SUPPLEMENTARY INFORMATION

Enhancer hijacking drives oncogenic *BCL11B* expression in lineage ambiguous stem cell leukemia

Table of Contents

Supplementary Table Legends 1-15 (See Supplementary Table Microsoft Excel workbook)

Supplementary Table 1. Cohort overview. All samples included in this study are listed along with their method of RNA-seq library preparation, identified fusion gene, diagnosis, subtype, tSNE coordinates, and the reference and accession number for data that are publicly available. *Note that samples from the MLL Munich Leukemia Laboratory are not publicly available; for information, contact Prof. Claudia Haferlach.

Supplementary Table 2. Non-B-ALL samples included in this study. Patient sample IDs are listed along with RNA-seq library preparation method, identified fusion gene, diagnosis, and subtype, *FLT3* mutation status, and tSNE coordinates.

Supplementary Table 3. *BCL11B* group leukemia samples. The *BCL11B* group cohort is listed with all relevant *BCL11B* structural variation results, sequencing data generated, mutation status and clinical characteristics.

Supplementary Table 4. Analysis of T cell receptor (TCR) rearrangements in the *BCL11B* group, and ETP-ALL, MPAL, and T-ALL samples not belonging to the *BCL11B* group. Both WGS and RNA-seq data were used for this analysis and the total number of samples with evidence of any TCR rearrangement (alpha, beta, gamma or delta TCR rearrangement) is shown.

Supplementary Table 5. Allele-specific expression analysis. All *BCL11B* group samples and a subset of T-ALL samples with matched WGS and RNA-seq data are listed.

Supplementary Table 6. Mutational analysis of the *BCL11B* group samples. All non-silent mutations are listed for each *BCL11B* group patient sample in the discovery cohort.

Supplementary Table 7. Mutation analysis (SNVs and indels) for the entire non-B ALL cohort for 5 key genes (*BCL11B*, *FLT3*, *WT1*, *RUNX1* and *ETV6*). Mutations in this table were identified from DNA-based sequencing (whole exome or whole genome sequencing).

Supplementary Table 8. Mutation analysis (SNVs and indels) for the entire non-B-ALL cohort for 5 key genes (*BCL11B*, *FLT3*, *WT1*, *RUNX1* and *ETV6*). Mutations in this table were identified from RNA-seq and do not overlap mutations identified from DNA-based sequencing.

Supplementary Table 9. CNV analysis for 5 key genes (*BCL11B*, *FLT3*, *WT1*, *RUNX1* and *ETV6*) in the non-B ALL cohort.

Supplementary Table 10. WGS analyses conducted. All samples with WGS from the Munich Leukemia Laboratory (MLL) are listed. These samples were analyzed for the presence of *BCL11B* structural alterations identified in the *BCL11B* leukemia subgroup.

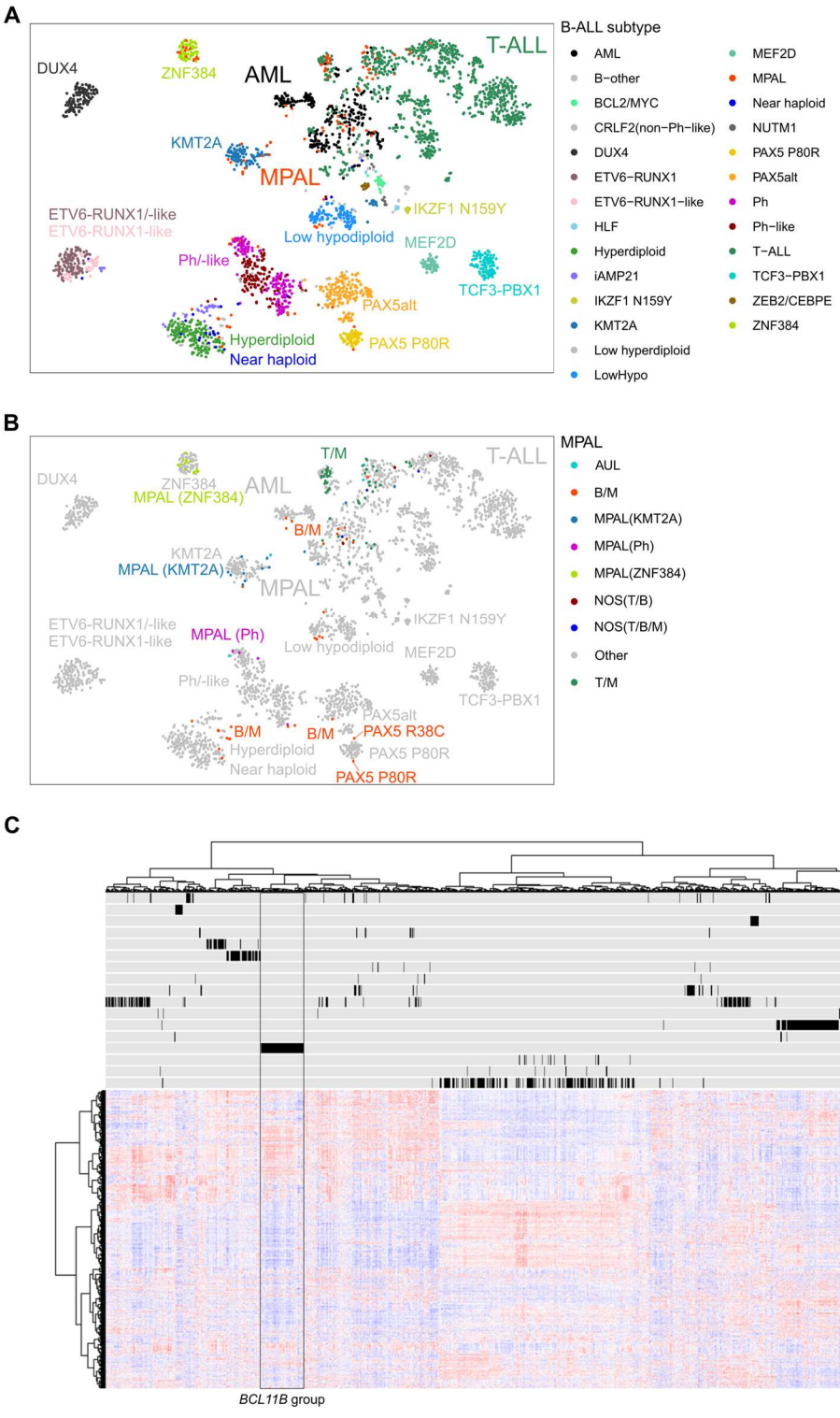
Supplementary Table 11. Sample and clinical information for the *BCL11B* validation cohort.

Supplementary Table 12. Outcome and survival data for 8 *BCL11B* group cases from ECOG with available data.

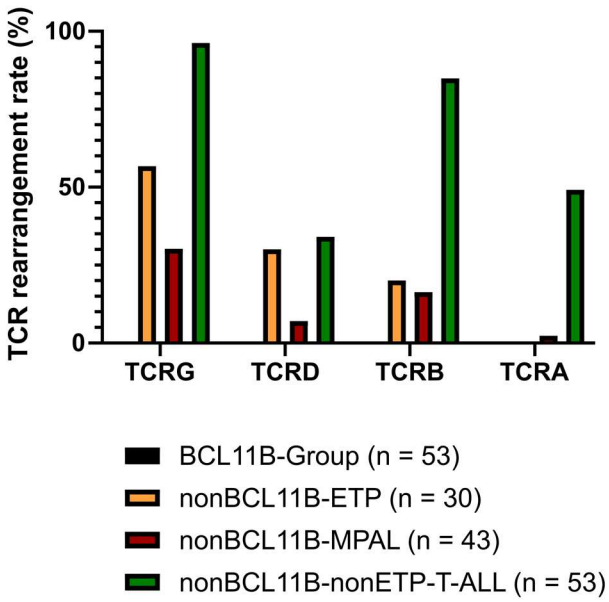
Supplementary Table 13. Long read sequencing analysis of BETA. All reads overlapping the BETA element are listed for SJTALL005006 and SJMPAL011911. The chromosome start and end are indicated along with the start relative to BETA ("unit start" and "unit end"), the number of repeat units ("N BETA units"), and the total read length, are indicated.

Supplementary Table 14. *BCL11B* overexpression in cbCD34+ cells. Expression matrix for human cbCD34+ cells overexpressing *BCL11B* or empty vector.

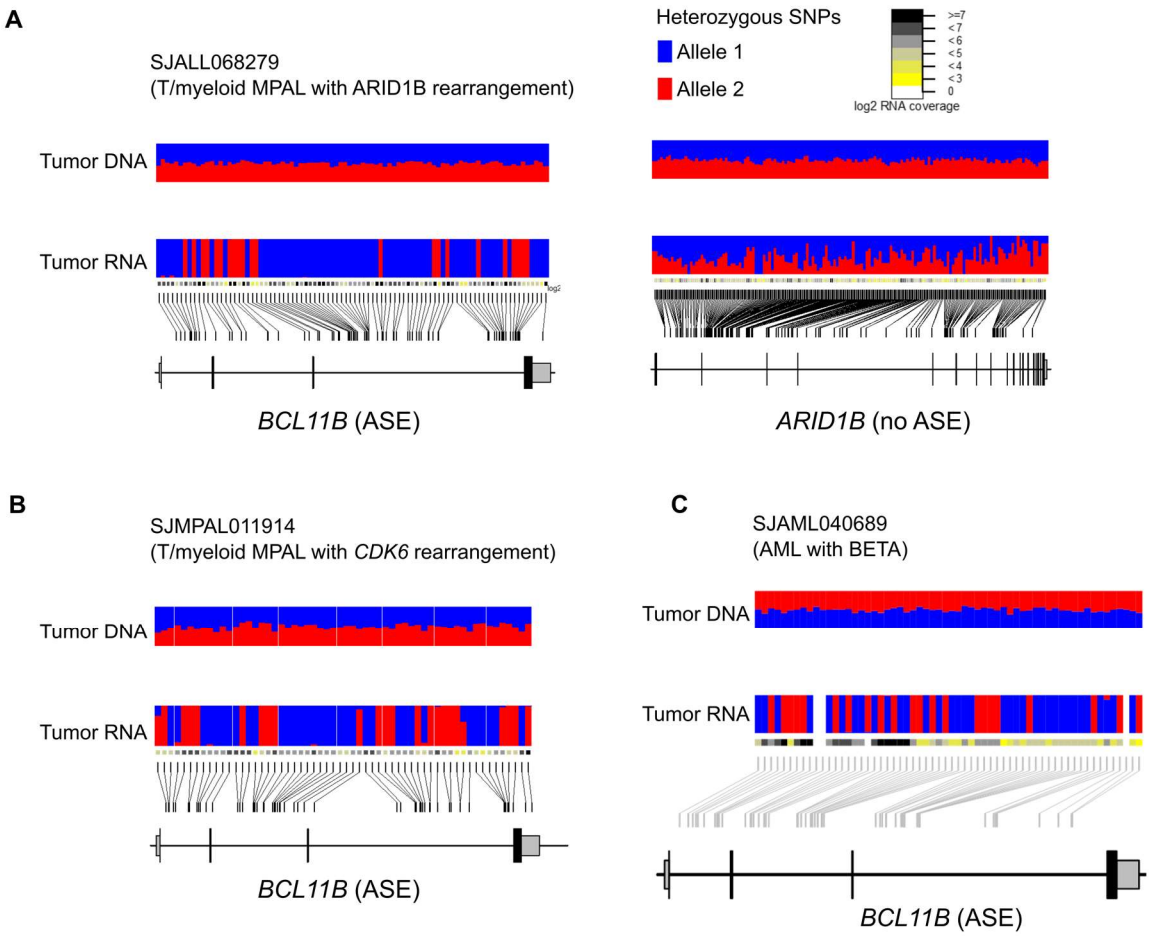
Supplementary Table 15. GSEA analysis of *BCL11B*-transduced cbCD34+ cells. Complete results of GSEA analysis using the ranking of the log2 fold change values of *BCL11B* compared to empty vector as the input.


A
DUX4
ZNF384
AML
T-ALL
KMT2A
MPAL
ETV6-RUNX1/-like
ETV6-RUNX1-like
IKZF1 N159Y
Low hypodiploid
MEF2D
Ph/-like
TCF3-PBX1
PAX5alt
Hyperdiploid
Near haploid
PAX5 P80R
B-ALL subtype
AML
B−other
BCL2/MYC
CRLF2(non−Ph−like)
DUX4
ETV6−RUNX1
ETV6−RUNX1−like
HLF
Hyperdiploid
iAMP21
IKZF1 N159Y
KMT2A
Low hyperdiploid
LowHypo
MEF2D
MPAL
Near haploid
NUTM1
PAX5 P80R
PAX5alt
Ph
Ph−like
T−ALL
TCF3−PBX1
ZEB2/CEBPE
ZNF384
B
DUX4
ZNF384
MPAL (ZNF384)
AML
T/M
T-ALL
B/M
KMT2A
MPAL (KMT2A)
MPAL
IKZF1 N159Y
ETV6-RUNX1/-like
ETV6-RUNX1-like
MPAL (Ph)
Low hypodiploid
MEF2D
Ph/-like
TCF3-PBX1
PAX5alt
B/M
B/M
PAX5 R38C
Hyperdiploid
PAX5 P80R
Near haploid
PAX5 P80R
MPAL
AUL
B/M
MPAL(KMT2A)
MPAL(Ph)
MPAL(ZNF384)
NOS(T/B)
NOS(T/B/M)
Other
T/M
C
NUP98
SPI1
nonTCR−LMO2
SET−NUP214
CBFB−MYH11
RUNX1−RUNX1T1
HOXA
DDX3X−MLLT10
PICALM−MLLT10
KMT2A
NKX2_1
TLX3
TLX1
BCL11B
LMO1/2
TAL2
TAL1
6
4
2
0
−2
−4
−6
z-score
BCL11B group

