# Dynamic CTCF binding directly mediates interactions among *cis*-regulatory elements essential for hematopoiesis

Qian Qi[1#], Li Cheng[1#], Xing Tang[1], Yanghua He[1], Yichao Li[1], Tiffany Yee[1], Dewan Shrestha[1], Ruopeng Feng[1], Peng Xu[1], Xin Zhou[2], Shondra Pruett-Miller[3], Ross C. Hardison[4]，Mitchell J. Weiss[1], Yong Cheng[1,2*]

1. Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
2. Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
3. Center for Advanced Genome Engineering, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
4. Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA.

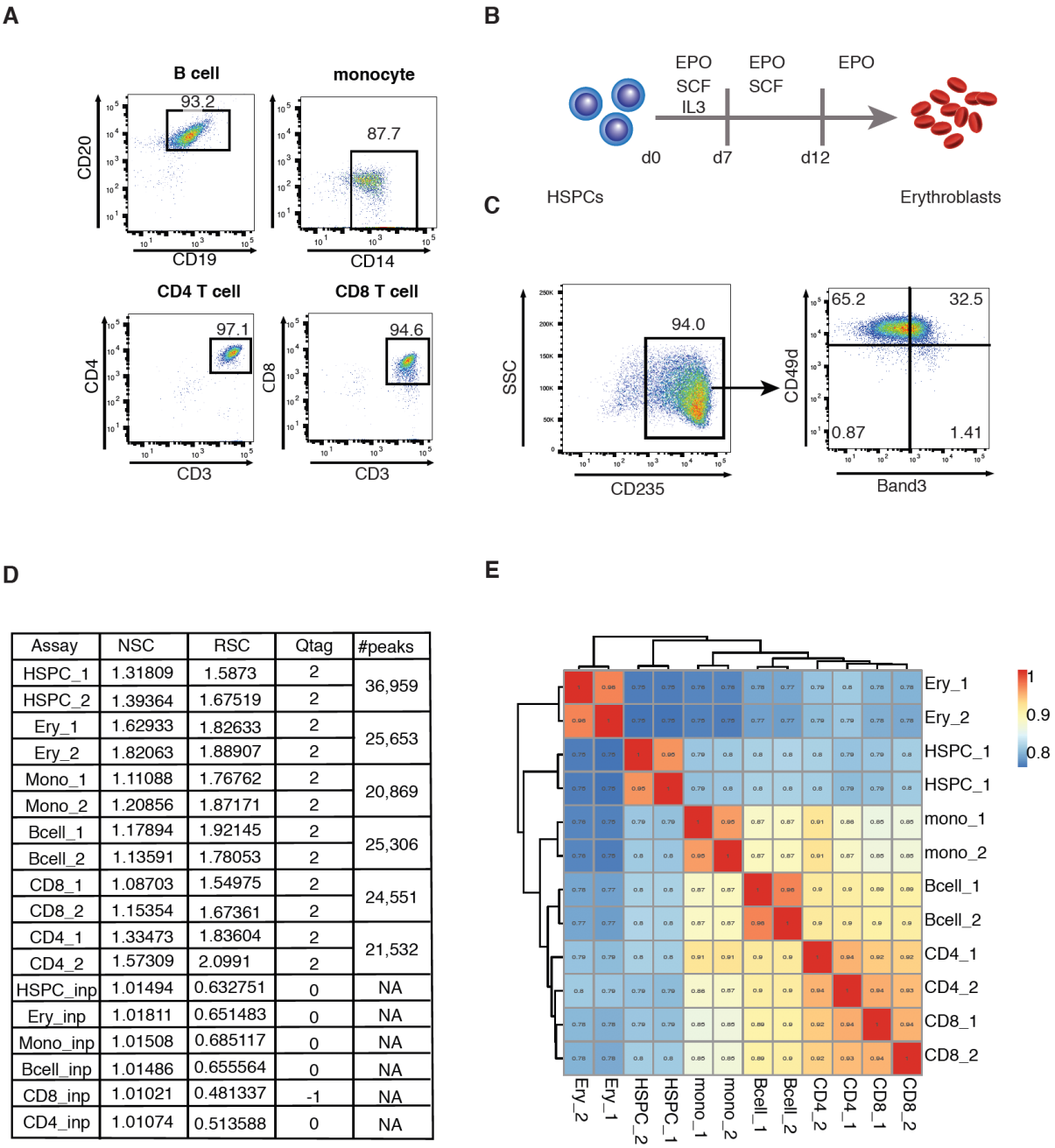
*Correspondence:
Yong Cheng, PhD
St. Jude Children's Research Hospital
262 Danny Thomas Place, MS-355
Memphis, TN 38105
yong.cheng@stjude.org

# equal contribution

## Supplementary Material

## Supplementary Figures

**Figure S1**




D

| Assay | NSC | RSC | Qtag | #peaks |
|---|---|---|---|---|
| HSPC_1 | 1.31809 | 1.5873 | 2 | 36,959 |
| HSPC_2 | 1.39364 | 1.67519 | 2 | |
| Ery_1 | 1.62933 | 1.82633 | 2 | 25,653 |
| Ery_2 | 1.82063 | 1.88907 | 2 | |
| Mono_1 | 1.11088 | 1.76762 | 2 | 20,869 |
| Mono_2 | 1.20856 | 1.87171 | 2 | |
| Bcell_1 | 1.17894 | 1.92145 | 2 | 25,306 |
| Bcell_2 | 1.13591 | 1.78053 | 2 | |
| CD8_1 | 1.08703 | 1.54975 | 2 | 24,551 |
| CD8_2 | 1.15354 | 1.67361 | 2 | |
| CD4_1 | 1.33473 | 1.83604 | 2 | 21,532 |
| CD4_2 | 1.57309 | 2.0991 | 2 | |
| HSPC_inp | 1.01494 | 0.632751 | 0 | NA |
| Ery_inp | 1.01811 | 0.651483 | 0 | NA |
| Mono_inp | 1.01508 | 0.685117 | 0 | NA |
| Bcell_inp | 1.01486 | 0.655564 | 0 | NA |
| CD8_inp | 1.01021 | 0.481337 | -1 | NA |
| CD4_inp | 1.01074 | 0.513588 | 0 | NA |

