Cell Reports, Volume *27*

Supplemental Information

Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis

Leif S. Ludwig, Caleb A. Lareau, Erik L. Bao, Satish K. Nandakumar, Christoph Muus, Jacob C. Ulirsch, Kaitavjeet Chowdhary, Jason D. Buenrostro, Narla Mohandas, Xiuli An, Martin J. Aryee, Aviv Regev, and Vijay G. Sankaran

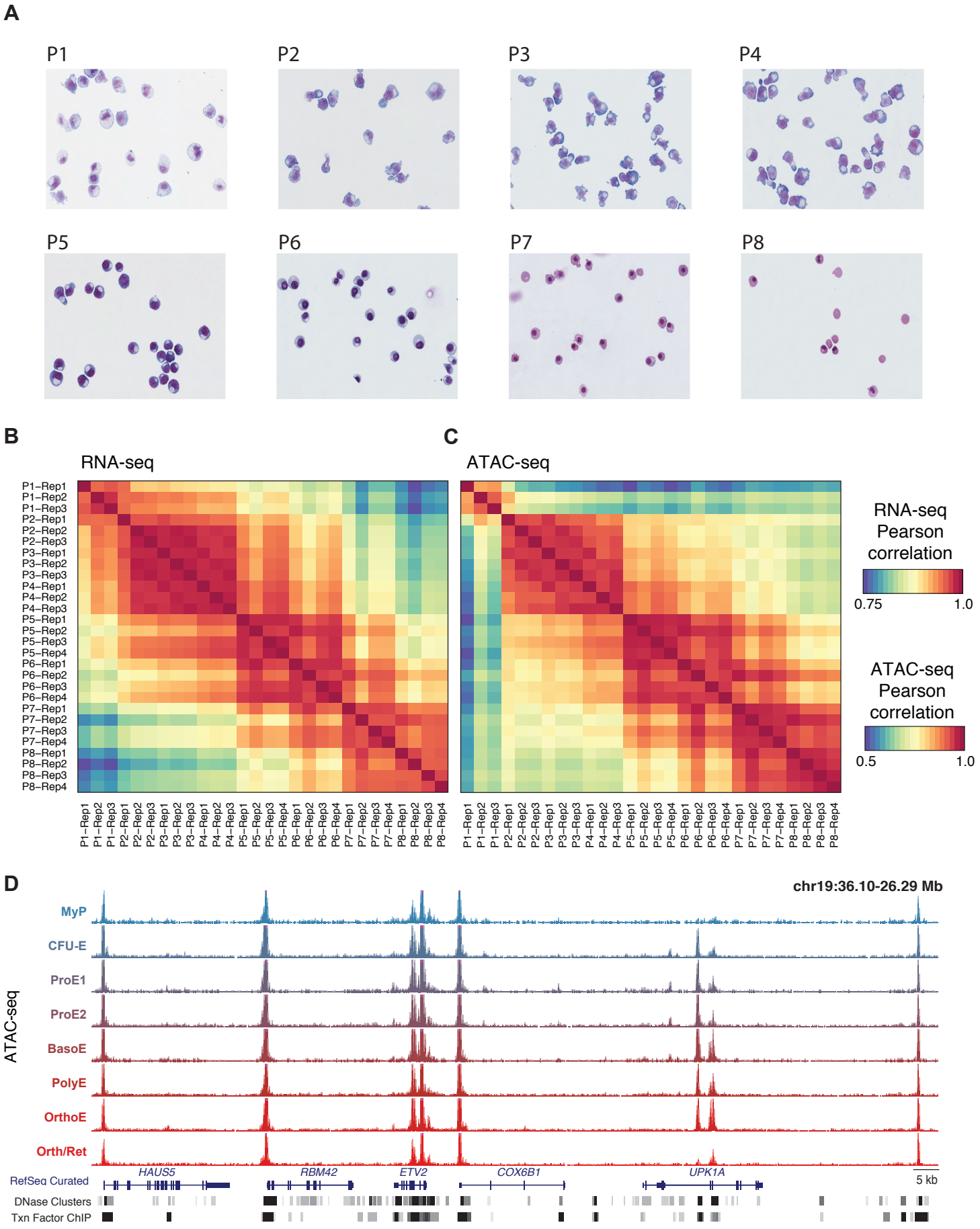
A
P1
P2
P3
P4
P5
P6
P7
P8
B
RNA-seq
C
ATAC-seq
P1–Rep1
P1–Rep2
P1–Rep3
P2–Rep1
P2–Rep2
P2–Rep3
P3–Rep1
P3–Rep2
P3–Rep3
P4–Rep1
P4–Rep2
P4–Rep3
P5–Rep1
P5–Rep2
P5–Rep3
P5–Rep4
P6–Rep1
P6–Rep2
P6–Rep3
P6–Rep4
P7–Rep1
P7–Rep2
P7–Rep3
P7–Rep4
P8–Rep1
P8–Rep2
P8–Rep3
P8–Rep4
RNA-seq Pearson correlation
0.75
1.0
ATAC-seq Pearson correlation
0.5
1.0
D
chr19:36.10-26.29 Mb
ATAC-seq
MyP
CFU-E
ProE1
ProE2
BasoE
PolyE
OrthoE
Orth/Ret
RefSeq Curated
HAUS5
RBM42
ETV2
COX6B1
UPK1A
5 kb
DNase Clusters
Txn Factor ChIP

