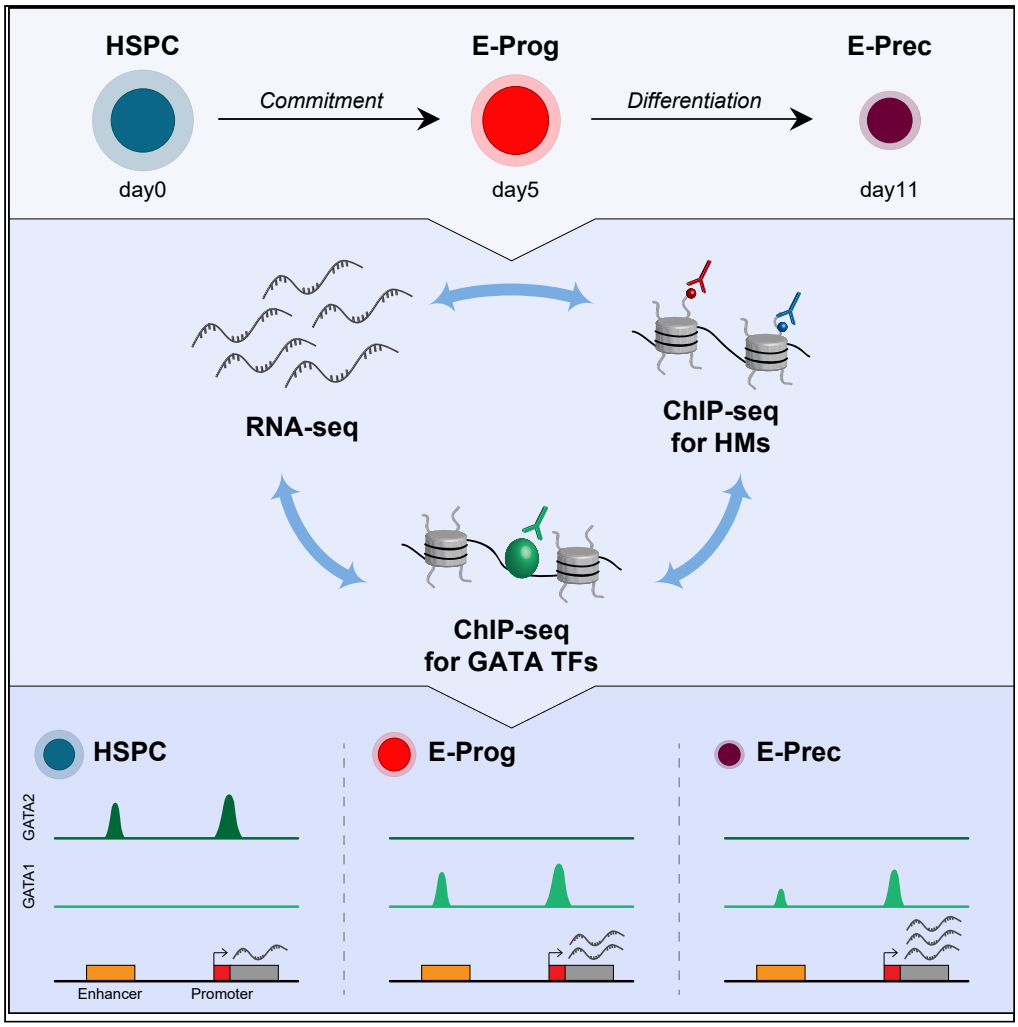
iScience

CellPress

Article

GATA Factor-Mediated Gene Regulation in Human Erythropoiesis



Oriana Romano, Luca Petiti, Tristan Felix, ..., Silvio Bicciato, Clelia Peano, Annarita Miccio

clelia.peano@humanitasresearch.it (C.P.)
annarita.miccio@institutimagine.org (A.M.)

HIGHLIGHTS

GATA2/1 binding to regulatory regions and transcriptional changes during erythropoiesis

GATA1 sustains KIT expression in human erythroid progenitors

DATA AND CODE AVAILABILITY

GSE124165

Romano et al., iScience *23*, 101018
April 24, 2020
https://doi.org/10.1016/j.isci.2020.101018

iScience

CellPress

Article

GATA Factor-Mediated Gene Regulation in Human Erythropoiesis

Oriana Romano,[1] Luca Petiti,[2] Tristan Felix,[3] Vasco Meneghini,[3] Michel Portafax,[3] Chiara Antoniani,[3] Mario Amendola,[4] Silvio Bicciato,[1] Clelia Peano,[2,6,7,8,*] and Annarita Miccio[3,5,8,9,*]

SUMMARY

Erythroid commitment and differentiation are regulated by the coordinated action of a host of transcription factors, including GATA2 and GATA1. Here, we explored GATA-mediated transcriptional regulation through the integrative analysis of gene expression, chromatin modifications, and GATA factors' binding in human multipotent hematopoietic stem/progenitor cells, early erythroid progenitors, and late precursors. A progressive loss of H3K27 acetylation and a diminished usage of active enhancers and super-enhancers were observed during erythroid commitment and differentiation. GATA factors mediate transcriptional changes through a stage-specific interplay with regulatory elements: GATA1 binds different sets of regulatory elements in erythroid progenitors and precursors and controls the transcription of distinct genes during commitment and differentiation. Importantly, our results highlight a pivotal role of promoters in determining the transcriptional program activated upon erythroid differentiation. Finally, we demonstrated that GATA1 binding to a stage-specific super-enhancer sustains the expression of the KIT receptor in human erythroid progenitors.

INTRODUCTION

The acquisition of cellular identity during stem cell commitment and differentiation relies on a combination of genetic and epigenetic information that ultimately determines cellular transcriptional outputs. Master transcription factors are responsible for the selection of unique enhancer repertoires and activate a cascade of epigenetic events (e.g., modification of histone tails and loss of DNA methylation) that could lead to cell-specific modulation of gene expression (Heinz et al., 2015). Recent advances in genome-wide technologies and bioinformatic data integration allow the elucidation of the molecular mechanisms underlying cell fate decision and lineage development with unprecedented levels of detail by analyzing transcriptional and epigenetic changes occurring at different stages of lineage progression.

Human erythropoiesis is an ideal model for studying mechanisms regulating cell commitment and differentiation as the individual developmental cell stages can be isolated and surface markers as well as the master transcription factors controlling this process are largely known. Erythropoiesis is a multi-step process that includes early erythroid commitment of hematopoietic stem cells (HSC), terminal erythroid differentiation, and reticulocyte maturation (Dzierzak and Philipsen, 2013). During early erythroid commitment, HSC give rise to highly proliferating committed erythroid progenitors, erythroid burst-forming units (BFU-Es), and then erythroid colony-forming units (CFU-Es). Erythroid progenitors subsequently undergo terminal differentiation, sequentially producing different populations of erythroid precursors (proerythroblasts, basophilic, polychromatic, and orthochromatic erythroblasts). During this process, cell size is progressively reduced and the cell membrane is reorganized; the cytoplasm first becomes basophilic, as ribosomes accumulate, and then eosinophilic, due to massive production of hemoglobin, whereas the nucleus becomes smaller, as a result of the progressive chromatin condensation. Finally, orthochromatic erythroblasts extrude their nuclei, endoplasmic reticulum, and mitochondria, generating reticulocytes. During maturation, reticulocytes lose the ribosomes and reorganize the cytoskeleton and cell membrane to acquire the distinctive biconcave shape of red blood cells.

Each developmental stage is characterized by a distinct transcriptional program, with a burst of erythroid-specific genes' expression occurring at the early stage of development, followed by the gradual silencing of the transcriptome in late erythroid precursors (An et al., 2014; Li et al., 2014). These transcriptional changes are governed by complex regulatory networks, consisting of the functional interplay between genomic regulatory regions (i.e., promoters and enhancers) and master transcription factors. In particular,

[1]Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

[2]Institute of Biomedical Technologies, CNR, Milan, Italy

[3]Laboratory of Chromatin and Gene Regulation during Development, Imagine Institute, INSERM UMR, 1163 Paris, France

[4]Genethon, INSERM UMR951, Evry, France

[5]Paris Descartes, Sorbonne Paris Cité University, Imagine Institute, Paris, France

[6]Institute of Genetic and Biomedical Research, UOS Milan, National Research Council, Rozzano, Milan, Italy

[7]Genomic Unit, Humanitas Clinical and Research Center, IRCCS, Rozzano, Milan, Italy

[8]These authors contributed equally

[9]Lead Contact

*Correspondence: clelia.peano@humanitasresearch.it (C.P.), annarita.miccio@institutimagine.org (A.M.)

https://doi.org/10.1016/j.isci.2020.101018

enhancers are the primary determinants of the gene expression program at the early stage of erythropoiesis (Huang et al., 2016; Romano et al., 2016; Xu et al., 2012).

GATA2 and GATA1 transcription factors are essential for hematopoietic development and recognize similar GATA DNA motifs. GATA2 has a fundamental role in the expansion and survival of hematopoietic stem and progenitor cells (HSPC) and has been mainly described as a positive regulator of gene expression (Vicente et al., 2012). GATA1 is the master regulator of erythropoiesis and functions as an activator or repressor depending on the chromatin context and cofactors (Ferreira et al., 2005). During erythropoiesis, the *GATA2* locus is shut down, whereas GATA1 levels increase, and this transcriptional change (known as *GATA factor switching*) is essential for survival and terminal differentiation of erythroid cells (Bresnick et al., 2010; Moriguchi and Yamamoto, 2014; Suzuki et al., 2013). As an example, in murine erythroid progenitors, GATA2 enhances the expression of the stem cell factor receptor KIT, which is essential for their proliferation, whereas GATA1 is responsible for KIT down-regulation, which is required to achieve terminal differentiation (Hong et al., 2005; Jing et al., 2008; Munugalavadla et al., 2005). A GATA2-to-GATA1 exchange takes place at specific genomic sites containing GATA DNA motifs ("GATA switching sites") and is critical to determine changes in the expression of target genes during erythroid development (Bresnick et al., 2010; Dore et al., 2012; Huang et al., 2016; Moriguchi and Yamamoto, 2014). Several studies investigated GATA2 and GATA1 genome-wide occupancy in both mouse (Cheng et al., 2009; Dore et al., 2012; May et al., 2013; Wu et al., 2011; Yu et al., 2009) and human hematopoietic cells (Beck et al., 2013; Fujiwara et al., 2009; Hu et al., 2011; Huang et al., 2016; Ludwig et al., 2019; Schulz et al., 2019; Su et al., 2013; Xu et al., 2012). However, a comprehensive analysis of the dynamics of GATA factor binding to regulatory elements during erythroid commitment and differentiation is still lacking.