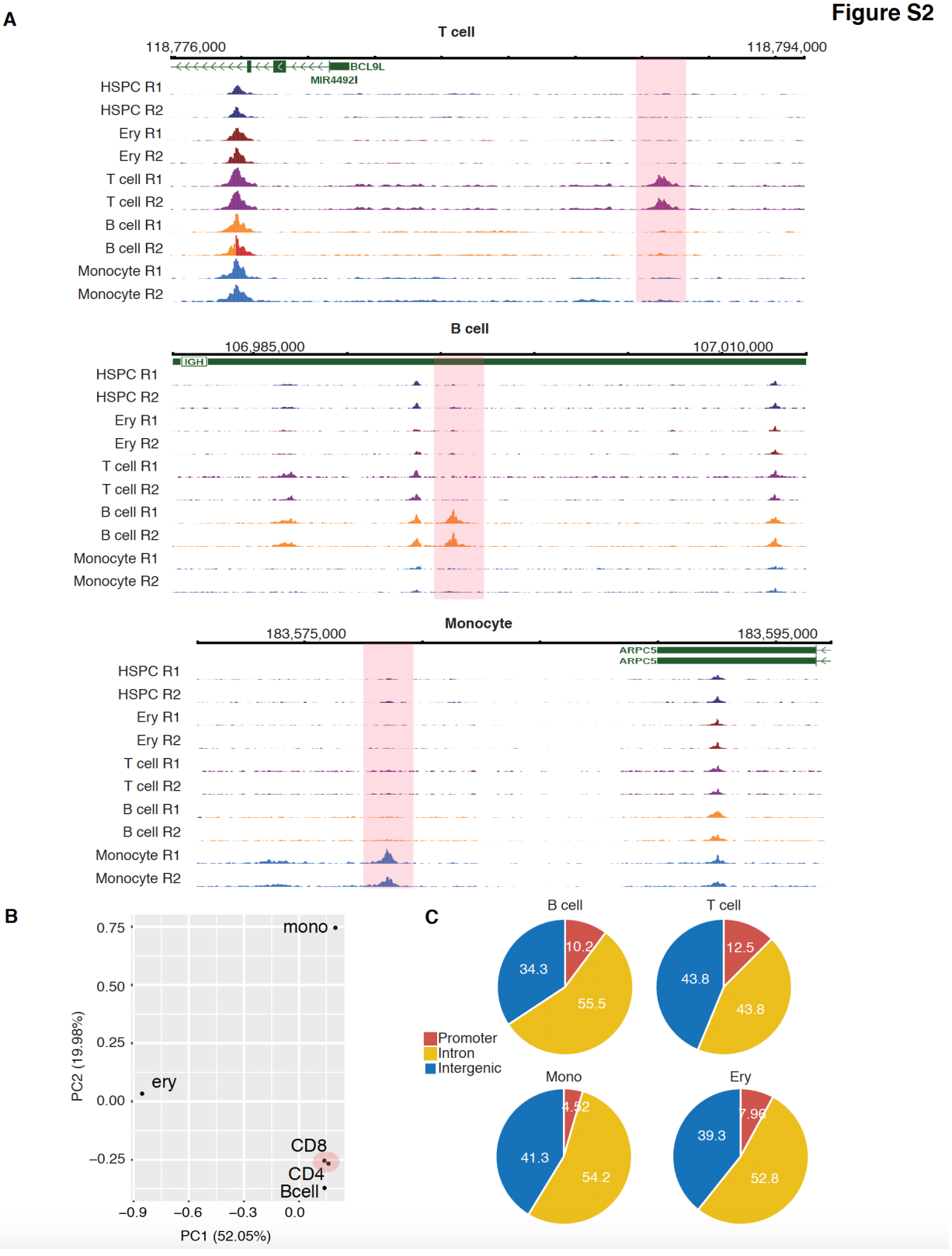
E




**Supplementary Fig.1. A**, B cells, monocytes, CD4$^+$ T cells, and CD8$^+$ T cells were isolated from CD34$^-$ PBMCs with their corresponding microbeads. Flow cytometric plots