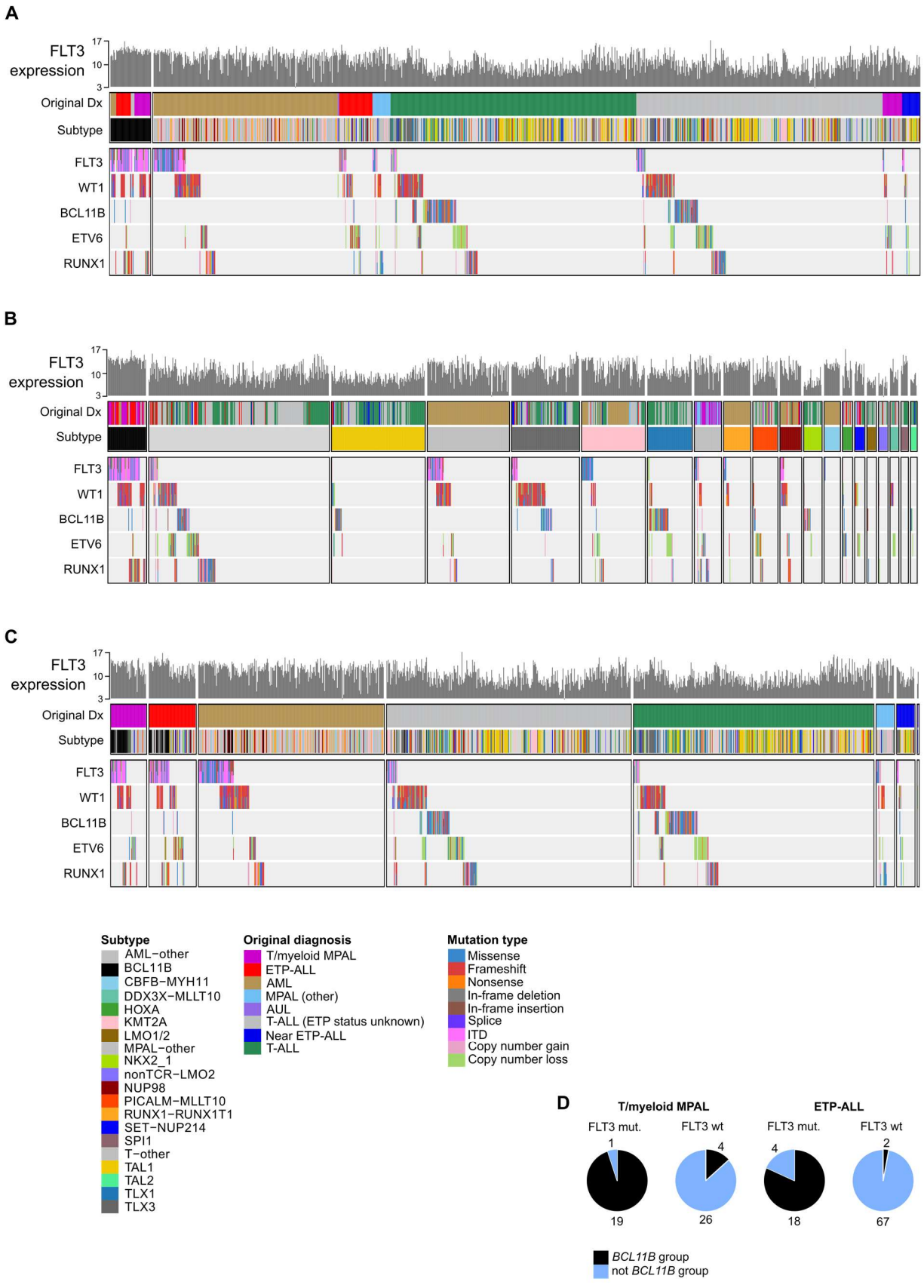
Supplementary Fig. 1. tSNE and hierarchical clustering of pan-leukemia samples. (**A**) tSNE projection analysis showing all leukemia samples analyzed in this study. B-ALL samples are colored based on genomic driver alteration. (**B**) Same projection as in (**A**) but highlighting the MPAL samples to illustrate clustering of B/myeloid MPAL with B-ALL with shared genomic alteration. (**C**) Pearson's correlation distance and Ward's clustering using the top 1000 genes with highest mean absolute deviation (MAD) of all non-B-ALL samples. Heatmap shows the Z-score for the normalized gene expression level. This analysis shows that the *BCL11B* group is stable under hierarchical clustering. iAMP21, intrachromosomal amplification of chromosome 21; AUL, acute undifferentiated leukemia; NOS, not otherwise specified.



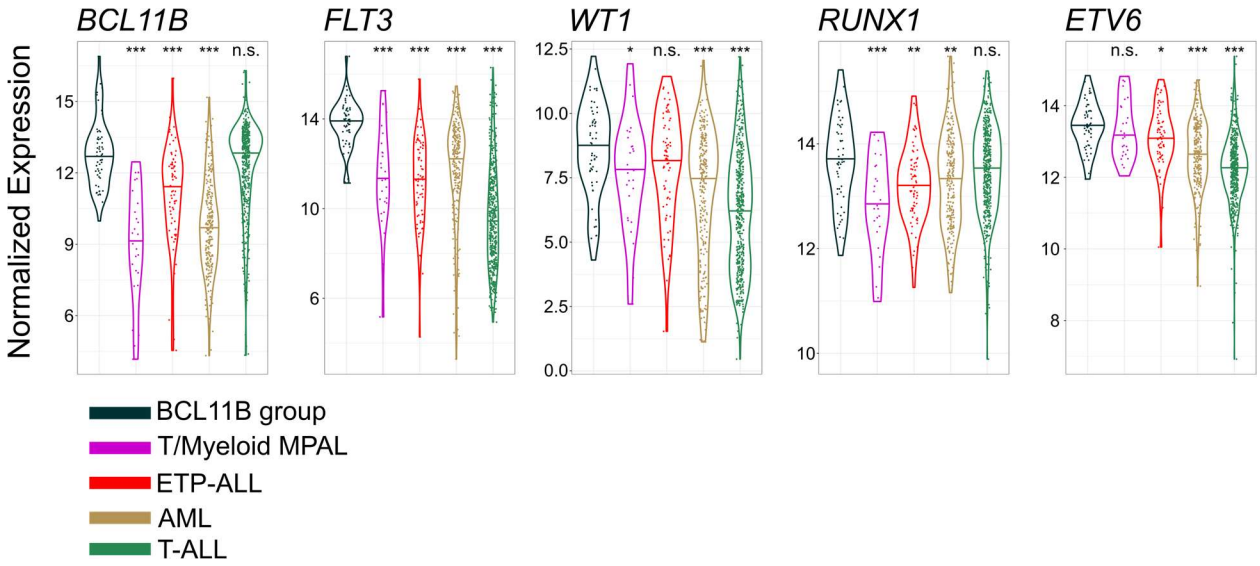
Supplementary Fig. 2. T cell receptor (TCR) gene rearrangement analysis. Bar plot summarizing the frequency of TCR rearrangements detected in different leukemia subgroups. See also Supplementary Table 4.



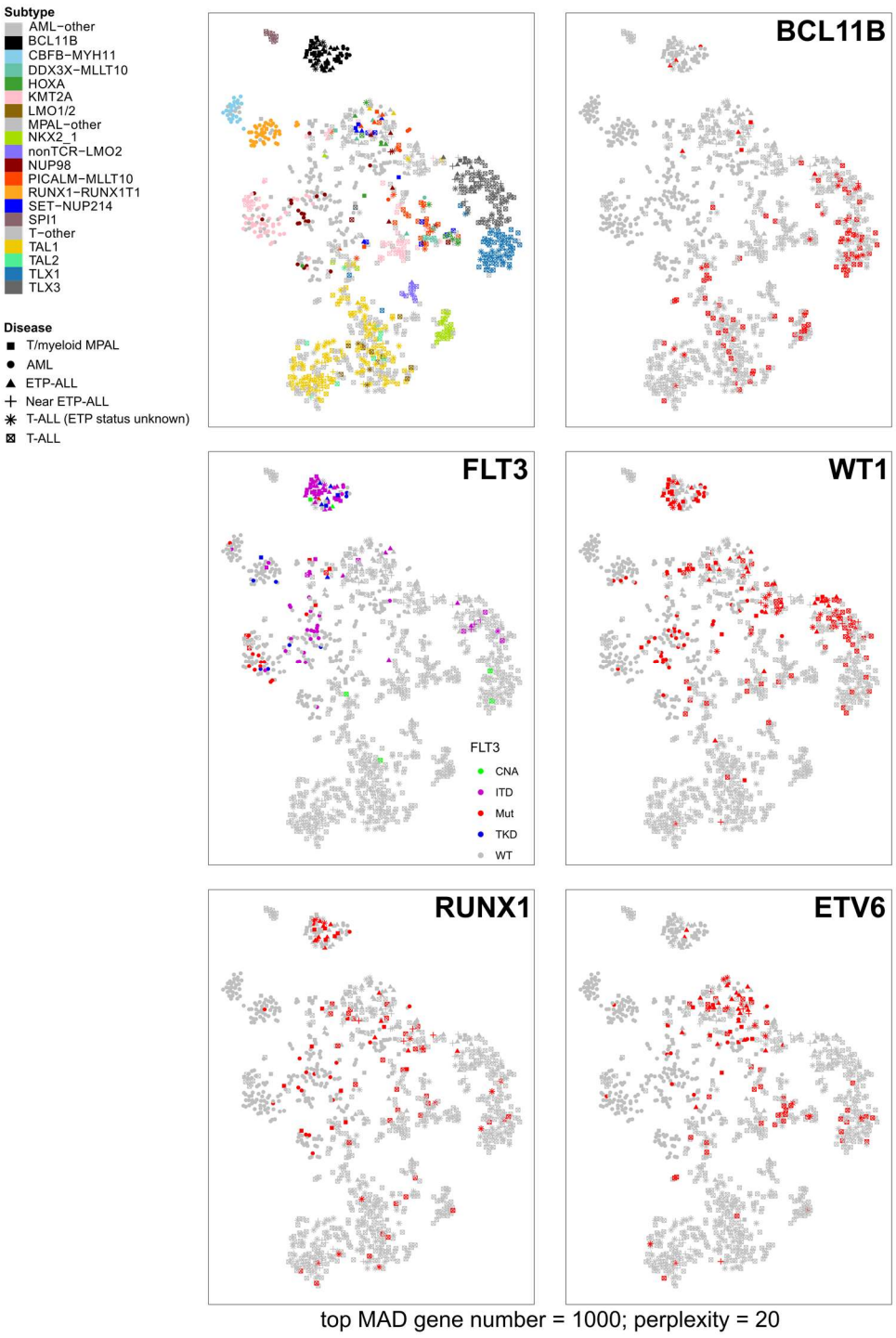
Supplementary Fig. 3. Allele specific expression (ASE) analysis of *BCL11B*. (**A**) ASE plots showing the allele frequency of heterozygous SNPs in the genome (Tumor DNA) compared to *BCL11B* mRNA from RNA-seq (Tumor RNA). Blue and red colors indicate the respective alleles of each heterozygous SNP identified in WGS data. For comparison the ASE data are shown for *ARID1B* which does not exhibit biased allele expression at the RNA level in comparison to *BCL11B*. (**B,C**) Same as in (**A**).


A
FLT3 expression
17
10
3
Original Dx
Subtype
FLT3
WT1
BCL11B
ETV6
RUNX1
B
FLT3 expression
17
10
3
Original Dx
Subtype
FLT3
WT1
BCL11B
ETV6
RUNX1
C
FLT3 expression
17
10
3
Original Dx
Subtype
FLT3
WT1
BCL11B
ETV6
RUNX1
Subtype
AML−other
BCL11B
CBFB−MYH11
DDX3X−MLLT10
HOXA
KMT2A
LMO1/2
MPAL−other
NKX2_1
nonTCR−LMO2
NUP98
PICALM−MLLT10
RUNX1−RUNX1T1
SET−NUP214
SPI1
T−other
TAL1
TAL2
TLX1
TLX3
Original diagnosis
T/myeloid MPAL
ETP-ALL
AML
MPAL (other)
AUL
T-ALL (ETP status unknown)
Near ETP-ALL
T-ALL
Mutation type
Missense
Frameshift
Nonsense
In-frame deletion
In-frame insertion
Splice
ITD
Copy number gain
Copy number loss
D
T/myeloid MPAL
FLT3 mut.
FLT3 wt
1
4
19
26
ETP-ALL
FLT3 mut.
FLT3 wt
4
2
18
67
BCL11B group
not BCL11B group

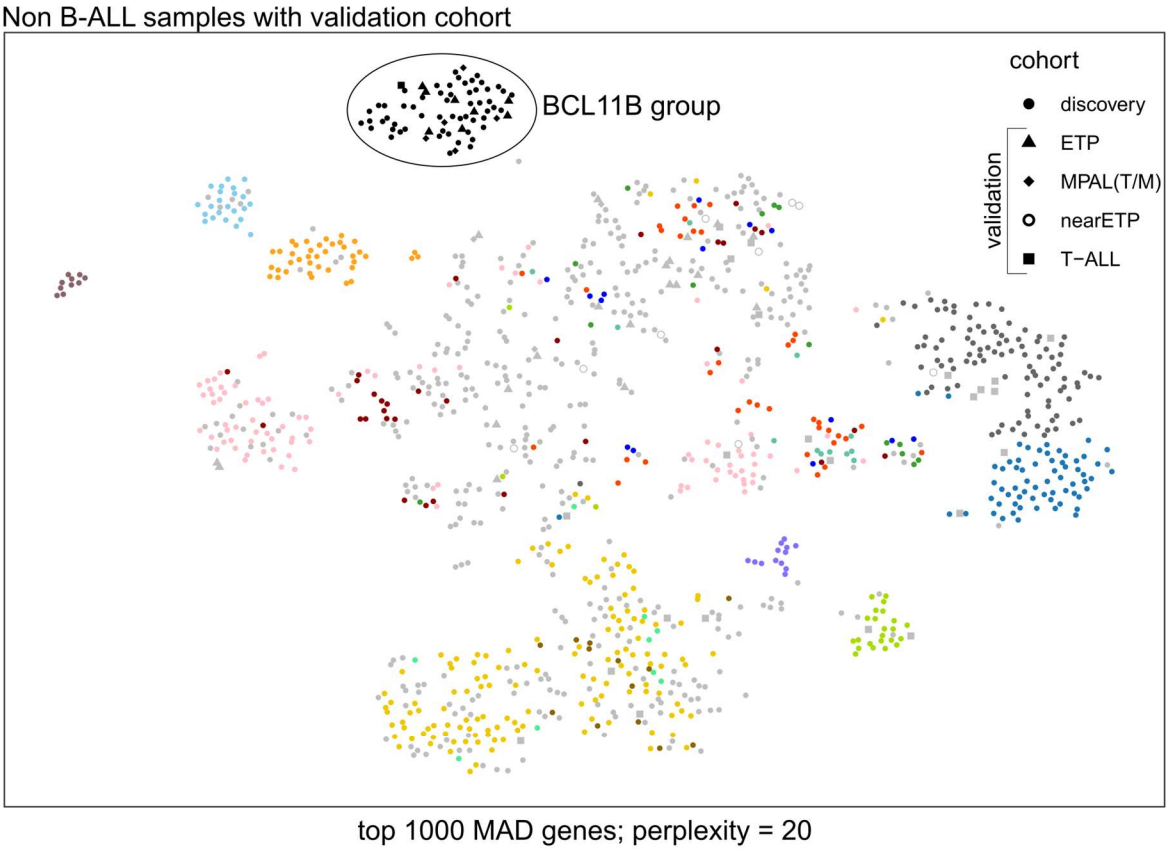
Supplementary Fig. 4. Oncoprint of key lesions across the entire non-B leukemia cohort. (**A**) Samples are grouped based on assignment to the *BCL11B* group, followed by original diagnosis. (**B**) Samples are grouped according to molecular subtype. (**C**) Samples are grouped according to original diagnosis. The top track shows variance stabilized expression values (vst) for *FLT3*. Only protein-coding mutations and copy number alterations impacting protein coding genes are shown; the *BCL11B*-deregulating SVs defining the *BCL11B* group leukemias are not shown in this oncoprint. (**D**) Pie charts showing number of T/myeloid MPAL and ETP-ALL samples belonging to the *BCL11B* group based on presence of activating (ITD or TKD) *FLT3* mutations. ($p < 0.0001$ for *FLT3*-mutated samples belonging to the *BCL11B* group, Fisher's exact test).