Here, we investigated the epigenetic and transcriptional changes occurring during human erythroid development. Integrating RNA sequencing (RNA-seq) data, chromation immunopreciptation sequencing (ChIP-seq) analysis of histone modifications typical of regulatory elements, and GATA factors' binding profiles in human HSPC, early erythroid progenitors, and late precursors, we shed light on the regulatory mechanisms controlling stage-specific transcriptional programs and on the distinct role of GATA1 in the early and late phases of human erythropoiesis. The novel key points of our study are as follows: (1) a progressive decrease of H3K27 acetylation, a histone mark typical of active regulatory regions, is a major epigenetic change during erythroid development and is associated with a reduction of active enhancers and super-enhancers (SEs) upon differentiation; (2) promoters are the primary determinants of the gene expression program at the late stage of erythropoiesis; (3) integration of the chromatin landscape and GATA1 occupancy revealed that GATA1 plays a crucial role in determining the global transcriptional changes occurring during erythroid development; (4) GATA1 exerts its transcriptional activity by occupying mainly promoters in late precursors; and (5) differently from mouse cells, GATA1 activates *KIT* gene expression in human erythroid progenitors.

RESULTS

Transcriptional Profiling Reveals Major Changes of Gene Expression upon Erythroid Differentiation

To gain a comprehensive view of gene expression changes occurring during erythroid development, we performed RNA-seq analysis of human HSPC differentiated *in vitro* into early committed erythroid progenitors (E-Prog; $CD34^{low}CD36^{+}GYPA^{low}$) and late erythroid precursors (E-Prec; $CD34^{-}CD36^{+}GYPA^{high}$) (Figure 1A and S1A–S1C). E-Prog contain BFU-E and CFU-E progenitors (Figure S1C and Romano et al., 2016), and E-Prec population consists mainly of polychromatic erythroid precursors (Figure S1C and data not shown). Overall, most genes displayed a lower expression level in E-Prec compared with HSPC and E-Prog (Figure S1D; An et al., 2014; Wong et al., 2011; Shi et al., 2014). Supervised analysis identified 2,485 differentially expressed genes (DEGs) during commitment (1,203 up-regulated and 1,282 down-regulated in HSPC-to-E-Prog transition; Figure S1E) and 6,496 DEGs upon differentiation (2,983 up-regulated and 3,513 down-regulated genes in E-Prog-to-E-Prec transition; Figure S1E). We classified all 7,469 unique DEGs in eight groups, according to their modulation during erythroid development (Figure 1B). Some DEGs were up- or down-regulated only in one transition (groups 1 and 5 upon commitment and groups 3 and 7 upon differentiation); others were progressively modulated during erythroid development (groups 2 and 6) or showed a non-monotonic expression pattern (groups 4 and 8). Different gene expression patterns reflected specific biological processes (Figure S1F). As reported in other studies (Li et al., 2014; Shi et al., 2014; Wong et al., 2011), genes progressively up-regulated during erythroid development (group

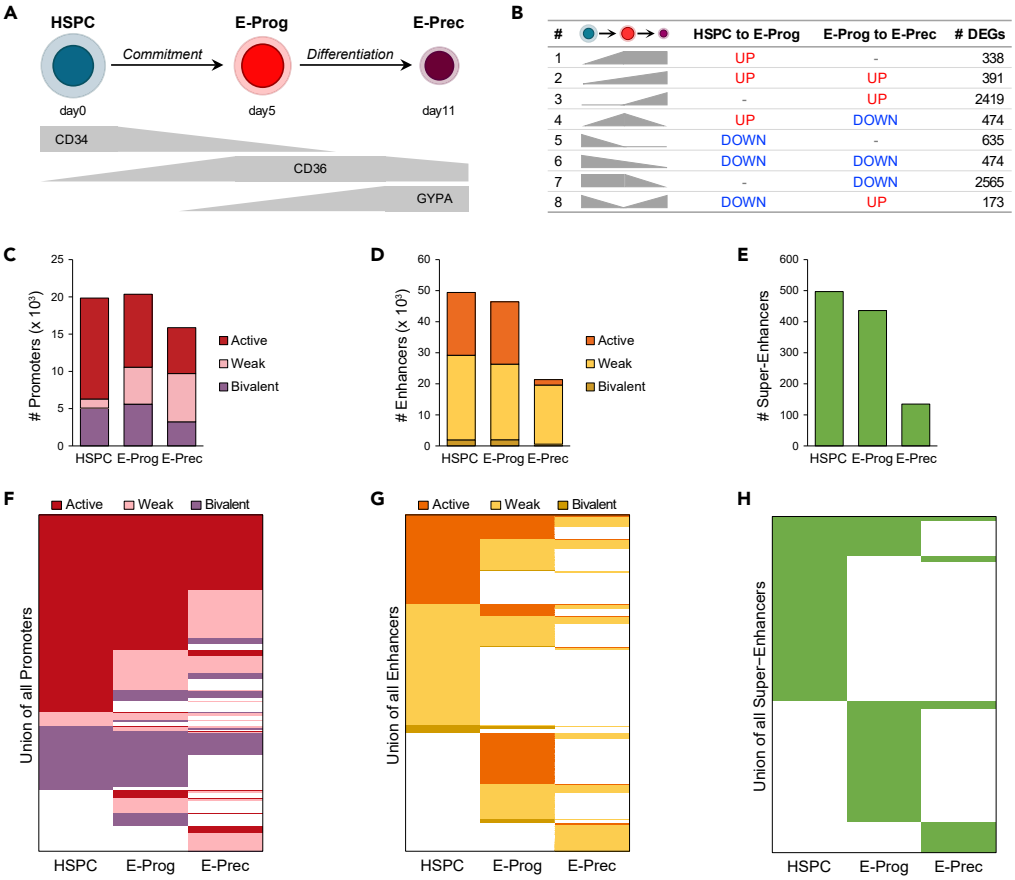


Figure 1. Changes in the Transcriptomic and Epigenomic Profile during Erythroid Development

(A) Schematic representation of erythroid development. HSPC are committed toward the erythroid lineage giving rise to E-Prog that then differentiate in E-Prec. Gray bars describe the expression of CD34, CD36, and GYPA surface markers during erythroid development. See also Figure S1.

(B) Different groups of DEGs, defined according to their modulation (up- or down-regulation) during erythroid commitment (HSPC to E-Prog) and differentiation (E-Prog to E-Prec). See also Figure S1.

(C–E) Bar plots showing the number of promoters (C), enhancers (D), and super-enhancers (E) identified in each cell population. Promoters and enhancers were classified as active ($H3K27ac^{+}H3K27me3^{-}$), weak ($H3K27ac^{-}H3K27me3^{-}$), or bivalent ($H3K27ac^{-}H3K27me3^{+}$). Active enhancers were used as constituent enhancers to identify super-enhancers. See also Figure S2 and Table S1.

(F–H) Heatmaps showing the dynamic usage of promoters (F), enhancers (G), and super-enhancers (H) during erythroid commitment and differentiation. Each row represents a regulatory region. The color code indicates active, weak, and bivalent regulatory regions at each stage of erythroid development. White color indicates absence of the regulatory region. See also Figures S3 and S4 and Tables S2, S3, and S4.

2) were related to hemoglobin synthesis and erythrocyte differentiation. Interestingly, we identified novel classes of genes specifically up-regulated during the differentiation phase (group 3), including genes involved in DNA packaging and chromosome condensation, in accordance with the heterochromatinization known to occur at late stages of erythropoiesis (Ji, 2015). Conversely, HSPC-specific markers (i.e., CD34 and CD133) or genes involved in other blood lineages' biology were down-regulated during commitment (group 5) or progressively down-regulated during erythroid development (group 6), whereas genes related to ribosome biogenesis were down-regulated in the differentiation phase (group 7), consistently with the ribosome loss occurring during terminal maturation (Moras et al., 2017). Interestingly, genes up-regulated upon erythroid commitment and down-regulated during differentiation (group 4) were involved in metal ion transport, cell morphogenesis, cytokine production, and signaling pathways, and included *KIT*, a gene essential for E-Prog survival and proliferation that must be down-regulated to achieve terminal erythroid maturation (Munugalavadla and Kapur, 2005).

Active Chromatin Regions Are Lost during Erythroid Commitment and Differentiation

To investigate the chromatin changes occurring during erythroid development, we analyzed the genome-wide distribution of histone modifications typically associated with promoters (H3K4me3) and enhancers (H3K4me1), active regulatory regions (H3K27ac), and Polycomb repression (H3K27me3). The fraction of the genome enriched in H3K4me3 was comparable across all stages, whereas genome coverage for all other histone modifications was reduced in E-Prog and E-Prec compared with HSPC. In particular, the coverage of H3K27ac and the amount of H3 histones harboring this modification progressively decreased during erythroid development (Figures S2A and S2B).