Figure S1, related to Figure 1. A time course of human erythroid differentiation. (A) Images of MayGrunwald stained cytospins of sorted populations P1-P8 (**Figure 1B**, 63x magnification). (**B,C**) Heatmap showing the correlation across individual populations and replicates of (**B**) RNA-seq and (**C**) ATAC-seq libraries generated in this study. Color bar: Pearson correlation. (**D**) *ETV2* locus with corresponding ATAC-seq data across indicated populations show high quality chromatin accessibility profiles at known promoter and gene regulatory regions.

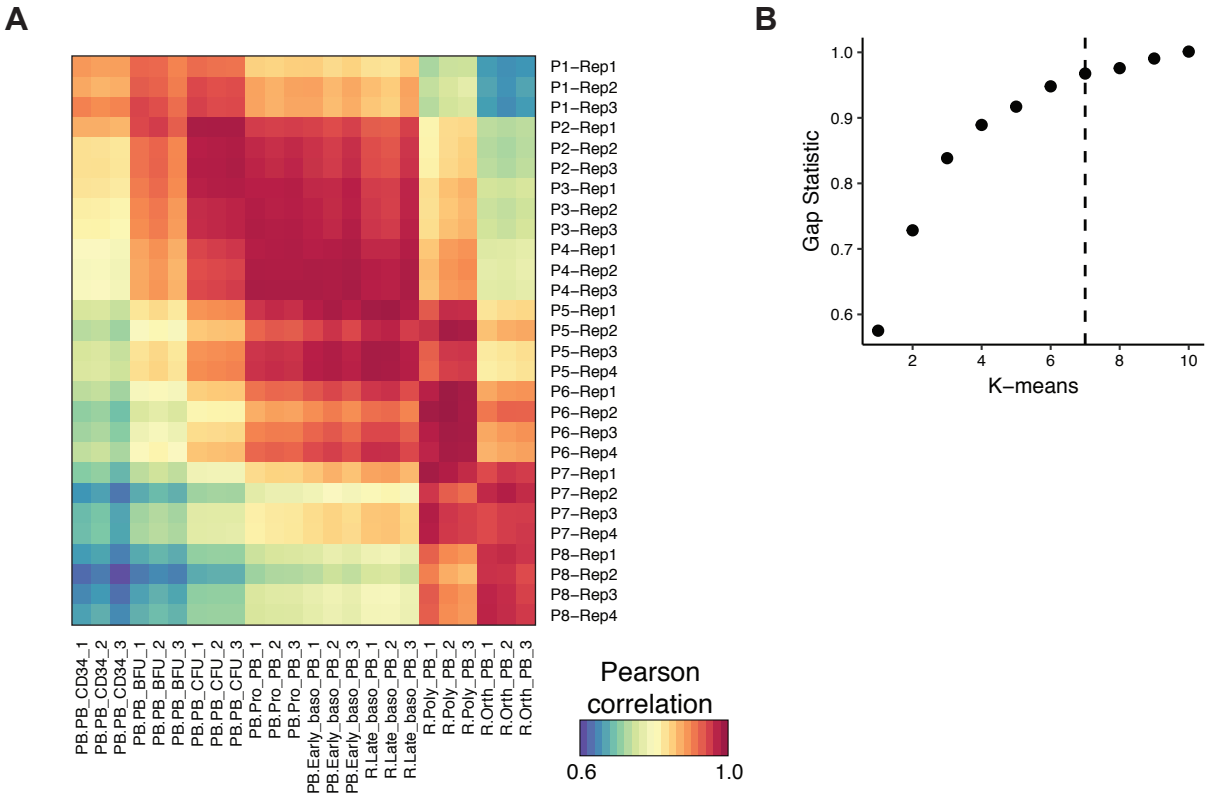
A
P1–Rep1
P1–Rep2
P1–Rep3
P2–Rep1
P2–Rep2
P2–Rep3
P3–Rep1
P3–Rep2
P3–Rep3
P4–Rep1
P4–Rep2
P4–Rep3
P5–Rep1
P5–Rep2
P5–Rep3
P5–Rep4
P6–Rep1
P6–Rep2
P6–Rep3
P6–Rep4
P7–Rep1
P7–Rep2
P7–Rep3
P7–Rep4
P8–Rep1
P8–Rep2
P8–Rep3
P8–Rep4
PB.PB_CD34_1
PB.PB_CD34_2
PB.PB_CD34_3
PB.PB_BFU_1
PB.PB_BFU_2
PB.PB_BFU_3
PB.PB_CFU_1
PB.PB_CFU_2
PB.PB_CFU_3
PB.Pro_PB_1
PB.Pro_PB_2
PB.Pro_PB_3
PB.Early_baso_PB_1
PB.Early_baso_PB_2
PB.Early_baso_PB_3
R.Late_baso_PB_1
R.Late_baso_PB_2
R.Late_baso_PB_3
R.Poly_PB_1
R.Poly_PB_2
R.Poly_PB_3
R.Orth_PB_1
R.Orth_PB_2
R.Orth_PB_3
Pearson correlation
0.6
1.0
B
Gap Statistic
1.0
0.9
0.8
0.7
0.6
2
4
6
8
10
K–means