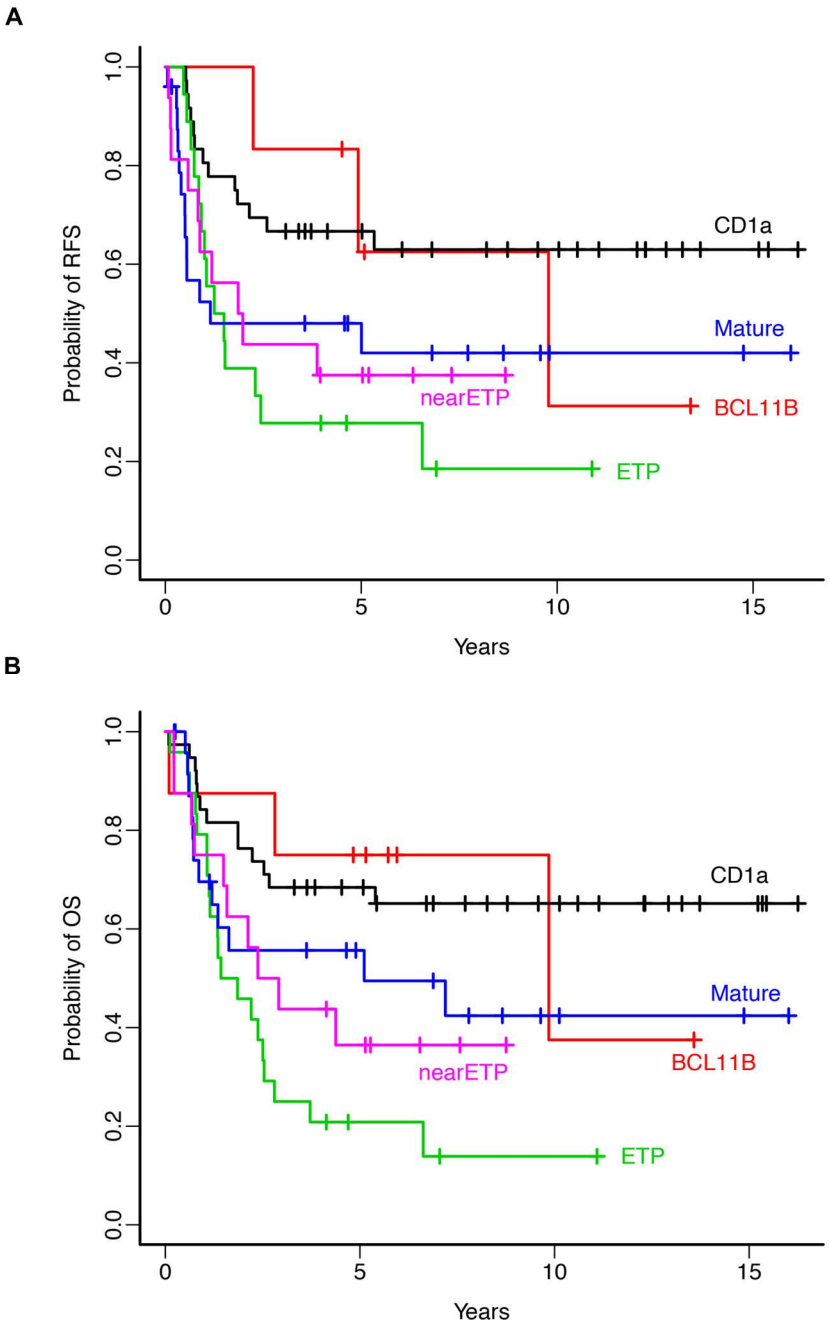
Supplementary Fig. 5. Expression of key genes in the *BCL11B* group compared to T/myeloid MPAL, ETP-ALL, AML and T-ALL samples that are not in the *BCL11B* expression group. Violin plots showing variance stabilized expression (vst) values for each group. Horizontal line indicates the median. Significance of expression differences between the *BCL11B* group and all other groups was assessed with a Wilcoxon rank-sum test (*p<0.05, **p<0.01, ***p<0.001).



Supplementary Fig. 6. Distribution of key lesions across the non-B leukemia cohort. Shapes correspond to original diagnosis. *FLT3* alterations are colored according to mutation type (internal tandem duplication (ITD), tyrosine kinase domain (TKD), other mutation (Mut), copy number alteration (CNA) or wildtype (WT). For *BCL11B*, *WT1*, *RUNX1*, and *ETV6*, samples are colored red if any alteration was identified (see also Supplementary Tables 6,7). Mutations called from DNA and/or RNA-based sequencing are shown.

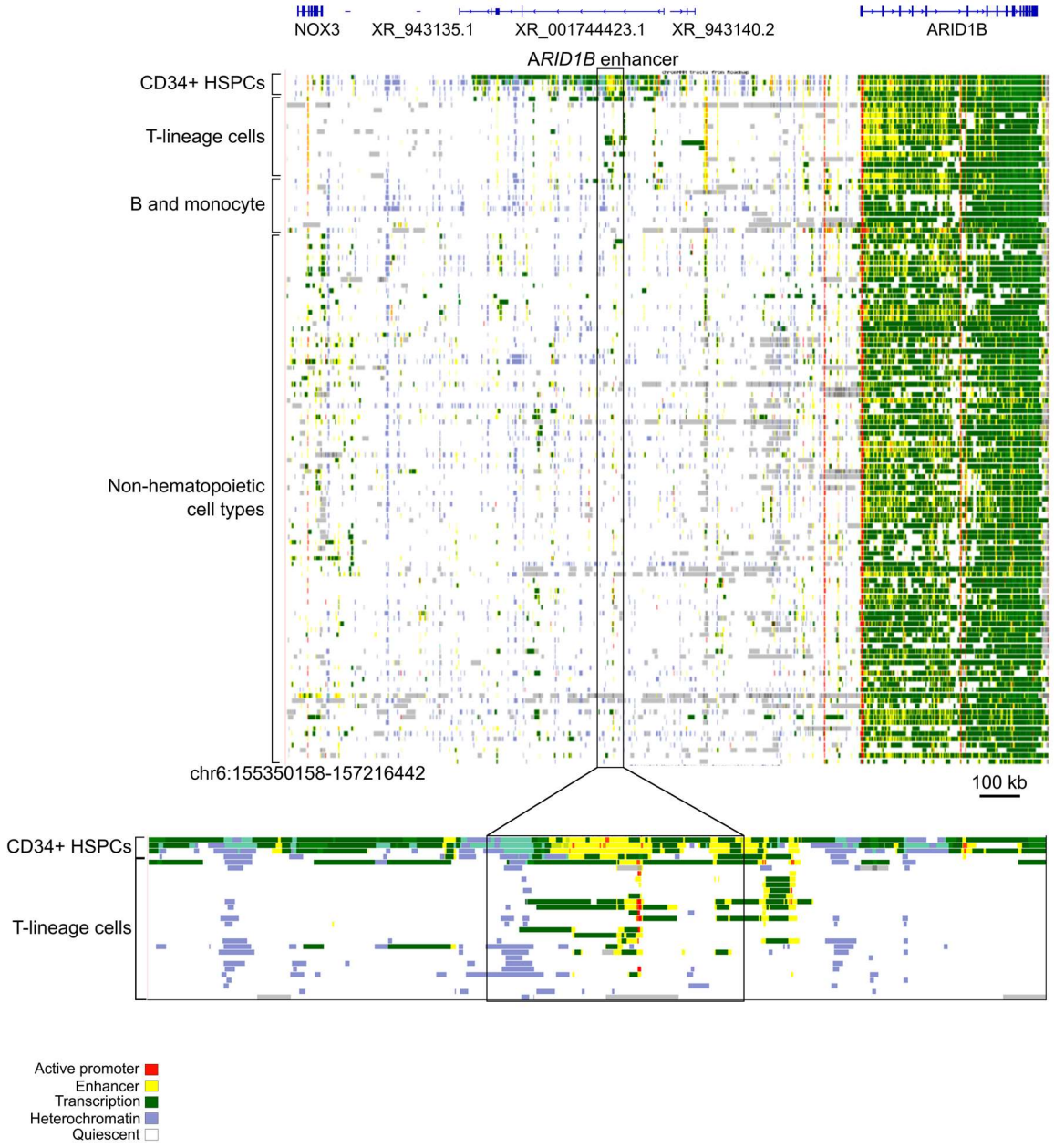


Supplementary Fig. 7. Validation of the *BCL11B* gene expression group. tSNE projection analysis of 70 T lineage leukemia samples from an independent cohort are shown together with the discovery cohort. 14 of 70 cases in this independent cohort cluster with the *BCL11B* group (discovery cohort samples are shown as filled circles; all other shapes correspond to the validation cohort). MAD=mean absolute deviation. See also Supplementary Table 11.

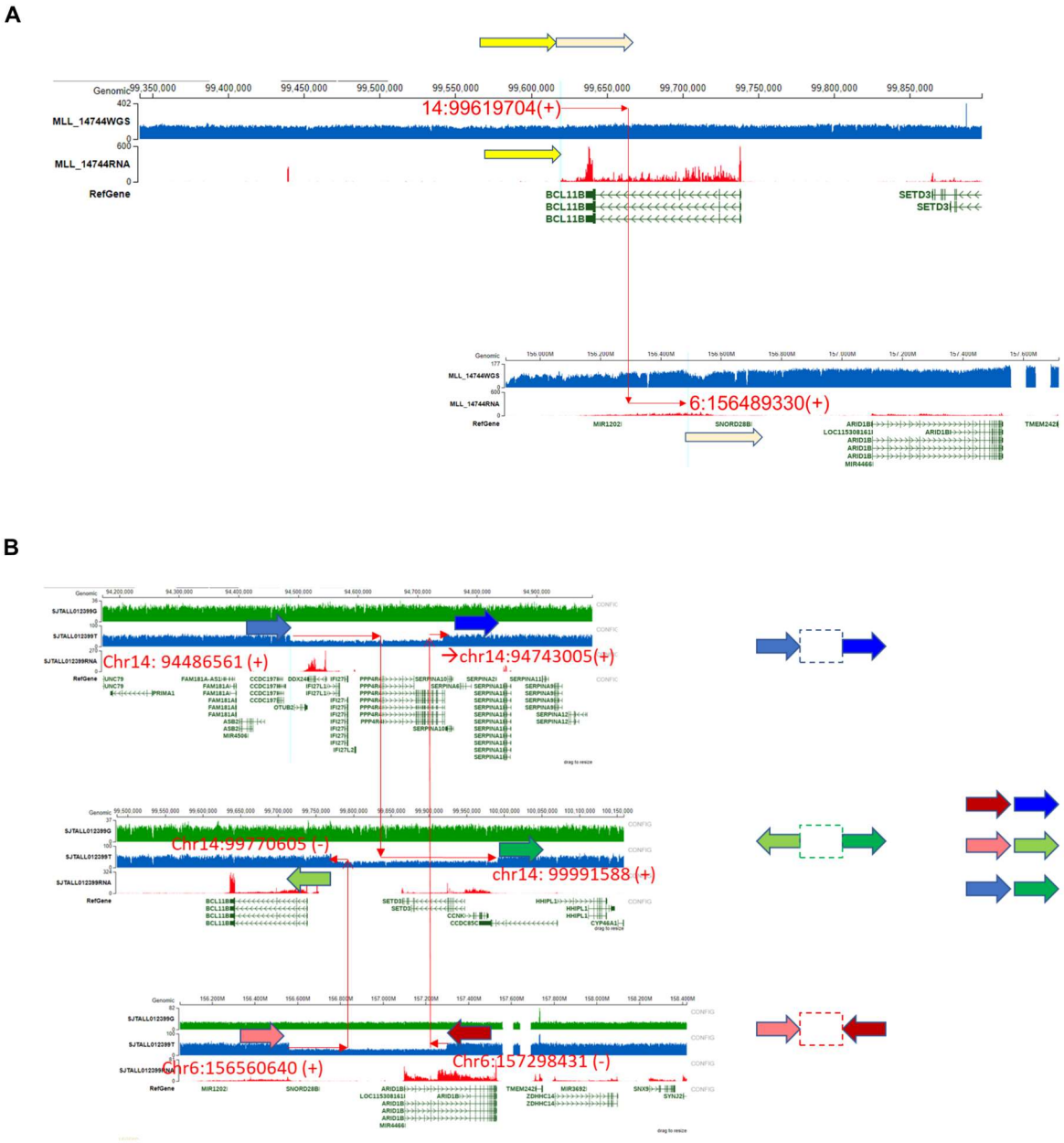


Supplementary Fig. 8. Kaplan-Meier survival curves for the *BCL11B* group compared to other T-ALL groups. (**A**) Relapse-free survival (RFS) and (**B**) overall survival (OS) of 8 *BCL11B* group cases compared to non-*BCL11B* group T-ALL from the ECOG-ACRIN/CALGB validation cohort. CD1a: CD1a positive T-ALL irrespective of other antigens; Mature: T-ALL with surface CD3 and negative for CD1a or surface CD3, but positive for gamma/delta or alpha/beta receptor; nearETP: T-ALL cases negative for surface CD3, CD4, CD8 but positive for CD5 and myeloid antigens; ETP: as for nearETP, but negative/dim for CD5.

