seq on early T cells in Figure 1A is marked.

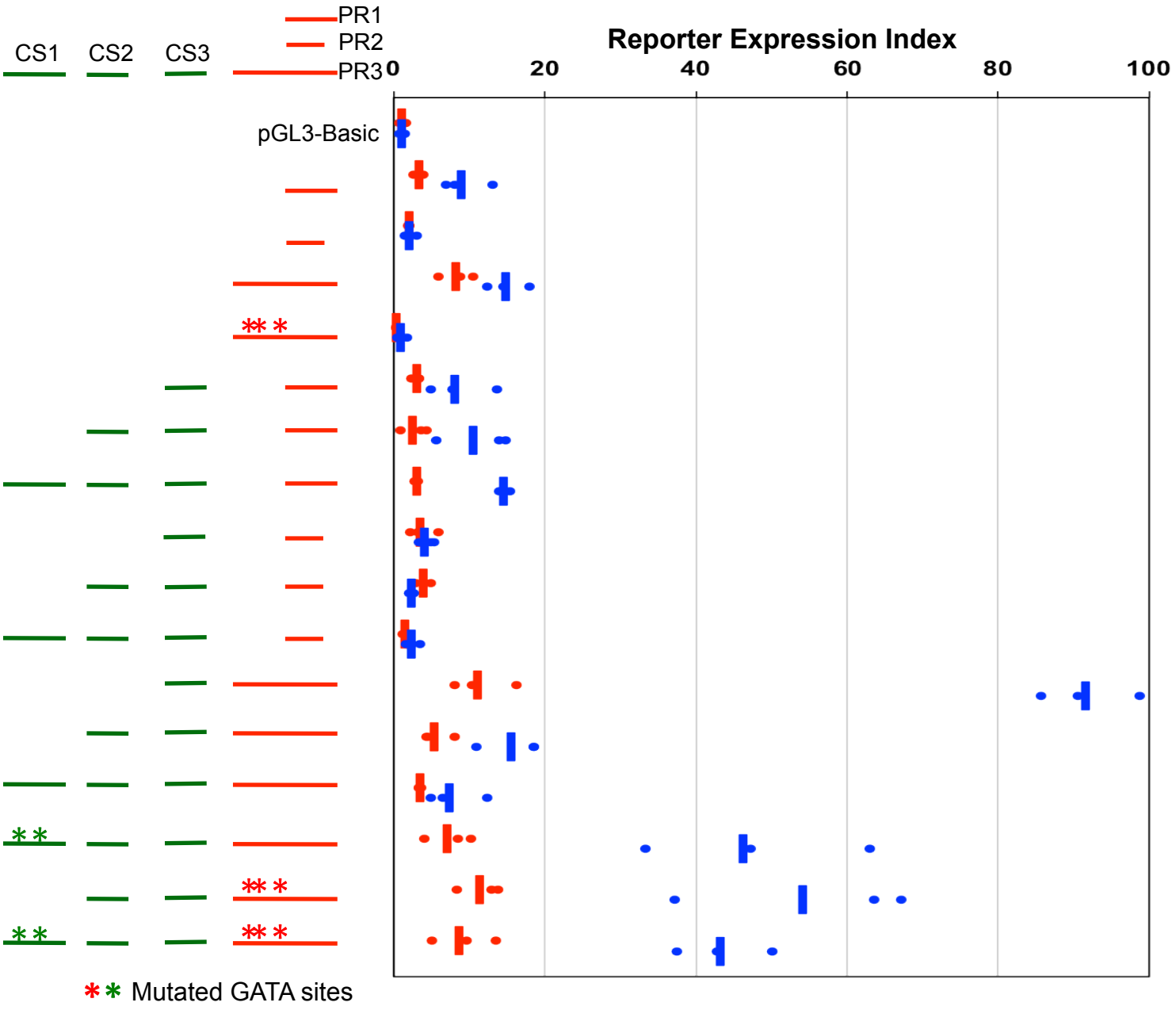


Figure S2. Transient transfection of CS-PR-pGL3-Basic constructs into P2C2 and Raw264.7 cells. PR sequences were cloned into pGL3-Basic vector. The CS (Conserved Sequences) sequences were then cloned into PR-pGL3-Basic constructs upstream of PR sequences to mimic the relative positions of these sequences in the genome. In certain constructs, GATA binding sites were mutated. Red, P2C2 (T lineage cells) samples. Blue, Raw264.7 (macrophages) samples. •and•, index of relative firefly luciferase activity normalized by renilla luciferase activity. −and−, geomean (the same arrangements of colors and symbols in Figure S3 and S4.). The Reporter Expression Index was calculated by designation of the geomean of pGL3-Basic as 1 unit.