To define the epigenetic landscape in a systematic manner, we exploited a machine learning approach to identify chromatin states using these four histone marks with a resolution of 200 bp. We defined five promoter states ($H3K4me3^{+}$), four enhancer states ($H3K4me1^{+}$), a polycomb-repressed state ($H3K27me3^{+}$), and a quiescent state devoid of any histone mark (Figures S2C and S2D). Based on the presence of H3K27ac or H3K27me3, we classified promoter and enhancer states as active ($H3K27ac^{+}H3K27me3^{-}$), weak ($H3K27ac^{-}H3K27me3^{-}$), or bivalent ($H3K27ac^{-}H3K27me3^{+}$; Ernst et al., 2011; Roadmap Epigenomics Consortium et al., 2015). Finally, we merged contiguous genomic segments (see Methods) to define different sets (active, weak, and bivalent) of promoter and enhancer regions (with a minimum size of 400 bp) for each stage of erythroid development (Tables S1, S2, and S3). Moreover, we defined SEs as clusters of active enhancers (Tables S1 and S4; Whyte et al., 2013; Hnisz et al., 2013). To validate our approach, we analyzed the expression levels of genes associated with different classes of regulatory elements. Genes associated with active regulatory regions showed higher expression levels compared with genes associated with weak or bivalent regions, which were expressed at medium and low levels, respectively (Figures S2E and S2F). We identified around 20,000 promoters in HSPC and E-Prog and 15,868 in E-Prec (Figure 1C), more than 45,000 enhancers in HSPC and E-Prog and only 21,337 in E-Prec Figure 1D, and 497 SEs in HSPC, 436 in E-Prog, and only 135 in E-Prec (Figure 1E). Most of the regulatory elements were active in HSPC, whereas the fraction of active promoters (Figures 1C and S2D), enhancers (Figures 1D and S2D), and SEs (Figure 1E) diminished in E-Prog and E-Prec in accordance with the progressive loss of H3K27ac during erythroid development.

Interestingly, most promoters were shared during erythroid development, with only 10% being stage specific (Figures 1F and S3A). Promoters that maintained an active state throughout erythroid development drove the expression of genes involved in common cell functions, as cell metabolism, cell cycle, and chromatin organization (Figure S3B). Bivalent promoters that maintained their state in both commitment and differentiation or completely lost H3K4me3 were associated with genes involved in non-hematopoietic tissue and organ development that need to be repressed to maintain lineage fidelity (Figure S3B). Interestingly, genes related to erythropoiesis were associated with either active promoters that maintained their state during the entire erythroid development or E-Prec-specific promoters.

Enhancer usage dramatically changed during commitment and differentiation, resulting in almost 40% enhancers being stage-specific in each cell type (Figures 1G and S3C). Genes targeted by active and weak stage-specific enhancers in HSPC were mainly involved in leukocyte differentiation, whereas in E-Prog they were related to chromosome organization, cytoskeleton and plasma membrane organization, and cell cycle, and in E-Prec they were related to erythrocyte differentiation, chromatin organization, and autophagy (Figure S3D).

SEs' usage was almost completely stage specific, with less than 30% SEs identified in a single cell type shared with the other stages (Figures 1H and S4A). HSPC-specific SEs were annotated to genes involved in leukocyte biology (Figure S4B) and to stem cell markers as *DNMT3A*, *CD34*, *PROM1* (*CD133*), *RUNX2/RUNX3*, *FLI1*, *ERG*, and *GFI1* (Figure S4C). Instead, genes targeted by E-Prog- and E-Prec-specific SEs were mostly related to erythrocyte biology, such as *CD55*, *RHAG*, *HBS1L*, *CD36*, *SLC44A1*, *SLC40A1*, and *SPTA1* in E-Prog and *RHD RBM38*, *HEMGN*, *TMEM56*, *SLC2A1*, *SLC22A23*, *SLC25A37*, *SLC22A4*, and *HBE1* (beta-globin locus control region) in E-Prec (Figures S4D and S4E). Only few loci (n = 60; e.g., *KIT* gene) displayed a combination of common and stage-specific SEs, suggesting a fine modulation of the expression of these genes during erythroid differentiation (see Figure 4).

GATA Transcription Factors' Occupancy during Erythroid Commitment and Differentiation

To elucidate how GATA factors regulate gene expression during erythroid development, we analyzed GATA2 and GATA1 binding profiles in HSPC, E-Prog, and E-Prec. Although GATA2 mRNA levels were

comparable in HSPC and E-Prog (Figure 2A), GATA2 protein was present at high levels in HSPC and decreased during erythroid development (Figure 2B). GATA1 protein was mainly present in E-Prog and substantially reduced in E-Prec (Figures 2B and S5A and Ludwig et al., 2019) despite the progressive increase of *GATA1* mRNA expression during differentiation (Figure 2A). The dynamic expression of GATA proteins clearly emerged when analyzing GATA factors' genome-wide occupancy. Indeed, the large number of GATA2-binding sites (BS) identified in HSPC (n = 15,171) drastically decreased in E-Prog cells (n = 419) and was reduced to zero in E-Prec (Figure 2C, Table S5 and Figures S5B and S5C). Similarly, we identified more than 23,000 GATA1 BS in E-Prog and almost half of these in E-Prec (n = 11,005), whereas no GATA1 BS was recovered in HSPC (Figure 2C, Table S5 and Figures S5B and S5C). Virtually all GATA2 BS were HSPC specific, and one-third of them underwent a GATA2-to-GATA1 switch during HSPC-to-E-Prog transition and 40% of these remained bound by GATA1 in E-Prec (Figure 2C). E-Prog and E-Prec shared ~7,000 GATA1 BS, whereas ~16,000 and ~4,000 stage-specific GATA1 BS were identified in E-Prog and E-Prec, respectively (Figure 2C). These data indicate that GATA1 occupies and regulates common and different genes during the early and late stages of erythroid development.

Genomic regions targeted by GATA2 were enriched in DNA motifs for ETS and RUNX factors, preferentially expressed in earlier stages of erythroid development, whereas GATA1-targeted regions contained motifs for the erythroid-specific KLF1 factor, and, in E-Prec, for the ubiquitous SP1 and NFY factors that occupy preferentially promoter elements (Figure S5D).

Mapping GATA BS to regulatory regions revealed that the two factors targeted only a small fraction of promoters and enhancers, whereas GATA2 and GATA1 occupied the majority of SEs in HSPC and in E-Prog/E-Prec, respectively (Figures S6A–S6C and Tables S2, S3, and S4). Among BS mapped to regulatory regions, 72% HSPC GATA2 BS were located within enhancers and SEs. GATA1 occupies both enhancers/SEs (58%) and promoters (42%) in E-Prog, whereas in E-Prec, GATA1 mainly targeted promoter regions (72%, Figure 2D). These results indicate that GATA1 may exert its transcriptional activity by occupying both enhancers and promoters in committed erythroid progenitors and mainly promoters in more differentiated precursors.

We then associated GATA-occupied regulatory regions with their target genes. Around half of GATA2-targeted genes in HSPC were targeted by GATA1 in E-Prog (Figure 2E) and more than 75% of them presented at least one GATA switching site. Interestingly, GATA1 occupied common and different sets of genes in E-Prog and E-Prec (Figure 2E). Genes targeted only by GATA2 in HSPC (group A) were down-regulated during erythroid development (Figure 2F) and are involved in immune system biological processes (Figure 2G). This analysis suggests that these genes are activated by GATA2 in HSPC, whereas loss of GATA2 binding during erythroid development leads to their down-regulation. Genes undergoing GATA2-to-GATA1 exchange upon erythroid commitment were either up-regulated or their expression remained stable, suggesting that GATA1 increases or sustains their expression in E-Prog, whereas GATA2 might repress them or maintain their low expression in HSPC. Then, these genes were either turned off during differentiation, when losing GATA1 binding in E-Prec (group B; Figures S6D and S6E), or up-regulated during erythroid development, if still targeted by GATA1 in E-Prec (group C; Figures 2F, S6D, and S6E). Group B was functionally enriched in immune cell activation and hematopoiesis, whereas group C was enriched in erythropoiesis, histone modifications, cell cycle, and autophagy (Figure 2G). In E-Prog and E-Prec (group F), GATA1 occupied genes that were up-regulated during erythroid development (Figures 2F, S6D, and S6E) and related to heme metabolic process, membrane lipid distribution, and chromatin organization (Figure 2G). Finally, genes specifically targeted by GATA1 in E-Prec (group G) were up-regulated upon differentiation and involved in cell cycle and autophagy (Figures 2F and 2G). These results suggest that GATA1 exerts mainly activating functions during human erythroid development.

Transcriptional Changes in Erythropoiesis Are Associated with a Complex Interplay between Regulatory Elements and GATA Factors' Binding

To unveil how the interplay between regulatory elements and GATA factors influences transcriptional regulation during erythropoiesis, we assigned the different regulatory elements to each DEG and evaluated the enrichment in GATA factor BS at each stage of erythroid development (Figure 3). In particular, we generated a gene-centered matrix that contains the total coverage of all promoters and