Figure S2, related to Figure 2. Transcriptomic landscape of human erythropoiesis. **(A)** Heatmap showing the correlation across individual replicates of RNA-seq libraries generated in this study (y-axis) compared to previously published reports (x-axis). Color bar: Pearson correlation. (**B**) Gap statistic associated with the choice of k=7 clusters for the RNA-seq data.

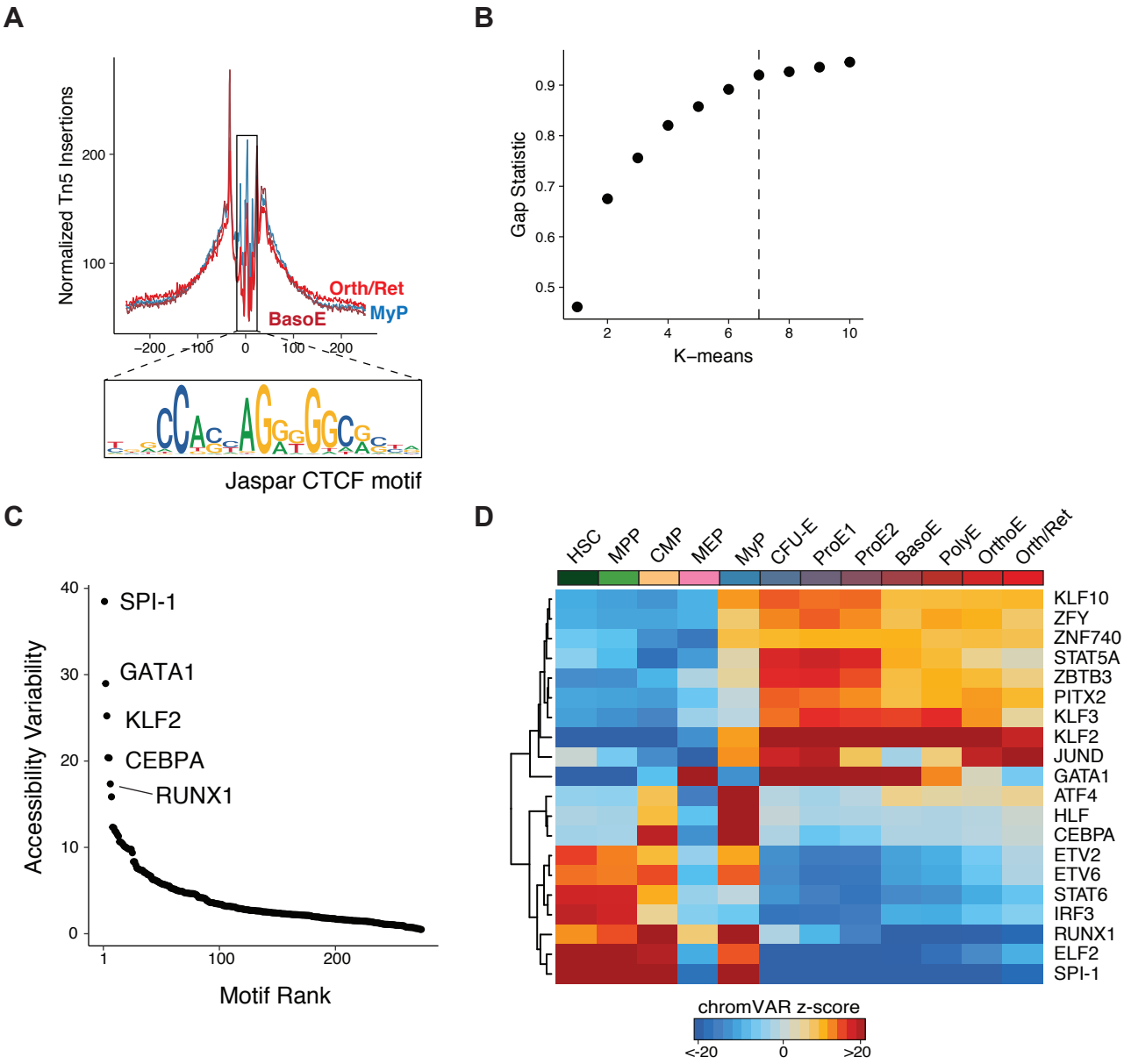
A
Normalized Tn5 Insertions
200
100
Orth/Ret
MyP
BasoE
−200
−100
0
100
200
Jaspar CTCF motif
B
Gap Statistic
0.9
0.8
0.7
0.6
0.5
2
4
6
8
10
K−means
C
Accessibility Variability
40
30
20
10
0
SPI-1
GATA1
KLF2
CEBPA
RUNX1
1
100
200
Motif Rank
D
HSC
MPP
CMP
MEP
MyP
CFU-E
ProE1
ProE2
BasoE
PolyE
OrthoE
Orth/Ret
KLF10
ZFY
ZNF740
STAT5A
ZBTB3
PITX2
KLF3
KLF2
JUND
GATA1
ATF4
HLF
CEBPA
ETV2
ETV6
STAT6
IRF3
RUNX1
ELF2
SPI-1
chromVAR z-score
<-20
0
>20