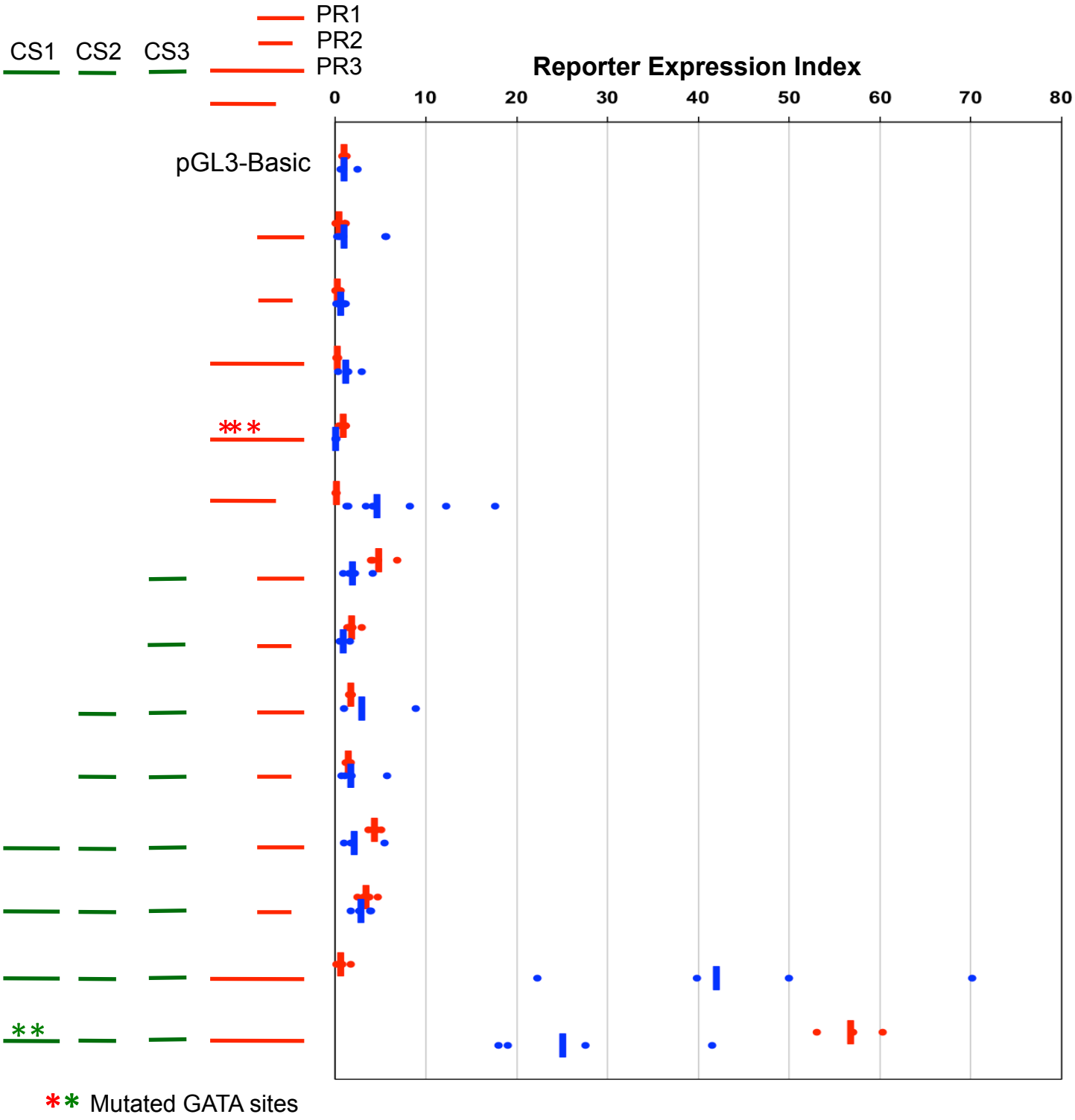


Figure S3. Stable transfection of CS-PR-pGL3-Basic constructs into P2C2 and Raw264.7 cells. PR sequences were cloned into pGL3-Basic vector. The CS (Conserved Sequences) sequences were then cloned into PR-pGL3-Basic vectors upstream of PR sequences to mimic the relative positions of these sequences in the genome. In certain constructs, GATA binding sites were mutated. The transfected cultures were selected by blasticidine for two weeks. Red, P2C2 (T cells) samples. Blue, Raw264.7 (macrophages) samples.