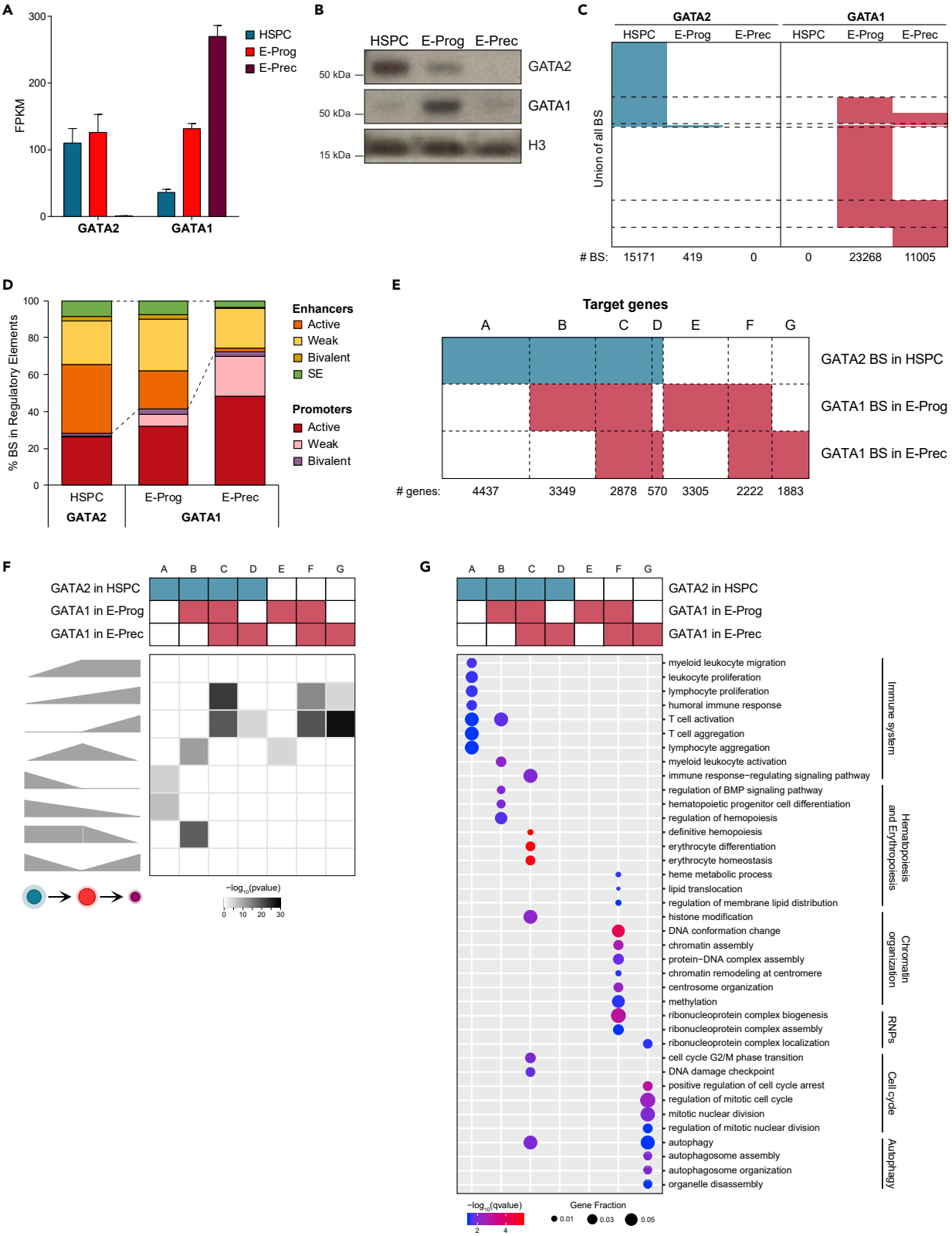

A
FPKM
HSPC
E-Prog
E-Prec
GATA2
GATA1
B
HSPC E-Prog E-Prec
50 kDa
GATA2
50 kDa
GATA1
15 kDa
H3
C
GATA2
GATA1
HSPC E-Prog E-Prec HSPC E-Prog E-Prec
Union of all BS
BS: 15171 419 0 0 23268 11005
D
% BS in Regulatory Elements
Enhancers
Active
Weak
Bivalent
SE
Promoters
Active
Weak
Bivalent
HSPC E-Prog E-Prec
GATA2 GATA1
E
Target genes
A B C D E F G
GATA2 BS in HSPC
GATA1 BS in E-Prog
GATA1 BS in E-Prec
genes: 4437 3349 2878 570 3305 2222 1883
F
A B C D E F G
GATA2 in HSPC
GATA1 in E-Prog
GATA1 in E-Prec
−log10(pvalue)
0 10 20 30
G
A B C D E F G
GATA2 in HSPC
GATA1 in E-Prog
GATA1 in E-Prec
myeloid leukocyte migration
leukocyte proliferation
lymphocyte proliferation
humoral immune response
T cell activation
T cell aggregation
lymphocyte aggregation
myeloid leukocyte activation
immune response−regulating signaling pathway
regulation of BMP signaling pathway
hematopoietic progenitor cell differentiation
regulation of hemopoiesis
definitive hemopoiesis
erythrocyte differentiation
erythrocyte homeostasis
heme metabolic process
lipid translocation
regulation of membrane lipid distribution
histone modification
DNA conformation change
chromatin assembly
protein−DNA complex assembly
chromatin remodeling at centromere
centrosome organization
methylation
ribonucleoprotein complex biogenesis
ribonucleoprotein complex assembly
ribonucleoprotein complex localization
cell cycle G2/M phase transition
DNA damage checkpoint
positive regulation of cell cycle arrest
regulation of mitotic cell cycle
mitotic nuclear division
regulation of mitotic nuclear division
autophagy
autophagosome assembly
autophagosome organization
organelle disassembly
Immune system
Hematopoiesis and Erythropoiesis
Chromatin organization
RNPs
Cell cycle
Autophagy
−log10(qvalue)
2 4
Gene Fraction
0.01 0.03 0.05

Figure 2. GATA Factors Occupancy during Erythroid Development

(A) Expression levels of GATA2 and GATA1 in HSPC, E-Prog, and E-Prec by RNA-seq. Data were plotted as mean with SEM.

(B) Western blot analysis of GATA2 and GATA1 in the nuclear fraction of HSPC, E-Prog, and E-Prec. Total H3 was used for normalization. See also Figure S5.

(C) Heatmap showing GATA2 and GATA1 BS dynamics during erythroid commitment and differentiation. Each row represents a GATA factor BS. The number of BS identified in each population is reported below the heatmap. Dashed lines separate common and stage-specific BS. See also Figure S5 and Table S5.

(D) GATA2 and GATA1 BS distribution within regulatory regions in HSPC, E-Prog, and E-Prec. See also Figure S6.

(E) Heatmap showing different groups of GATA2-and GATA1-targeted genes in HSPC, E-Prog, and E-Prec. We defined GATA2- and GATA1-targeted genes as the three nearest genes (with a maximum distance of 100 kb) to each GATA-targeted regulatory region. Each target gene can be associated with one or more GATA BS.

(F) Heatmap showing the enrichment of the different groups of GATA factor target genes, as defined in (E), in genes with different expression patterns (as defined in Figure 1B). Gray-scale indicates enrichment p value. See also Figure S6.

(G) Gene ontology enrichment analysis of GATA2- and GATA1-targeted genes. Enriched Biological Process (BP) terms are shown on the y axis; different groups of GATA factor target genes, as defined in (E), are shown on the x axis. Dots are color coded based on the enrichment q-values; dot size indicates the fraction of genes in each BP term.

enhancers assigned to each DEG at each cell stage. This procedure allowed grouping of DEGs in 19 clusters based on their different chromatin landscapes. These clusters were either controlled by promoters only (clusters 1, 2, 8, 13, and 17) or by the combined action of promoters and enhancers (Figure 3A). Interestingly, the chromatin landscape of DEGs was more heterogeneous than their transcriptional status, as genes associated with different chromatin landscapes showed the same expression pattern during erythroid development (Figures 3A and 3B). In particular, clusters from 1 to 7 comprised genes that were mostly up-regulated during erythroid commitment and differentiation (e.g., genes involved in erythrocyte biology and chromatin reorganization; Figures 3B and 3C) despite their different chromatin landscapes (Figure 3A). Promoters of these genes were already marked by H3K27ac in HSPC, before erythroid induction, and maintained an active state during the entire development (Figure 3A). These data indicate that promoters of up-regulated genes are bookmarked in earlier stages of erythroid development. Active enhancers, if present, either did not change their state during erythroid development or were even lost, particularly upon terminal differentiation (Figure 3A), indicating that enhancers play a minor role in the transcriptional up-regulation occurring during E-Prog-to-E-Prec transition. On the contrary, genes in clusters from 8 to 16 (e.g., genes related to leukocyte biology and non-hematopoietic development) showed a sharp change in their chromatin landscape, characterized by the loss of active regulatory elements during erythroid development accompanied by the concomitant decrease in the expression levels (Figures 3A–3C).

To investigate this partial dichotomy between epigenetic and transcriptional profiles, we evaluated the enrichment in GATA factors' BS within promoters and enhancers associated with each cluster of DEGs. As expected, GATA2 targeted mostly enhancers in HSPC, whereas GATA1 binds both promoters and enhancers in E-Prog and mainly promoters in E-Prec (Figures 3D and 3E). Considering clusters of up-regulated genes (from 1 to 7), GATA2 BS were enriched at promoters or enhancers of few clusters in HSPC, whereas in E-Prog and E-Prec, GATA1 binding was enriched within active promoters of virtually all clusters and within active and weak enhancers of some of them (Figures 3D and 3E). These findings indicate that GATA1 plays a role in up-regulating these genes during erythroid development mainly through the binding of active promoters, especially in late precursors, whereas GATA2 likely maintains their low activity in HSPC. Clusters of down-regulated genes (from 8 to 16) were characterized by an enrichment of GATA2 BS at active and weak enhancers in HSPC and by an enrichment of GATA1 BS at weak promoters and active and weak enhancers in E-Prog and at weak and bivalent promoters and weak enhancers in E-Prec (Figures 3D and 3E). This suggests that, in HSPC, GATA2 binding at enhancers might play a role in boosting the expression of these genes, which is still sustained by GATA1 in E-Prog. Then, the loss of GATA1 binding at active enhancers or the binding of GATA1 to weak or bivalent promoters might induce the transcriptional silencing of these genes in E-Prec. In E-Prog, GATA1 BS were also enriched at bivalent promoters of lowly expressed genes involved in the development of unrelated tissues (clusters 17–19; Figure 3C) indicating that, when binding at bivalent promoters, GATA1 might act as a transcriptional repressor.

Overall, virtually all the clusters enriched in GATA2 BS within enhancers in HSPC showed enrichment in GATA1 BS in E-Prog and E-Prec, suggesting that the transition from GATA2- to GATA1-mediated gene regulation occurs mainly at enhancer regions (Figure 3E). Of note, 37% enhancers targeted by GATA2 in HSPC contain GATA switching sites.