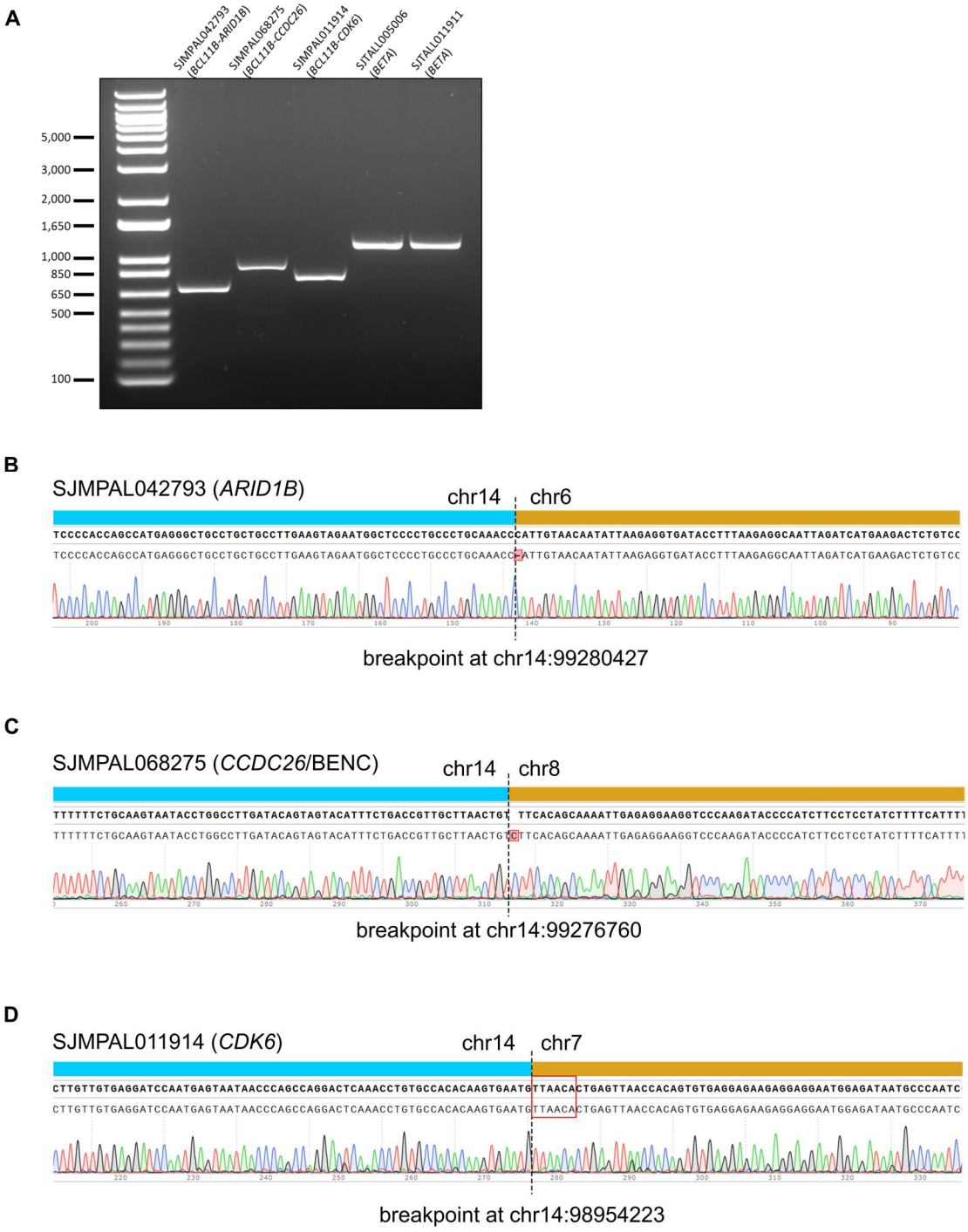
A


NOX3
XR_943135.1
XR_001744423.1
XR_943140.2
ARID1B
ARID1B enhancer
CD34+ HSPCs
T-lineage cells
B and monocyte
Non-hematopoietic cell types
chr6:155350158-157216442
100 kb
CD34+ HSPCs
T-lineage cells
Active promoter
Enhancer
Transcription
Heterochromatin
Quiescent

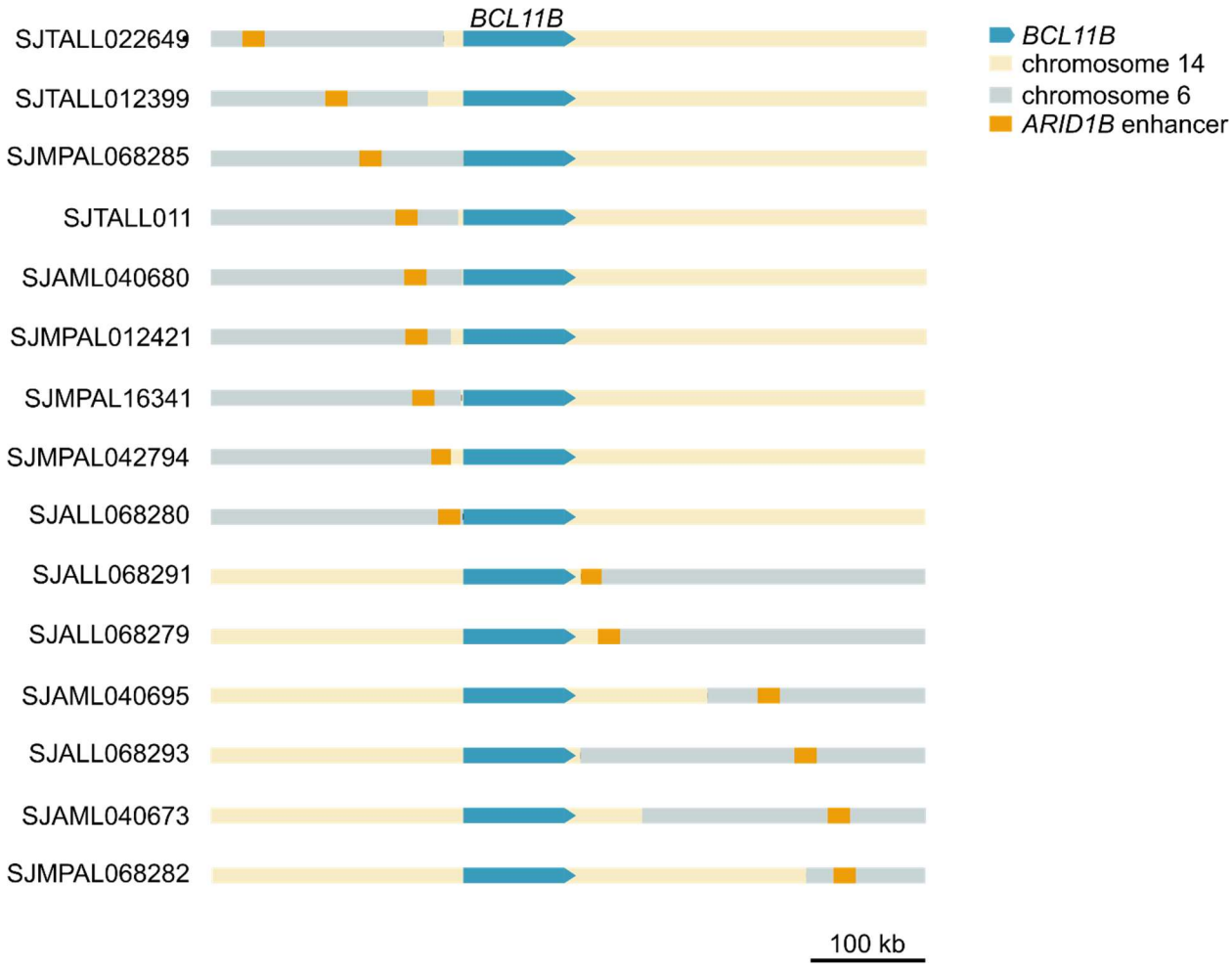
Supplementary Fig. 9. Chromatin state of the *ARID1B* and *CCDC26* genomic regions involved in *BCL11B* rearrangements. chromHMM (1) annotations are indicated and show that the *ARID1B* enhancer (**A**) is specific to CD34+ HSPC cell types (yellow), whereas it is greatly diminished in more differentiated T lineage and other hematopoietic cell types. Similarly, the BENC enhancer (**B**) is most prominent in CD34+ HSPC cells and is completely absent in other T lineage cells. All cell types shown are from the Roadmap Epigenomics Project (2).



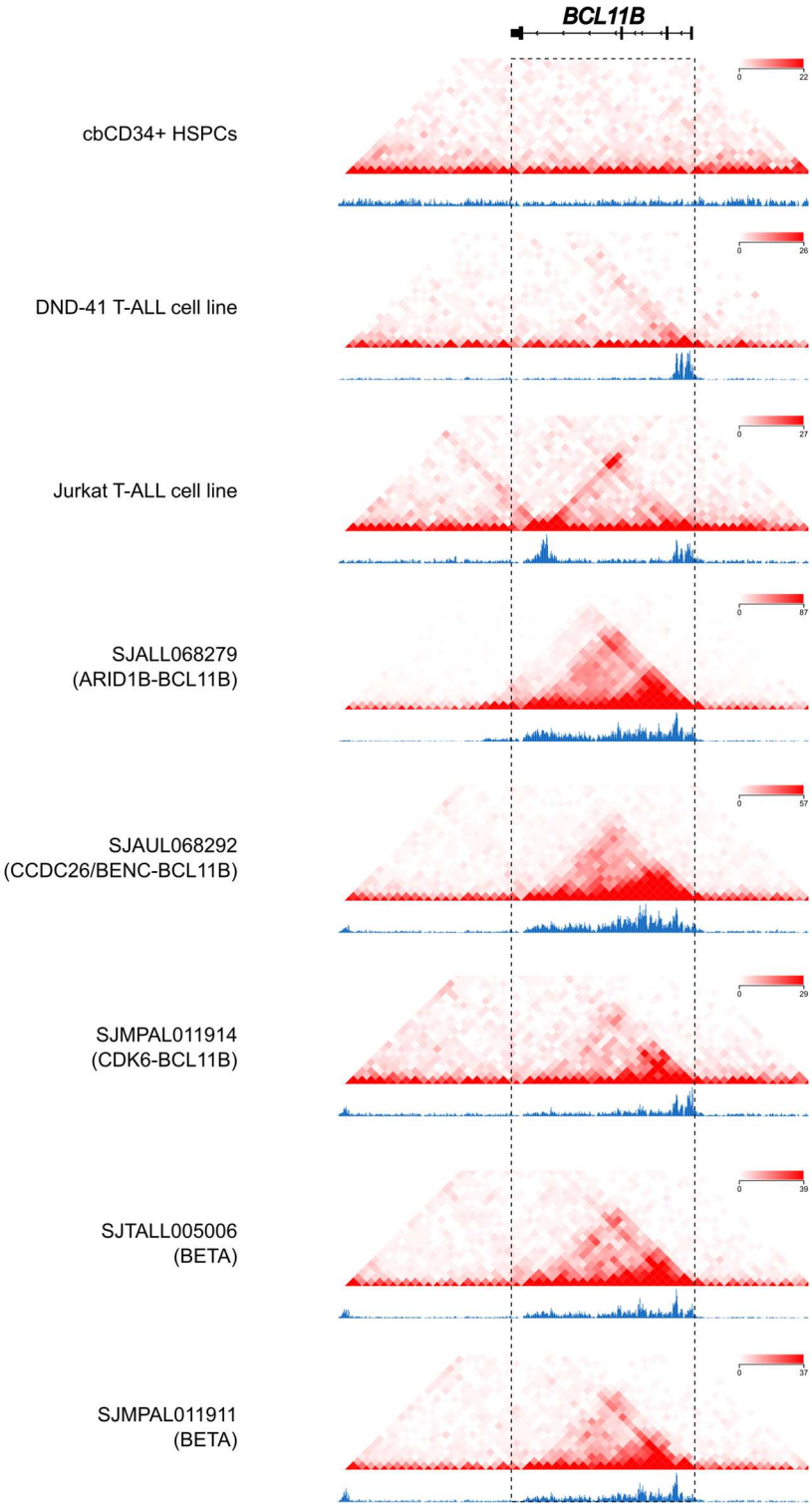
Supplementary Fig. 10. Detailed representation of SV rearrangements from two samples. (**A**) Sample SJALL068279 (*ARID1B-BCL11B*) WGS data are shown as coverage tracks in blue. Red lines indicate the orientation of the chr6-chr14 translocation. Yellow arrowheads indicate the resulting configuration relative to *BCL11B*. (**B**) Same as in (**A**) but for sample SJTALL012399 (*ARID1B-BCL11B*) with a complex SV and resulting genomic configuration.



Supplementary Fig. 11. Sanger sequencing of breakpoint junctions. (**A**) Agarose gel electrophoresis of the PCR product of breakpoint junctions in each sample. See Methods for primer sequences. (**B-D**) Sanger sequencing chromatogram showing the PCR product spanned the predicted SV junction. Blue bars indicate the *BCL11B*/chr14 side of each junction; gold bars indicate the enhancer donor side of each junction. Red box in (**D**) highlights a somatically acquired insertion.