Figure S3, related to Figure 3. Open chromatin landscape of human erythropoiesis. **(A)** ATAC-seq footprinting plot, showing that Tn5 insertion density near the CTCF motif is relatively constant throughout erythroid differentiation. Normalized insertions of MyP, BasoE and Orth/Ret are shown. (**B**) Gap statistic associated with the choice of k=7 clusters for the ATAC-seq peaks. (**C**) Rank order plot of transcription factor binding sites with greatest variability in chromatin accessibility across sampled cell populations as also indicated in panel D. (**D**) Heatmap showing temporal changes in chromatin accessibility for the top 20 TFs with greatest accessibility variability between populations. Color bar: chromVAR deviation z-score.

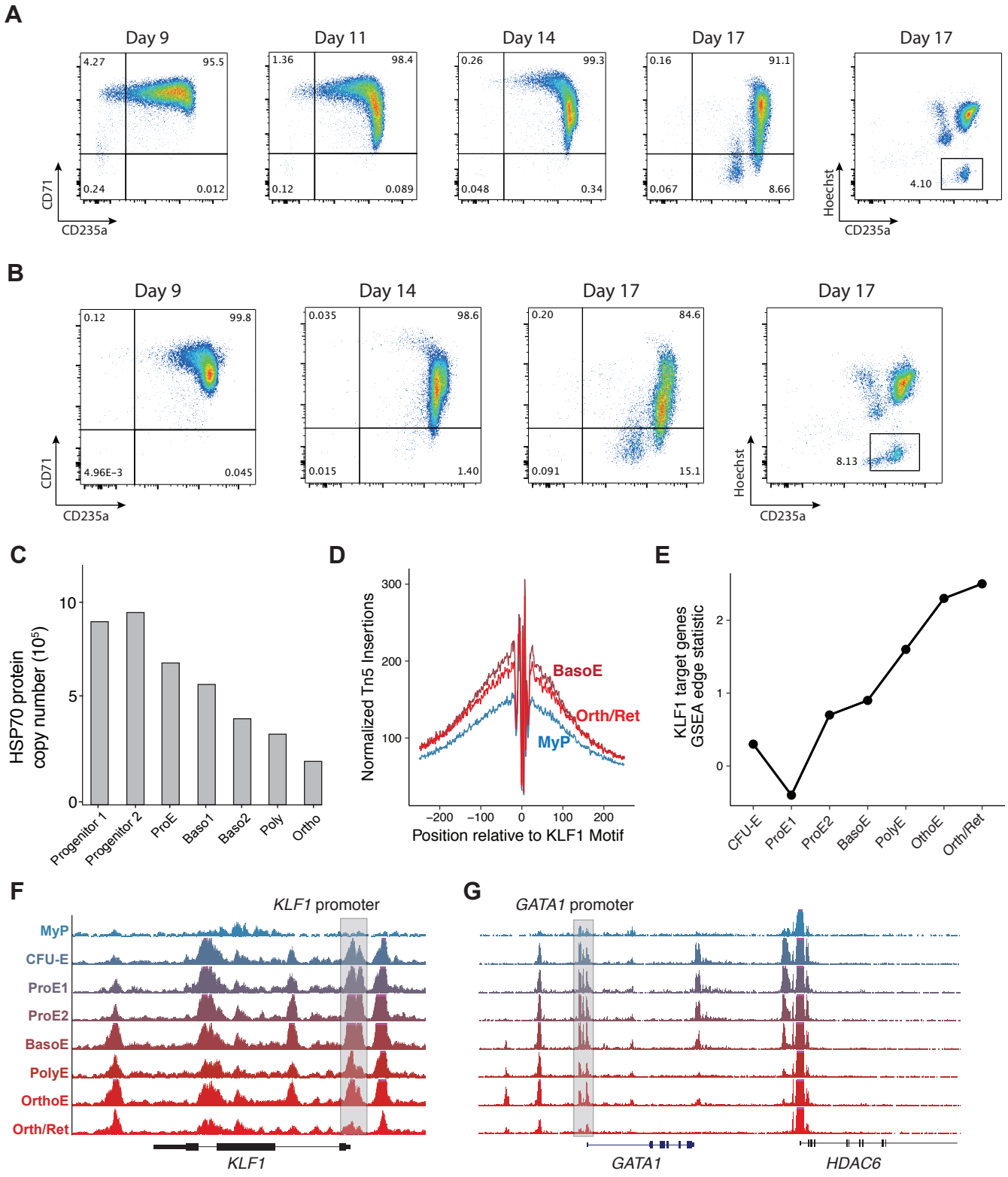
A
Day 9
Day 11
Day 14
Day 17
Day 17
4.27
95.5
0.24
0.012
1.36
98.4
0.12
0.089
0.26
99.3
0.048
0.34
0.16
91.1
0.067
8.66
4.10
CD71
CD235a
Hoechst
B
Day 9
Day 14
Day 17
Day 17
0.12
99.8
4.96E−3
0.045
0.035
98.6
0.015
1.40
0.20
84.6
0.091
15.1
8.13
C
HSP70 protein copy number (10^5)
Progenitor 1
Progenitor 2
ProE
Baso1
Baso2
Poly
Ortho
D
Normalized Tn5 Insertions
BasoE
Orth/Ret
MyP
Position relative to KLF1 Motif
E
KLF1 target genes GSEA edge statistic
CFU-E
ProE1
ProE2
BasoE
PolyE
OthoE
Orth/Ret
F
KLF1 promoter
MyP
CFU-E
ProE1
ProE2
BasoE
PolyE
OrthoE
Orth/Ret
KLF1
G
GATA1 promoter
GATA1
HDAC6