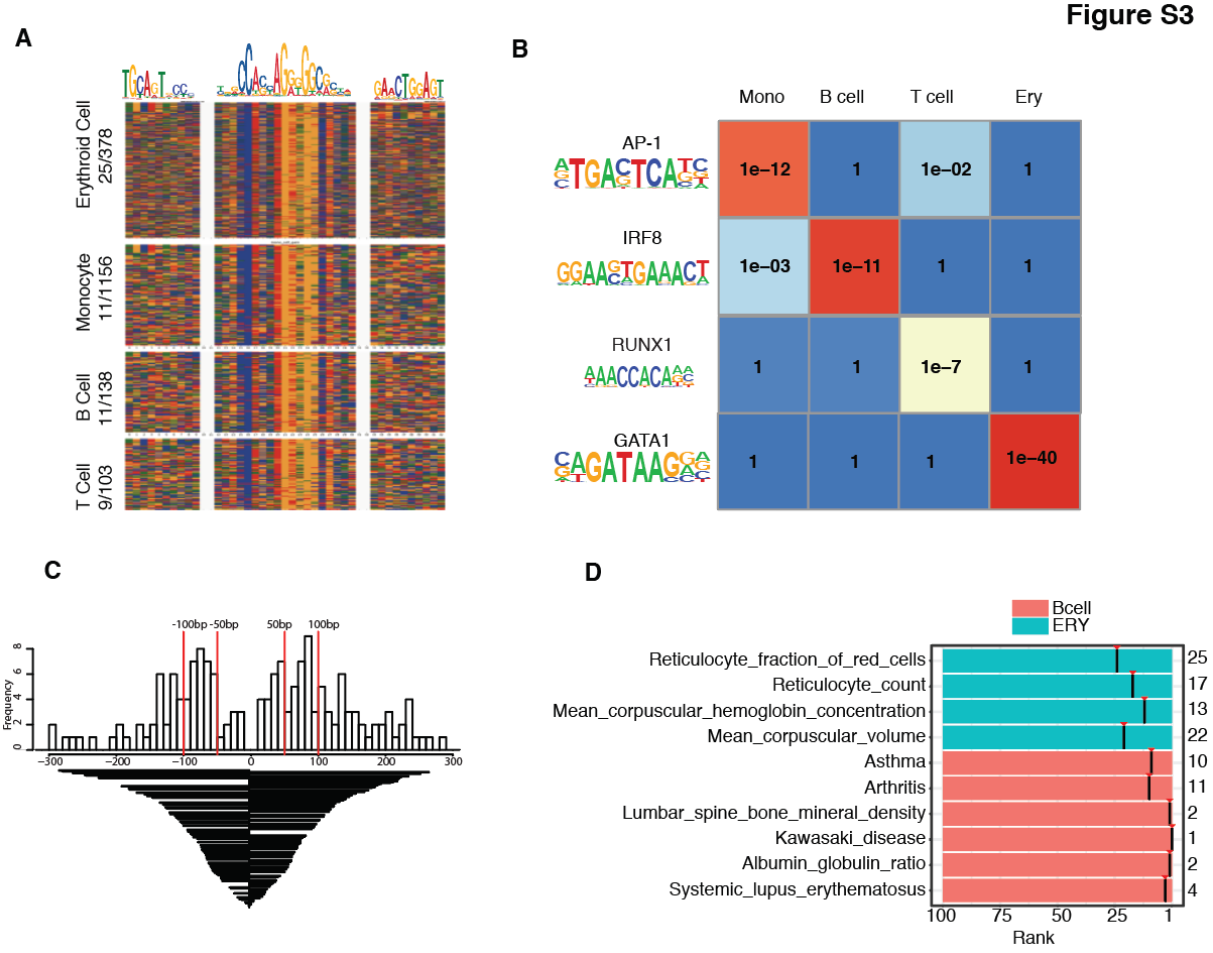
show the purity of B cells, monocytes, $CD4^+$ T cells, and $CD8^+$ T cells were 93.2%, 87.7%, 97.1%, and 94.6%, respectively. **B**, The 3-phase in vitro erythropoiesis protocol for human HSPCs. **C**, Flow cytometric plots of erythroid marker CD235a (glycophorin A) levels after $CD34^+$ HSPCs were maintained in erythroid differentiation medium for 12 d (left panel). The levels of Band3 and CD49d in the gated $CD235a^+$ cells were measured by flow cytometry (right panel). **D,** ChIP-seq quality measured by Cross-correlation analysis and total number of peaks detected. **E**, Heatmap showing the unsupervised hierarchical clusters based on Pearson correlation of CTCF ChIP-seq signals in different blood cell lineages. Numbers represent the value of Pearson correlation coefficient.

**Figure S2**


A
T cell
118,776,000
118,794,000
BCL9L
MIR4492I
HSPC R1
HSPC R2
Ery R1
Ery R2
T cell R1
T cell R2
B cell R1
B cell R2
Monocyte R1
Monocyte R2
B cell
106,985,000
107,010,000
IGH
Monocyte
183,575,000
183,595,000
ARPC5
ARPC5
B
mono
ery
CD8
CD4
Bcell
0.75
0.50
0.25
0.00
−0.25
−0.9
−0.6
−0.3
0.0
PC1 (52.05%)
PC2 (19.98%)
C
B cell
T cell
Mono
Ery
Promoter
Intron
Intergenic
10.2
34.3
55.5
12.5
43.8
43.8
4.52
41.3
54.2
7.96
39.3
52.8

**Supplementary Fig. 2. A**, Genome browser tracks showing dynamic CTCF-binding sites (highlighted in pink) in T cell, B cell, and Monocyte respectively. **B,** PCA analysis of CTCF ChIP-seq in different blood lineages. The PCA is based on the raw ChIP-seq signal at CTCF binding sites. **C,** Pie charts showing percent of dynamic CTCF sites in different genomic locations. Promoters are defined as +/-2000bp of transcription start sites (TSSs).