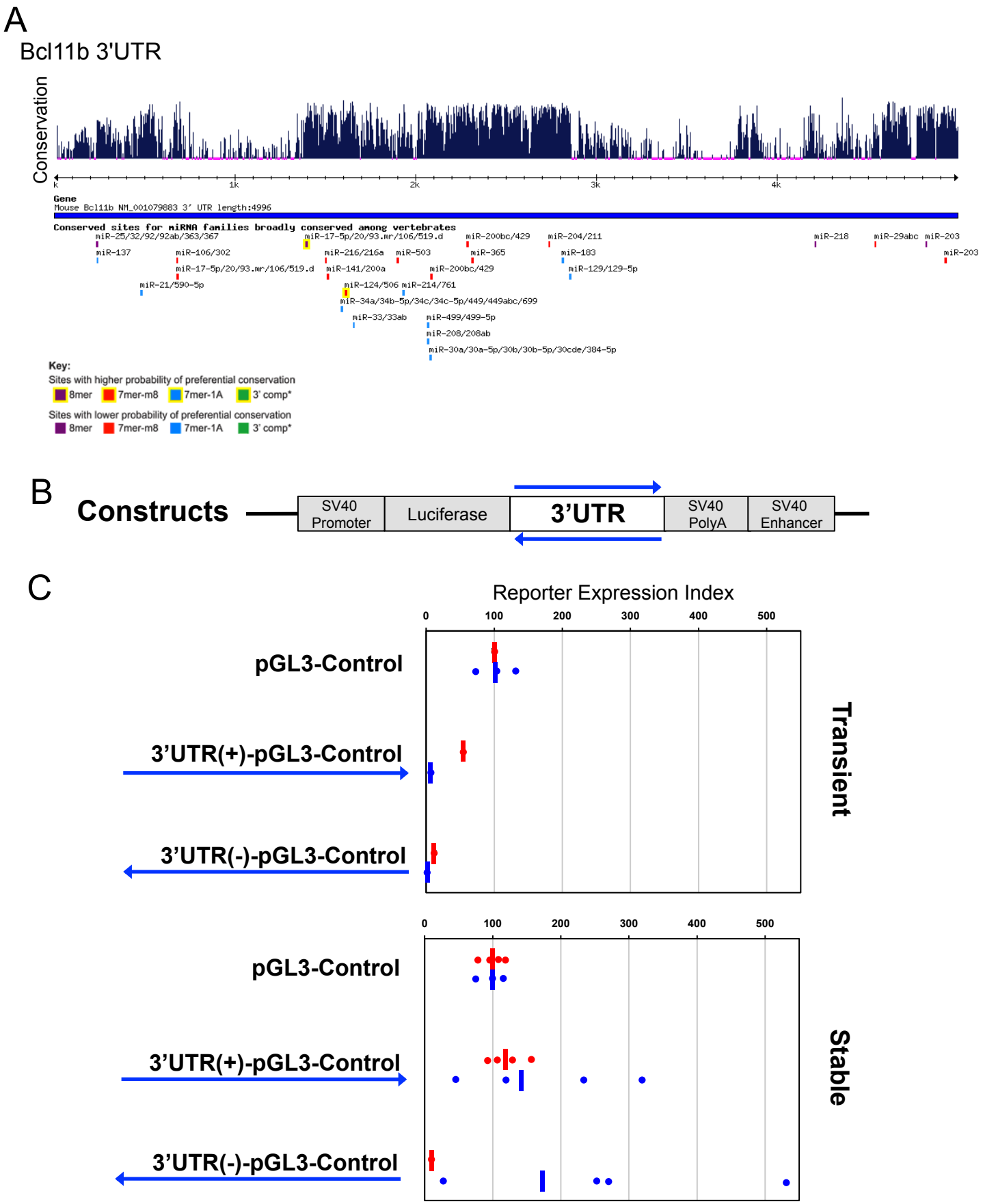


Figure S4. Transfection of *Bcl11b*-3'UTR-pGL3-Control constructs into P2C2 and Raw264.7 cells. A, predicted microRNA targets in *Bcl11b*-3'UTR (www.targetscan.org/cgi-bin/targetscan/mmu_50/view_gene.cgi?taxid=10090&gs=Bcl11b&showcnc=0&shownc=0#miR-218). B, the 5-kb *Bcl11b*-3'UTR in two orientations was cloned into pGL3-Control vector between luciferase coding sequence and SV40-poly(A) sequence. C, the constructs were transfected into P2C2 and Raw264.7 cells. The Reporter Expression Index was calculated by designation of the geomean of pGL3-Control as 100 units. Red, P2C2 (T cells) samples. Blue, Raw264.7 (macrophages) samples.