Figure S4, related to Figure 4. Regulatory dynamics of erythroid transcription factors. (**A, B**) Flow cytometry plots assessing erythroid differentiation using surface markers CD71 and CD235 at indicated time points of *in vitro* culture of primary hematopoietic cells. Hoechst staining was used to measure enucleation frequency. Plots from two cultures are shown. Samples shown in (**A**) correspond to the western blot shown in Figure 4C. Samples shown in (**B**) correspond to the western blot shown in Figure 4D. **(C)** Protein copy number for HSP70 across erythroid differentiation. **(D)** ATAC-seq footprinting plot, showing that Tn5 insertion density near the KLF1 motif is most "active" in BasoE, but remains high in Ortho/Ret. (**E**) *KLF1* transcriptional activity using GSEA edge statistics across differential gene expressions across indicated populations. (**F,G**) ATAC-seq peaks are shown in the *KLF1* and *GATA1* loci across indicated populations. Shaded regions highlight differential peaks in the promoter regions of both genes.

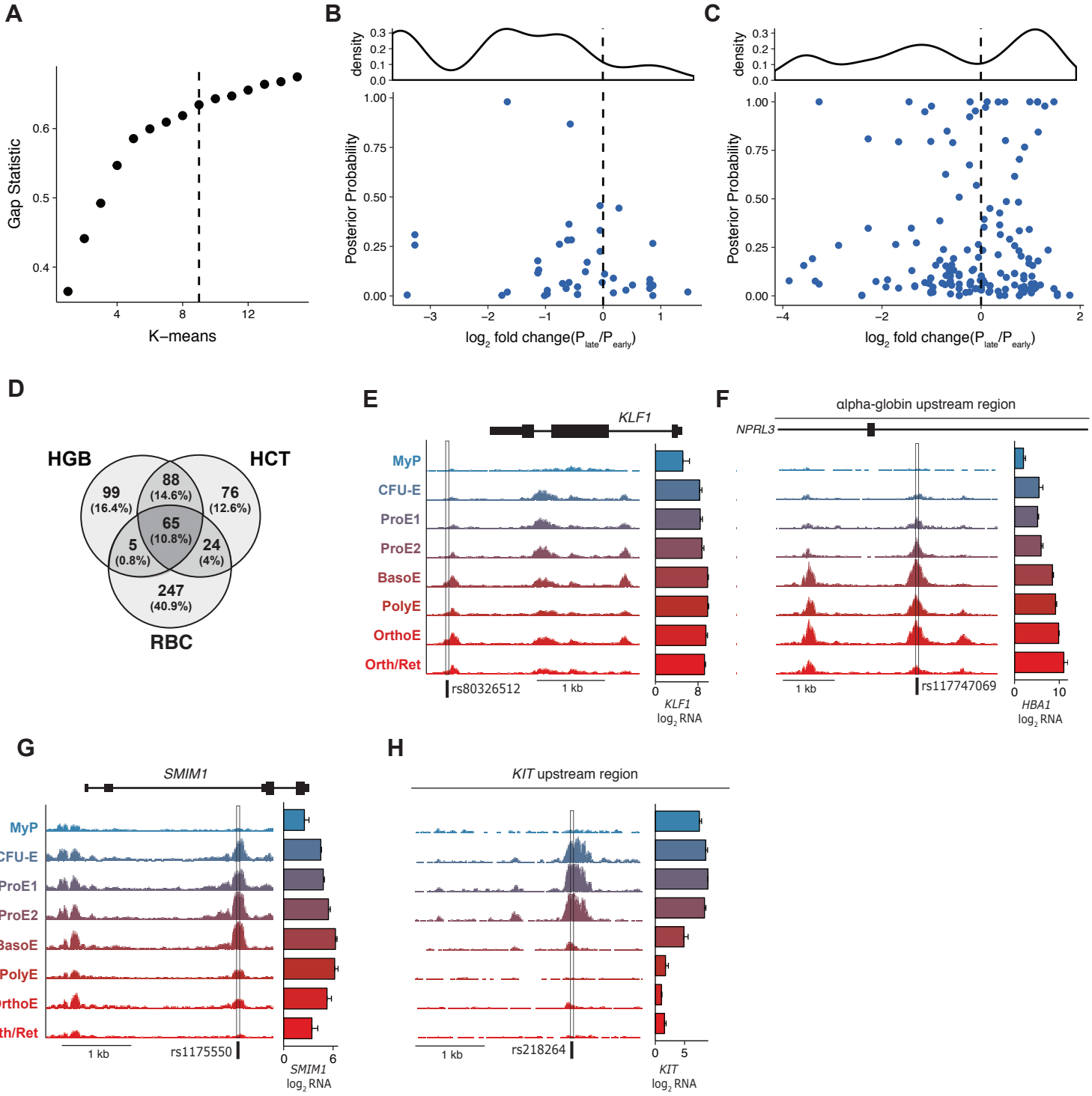
A
Gap Statistic
K−means
B
density
Posterior Probability
log2 fold change(Plate/Pearly)
C
density
Posterior Probability
log2 fold change(Plate/Pearly)
D
HGB
HCT
RBC
99 (16.4%)
88 (14.6%)
76 (12.6%)
65 (10.8%)
5 (0.8%)
24 (4%)
247 (40.9%)
E
KLF1
MyP
CFU-E
ProE1
ProE2