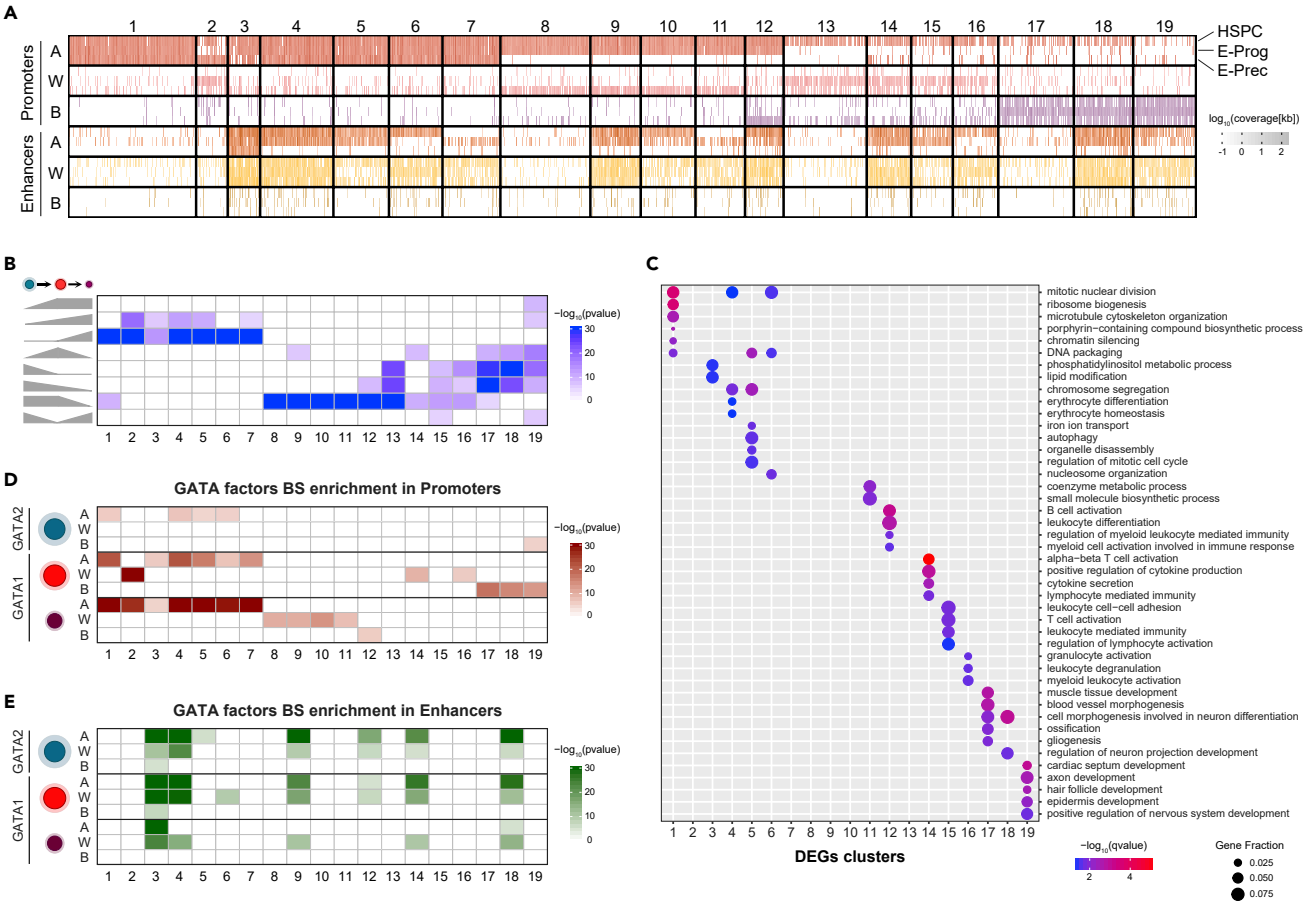


Figure 3. Chromatin Landscape Defines Clusters of DEGs

(A) Heatmap representing the clustering of the gene-centered matrix containing the total coverage of the promoters and enhancers annotated to each DEG. DEGs are clustered according to the total extension of the associated regulatory regions. Each column represents a DEG, and each row, the extension (in kb) of the associated regulatory elements in each cell stage. Color scale indicates the total extension as $\log_{10}$(coverage [kb]).

(B) Heatmap showing cluster enrichment in genes with different expression patterns (as defined in Figure 1B). Color scale indicates the statistical significance of the enrichment calculated by a Fisher's exact test.

(C) Biological process terms (BPs) enriched in each group are shown on the y axis; DEG clusters, as defined in Figure 3A, are shown on the x axis. Dots are color coded based on the enrichment q-values; dot size indicates the fraction of genes in each Gene Ontology term. DEGs of clusters 1–7, mostly up-regulated during erythroid commitment and differentiation, are functionally annotated to erythropoiesis, chromatin organization, cell cycle, and autophagy BPs. Genes within clusters from 8 to 19, mostly down-regulated during erythroid development, were related to immune cell biology (12, 14–16) or to the development of unrelated tissues (17–19).

(D and E) Heatmaps showing cluster enrichment in GATA factors BS within promoters (D) and in GATA factor BS within enhancers (E). Color scale indicates the statistical significance of the enrichment calculated by a Fisher's exact test.

See also Figure S7.

GATA1-bound regions associated with induced genes were enriched in motifs for KLF1 and E2F4 factors (Figure S7), which are up-regulated during erythroid development and play a fundamental role in erythroid differentiation and proliferation (Kinross et al., 2006; Siatecka and Bieker, 2011). Conversely, GATA1-occupied regions associated with down-regulated genes contained motifs for ETS factors (e.g., ERG, FLI1, and PU.1; Figure S7), mostly expressed in HSPC and down-regulated during erythroid development, and fundamental for the maintenance of HSC and the development of other hematopoietic lineages (Athanasiou et al., 2000; Calero-Nieto et al., 2010; Knudsen et al., 2015; Nishiyama et al., 1999; Rekhtman et al., 1999; Wontakal et al., 2012). Thus the low expression of ETS factors in erythroid cells and the consequent lack of ETS factor binding to these genes could contribute to the silencing of genes involved in non-erythroid functions (Wontakal et al., 2012). Finally, composite GATA:TAL1 motifs were enriched in GATA-occupied regulatory regions associated with both up- and down-regulated genes (Figure S7), suggesting a role for TAL1 in both gene activation and gene repression (Huang and Brandt, 2000; Pinello et al., 2014; Schuh et al., 2005; Van Handel et al., 2012).

A Novel GATA1-Dependent Super-enhancer Sustains KIT Expression in Erythroid Progenitors

To prove that this interplay between regulatory elements and GATA factors' binding is crucial to control gene expression during erythroid development, we focused on the *KIT* gene. A precise regulation of KIT expression is required for erythroid progenitor survival and proliferation and to achieve terminal erythroid maturation (Munugalavadla and Kapur, 2005). Indeed, we found that, in HSPC, *KIT* mRNA levels were relatively low and ~65% cells poorly expressed KIT on the cell surface (median fluorescence intensity [MFI] = 17) (Figures 4A and 4B). Upon commitment, *KIT* transcription significantly increased (~14 fold) and almost all E-Prog (~93%) expressed KIT at high levels (MFI = 189) (Figures 4A and 4B). In E-Prec, *KIT* gene was substantially down-regulated with less than 55% cells still expressing low levels of KIT (MFI = 65) (Figures 4A and 4B).

In mouse KIT^+ progenitor cells, GATA2 activates *KIT* expression through the binding of an enhancer located −114 kb upstream of the gene (Jing et al., 2008). Upon erythroid differentiation, this enhancer is occupied by GATA1, which induces *KIT* down-regulation (Jing et al., 2008). The analysis of regulatory elements during human erythroid development revealed that both *KIT* promoter and *KIT*-associated enhancers were active in HSPC and E-Prog and lost H3K27ac upon differentiation (Figure 4C; cluster 9 in Figure 3A). Moreover, *KIT* regulatory elements were targeted by GATA2 in HSPC and by GATA1 upon erythroid commitment (Figure 4C; group B in Figure 2F and cluster 9 of Figure 3D). In particular, we identified three SEs that could contribute to *KIT* transcriptional regulation during erythropoiesis, i.e., an SE located within *KIT* first intron in both HSPC and E-Prog, an HSPC-specific SE located +160/+180 kb downstream of *KIT* TSS (Aranda-Orgilles et al., 2016), and an E-Prog-specific SE located −118/-70 kb upstream of *KIT* TSS (Figure 4C). This latter SE comprised two main constituent active enhancers (Enhancer I and II) marked by high H3K27ac levels and bound by GATA1. Interestingly, upon differentiation, both Enhancer I and II lost H3K27ac and Enhancer II lost GATA1 binding (Figure 4C). In addition, the E-Prog-specific SE contains two additional GATA1-occupied active enhancers that show lower H3K27ac levels compared with Enhancers I and II: (1) Enhancer III that is homologous to the murine −114-kb enhancer but is not targeted by GATA2 in KIT^+ progenitors as in the mouse system and (2) Enhancer IV that is occupied by GATA2 in HSPC (as a weak enhancer) and is targeted by GATA1 upon erythroid commitment (Figure 4C). Of note, in HSPC, besides Enhancer IV, GATA2 occupies also a weak enhancer upstream of the E-Prog-specific SE, suggesting that it may contribute to sustain *KIT* expression in HSPC (Figure 4). Interestingly, two SNPs (rs2703485 and rs218264) associated with red blood cell phenotypes map to the GATA1 BS in Enhancer I and Enhancer IV, respectively (Figure 4C; Astle et al., 2016).

To demonstrate that GATA1 binding to the highly acetylated Enhancer I and II is essential to sustain KIT expression in human erythroid progenitors, we performed chromatin conformation capture (3C) and Cas9-mediated genome editing in HUDEP-2 cells. HUDEP-2 is an immortalized erythroid progenitor cell line (Kurita et al., 2013) that, similarly to E-Prog, expresses high levels of KIT when undifferentiated (Day0) and that, as in the E-Prog-to-E-Prec transition, down-regulates KIT upon differentiation (Day 7) (Figures S8A and S8B). Moreover, undifferentiated HUDEP-2 cells present accessible chromatin regions mapping to the E-Prog-specific *KIT* enhancers (Figure S8C), suggesting that these regulatory elements are active in this cell line (Masuda et al., 2016). 3C experiments evidenced that, only in undifferentiated KIT^{high} HUDEP-2 cells, the *KIT* promoter interacts with Enhancer I and II (Figure 4D) and that Enhancer I and II interact with each other (Figures S8D and S8E). This indicates the presence, in human erythroid progenitors, of an active chromatin hub containing the *KIT* promoter and its active regulatory elements. We then transfected KIT^{high} HUDEP-2 cells with plasmids expressing a Cas9-GFP fusion protein and guide RNAs deleting Enhancer I or II or specifically disrupting their respective GATA1 BS (Figure 5A). Both the deletion of *KIT* Enhancer I or II and the disruption of GATA1 BS strongly reduced *KIT* gene expression, the percentage of KIT-expressing cells, and KIT MFI in the GFP^+-edited fraction (Figures 5B–5E and S9), indicating that GATA1 binding at Enhancer I and II plays a crucial role in boosting *KIT* expression in erythroid progenitor cells. Notably, the lower expression of KIT was accompanied by a marked increase in the expression of the erythroid differentiation marker GYPA, with a trend in decreased GATA2 and increased GATA1 mRNA levels (Figure 5F).