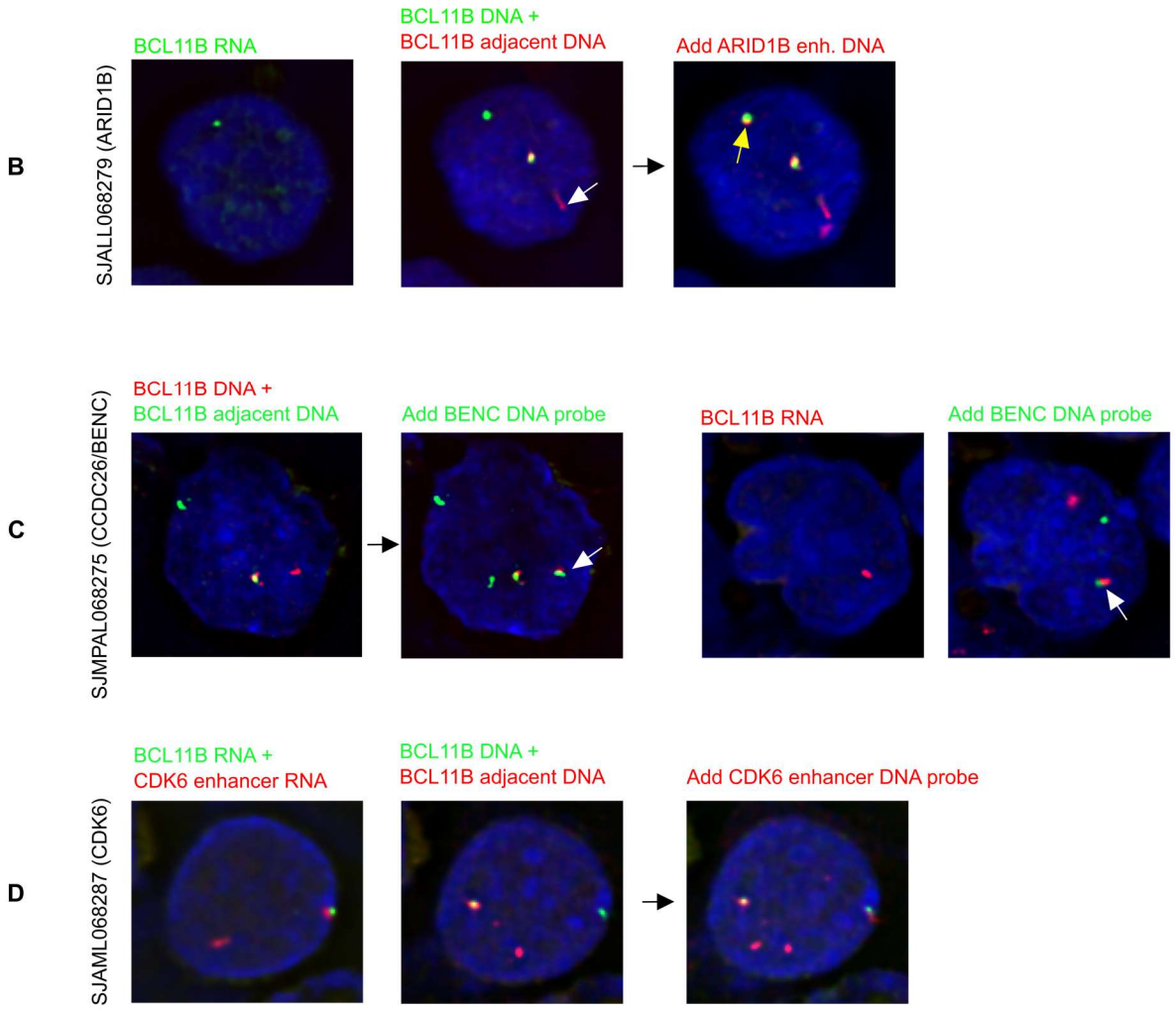
Supplementary Fig. 12. Genomic organization of the *BCL11B* locus and its upstream and downstream region following chromosome 6 (*ARID1B*) rearrangements. *ARID1B* rearrangement was manually inspected in 15 cases to confirm the orientation and allelic configuration of the rearranged *BCL11B* allele. All genomic features are shown relative to the 5' position of the *BCL11B* coding sequence.



Supplementary Fig. 13. Local chromatin organization at the *BCL11B* locus. H3K27ac HiChIP data are displayed as a heatmap of pair-wise 5kb interacting bins for healthy normal cbCD34+ HSPCs, two T-ALL cell lines, and 5 primary *BCL11B*-deregulated leukemia samples. The 1D coverage signal is shown below in blue. These images show that the *BCL11B* locus is uniquely organized into a small topological domain mediated by H3K27ac in primary leukemia cells with *BCL11B* deregulating SVs.

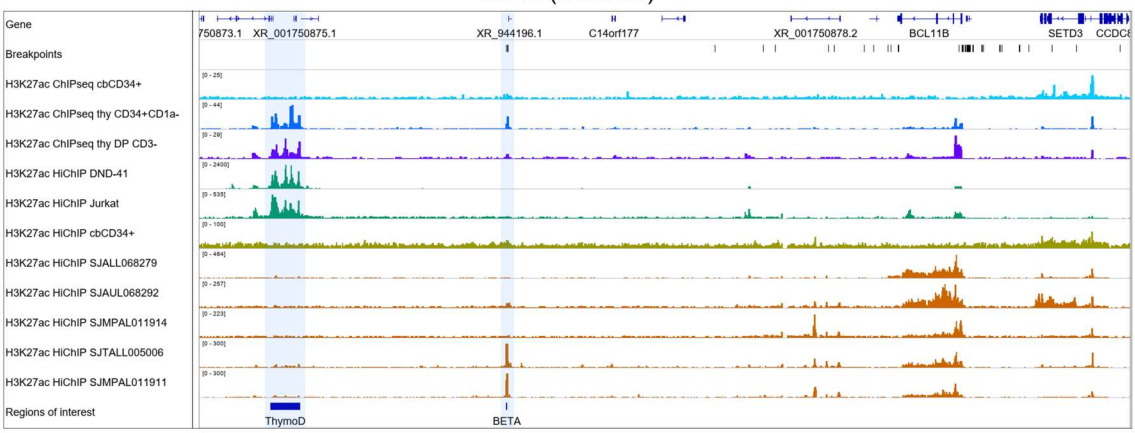
A

Patient Sample ID	Total cells analyzed	Cells with SV	Cells expressing hijacked *BCL11B* allele	Cells expressing wildtype *BCL11B* allele	Cells lacking hijacked enhancer
SJALL068279	111	83	37	2	28
SJMPAL068275	165	140	93	0	25
SJAML068287	104	88	51	0	16

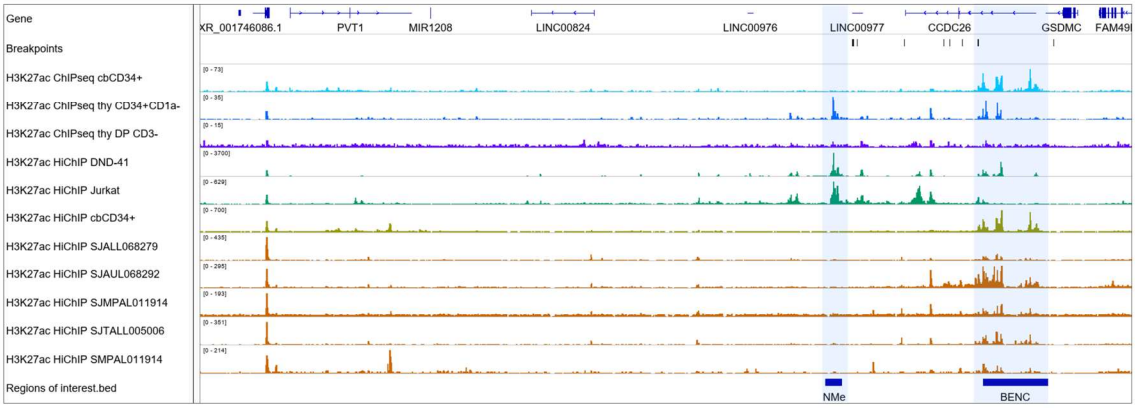


Supplementary Fig. 14. DNA and RNA FISH of representative *BCL11B* group samples. (**A**) Summary of the number of cells showing evidence of enhancer hijacking for each of the following samples (**B-E**). (**B**) An *ARID1B-BCL11B* rearranged case showing *BCL11B* mRNA FISH signal in only one allele (left). Break-away fosmid clones were used to demonstrate that the immediate upstream region of the expressed *BCL11B* allele is no longer linked (white arrow). Additional of an *ARID1B* enhancer fosmid shows this region in proximity to the expressed *BCL11B* allele (yellow arrow).(**C**) A *CCDC26*/BENC rearranged case showing that the BENC enhancer DNA probe colocalizes with the *BCL11B* DNA and mRNA probes. (**D**) A *CDK6* rearranged case showing that the expressed allele of *BCL11B* colocalizes with *CDK6*-derived enhancer RNA.

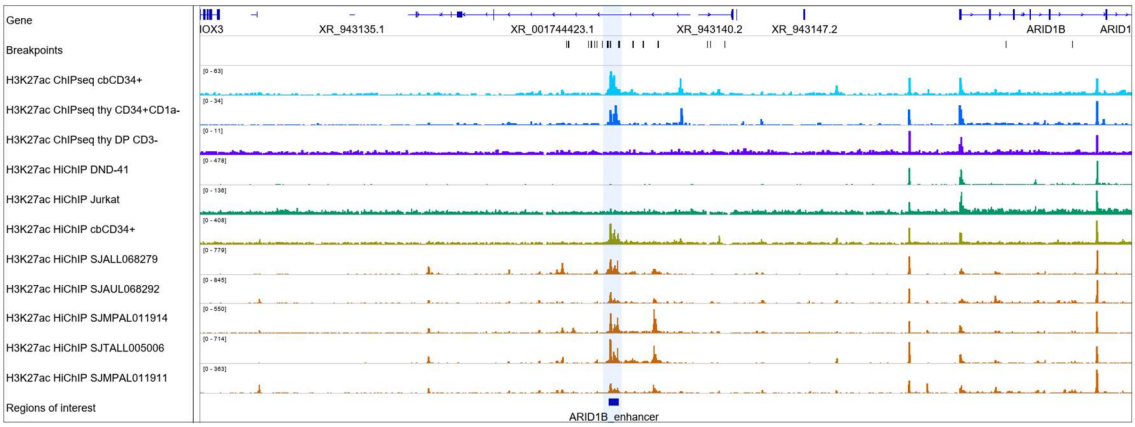
A


chr14 (*BCL11B*)
Gene
Breakpoints
H3K27ac ChIPseq cbCD34+
H3K27ac ChIPseq thy CD34+CD1a-
H3K27ac ChIPseq thy DP CD3-
H3K27ac HiChIP DND-41
H3K27ac HiChIP Jurkat
H3K27ac HiChIP cbCD34+
H3K27ac HiChIP SJALL068279
H3K27ac HiChIP SJAUL068292
H3K27ac HiChIP SJMPAL011914
H3K27ac HiChIP SJTALL005006
H3K27ac HiChIP SJMPAL011911
Regions of interest
750873.1
XR_001750875.1
XR_944196.1
C14orf177
XR_001750878.2
BCL11B
SETD3
ThymoD
BETA

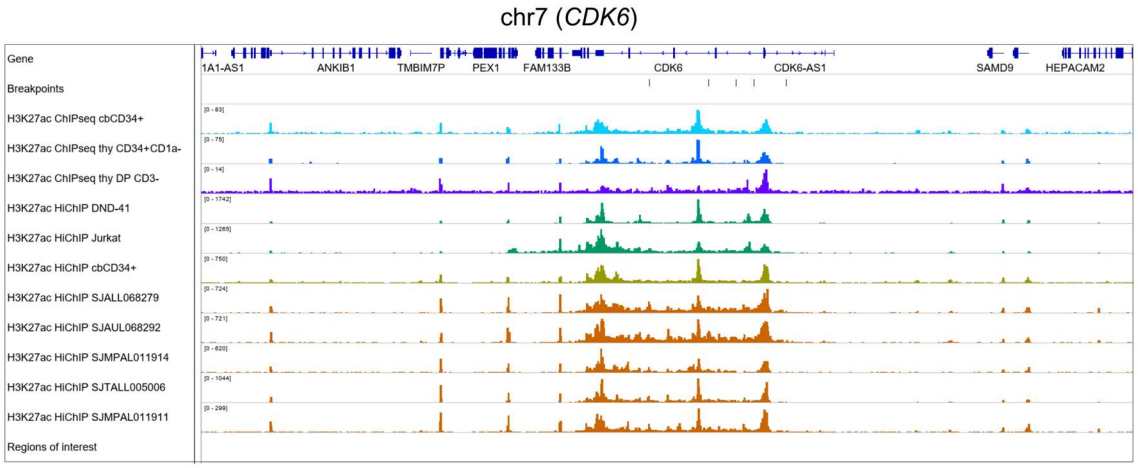
B


chr8 (*CCDC26*/BENC)
Gene
Breakpoints
H3K27ac ChIPseq cbCD34+
H3K27ac ChIPseq thy CD34+CD1a-
H3K27ac ChIPseq thy DP CD3-
H3K27ac HiChIP DND-41
H3K27ac HiChIP Jurkat
H3K27ac HiChIP cbCD34+
H3K27ac HiChIP SJALL068279
H3K27ac HiChIP SJAUL068292
H3K27ac HiChIP SJMPAL011914
H3K27ac HiChIP SJTALL005006
H3K27ac HiChIP SMPAL011914
Regions of interest.bed
XR_001746086.1
PVT1
MIR1208
LINC00824
LINC00976
LINC00977
CCDC26
GSDMC
NMe
BENC