BasoE
PolyE
OrthoE
Orth/Ret
rs80326512
1 kb
KLF1 log2 RNA
F
alpha-globin upstream region
NPRL3
1 kb
rs117747069
HBA1 log2 RNA
G
SMIM1
MyP
CFU-E
ProE1
ProE2
BasoE
PolyE
OrthoE
Orth/Ret
1 kb
rs1175550
SMIM1 log2 RNA
H
KIT upstream region
1 kb
rs218264
KIT log2 RNA

Figure S5, related to Figure 5. *Cis*-regulatory variation and dynamics in human erythropoiesis. (A) Gap statistic associated with the choice of k=9 clusters for the ATAC-seq peaks overlapping fine-mapped genetic variants with PP > 0.10. Fine-mapped GWAS variants associated with (**B**) HGB/HCT and (**C**) MCV/MCH/MCHC/RBC, plotted by posterior probability of causal association vs. log2 fold-change in ATAC-seq counts per million from CFU-E to ProE2 (P_{early}) to BasoE/PolyE (P_{late}). Density plots of variants are shown above, weighted by PP*log2 fold-change. (**D**) Venn diagram showing the overlap of fine mapped variants associated with HGB, HCT and RBC with PP > 0.10. (**E-H**) ATAC-seq peaks are shown at indicated loci across indicated populations. Boxes indicate open chromatin regions overlapping with erythroid-trait associated variants. These open chromatin regions show different accessibility dynamics throughout differentiation, suggesting potential windows of active gene regulation. Bar graphs indicate mean (+/- SEM) log_2 counts-per-million RNA-seq reads per-population.