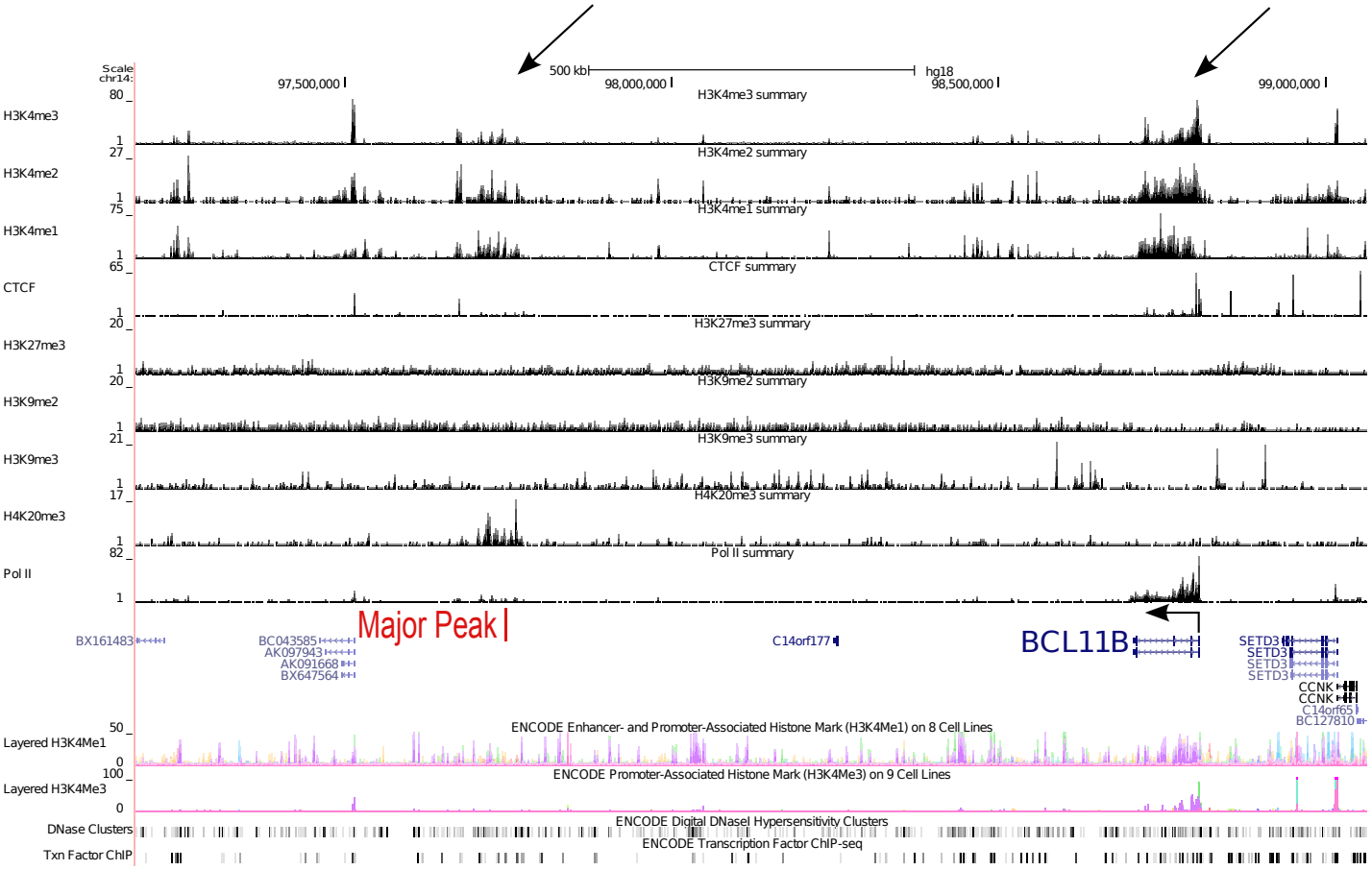


Figure S5. Histone modifications of *BCL11B* and Major Peak loci in human cells. Mouse Major Peak sequence was aligned to chr14:97,744,724-97,746,607 in human genome (NCBI36/hg18) at 86.9% identity by UCSC Genome Browser BLAT. Histone marks, CTCF and RNA Pol II binding patterns in human CD4+ T cells are shown in the upper panel (from ref. 34). The lower panel shows ENCODE Enhancer- and Promoter-Associated Histone Mark H3K4Me1 on 8 human cell lines (Gm12878, H1 ES, HMEC, HSMM, HUVEC, K562, NHEK and NHLF), ENCODE Promoter-Associated Histone Mark H3K4Me3 on 9 human cell lines (eight cell lines mentioned above and HepG2), ENCODE Digital DNaseI Hypersensitivity Clusters and ENCODE Transcription Factor ChIP-seq (from ref. 35).