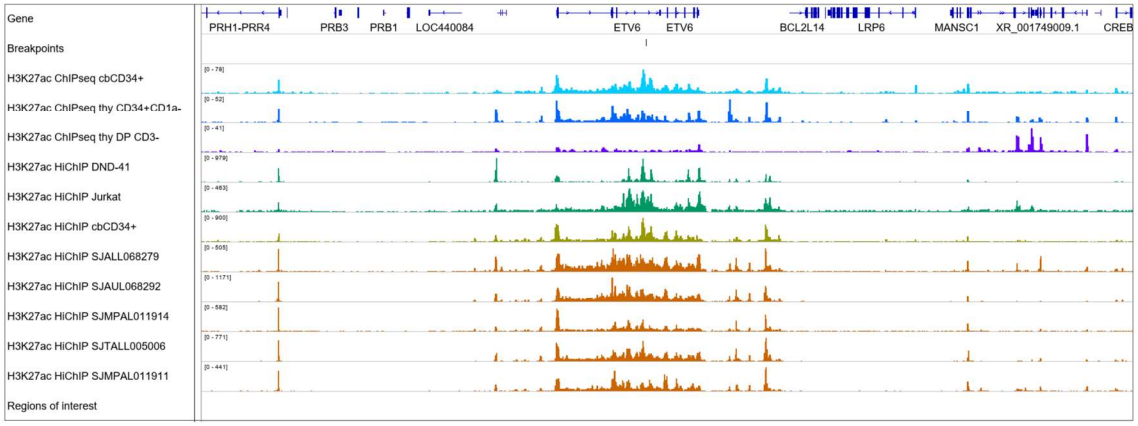
C


chr6 (*ARID1B*)
Gene
Breakpoints
H3K27ac ChIPseq cbCD34+
H3K27ac ChIPseq thy CD34+CD1a-
H3K27ac ChIPseq thy DP CD3-
H3K27ac HiChIP DND-41
H3K27ac HiChIP Jurkat
H3K27ac HiChIP cbCD34+
H3K27ac HiChIP SJALL068279
H3K27ac HiChIP SJAUL068292
H3K27ac HiChIP SJMPAL011914
H3K27ac HiChIP SJTALL005006
H3K27ac HiChIP SJMPAL011911
Regions of interest
IOX3
XR_943135.1
XR_001744423.1
XR_943140.2
XR_943147.2
ARID1B
ARID1B_enhancer

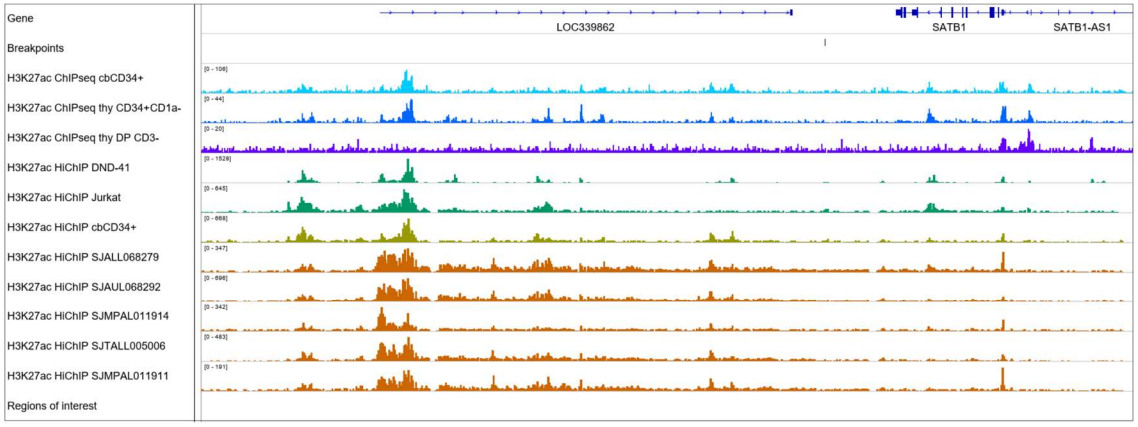
D

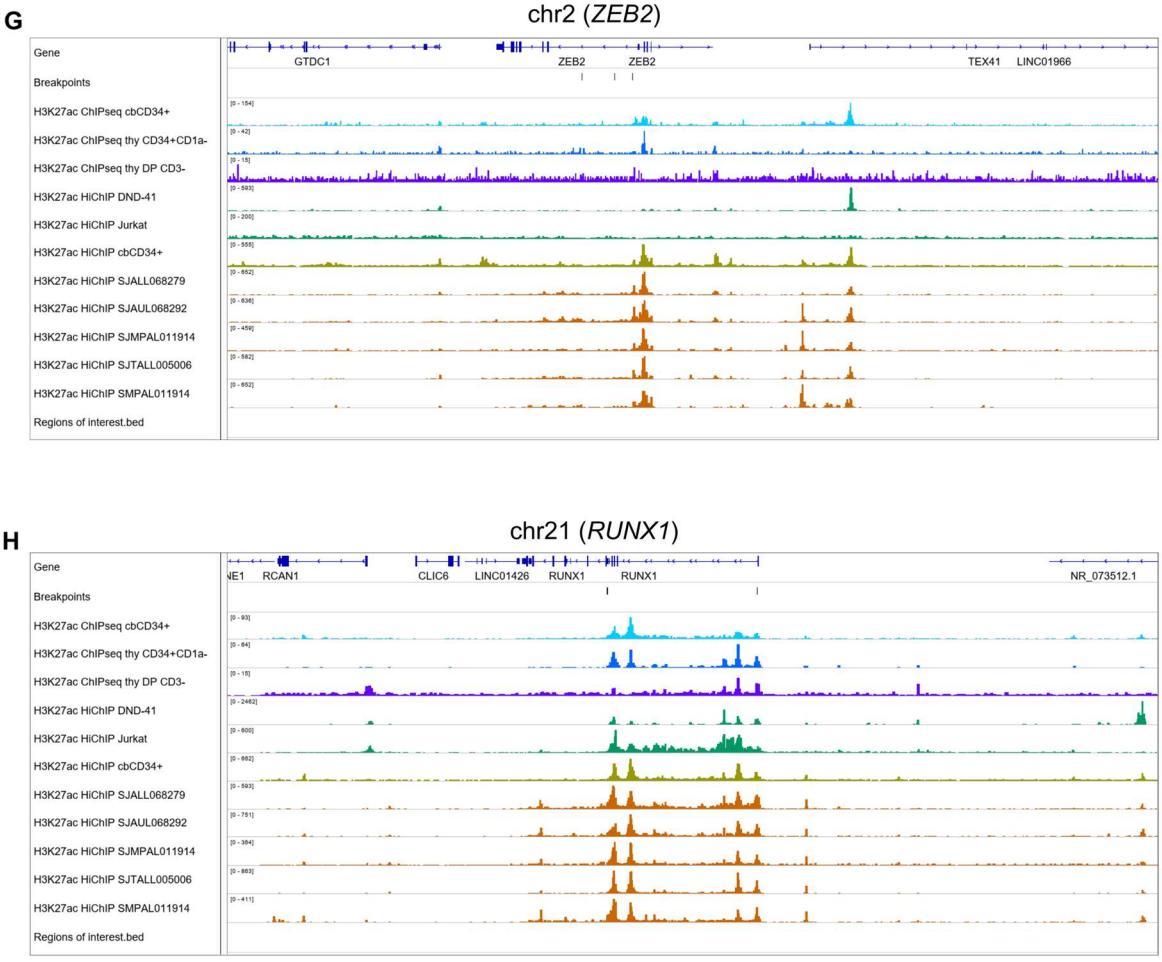

chr7 (CDK6)
Gene
Breakpoints
H3K27ac ChIPseq cbCD34+
H3K27ac ChIPseq thy CD34+CD1a-
H3K27ac ChIPseq thy DP CD3-
H3K27ac HiChIP DND-41
H3K27ac HiChIP Jurkat
H3K27ac HiChIP cbCD34+
H3K27ac HiChIP SJALL068279
H3K27ac HiChIP SJAUL068292
H3K27ac HiChIP SJMPAL011914
H3K27ac HiChIP SJTALL005006
H3K27ac HiChIP SJMPAL011911
Regions of interest
1A1-AS1
ANKIB1
TMBIM7P
PEX1
FAM133B
CDK6
CDK6-AS1
SAMD9
HEPACAM2

E

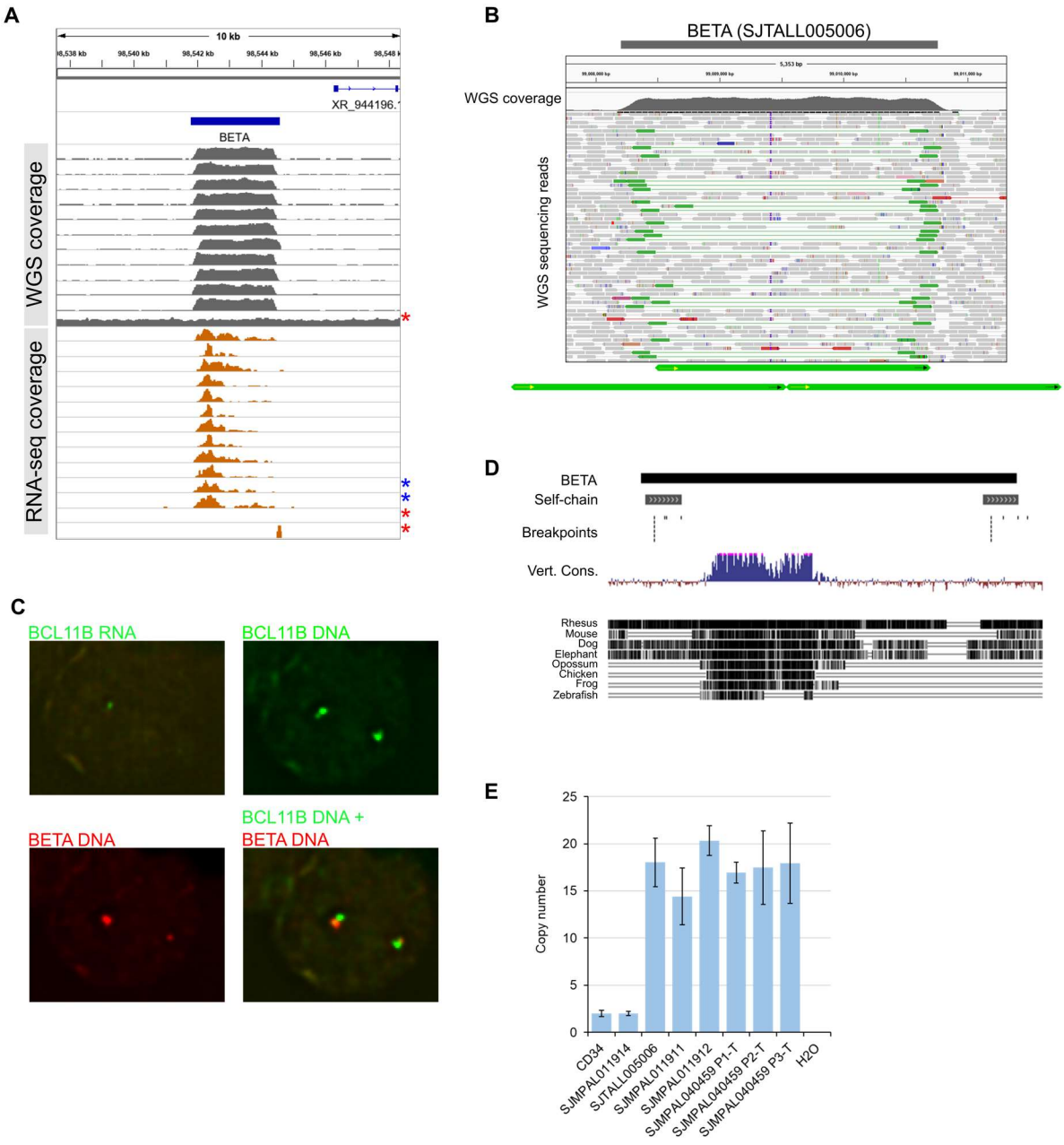

chr12 (ETV6)
Gene
Breakpoints
H3K27ac ChIPseq cbCD34+
H3K27ac ChIPseq thy CD34+CD1a-
H3K27ac ChIPseq thy DP CD3-
H3K27ac HiChIP DND-41
H3K27ac HiChIP Jurkat
H3K27ac HiChIP cbCD34+
H3K27ac HiChIP SJALL068279
H3K27ac HiChIP SJAUL068292
H3K27ac HiChIP SJMPAL011914
H3K27ac HiChIP SJTALL005006
H3K27ac HiChIP SJMPAL011911
Regions of interest
PRH1-PRR4
PRB3
PRB1
LOC440084
ETV6
ETV6
BCL2L14
LRP6
MANSC1
XR_001749009.1
CREB

F

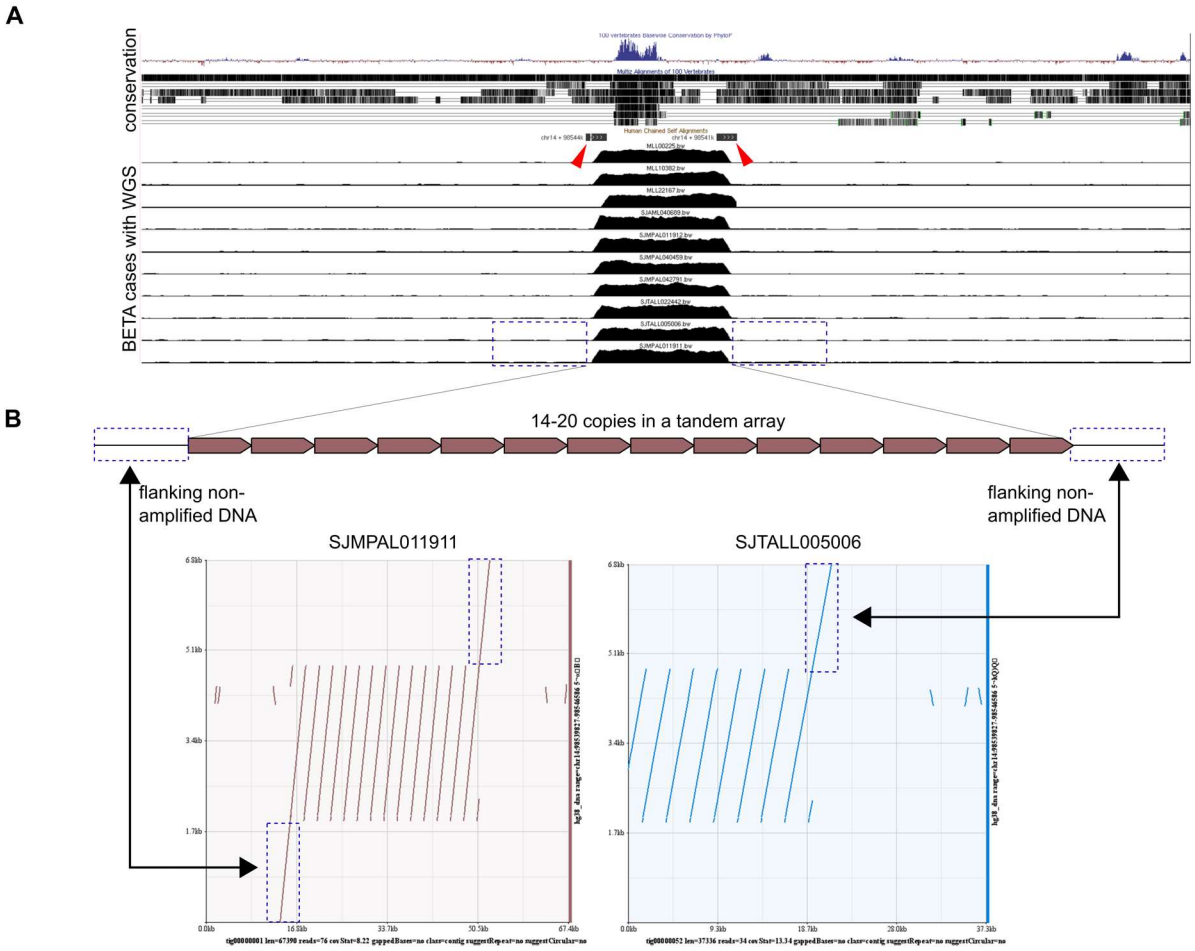

chr3 (SATB1)
Gene
Breakpoints
H3K27ac ChIPseq cbCD34+
H3K27ac ChIPseq thy CD34+CD1a-
H3K27ac ChIPseq thy DP CD3-
H3K27ac HiChIP DND-41
H3K27ac HiChIP Jurkat
H3K27ac HiChIP cbCD34+
H3K27ac HiChIP SJALL068279
H3K27ac HiChIP SJAUL068292
H3K27ac HiChIP SJMPAL011914
H3K27ac HiChIP SJTALL005006
H3K27ac HiChIP SJMPAL011911
Regions of interest
LOC339862
SATB1
SATB1-AS1