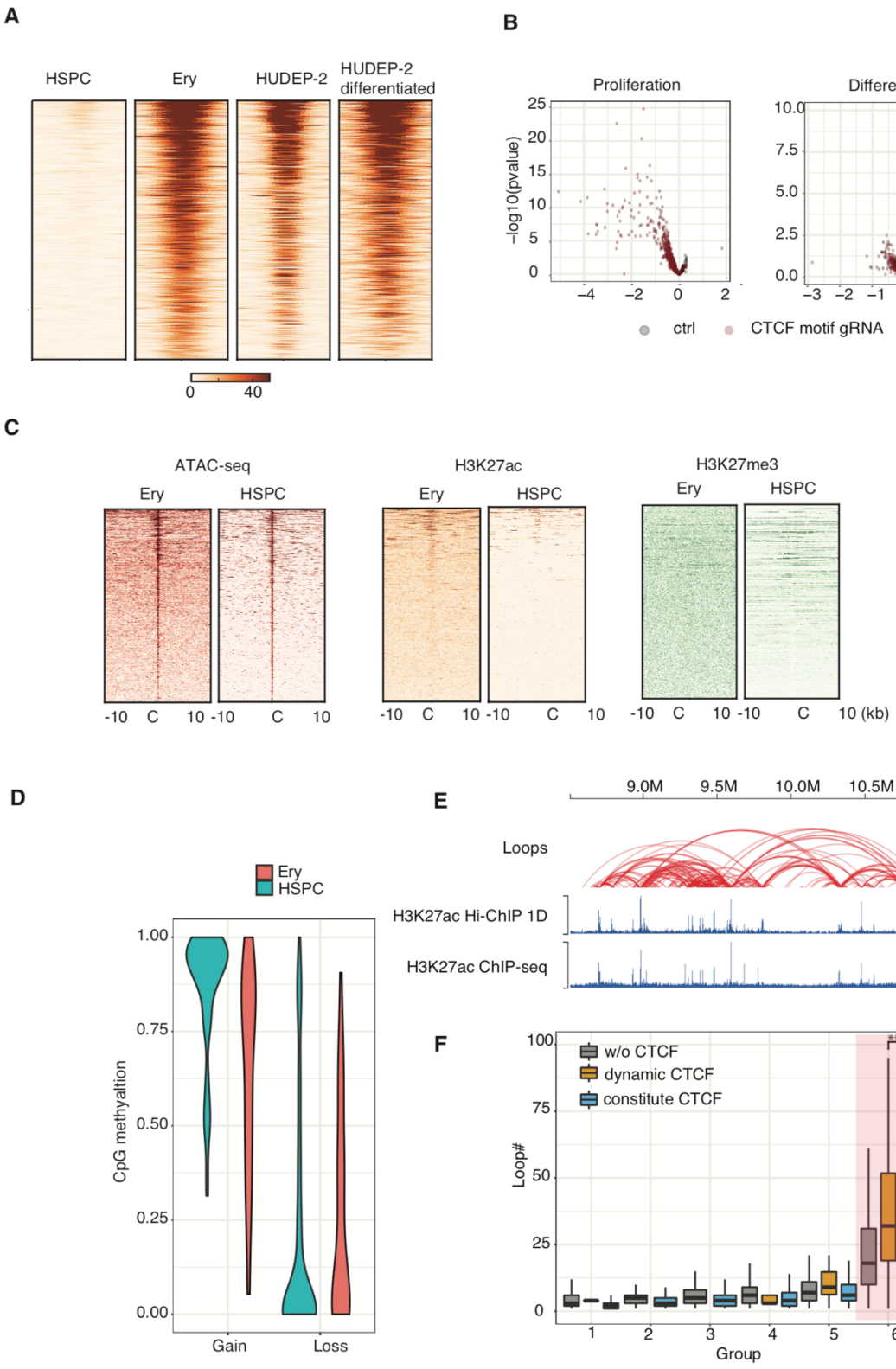
Figure S3




**Supplementary Fig. 3. A,** The color chart shows the DNA sequence of core CTCF motif and upstream motifs. Red represents A, blue represents C, green represents T and yellow represents G. The four groups are dynamic CTCF sites in different cell types. **B,** p-values for enrichment of tissue-specific TF motifs in different blood cell types, calculated by HOMER, is shown as a heatmap. **C,** Histogram in the top panel represents the distribution of distance between CTCF and GATA1 motif instances in erythroid dynamic CTCF sites. Negative distance indicates that GATA1 motifs are located in the 3'end of CTCF motifs. Positive distance indicates that GATA1 motifs are located in the 5'end of CTCF motifs. Each row in the bottom plot represents the distance between GATA1 and CTCF within one dynamic CTCF site. Distances were sorted