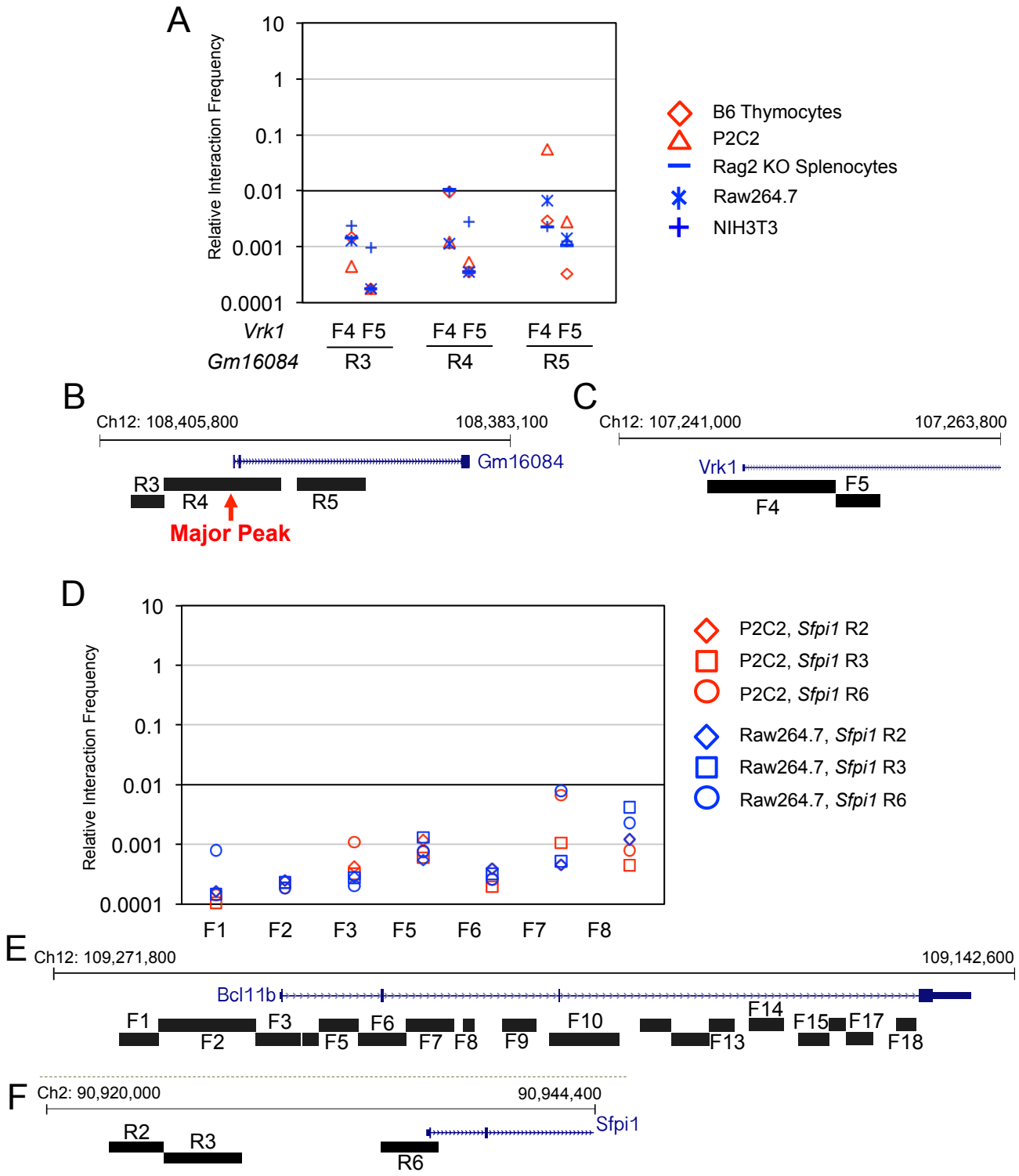


Figure S6. Chromatin interactions between Major Peak (*Gm16084*) and *Vrk1*, and between *Bcl11b* and *Sfpi1*. A, chromatin interactions between the Major Peak region and the promoter region of *Vrk1* in five cell types. The 3C assays did not detect any significant cell type specific interactions between the two regions. B and C, 3C fragments. D, chromatin interactions between *Bcl11b* and *Sfpi1* in P2C2 cells and Raw264.7 cells. The 3C assays did not detect any significant cell type specific interactions between the two regions. E and F, 3C fragments. The *Sfpi1* R2 fragment covers the Upstream Regulatory Element and additional myeloid-specific regulatory elements (M. Leddin et al., *Blood*, 2011, 117: 2827-2838); *Sfpi1* R3 fragment covers the internal regulatory elements (ref. 17) and the R6 fragment covers the promoter. *P*-value of populations in all groups is greater than 0.02 by the Mann-Whitney *U* test.