DISCUSSION

In this study, we analyzed the usage of regulatory elements and GATA factor dynamic occupancy in multipotent human primary HSPCs and in their erythroid progeny, including human early erythroid progenitors (Romano et al., 2016) and late polychromatic erythroid precursors.

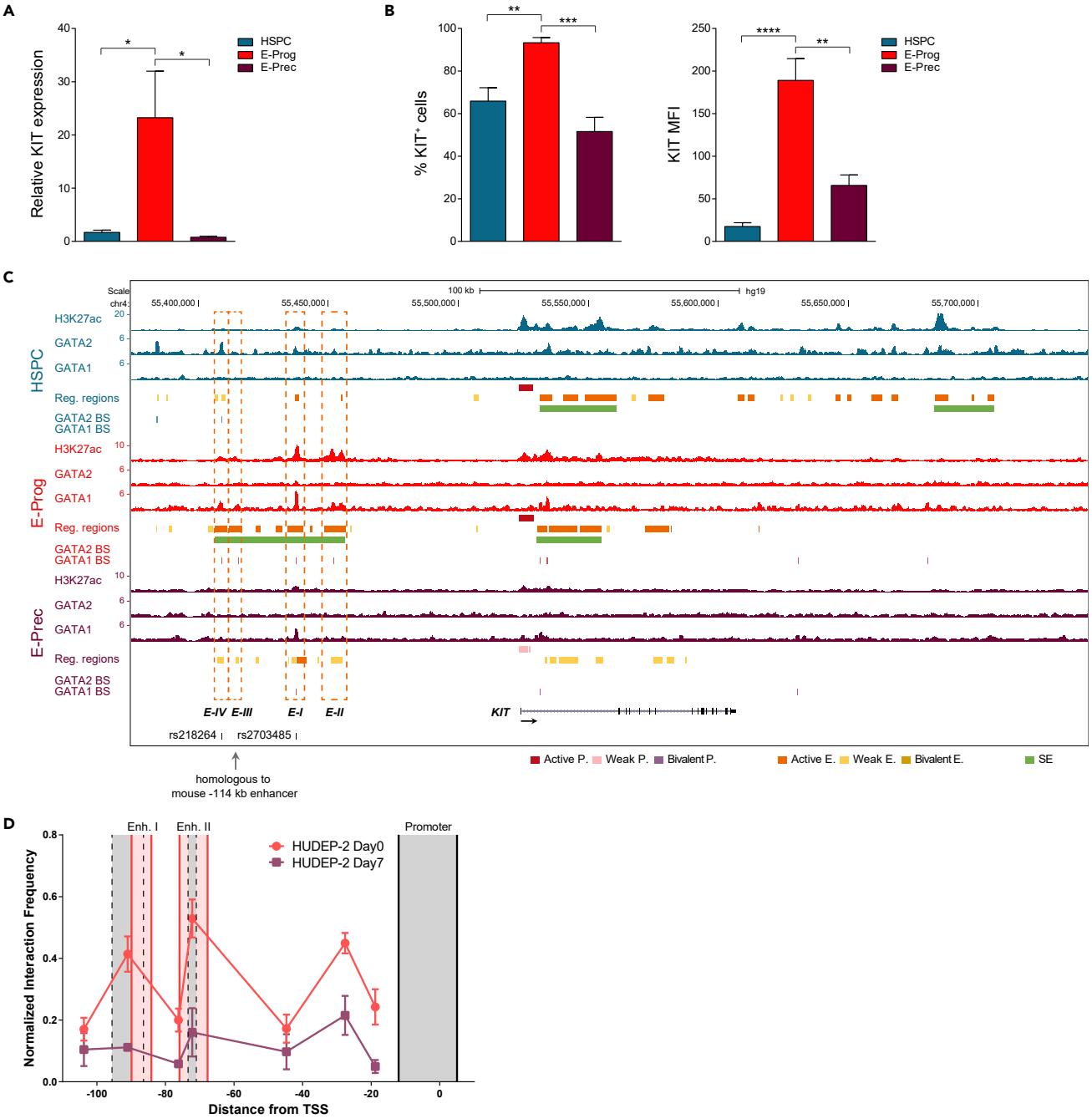


Figure 4. Epigenetic Regulation and Chromatin Interactions in the Human KIT Locus during Erythroid Development

(A) *KIT* mRNA expression levels in HSPC, E-Prog, and E-Prec, as determined by RT-qPCR. Data were plotted as mean with SEM. *p < 0.05 (unpaired t test).

(B) Flow cytometry analysis of KIT expression during erythroid commitment and differentiation. The percentage of KIT^+ cells and the MFI (median fluorescence intensity) are shown. Data were plotted as mean with SEM. **p < 0.01; ***p < 0.001; ****p < 0.0001 (unpaired t test).

(C) Regulatory elements and GATA factors BS within the *KIT* locus. Green boxes indicate SEs. Orange dashed boxes indicate the constituent enhancers within the E-Prog-specific super-enhancer. Enhancer I and II (E-I and E-II) show the highest H3K27ac peaks. Enhancer III (E-III) is homologous to the mouse −114 enhancer. Enhancer IV (E-IV) undergoes GATA2-to-GATA1 switch. SNPs mapping to the E-Prog-specific super-enhancer are indicated.

(D) Chromatin interactions within the *KIT* locus in undifferentiated (Day 0) and differentiated (Day 7) HUDEP-2 cells. We used as anchor a genomic fragment containing the *KIT* promoter (flanked by solid black lines). HindIII digestion fragments of interest are flanked by dashed black lines. Distances on the x axis are in kb counting from the transcription start site (TSS) of the *KIT* gene. *KIT* promoter interacts with Enhancer I and II only in KIT^{high} HUDEP-2 undifferentiated cells. See also Figure S8.

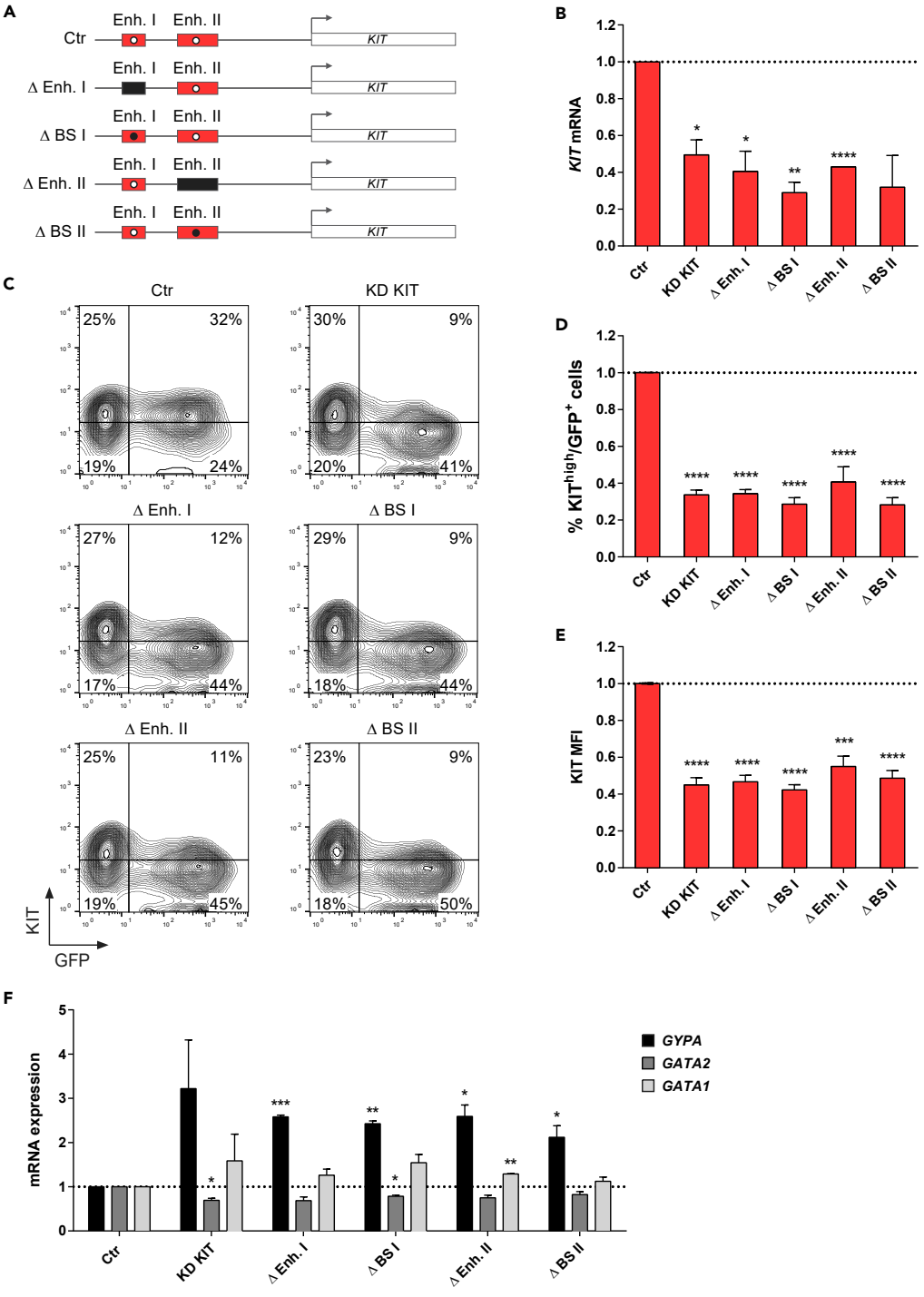


Figure 5. CRISPR/Cas9-Mediated Disruption of E-Prog-Specific *KIT* Regulatory Elements

(A) Schematic representation of the genomic regions and GATA1 BS in the *KIT* locus targeted using the CRISPR/Cas9 system. Red boxes indicate unedited wild-type Enhancer I or II, and white dots indicate unedited wild-type GATA1 BS within the enhancers (Ctr). Black boxes indicate Enhancer I or II deletion (Δ Enh. I and Δ Enh. II). Black dots indicate GATA1 BS disruption (Δ BS I and Δ BS II).