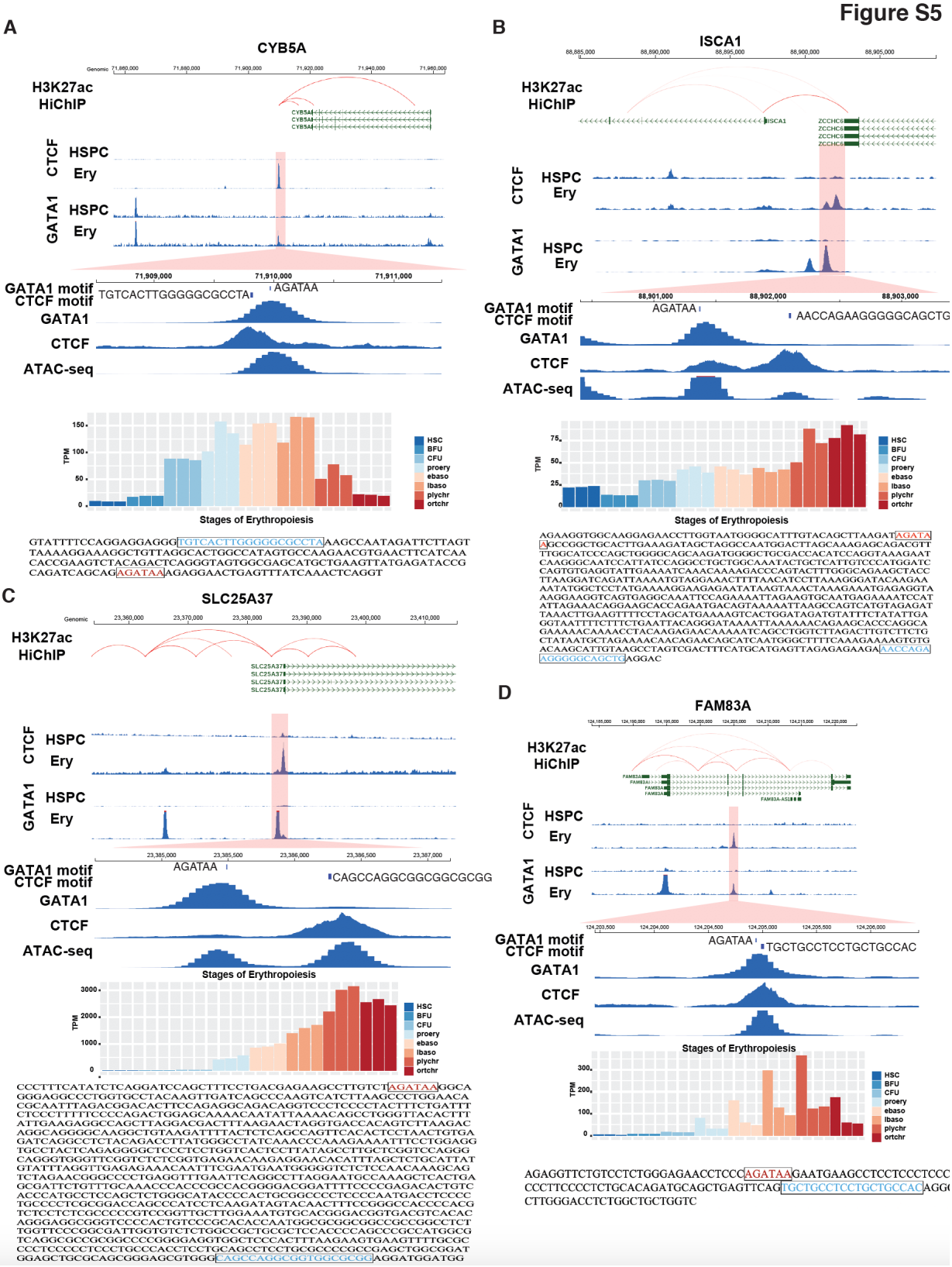
based on absolute value. **D,** GWAS hits enrichment analysis were conducted using randomly (100X) selected GATA1 and PU.1 peaks in erythroid and B-cells that have the same total numbers and length distribution as the dynamic CTCF sites. If the q-value of the dynamic CTCF sites was more significant than all the 100 random ones, the rank was 1st. In contrast, if the q-value was less significant than the 100 random ones, the rank was 100th. The actual ranking of dynamic CTCF sites for each trait is plotted.

Figure S


A
HSPC
Ery
HUDEP-2
HUDEP-2 differentiated
0
40
B
Proliferation
Differentiation
−log10(pvalue)
ctrl
CTCF motif gRNA
C
ATAC-seq
H3K27ac
H3K27me3
Ery
HSPC
-10
C
10
10 (kb)
D
Ery
HSPC
CpG methyaltion
Gain
Loss
E
9.0M
9.5M
10.0M
10.5M
11.0M
Loops
H3K27ac Hi-ChIP 1D
H3K27ac ChIP-seq
F
w/o CTCF
dynamic CTCF
constitute CTCF
Loop#
Group
***