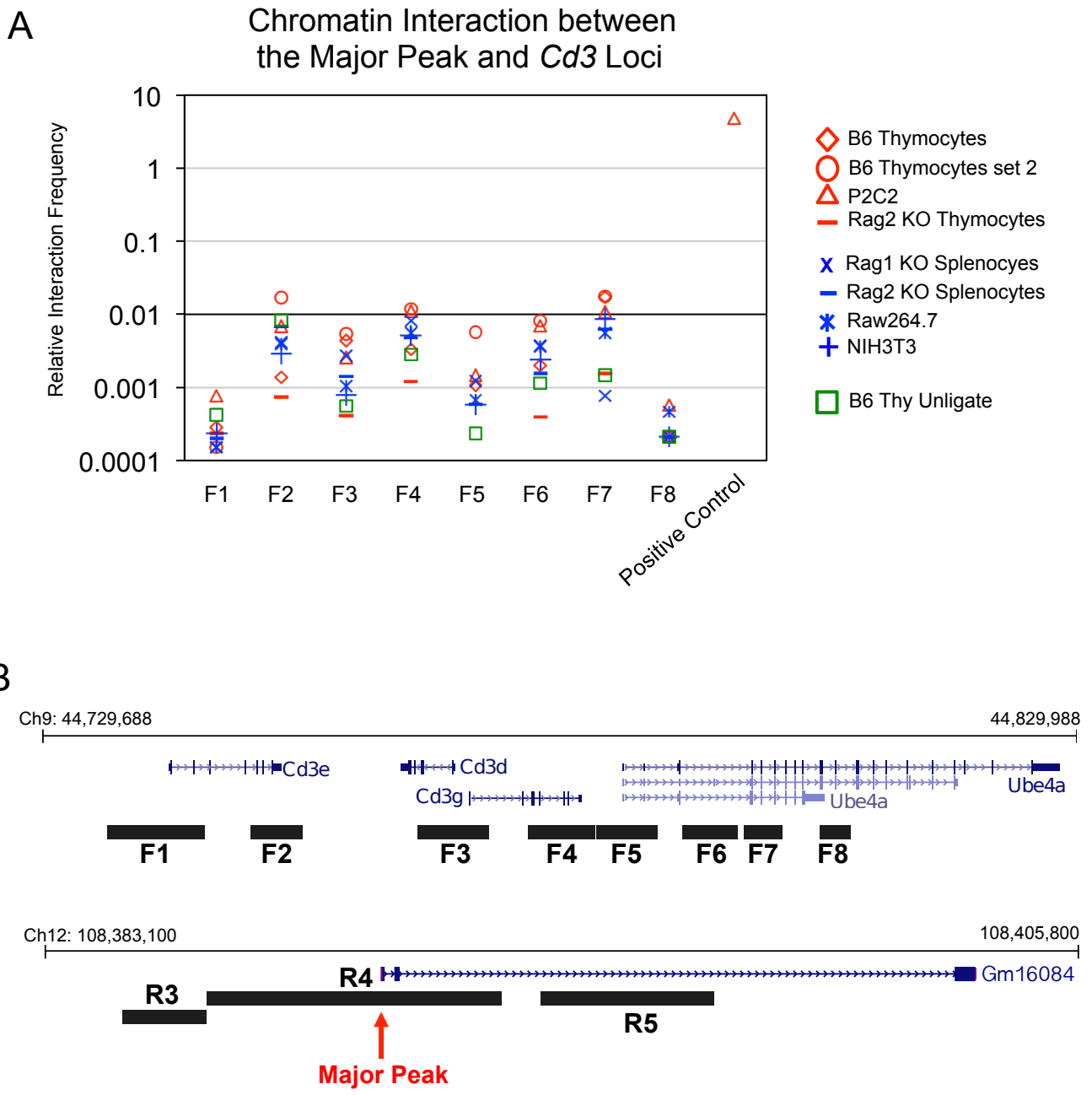
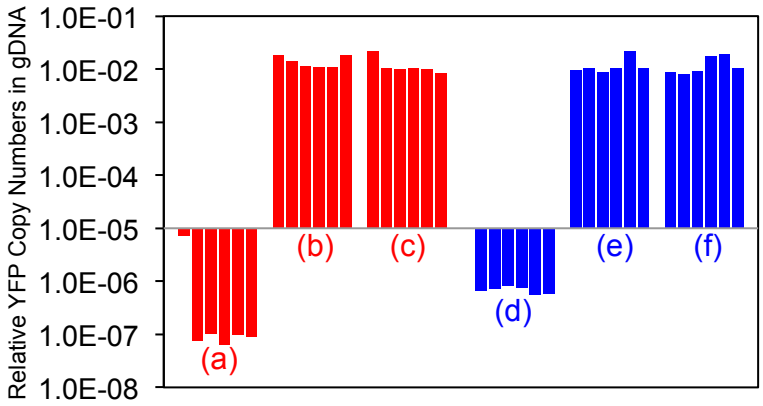


Figure S7. Chromatin interactions between Major Peak (*Gm16084*) and *Cd3* cluster A, chromatin interactions between the Major Peak region on Ch12 and the *Cd3* cluster on Ch9. The 3C assays did not detect any significant cell type specific interactions between the two loci. The value of 0.01 indicates the threshold of confidence. The last data point in the figure shows a positive control, which was qPCR by primers that detect interactions between *Bcl11b* F3 and Major Peak R4 on a sample from P2C2 cells. B, positions of the 3C fragments shown in A. The 3C fragments on the *Cd3* cluster cover T-cell specific genes *Cd3d*, *Cd3e* and *Cd3g*, and a non-specific gene *Ube4a*. The results confirm that Major Peak is a *Bcl11b*-specific enhancer in cells of T-lineage.