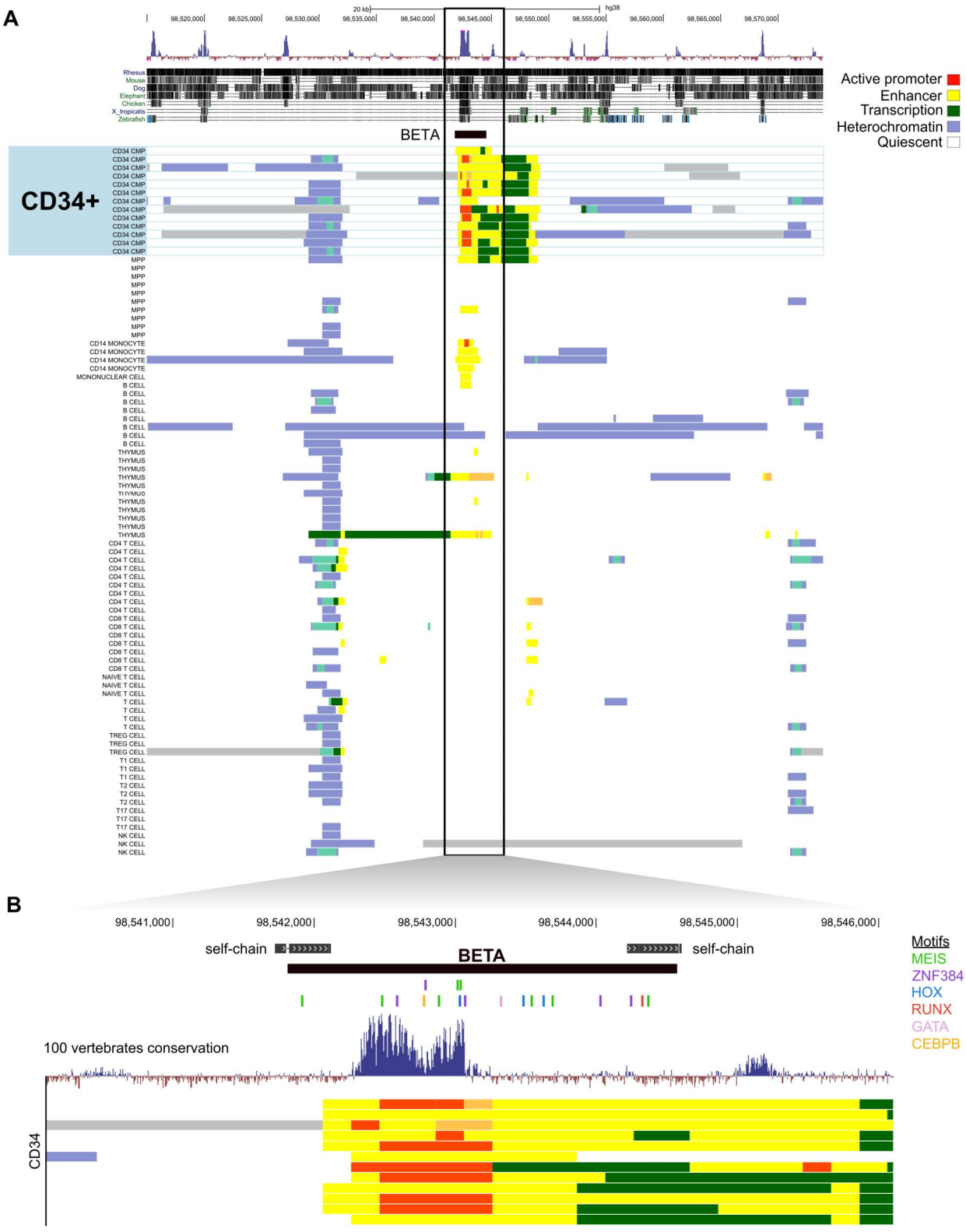
Supplementary Fig. 15. H3K27ac ChIP-seq and HiChIP of all samples showing the *BCL11B* rearrangement partner loci. (**A-H**) Each panel shows H3K27ac ChIP-seq and H3K27ac HiChIP coverage tracks for all leukemia samples analyzed, DND-41 and Jurkat T-ALL cell lines, normal cbCD34+ HSPCs, and two normal thymic populations (CD34+CD1a- progenitors and committed double positive (DP) CD3- thymocytes (3). Note the presence of the N-Me T cell enhancer in CD34+CD1a- thymocytes and the T-ALL cell lines DND-41, Jurkat, but not in cbCD34+ HSPCs or BCL11B-deregulated leukemias (panel **C**).



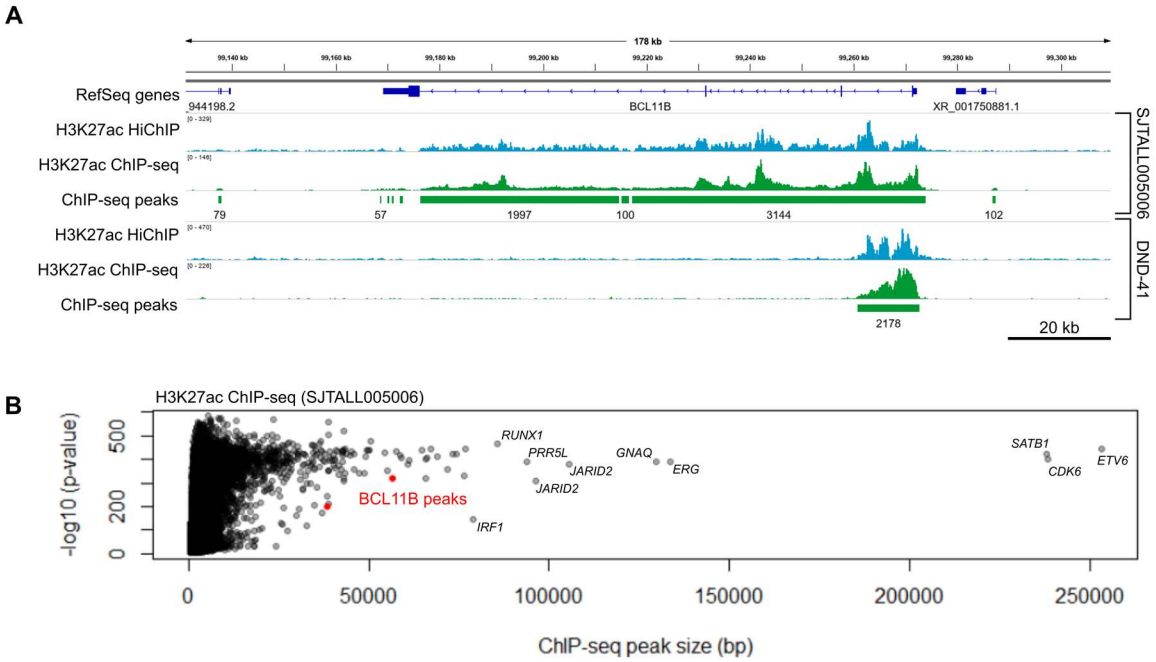
Supplementary Fig. 16. Tandem amplification of a non-coding element (BETA) occurs in ~20% of *BCL11B* group cases. (**A**) WGS and RNA-seq coverage tracks in all cases with BETA. Non-BETA cases are indicated with red asterisks. BETA cases identified from RNA-seq but lacking WGS data are indicated with blue asterisks. (**B**) IGV snapshot showing WGS read pairs at the region corresponding to BETA. Tandem amplification of this region was inferred based on the orientation of read-pairs mapping to the left and right breakpoints. (**C**) Representative image of sequential RNA-DNA FISH performed on sample SJTALL005006. The RNA signal indicates *BCL11B* is only expressed from one allele which is in proximity to the larger BETA signal, due to incraesed copy number of the 2.5kb element. 471 cells were analyzed, of which 199 showed evidence of *BCL11B* mRNA expression on the allele with amplified BETA signal. (**D**) UCSC Genome Browser snapshot showing the genomic region corresponding to BETA with the 100 vertebrates evolutionary conservation track below. (**E**) Genomic qPCR results for 4 BETA cases showing 15-20 copies in each case. Primers and probes used can be found in Methods.



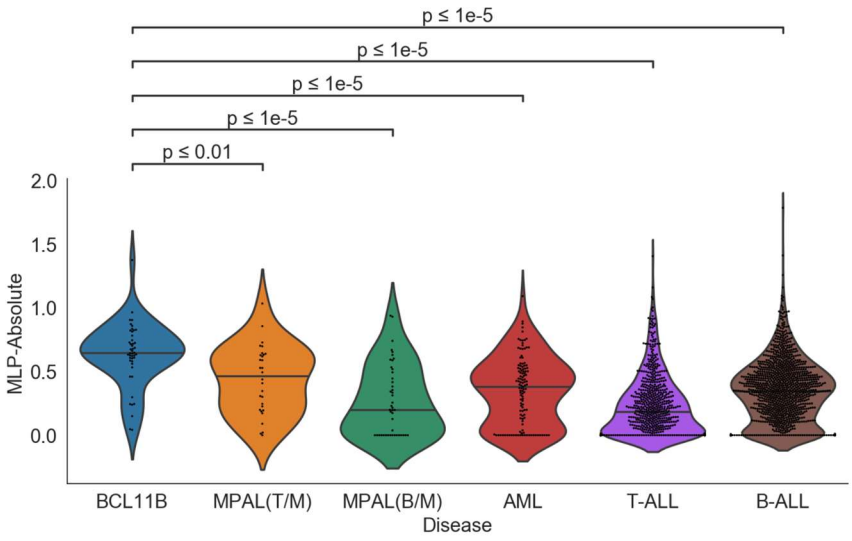
Supplementary Fig. 17. Long-read sequencing of two *BCL11B* group cases harboring BETA. (**A**) Genome browser snapshot of the genomic region harboring the amplified element showing the ~1kb region of evolutionary conservation, flanking sequences of self-homology (red arrowheads), and WGS coverage tracks for all cases with available data. (**B**) An individual PacBio subread from each case is displayed in alignment to the reference genome. Horizontal repeat units indicate tandem copies of the element. Dashed blue box indicates flanking, non-amplified DNA sequence.


A
20 kb
hg38
98,520,000
98,525,000
98,530,000
98,535,000
98,540,000
98,545,000
98,550,000
98,555,000
98,560,000
98,565,000
98,570,000
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
BETA
Active promoter
Enhancer
Transcription
Heterochromatin
Quiescent
CD34+
CD34 CMP
MPP
CD14 MONOCYTE
MONONUCLEAR CELL
B CELL
THYMUS
CD4 T CELL
CD8 T CELL
NAIVE T CELL
T CELL
TREG CELL
T1 CELL
T2 CELL
T17 CELL
NK CELL
B
98,541,000
98,542,000
98,543,000
98,544,000
98,545,000
98,546,000
self-chain
BETA
self-chain
Motifs
MEIS
ZNF384
HOX
RUNX
GATA
CEBPB
100 vertebrates conservation
CD34