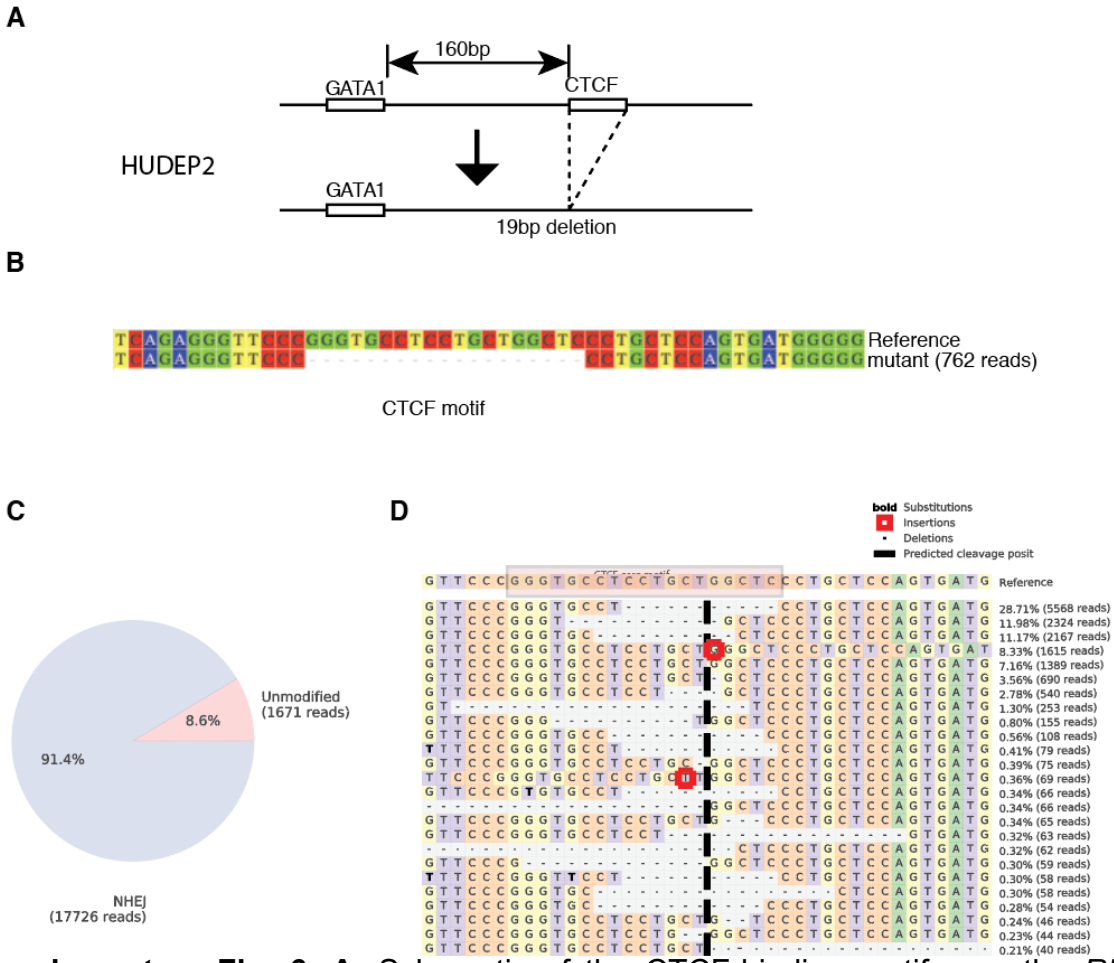
**Supplementary Fig. 4. A**, Heatmap represents the ChIP-seq signals distribution of 378 erythroid dynamic CTCF sites within (+/- 500bp window) centered by the ChIP-seq peaks. **B,** Cas9-expressing HUDEP-2 cells were transduced with a lentiviral vector library containing 1188 sgRNA targeting erythroid dynamic CTCF-binding sites and 100 control nontarget sgRNAs. Left panel: A volcano plot is depicted with the log2 fold change for each gRNA between cells that grow in expansion medium on 0 and 8 days. Right panel: A volcano plot is depicted with the log2 fold change for each sgRNA between $Band3^{+}$ and $Band3^{-}$ cells after 5 days of induced erythroid maturation. **C,** Heatmap showing the signals of ATAC-seq, H3K27ac, and H3K27me3 ChIP-seq near the 20-kb windows of constitutive CTCF-binding sites in HSPCs and erythroblasts. **D**, Violin box plots showing the CpG methylation levels of GOSs and LOSs in HSPCs and erythroblasts. **E**, Screenshot of loops identified by H3K27ac HiChIP (top), H3K27ac HiChIP signal (middle), and H3K27ac ChIP signal tracks (bottom) in $CD34^{+}$ HSPCs derived erythroblasts from the same donor in a 3-Mb window. **F,** Each boxplot represents the distribution of the number of loops per anchor region (y-axis). The different groups of loop-anchors based on the enrichment of H3K27ac signals from low (group 1) to high (group 6) are presented on the x-axis. *** marks the comparison with a significant difference (p-value < 2.2e-16) in the number of loops per anchor between dynamic CTCF sites and constitute CTCF sites in group 6 (highlighted), using a Wilcoxon-test.