(a): Subcultures of P2C2 transfected with pTracer;
(b): Subcultures of P2C2 transfected with pTracer + BAC-*Bcl11b*-YFP;
(c): Subcultures of P2C2 transfected with pTracer + BAC-*Bcl11b*-YFP-Major Peak;
(d): Subcultures of Raw264.7 transfected with pTracer;
(e): Subcultures of Raw264.7 transfected with pTracer + BAC-*Bcl11b*-YFP;
(f): Subcultures of Raw264.7 transfected with pTracer + BAC-*Bcl11b*-YFP-Major Peak.

Figure S8. Integration of BAC-*Bcl11b*-YFP into the genomes of the stably selected subcultures. The levels of YFP coding sequences in the genomic DNA (gDNA) from the subcultures in Figure 7 were measrued by qPCR, and normalized by GAPDH.

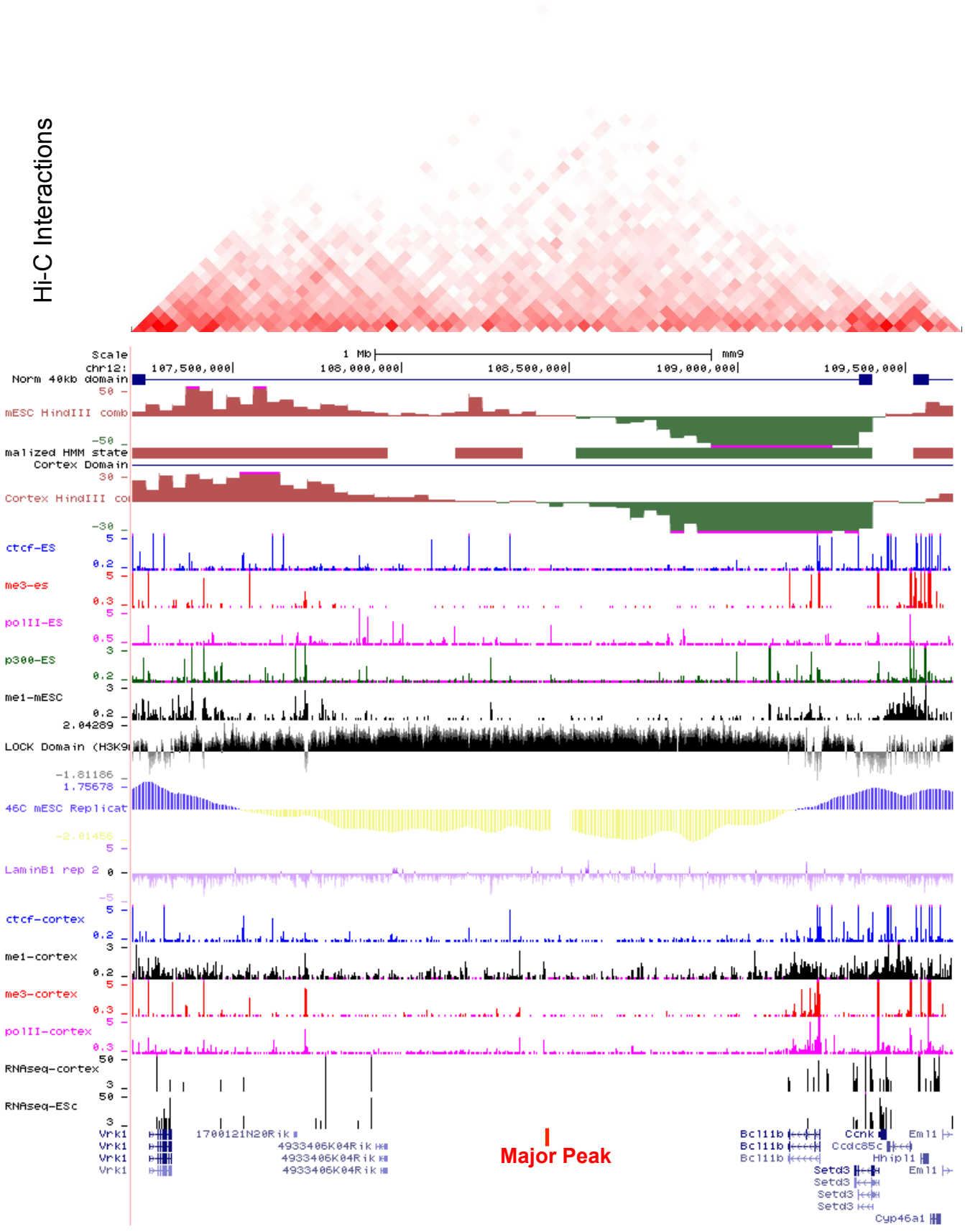


Figure S9. Hi-C interactions of Major Peak and *Bcl11b* loci. The region shown is Chr12:107,200,000 to 109,600,000 from mouse ES cells (http://chromosome.sdsc.edu/mouse/hi-c/myform.php) (data from ref. 41). The 2.4-Mb region covers chromosome 12 from *Vrk1* to about 200-kb upstream of *Bcl11b*.