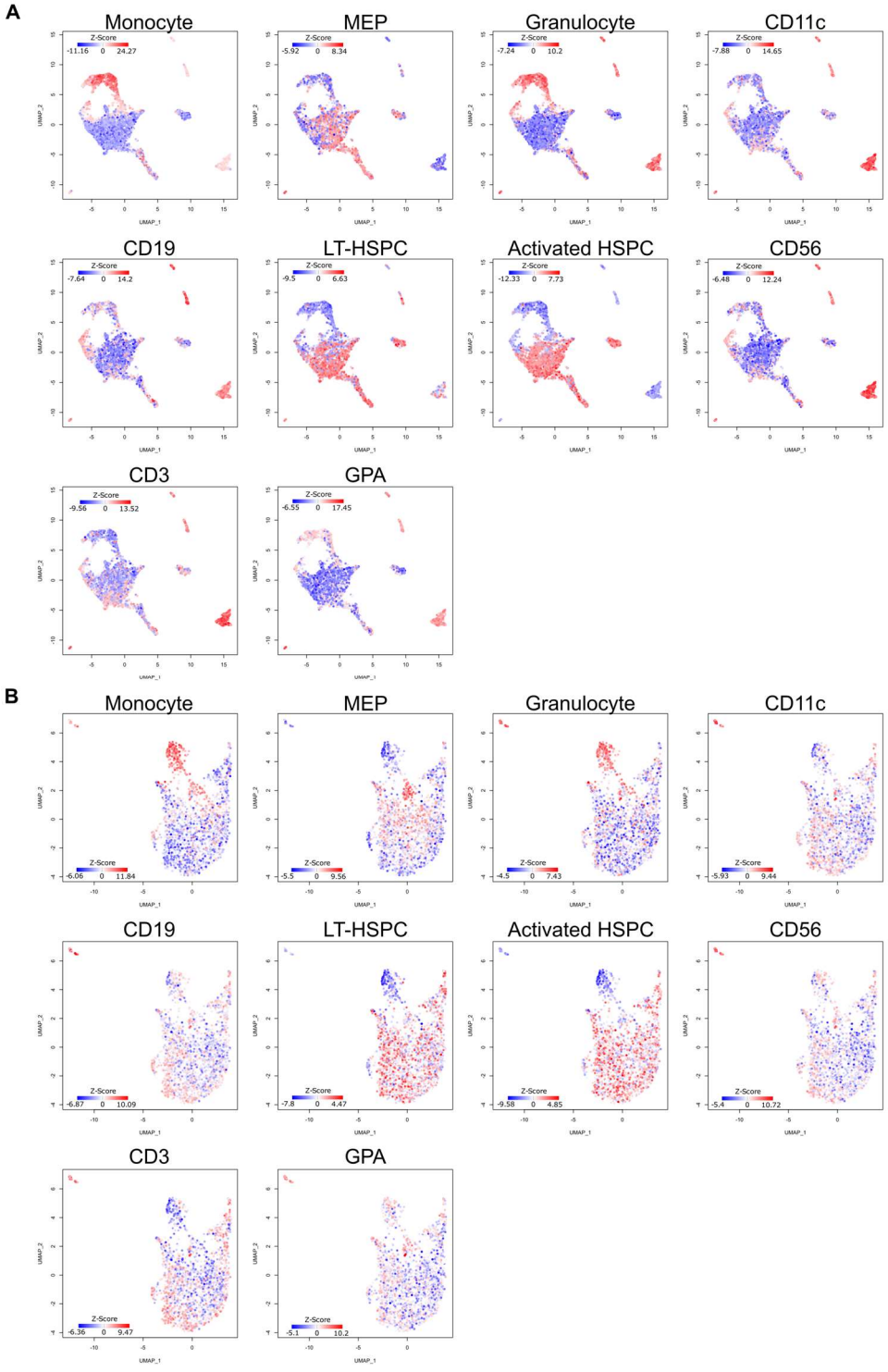
Supplementary Fig. 18. Chromatin state of BETA in normal hematopoietic cell types and motif enrichment analysis. (**A**) ChromHMM data are shown for all hematopoietic cell types available from the most recent release of the EpiMap (4)/Epigenome Roadmap (2) database. CD34+ HSPCs are highlighted in blue which demonstrates the presence of active enhancer (yellow) and promoter (red) chromatin states at this element in CD34+ cells. This chromatin state is less pronounced or absent in other progenitor cells and more differentiated cell types. (**B**) Zoomed in view of the region containing BETA with instances of TF motifs for MEIS, ZNF384, HOX, RUNX, GATA, and CEBPB family proteins shown as colored ticks. Coordinates are hg38.



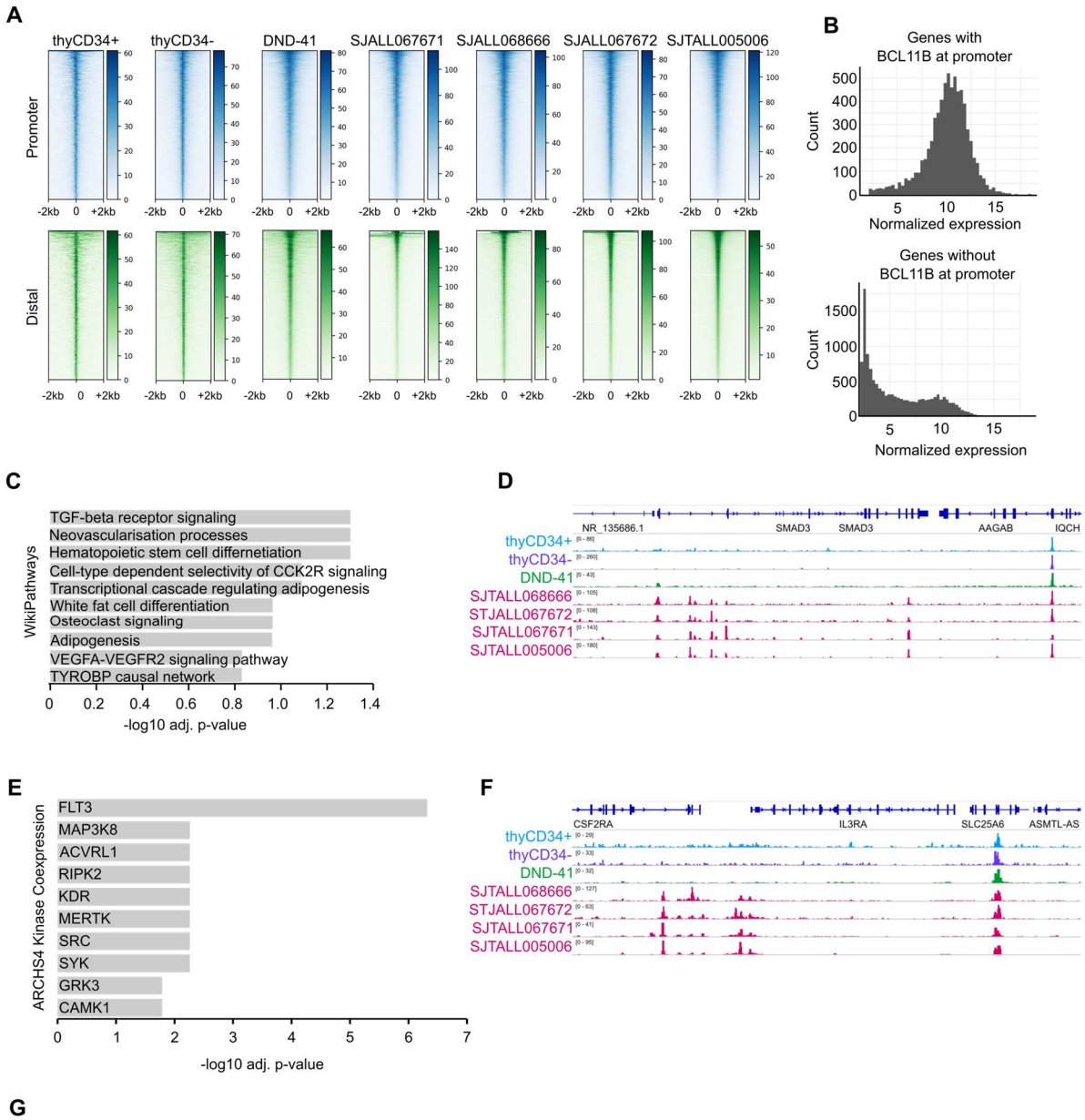
Supplementary Fig. 19. H3K27ac ChIP-seq for BETA case SJTALL005006. (**A**) H3K27ac ChIP-seq data in the DND-41 T-ALL cell line and BETA case SJTALL005006 are shown along with matched H3K27ac HiChIP data. Note the broad H3K27ac domain apparent in both ChIP-seq and HiChIP in the BETA case. (**B**) All significant (FDR < 0.01) H3K27ac ChIP-seq peaks for SJTALL005006 are plotted to highlight that the H3K27ac peaks along the *BCL11B* gene (shown in red) are among the largest (in bp) in the genome.



Supplementary Fig. 20. Enrichment of *BCL11B* group transcriptional signature with the MLP hematopoietic cell signature. Violin plots showing multi-lymphoid progenitor (MLP) signature enrichment ("MLP-Absolute") in individual *BCL11B* group leukemia samples compared with other leukemia subtypes.



Supplementary Fig. 21. Enrichment for normal hematopoietic open chromatin signatures. Using single-cell ATAC-seq data from the multimodal ATAC-seq/RNA-seq assay, z-scores were calculated for the correspondence between each single cell of SJMPAL011913 (**A**) and SJTALL005006 (**B**) with open chromatin signatures of normal hematopoietic cell types. MEP, myeloid-erythroid progenitor; GPA, glycophorin-A (i.e. erythroid).

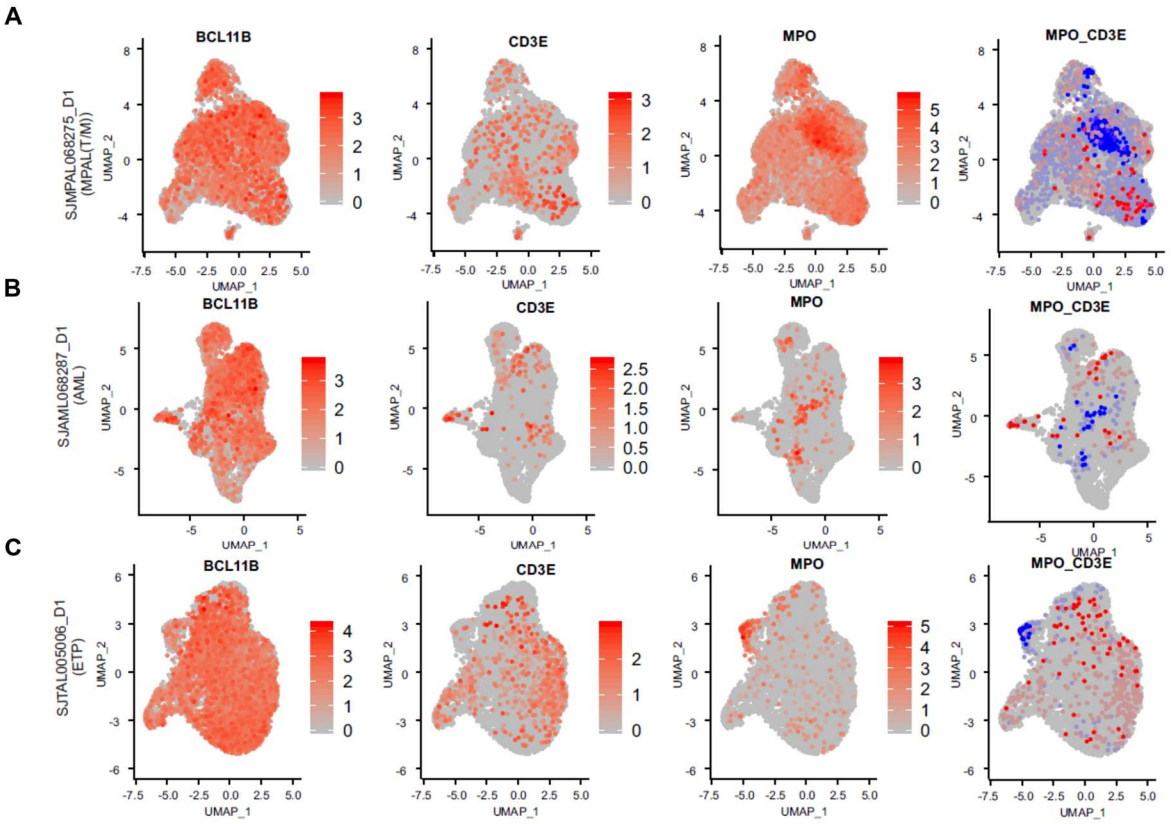


G

Sample	Rank	N peaks	%peaks with motif	%background with motif	Fold change	P-value
thyCD34+	1	1480	13.72%	2.23%	6.15	1x10-92
thyCD34-	2	4233	9.40%	2.22%	4.23	1x10-123
DND-41	1	5572	9.98%	1.75%	5.70	1x10-230
SJTALL005006	1	22177	13.50%	2.40%	5.63	1x10-1235
SJALL067671	1	10456	18.65%	3.15%	5.92	1x10-856
SJALL068666	1	6146	14.25%	2.55%	5.59	1x10-355
SJALL067672	1	7023	11.46%	2.48%	4.62	1x10-272

RUNX1 motif

Supplementary Fig. 22. ChIP-seq analysis of BCL11B occupancy in primary *BCL11B* group samples. (**A**) ChIP-seq for BCL11B in primary human CD34+ and CD34- thymocytes (from Ha et al Leukemia 2017), the DND-41 T-ALL cell line, and four primary BCL11B group leukemia samples. Tornado plots show the distribution of BCL11B occupancy over the 8,437 promoters (+/- 2kb of the TSS) with BCL11B binding in all 4 *BCL11B* group leukemia samples (blue), and at all distal (>10kb from a TSS) binding sites (green). (**B**) Histograms depicting the average expression levels of genes in the *BCL11B* subgroup based on whether BCL11B occupies each gene's promoter. Left, 8,437 genes with evidence of BCL11B binding in all 4 *BCL11B* group samples profiled. Right, the remaining genes not shown in the left panel, but which passed expression cutoffs for differential expression analysis. (**C**) WikiPathways enrichment of the 387 genes specifically bound by BCL11B and not normal thymocytes or DND-41 cells. (**D**) Genome browser track of BCL11B binding profile at the *SMAD3* and *GATA2* loci. (**E**) ARCHS4 (6) analysis of the 78 genes specifically bound by BCL11B in *BCL11B* group leukemia, and significantly upregulated in this subgroup compared to non-*BCL11B* group leukemias. These results highlight that genes co-expressed with *FLT3* are enriched in these 78 genes, for example the *IL3RA* gene shown in panel (**F**). (**G**) Motif enrichment analysis results for RUNX1.



Supplementary Fig. 23. Single cell RNA sequencing of *BCL11B* deregulated leukemia samples. (**A-C**) Each row corresponds to an individual scRNA-seq experiment and contains expression data for *BCL11B*, *CD3E*, *MPO*, and the overlap between *MPO* and *CD3E* (log2 expression). Note that *CD3E* and *MPO* are variably expressed and frequently occur in distinct subpopulations within each leukemia sample, whereas *BCL11B* is expressed homogenously throughout the sample.

SUPPLEMENTARY REFERENCES

1. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods **2012**;9(3):215-6
2. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A*, et al.* Integrative analysis of 111 reference human epigenomes. Nature **2015**;518(7539):317-30
3. Roels J, Kuchmiy A, De Decker M, Strubbe S, Lavaert M, Liang KL*, et al.* Distinct and temporary-restricted epigenetic mechanisms regulate human alphabeta and gammadelta T cell development. Nat Immunol **2020**;21(10):1280-92
4. Boix CA, James BT, Park YP, Meuleman W, Kellis M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. Nature **2021**;590(7845):300-7
5. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA*, et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res **2020**;48(D1):D87-d92
6. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L*, et al.* Massive mining of publicly available RNA-seq data from human and mouse. Nature communications **2018**;9(1):1366