Figure S5


A
CYB5A
H3K27ac HiChIP
CTCF
HSPC
Ery
GATA1
HSPC
Ery
GATA1 motif
CTCF motif
TGTCACTTGGGGGCGCCTA
AGATAA
GATA1
CTCF
ATAC-seq
TPM
Stages of Erythropoiesis
HSC
BFU
CFU
proery
ebaso
lbaso
plychr
ortchr
GTATTTTCCAGGAGGAGGGTGTCACTTGGGGGCGCCTAAAGCCAATAGATTCTTAGT
TAAAAGGAAAGGCTGTTAGGCACTGGCCATAGTGCCAAGAACGTGAACTTCATCAA
CACCGAAGTCTACAGACTCAGGGTAGTGGCGAGCATGCTGAAGTTATGAGATACCG
CAGATCAGCAGAGATAAAGAGGAACTGAGTTTATCAAACTCAGGT
B
ISCA1
H3K27ac HiChIP
CTCF
HSPC
Ery
GATA1
HSPC
Ery
GATA1 motif
CTCF motif
AGATAA
AACCAGAAGGGGGCAGCTG
GATA1
CTCF
ATAC-seq
TPM
Stages of Erythropoiesis
AGAAGGTGGCAAGGAGAACCTTGGTAATGGGGCATTTGTACAGCTTAAGATAGATA
AGCCGGCTGCACTTGAAAGATAGCTAGGCCAATGGACTTAGCAAAGAGCAGACGTT
TTGGCATCCCAGCTGGGGCAGCAAGATGGGGCTGCGACCACATCCAGGTAAAGAAT
CAAGGGCAATCCATTATCCAGGCCTGCTGGCAAATACTGCTCATTGTCCCATGGATC
CAGTGTGAGGTATTGAAAATCAAACAAAAGACCCAGTACTTTGGGCAGAAGCTACC
TTAAGGATCAGATTAAAATGTAGGAAACTTTTAACATCCTTAAAGGGATACAAGAA
AATATGGCTCCTATGAAAAGGAAGAGAATATAAGTAAACTAAAGAAATGAGAGGTA
AAGGAAGGTCAGTGAGGCAAATTCCAGAAAATTAGAAGTGCAATGAGAAAATCCAT
ATTAGAAACAGGAAGCACCAGAATGACAGTAAAAATTAAGCCAGTCATGTAGAGAT
TAAACTTGAAGTTTTCCTAGCATGAAAAGTCACTGGATAGATGTATTTCTATATTGA
GGTAATTTTCTTTCTGAATTACAGGGATAAAATTAAAAAACAGAAGCACCCAGGCA
GAAAAACAAAACCTACAAGAGAACAAAAATCAGCCTGGTCTTAGACTTGTCTTCTG
CTATAATGCTAGAAAACAACAGAACAGCATCAATGGGCTTTTCAAAGAAAAGTGTG
ACAAGCATTGTAAGCCTAGTCGACTTTCATGCATGAGTTAGAGAGAAGAAACCAGA
AGGGGGCAGCTGAGGAC
C
SLC25A37
H3K27ac HiChIP
CTCF
HSPC
Ery
GATA1
HSPC
Ery
GATA1 motif
CTCF motif
AGATAA
CAGCCAGGCGGCGGCGCGG
GATA1
CTCF
ATAC-seq
Stages of Erythropoiesis
TPM
HSC
BFU
CFU
proery
ebaso
lbaso
plychr
ortchr
CCCTTTCATATCTCAGGATCCAGCTTTCCTGACGAGAAGCCTTGTCTAGATAAGGCA
GGGAGGCCCTGGTGCCTACAAGTTGATCAGCCCAAGTCATCTTAAGCCCTGGAACA
CGCAATTTAGACGGACACTTCCAGAGGCAGACAGGTCCCTCCCCTACTTTCTGATTT
CTCCCTTTTCCCCAGACTGGAGCAAAACAATATTAAAACAGCCTGGGTTACACTTT
ATTGAAGAGGCCAGCTTAGGACGACTTTAAGAACTAGGTGACCACAGTCTTAAGAC
AGGCAGGGGCAAGGCTGTAAGATTTTACTCTCAGCCAGTTCACACTCCTAACTGTGA
GATCAGGCCTCTACAGACCTTATGGGCCTATCAAACCCAAAGAAAATTTCCTGGAGG
TGCCTACTCAGAGGGGCTCCCTCCTGGTCACTCCTTATAGCCTTGCTCGGTCCAGGG
CAGGGTGGGTTCGGTCTCTCGGTGAGAACAAGAGGAACACATTTAGCTCTGCATTAT
GTATTTAGGTTGAGAGAAACAATTTCGAATGAATGGGGGTCTCTCCAACAAAGCAG
TCTAGAACGGGCCCCTGAGGTTTGAATTCAGGCCTTAGGAATGCCAAAGCTCACTGA
GCGATTCTGTTTGCAAACCCACCCGCCACGGGGACGGATTTTCCCCGAGACACTGTC
ACCCATGCCTCCAGCTCTGGGCATACCCCACTGCGGCCCCTCCCCAATGACCTCCCC
TGCCCCTCGCGGACCAGCCCATCCTCAAGATAGTACAACTTTCCGGGCCACCCCACG
TCTCCTCTCGCCCCCGTCCGGTTGCTTGGAAATGTGCACGGGACGGTGACGTCACAC
AGGGAGGCGGGTCCCCACTGTCCCGCACACCAATGGCGCGGCGGCCGCCGGCCTCT
TGGTTCCCGGCGATTGGTGTCTCTGGCCGCTGCGCTCCACCCCAGCCCGCCATGGCG
TCAGGCGCCGCGGCCCCGGGGAGGTGGCTCCCACTTTAAGAAGTGAAGTTTTGCGC
CCCTCCCCCTCCCTGCCCACCTCCTGCAGCCTCCTGCGCCCCGCCGAGCTGGCGGAT
GGAGCTGCGCAGCGGGAGCGTGGGCAGCCAGGCGGTGGCGCGGAGGATGGATGG
D
FAM83A
H3K27ac HiChIP
CTCF
HSPC
Ery
GATA1
HSPC
Ery
GATA1 motif
CTCF motif
AGATAA
TGCTGCCTCCTGCTGCCAC
GATA1
CTCF
ATAC-seq
Stages of Erythropoiesis
TPM
HSC
BFU
CFU
proery
ebaso
lbaso
plychr
ortchr
AGAGGTTCTGTCCTCTGGGAGAACCTCCCAGATAAGAATGAAGCCTCCTCCCTCCC
CCCTTCCCCTCTGCACAGATGCAGCTGAGTTCAGTGCTGCCTCCTGCTGCCACAGGC
CTTGGGACCTCTGGCTGCTGGTC