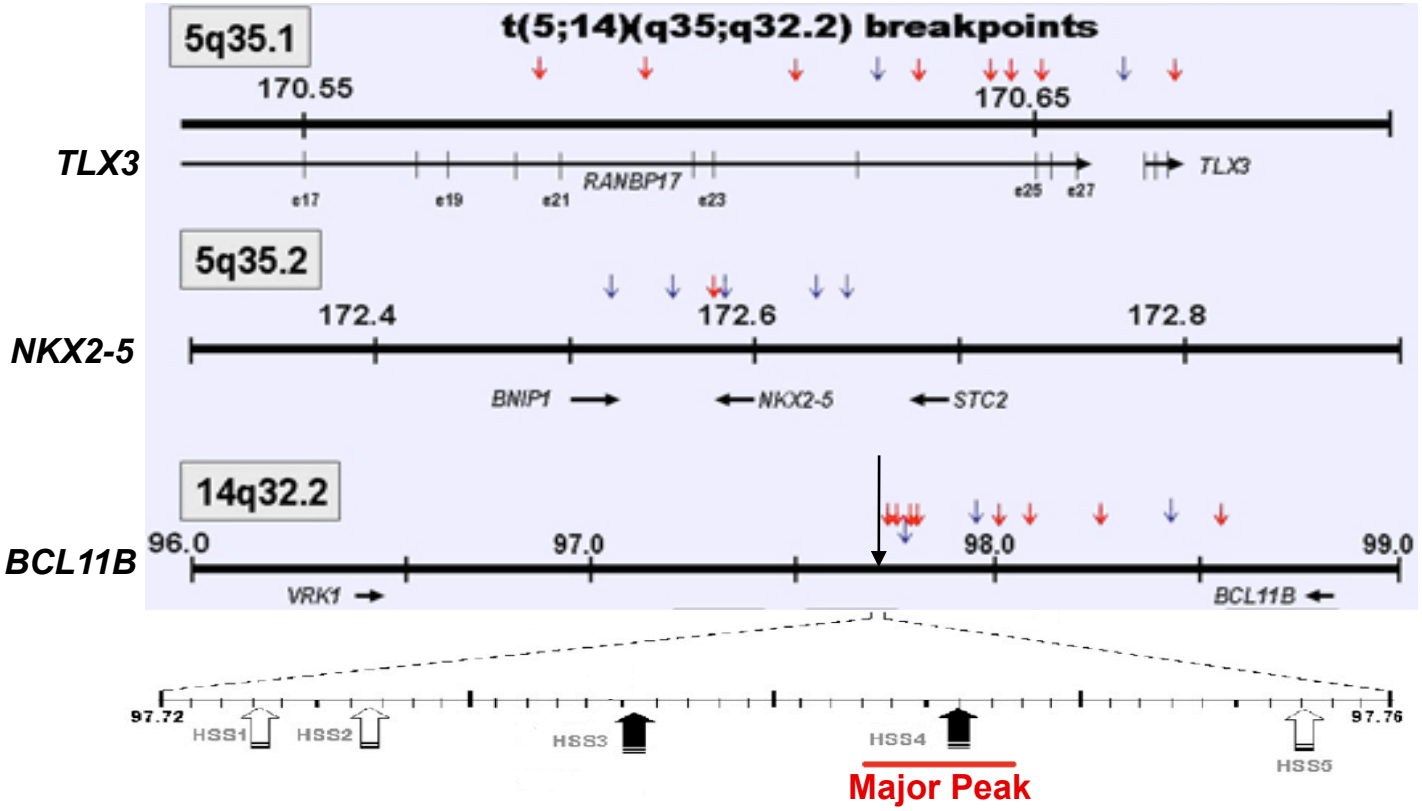


Figure S10. Analysis of t(5;14)(q35;q32.2) breakpoint cluster downstream of *BCL11B* in human T-ALL. The translocation was found in many cases of T-ALL, juxtaposing *TLX3* or *NKX2-5* to *BCL11B* locus. The chromosome breakpoints on *BCL11B* locus spread over 800-kb downstream of *BCL11B*. There are two DNase hypersensitive sites HSS3 and HSS4 on the boundary of the breakpoint cluster that may activate promoter of *TLX3*. The Major Peak sequence identified by us aligned to HSS4 by UCSC Genome Browser BLAT. The figure was adopted from http://atlasgeneticsoncology.org/Anomalies/t0514q35q32ID1386.html with modifications based on data from Ref. 9, Ref. 14 and Ref. 15.