(B) *KIT* expression levels in control and edited cells, as determined by RT-qPCR. Control samples (Ctr) were transfected with plasmids encoding for Cas9-GFP and a guide RNA targeting the luciferase gene. mRNA levels were expressed as fold change versus control cells. Data were plotted as mean with SEM.

Figure 5. *Continued*

(C–E) Flow cytometry analysis of GFP and KIT expression in control and edited cells. For all the edited samples, both percentage of KIT^{high} cells (C and D) and KIT MFI (C and E) in the GFP^{+} populations were reduced compared with control cells. Percentage of KIT^{high} cells and KIT MFI were expressed as fold change versus control cells. Data were plotted as mean with SEM.

(F) *GYPA, GATA2, and GATA1* expression levels by RT-qPCR in control and edited cells. mRNA levels were expressed as fold change versus control cells. Data were plotted as mean with SEM.

(B and D–F) *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001 (unpaired t test).

During erythropoiesis, inhibition of histone acetylation and increase in histone deacetylation play a critical role in chromatin condensation and enucleation of erythroid precursors (Ji et al., 2010, 2011; Popova et al., 2009). Our analysis of the epigenetic landscapes of human HSPC, early erythroid progenitors, and late precursors revealed for the first time a progressive decrease of H3K27ac abundance and genome-wide coverage during erythroid development, accompanied by the up-regulation of genes involved in chromatin condensation and by the overall reduction in gene expression upon differentiation. These results suggest that the decrease of H3K27ac during erythropoiesis contributes to the heterochromatin formation and the consequent global gene silencing at late stage of erythroid development. This extensive epigenetic change influences the regulatory mechanisms controlling transcription, in particular in late erythroid precursors. Indeed, we found a reduced number of active $H3K27ac^{+}$ regulatory elements, in particular enhancers and SEs, upon differentiation, suggesting that enhancer decommissioning occurs at the late stage of erythropoiesis. In accordance, gene expression in late precursors is mainly sustained by promoters and weak $H3K27ac^{-}$ enhancers, and the few SEs identified in E-Prec control the transcription of only a subset of stage-specific highly expressed genes. Our findings are consistent with the decreased number of chromatin accessibility peaks in polychromatic human erythroid precursors (Ludwig et al., 2019; Schulz et al., 2019).

The coordinate action of GATA factors has a fundamental role in the transcriptional regulation of erythropoiesis. The "GATA factor switching" (Bresnick et al., 2010; Moriguchi and Yamamoto, 2014; Suzuki et al., 2013) occurred in E-Prog, where GATA2 protein levels decreased, whereas the amount of GATA1 protein increased, resulting in a prevalent and specific chromatin occupancy of GATA2 in HSPC and GATA1 in E-Prog. Several genes regulated by GATA2 in HSPC were targeted by GATA1 in E-Prog, and the transition from GATA2 to GATA1-mediated gene regulation can occur via a GATA2-to-GATA1 exchange at the same BS (GATA switching sites) and/or through the alternative binding of GATA2 and GATA1 to different sites within the regulatory regions associated with the target gene. Interestingly, for differentially expressed genes, GATA switch occurs mostly at enhancer regions, and either leads to gene up-regulation or does not significantly change gene expression in E-Prog.

However, most GATA1 functions during erythropoiesis occur via *de novo* binding at open chromatin regions. Interestingly, GATA1 targeted common and distinct sets of genes in E-Prog and E-Prec, displaying different binding preferences. Indeed, in E-Prog, GATA1 occupied both promoters and enhancers (or SEs), whereas in E-Prec GATA1 mainly bound promoters. This change in GATA1 binding preferences during E-Prog-to-E-Prec transition together with the diminished enhancer usage observed in E-Prec highlights a novel pivotal role of promoter regions as determinants of the transcriptional program activated during terminal erythroid differentiation. This novel finding is a unique feature of erythroid precursors, as previous studies indicate that enhancers play a main role in defining cell fate at the early stage of development (Huang et al., 2016; Romano et al., 2016; Xu et al., 2012).

To better investigate the interplay between regulatory elements and GATA factors in determining gene transcription during erythropoiesis, we used an integrative bioinformatics approach. Our results showed that investigating the chromatin landscape dynamic alone is not sufficient to infer the transcriptional modulations that occur during erythroid development, and that integrating the binding profiles of key master regulators, as GATA2 and GATA1, is essential to understand the epigenetic mechanisms governing the transcriptional changes. Moreover, our findings evidenced that GATA1 binding in different chromatin contexts consistently correlates with its dual activity as transcriptional activator or repressor. In fact, GATA1 binding at active regulatory regions (mainly promoters) is associated with a positive transcriptional regulation of the target gene (as for genes related to erythrocyte differentiation and homeostasis). Conversely, GATA1 binding at weak (in E-Prec) or bivalent (in both E-Prog and E-Prec) promoters correlates with the transcriptional silencing of the target genes (as for genes related to leukocyte differentiation or development of unrelated tissues).

Importantly, we performed validation studies of regulatory elements identified in the *KIT* locus. In mouse erythroid cells, several regulatory elements were identified upstream of the *KIT* gene or within its introns, and GATA2 is known to activate *KIT* expression in early progenitors, whereas GATA1 is responsible for *KIT* down-regulation upon differentiation (Cairns, 2003; Hong et al., 2005; Jing et al., 2008; Munugalavadla et al., 2005). However, little is known about the transcriptional regulation of *KIT* gene in human cells (Aranda-Orgilles et al., 2016; Romano et al., 2016). Here, we identified an E-Prog-specific SE upstream of *KIT* gene targeted by GATA1. Our results demonstrate that GATA1 binding at the E-Prog-specific SE is required for high-level *KIT* expression in human erythroid progenitors. Interestingly, the enhancer homologous to the murine −114-kb regulatory region (which was associated with high KIT expression when occupied by GATA2 in mouse progenitors and to KIT down-regulation when occupied by GATA1 in mouse erythroid precursors; Jing et al., 2008) is not targeted by GATA2 but only by GATA1 in KIT^+ E-Prog. Altogether, these data indicate that, differently from mouse erythroid cells, GATA1 activates *KIT* expression in human erythroid progenitors, supporting its up-regulation during erythroid commitment.

Several genes are known to be differentially regulated in human versus murine erythropoiesis (An et al., 2014). Notably, some genes upregulated during human erythropoiesis and down-regulated during murine erythropoiesis (e.g., RAPGEF2, MAP2K3, and RNF187; An et al., 2014) were indeed associated with active regulatory regions targeted by GATA1 in E-Prog and E-Prec (Figures S6D and S6E).

Overall, our stage-specific analysis of the transcriptional and epigenetic profiles and GATA factors' occupancy at key stages of human erythropoiesis provides new insights into the complex transcriptional regulatory mechanisms that control human erythroid commitment and differentiation, dissecting the pivotal role of GATA1 at both early and late stages of erythropoiesis.

Limitations of the Study

Given the high number of cells required to detect GATA factor binding by ChIP-seq, we analyzed mixed populations of erythroid progenitors (E-Prog: BFU-E, and CFU-E) and precursors (E-Prec: mainly polychromatic precursors). Therefore, our analyses may have a lower definition compared with transcriptomic and epigenomic studies analyzing highly purified erythroid progenitors and precursors (An et al., 2014; Li et al., 2014; Ludwig et al., 2019; Schulz et al., 2019).

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

The accession number for all RNA-seq and ChIP-seq data reported in this paper is GEO:GSE124165.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101018.

ACKNOWLEDGMENTS

This work was supported by grants from FIRB-Futuro in Ricerca (2010-RBFR10OS4G to A.M. and C.P.), EPIGEN Epigenomics Flagship Project (to O.R., S.B., C.P., and A.M.), and Agence Nationale de la Recherche (ANR-10-IAHU-01 "Investissements d'avenir" program). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement no. 670126-DENOVOSTEM to O.R. and S.B.). The authors thank Ryo Kurita and Yukio Nakamura for providing the HUDEP-2 cell line and Olivier Alibeau and Christine Bole for the RNA sequencing.

AUTHOR CONTRIBUTIONS

O.R. designed the experimental strategy and performed experiments, data analysis and interpretation. L.P. performed data analysis and interpretation. T.F., V.M., M.P., and C.A. performed experiments and data analysis. M.A. contributed to the design of the experimental strategy. S.B. contributed to data analysis and interpretation. C.P. designed the experimental strategy and performed experiments, data analysis and interpretation. A.M. conceived the study, designed the experimental strategy, and interpreted data. O.R., S.B., and A.M. wrote the paper.

DECLARATION OF INTERESTS

The Authors declare no competing financial interests.

Received: December 8, 2019
Revised: February 14, 2020
Accepted: March 24, 2020
Published: April 24, 2020

REFERENCES

An, X., Schulz, V.P., Li, J., Wu, K., Liu, J., Xue, F., Hu, J., Mohandas, N., and Gallagher, P.G. (2014). Global transcriptome analyses of human and murine terminal erythroid differentiation. Blood *123*, 3466–3477.

Aranda-Orgilles, B., Saldaña-Meyer, R., Wang, E., Trompouki, E., Fassl, A., Lau, S., Mullenders, J., Rocha, P.P., Raviram, R., Guillamot, M., et al. (2016). MED12 regulates HSC-specific enhancers independently of mediator kinase activity to control hematopoiesis. Cell Stem Cell *19*, 784–799.

Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., et al. (2016). The allelic landscape of human blood cell trait variation and links to common complex disease. Cell *167*, 1415–1429.e19.

Athanasiou, M., Mavrothalassitis, G., Sun-Hoffman, L., and Blair, D. (2000). FLI-1 is a suppressor of erythroid differentiation in human hematopoietic cells. Leukemia *14*, 439–445.