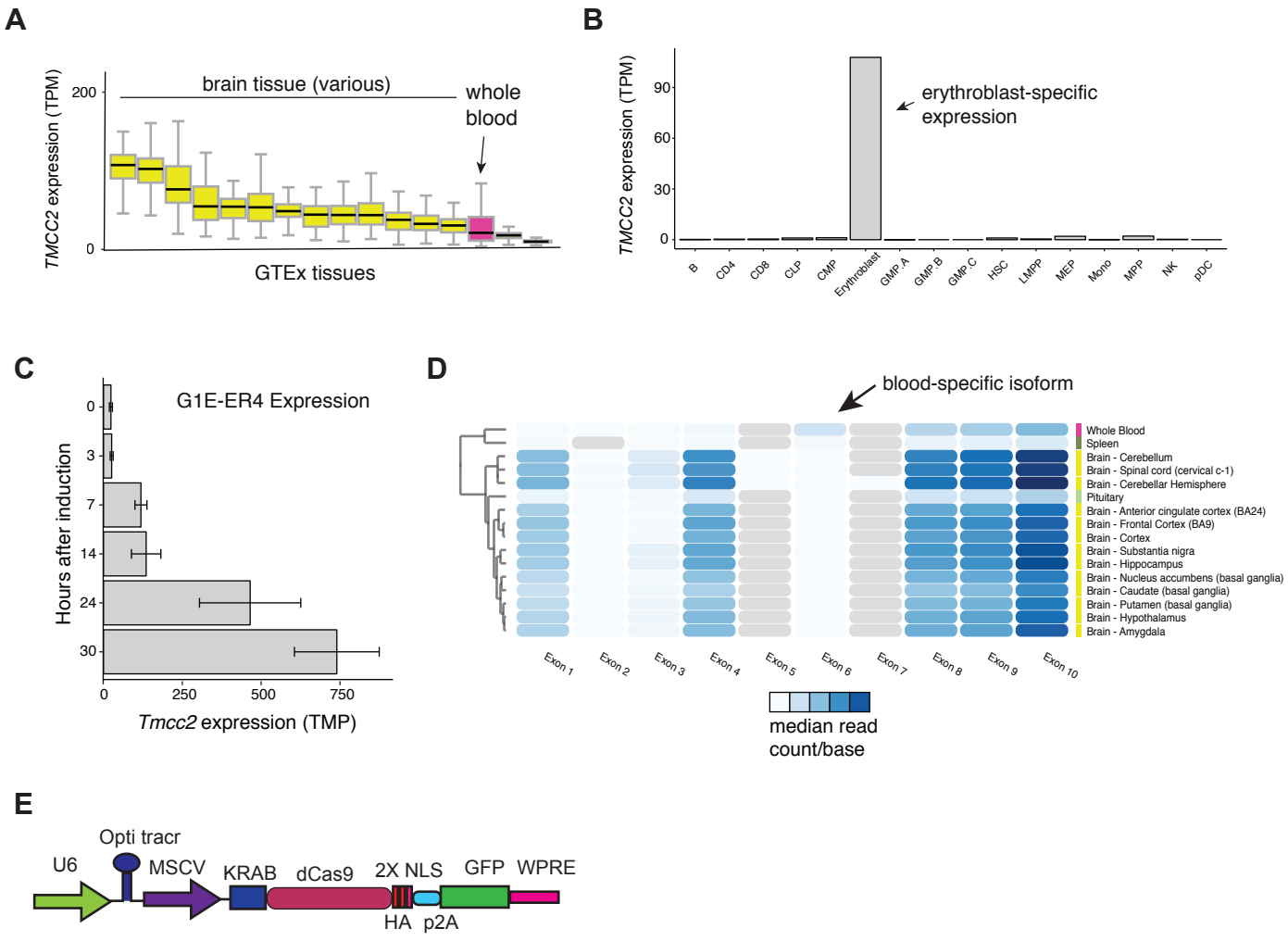
A
TMCC2 expression (TPM)
200
0
brain tissue (various)
whole blood
GTEx tissues
B
TMCC2 expression (TPM)
90
60
30
0
erythroblast-specific expression
B
CD4
CD8
CLP
CMP
Erythroblast
GMP.A
GMP.B
GMP.C
HSC
LMPP
MEP
Mono
MPP
NK
pDC
C
G1E-ER4 Expression
Hours after induction
0
3
7
14
24
30
0
250
500
750
Tmcc2 expression (TMP)
D
blood-specific isoform
Whole Blood
Spleen
Brain - Cerebellum
Brain - Spinal cord (cervical c-1)
Brain - Cerebellar Hemisphere
Pituitary
Brain - Anterior cingulate cortex (BA24)
Brain - Frontal Cortex (BA9)
Brain - Cortex
Brain - Substantia nigra
Brain - Hippocampus
Brain - Nucleus accumbens (basal ganglia)
Brain - Caudate (basal ganglia)
Brain - Putamen (basal ganglia)
Brain - Hypothalamus
Brain - Amygdala
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Exon 7
Exon 8
Exon 9
Exon 10
median read count/base
E
Opti tracr
U6
MSCV
KRAB
dCas9
2X NLS
GFP
WPRE
HA
p2A

Figure S6, related to Figure 6. Erythroid-specific isoform expression and regulation of *TMCC2*. (A) *TMCC2* shows high transcript expression in the human brain and whole blood. Center line: median; box limits: upper and lower quartiles; whisker: 1.5x interquartile range. (**B**) *TMCC2* is highly expressed in erythroblasts among displayed hematopoietic cell types. (**C**) *Tmcc2* is strongly induced in differentiating G1E-ER4 cells. Error bars represent +/- 1 standard deviation between replicates. (**D**) *TMCC2* shows a blood-specific isoform including exon 6, which is absent in different regions of the human brain. Color bar: Median read count per base. (**E**) Schematic of the lentiviral vector construct expressing a guide RNA and the KRAB domain fused to dCas9.