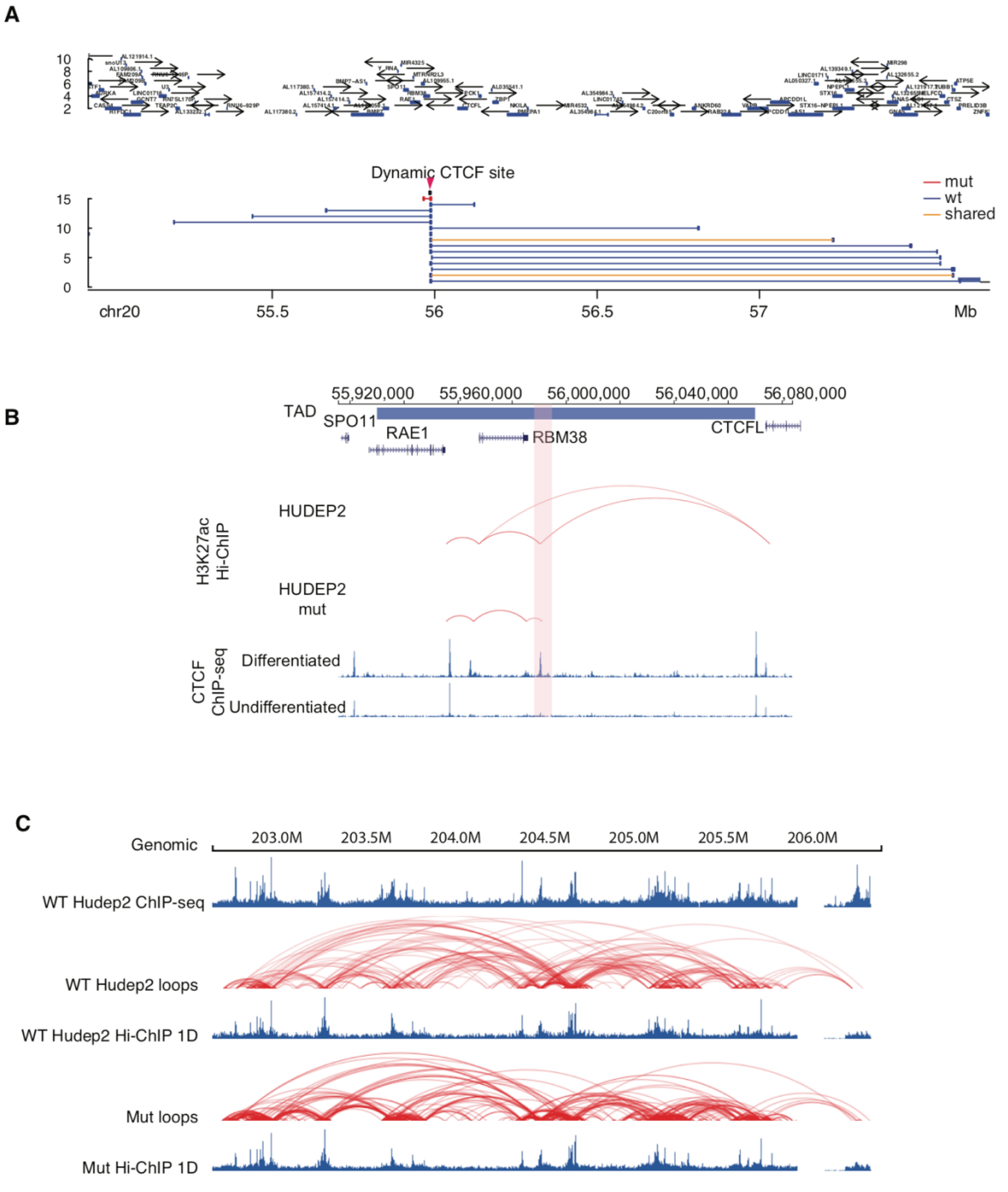
**Supplementary Fig. 5.** Examples of dynamic CTCF sites co-localized with GATA1 peaks and implicated in regulating gene expression during erythropoiesis. Within each locus, the top panel is the screenshot of genome browser that includes the CRE-loops detected by H3K27ac HiChIP, CTCF and GATA1 ChIP-seq in HSPC and erythroid cells. The middle panel shows a zoomed-in view (highlighted in Pink) of the peaks and consensus motifs of GATA1 and CTCF together with ATAC-seq signals in erythroid cells. The bottom panel shows the expression level (RNA-seq) of each gene at different stages of hematopoiesis. DNA sequences around CTCF and GATA1 motifs are shown in the bottom panel with motif sequences highlighted and boxed.

**Figure S6**




**Supplementary Fig. 6. A**, Schematic of the CTCF-binding motif near the *RBM38* gene deleted with homology-directed genome editing in HUDEP-2 cells. A GATA1 binding site is located 160-bp upstream of this CTCF-binding motif. **B**, Alignment of targeted deep sequencing data in a homozygous mut HUDEP-2 clone with the CTCF motif deleted. **C–D**, A Cas9 RNP with a gRNA targeting the CTCF motif was transiently transfected in CD34$^{+}$ HSPCs. HSPCs were grown in the erythroid Phase I differentiation medium for 7 d and then harvested for targeted deep sequencing. (C) Pie chart showing the percentage of modified (purple) and unmodified DNA sequences. (D) Alignment of targeted deep-sequencing data.

**Figure S7**


A
Dynamic CTCF site
mut
wt
shared
chr20
55.5
56
56.5
57
Mb
B
TAD
SPO11
RAE1
RBM38
CTCFL
55,920,000
55,960,000
56,000,000
56,040,000
56,080,000
H3K27ac Hi-ChIP
HUDEP2
HUDEP2 mut
CTCF ChIP-seq
Differentiated
Undifferentiated
C
Genomic
203.0M
203.5M
204.0M
204.5M
205.0M
205.5M
206.0M
WT Hudep2 ChIP-seq
WT Hudep2 loops
WT Hudep2 Hi-ChIP 1D
Mut loops
Mut Hi-ChIP 1D

**Supplementary Fig. 7. A.** Top panel shows the location and transcriptional direction of genes. The bottom panel represents the interactions involved with the dynamic CTCF-binding site measured by H3K27ac HiChIP in $CD34^+$ HSPCs differentiated erythroblasts. The color represents the different types of interactions. Blue, interactions in wildtype HUDEP2 cells only, orange, interactions shared between wildtype and mutant cells. Red, interactions in mutant cells only. **B,** Screenshot of chromatin interactions and related ChIP-seq signals within the TAD containing *RBM38* gene. The dynamic CTCF binding site is highlighted. **C**, Screenshot of 3D chromatin loops and ChIP-seq signals identified by H3K27ac-HiChIP with WT HUDEP-2 cells and a mut clone after 3 days of induced maturation.

**Supplementary Table Legends (separate files)**

**Supplementary Table 1**. List of dynamic CTCF sites in Erythroid cells, B cells, T cells and Monocytes. The coordinates and summits of the peaks; Log2 fold change, p-value and adjust p-value of Deseq2; the ChIP-seq signals and existence of TF motifs within +/-200bp of peak summits; the nearest gene and distance, together with the 3D interactions of promoters detected by promoter capture Hi-C and HiChIP against H3K27ac are shown.

**Supplementary Table 2.** DNA oligo sequences targeting RBM38 promoter for capture Hi-C. Two DNA sequences of biotinylated oligos used are shown.

**Supplementary Table 3.** List of functional genomic data used in this study. The cell type, culture condition, antibodies of newly generated are shown. The access ID of downloaded data are shown.