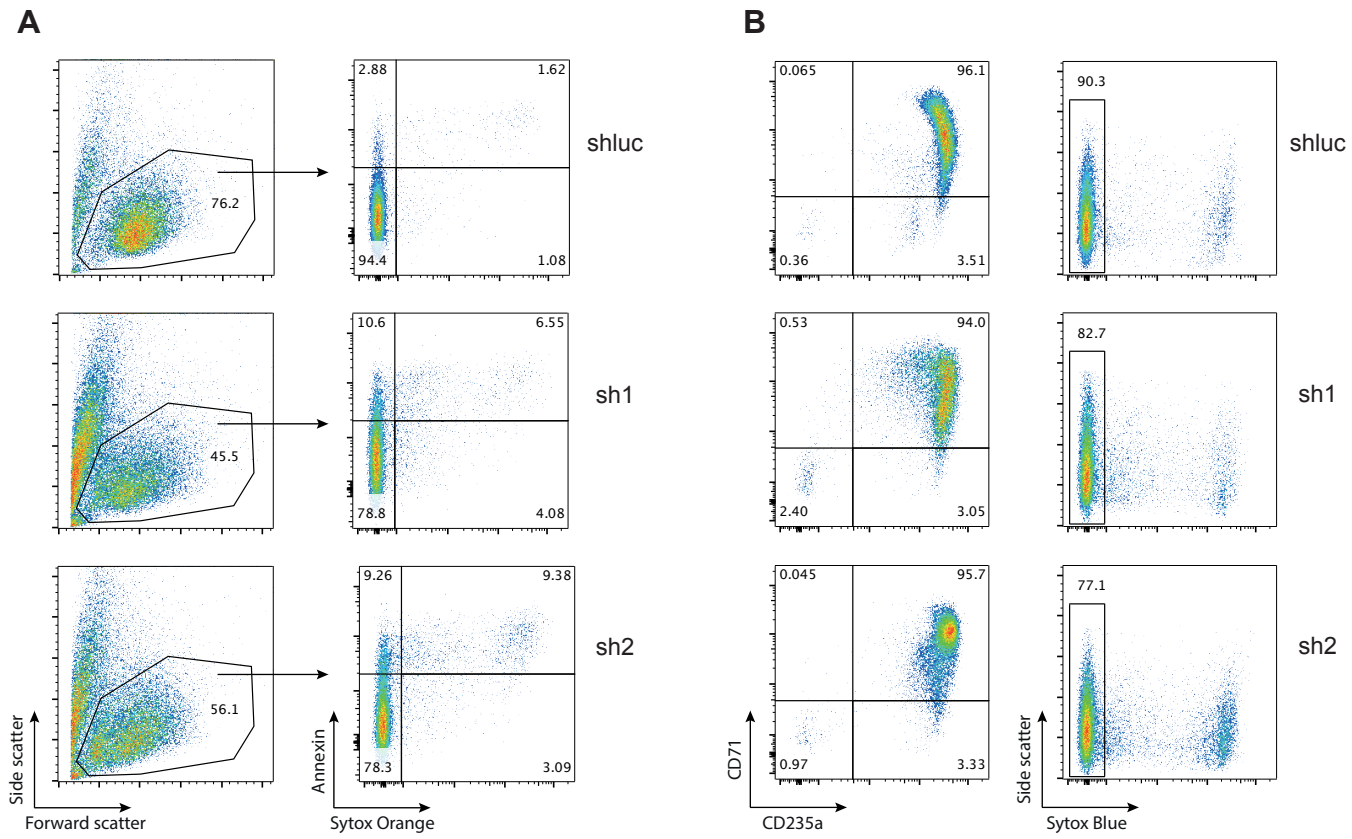
A
B
shluc
sh1
sh2
76.2
2.88
1.62
94.4
1.08
45.5
10.6
6.55
78.8
4.08
56.1
9.26
9.38
78.3
3.09
Side scatter
Forward scatter
Annexin
Sytox Orange
0.065
96.1
0.36
3.51
90.3
0.53
94.0
2.40
3.05
82.7
0.045
95.7
0.97
3.33
77.1
CD71
CD235a
Side scatter
Sytox Blue

Figure S7, related to Figure 7. *TMCC2* is an essential regulator of terminal human erythropoiesis. (**A**) Flow cytometry plots showing forward and side scatter distributions (left) and Annexin vs. Sytox Orange staining (right) in *TMCC2* knockdown cells (sh1, sh2) compared to control cells (shluc). (**B**) Flow cytometry plots showing CD71 and CD235 surface marker expression (left) and side scatter vs. Sytox Blue staining (right) in *TMCC2* knockdown cells (sh1, sh2) compared to control cells (shluc).