Beck, D., Thoms, J.A.I., Perera, D., Schutte, J., Unnikrishnan, A., Knezevic, K., Kinston, S.J., Wilson, N.K., O'Brien, T.A., Gottgens, B., et al. (2013). Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. Blood *122*, e12–e22.

Bresnick, E.H., Lee, H.-Y., Fujiwara, T., Johnson, K.D., and Keles, S. (2010). GATA switches as developmental drivers. J. Biol. Chem. *285*, 31087–31093.

Cairns, L.A. (2003). Kit regulatory elements required for expression in developing hematopoietic and germ cell lineages. Blood *102*, 3954–3962.

Calero-Nieto, F.J., Wood, A.D., Wilson, N.K., Kinston, S., Landry, J.-R., and Göttgens, B. (2010). Transcriptional regulation of Elf-1: locus-wide analysis reveals four distinct promoters, a tissue-specific enhancer, control by PU.1 and the importance of Elf-1 downregulation for erythroid maturation. Nucleic Acids Res. *38*, 6363–6374.

Cheng, Y., Wu, W., Ashok Kumar, S., Yu, D., Deng, W., Tripic, T., King, D.C., Chen, K.-B., Zhang, Y., Drautz, D., et al. (2009). Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome Res. *19*, 2172–2184.

Dore, L.C., Chlon, T.M., Brown, C.D., White, K.P., and Crispino, J.D. (2012). Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. Blood *119*, 3724–3733.

Dzierzak, E., and Philipsen, S. (2013). Erythropoiesis: development and differentiation. Cold Spring Harb. Perspect. Med. *3*, a011601.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. Mol. Cell. Biol. *25*, 1215–1227.

Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.-A., Choi, K., Farnham, P.J., and Bresnick, E.H. (2009). Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. Mol. Cell *36*, 667–681.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hong, W., Nakazawa, M., Chen, Y.-Y., Kori, R., Vakoc, C.R., Rakowski, C., and Blobel, G.A. (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. EMBO J. *24*, 2367–2378.

Hu, G., Schones, D.E., Cui, K., Ybarra, R., Northrup, D., Tang, Q., Gattinoni, L., Restifo, N.P., Huang, S., and Zhao, K. (2011). Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. Genome Res. *21*, 1650–1658.

Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T.V., Zon, L.I., Yuan, G., et al. (2016). Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. Dev. Cell *36*, 9–23.

Huang, S., and Brandt, S.J. (2000). mSin3A regulates murine erythroleukemia cell differentiation through association with the TAL1 (or SCL) transcription factor. Mol. Cell. Biol. *20*, 2248–2259.

Ji, P. (2015). New insights into the mechanisms of mammalian erythroid chromatin condensation and enucleation. Int. Rev. Cell Mol. Biol. *316*, 159–182.

Ji, P., Yeh, V., Ramirez, T., Murata-Hori, M., and Lodish, H.F. (2010). Histone deacetylase 2 is required for chromatin condensation and subsequent enucleation of cultured mouse fetal erythroblasts. Haematologica *95*, 2013–2021.

Ji, P., Murata-Hori, M., and Lodish, H.F. (2011). Formation of mammalian erythrocytes: chromatin condensation and enucleation. Trends Cell Biol. *21*, 409–415.

Jing, H., Vakoc, C.R., Ying, L., Mandat, S., Wang, H., Zheng, X., and Blobel, G.A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. Mol. Cell *29*, 232–242.

Kinross, K.M., Clark, A.J., Iazzolino, R.M., and Humbert, P.O. (2006). E2f4 regulates fetal erythropoiesis through the promotion of cellular proliferation. Blood *108*, 886–895.

Knudsen, K.J., Rehn, M., Hasemann, M.S., Rapin, N., Bagger, F.O., Ohlsson, E., Willer, A., Frank, A.-K., Søndergaard, E., Jendholm, J., et al. (2015). ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. Genes Dev. *29*, 1915–1929.

Kurita, R., Suda, N., Sudo, K., Miharada, K., Hiroyama, T., Miyoshi, H., Tani, K., and Nakamura, Y. (2013). Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. PLoS One *8*, e59890.

Li, J., Hale, J., Bhagia, P., Xue, F., Chen, L., Jaffray, J., Yan, H., Lane, J., Gallagher, P.G., Mohandas, N., et al. (2014). Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. Blood *124*, 3636–3645.

Ludwig, L.S., Lareau, C.A., Bao, E.L., Nandakumar, S.K., Muus, C., Ulirsch, J.C., Chowdhary, K., Buenrostro, J.D., Mohandas, N., An, X., et al. (2019). Transcriptional states and chromatin accessibility underlying human erythropoiesis. Cell Rep. *27*, 3228–3240.e7.

Masuda, T., Wang, X., Maeda, M., Canver, M.C., Sher, F., Funnell, A.P.W., Fisher, C., Suciu, M., Martyn, G.E., Norton, L.J., et al. (2016). Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. Science *351*, 285–289.

May, G., Soneji, S., Tipping, A.J., Teles, J., McGowan, S.J., Wu, M., Guo, Y., Fugazza, C., Brown, J., Karlsson, G., et al. (2013). Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. Cell Stem Cell *13*, 754–768.

Moras, M., Lefevre, S.D., and Ostuni, M.A. (2017). From erythroblasts to mature red blood cells: organelle clearance in mammals. Front. Physiol. *8*, 1–9.

Moriguchi, T., and Yamamoto, M. (2014). A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation. Int. J. Hematol. *100*, 417–424.

Munugalavadla, V., Dore, L.C., Tan, B.L., Hong, L., Vishnu, M., Weiss, M.J., and Kapur, R. (2005). Repression of c-kit and its downstream substrates by GATA-1 inhibits cell proliferation during erythroid maturation. Mol. Cell. Biol. *25*, 6747–6759.

Munugalavadla, V., and Kapur, R. (2005). Role of c-Kit and erythropoietin receptor in erythropoiesis. Crit. Rev. Oncol. Hematol. *54*, 63–75.

Nishiyama, C., Yokota, T., Okumura, K., and Ra, C. (1999). The transcription factors Elf-1 and GATA-1 bind to cell-specific enhancer elements of human high-affinity IgE receptor alpha-chain gene. J. Immunol. *163*, 623–630.

Pinello, L., Xu, J., Orkin, S.H., and Yuan, G.-C. (2014). Analysis of chromatin-state plasticity identifies cell-type-specific regulators of H3K27me3 patterns. Proc. Natl. Acad. Sci. U S A *111*, E344–E353.

Popova, E.Y., Krauss, S.W., Short, S.A., Lee, G., Villalobos, J., Etzell, J., Koury, M.J., Ney, P.A., Chasis, J.A., and Grigoryev, S.A. (2009). Chromatin condensation in terminally differentiating mouse erythroblasts does not involve special architectural proteins but depends on histone deacetylation. Chromosome Res. *17*, 47–64.

Rekhtman, N., Radparvar, F., Evans, T., and Skoultchi, A.I. (1999). Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. Genes Dev. *13*, 1398–1411.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Romano, O., Peano, C., Tagliazucchi, G.M., Petiti, L., Poletti, V., Cocchiarella, F., Rizzi, E., Severgnini, M., Cavazza, A., Rossi, C., et al. (2016). Transcriptional, epigenetic and retroviral signatures identify regulatory regions involved in hematopoietic lineage commitment. Sci. Rep. *6*, 24724.

Schuh, A.H., Tipping, A.J., Clark, A.J., Hamlett, I., Guyot, B., Iborra, F.J., Rodriguez, P., Strouboulis, J., Enver, T., Vyas, P., and Porcher, C. (2005). ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis. Mol. Cell. Biol. *25*, 10235–10250.

Schulz, V.P., Yan, H., Lezon-Geyda, K., An, X., Hale, J., Hillyer, C.D., Mohandas, N., and Gallagher, P.G. (2019). A unique epigenomic landscape defines human erythropoiesis. Cell Rep. *28*, 2996–3009.e7.

Shi, L., Lin, Y.-H., Sierant, M.C., Zhu, F., Cui, S., Guan, Y., Sartor, M.A., Tanabe, O., Lim, K.-C., and Engel, J.D. (2014). Developmental transcriptome analysis of human erythropoiesis. Hum. Mol. Genet. *23*, 4528–4542.

Siatecka, M., and Bieker, J.J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. Blood *118*, 2044–2054.

Su, M.Y., Steiner, L.A., Bogardus, H., Mishra, T., Schulz, V.P., Hardison, R.C., and Gallagher, P.G. (2013). Identification of biologically relevant enhancers in human erythroid cells. J. Biol. Chem. *288*, 8433–8444.

Suzuki, M., Kobayashi-Osaki, M., Tsutsumi, S., Pan, X., Ohmori, S., Takai, J., Moriguchi, T., Ohneda, O., Ohneda, K., Shimizu, R., et al. (2013). GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. Genes Cells *18*, 921–933.

Van Handel, B., Montel-Hagen, A., Sasidharan, R., Nakano, H., Ferrari, R., Boogerd, C.J., Schredelseker, J., Wang, Y., Hunter, S., Org, T., et al. (2012). Scl represses cardiomyogenesis in prospective hemogenic endothelium and endocardium. Cell *150*, 590–605.

Vicente, C., Conchillo, A., García-Sánchez, M.A., and Odero, M.D. (2012). The role of the GATA2 transcription factor in normal and malignant hematopoiesis. Crit. Rev. Oncol. Hematol. *82*, 1–17.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wong, P., Hattangadi, S.M., Cheng, A.W., Frampton, G.M., Young, R.A., and Lodish, H.F. (2011). Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. Blood *118*, e128–e138.

Wontakal, S.N., Guo, X., Smith, C., MacCarthy, T., Bresnick, E.H., Bergman, a., Snyder, M.P., Weissman, S.M., Zheng, D., and Skoultchi, a.I. (2012). A core erythroid transcriptional network is repressed by a master regulator of myelo-lymphoid differentiation. Proc. Natl. Acad. Sci. *109*, 3832–3837.

Wu, W., Cheng, Y., Keller, C.A., Ernst, J., Kumar, S.A., Mishra, T., Morrissey, C., Dorman, C.M., Chen, K.-B., Drautz, D., et al. (2011). Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. Genome Res. *21*, 1659–1671.

Xu, J., Shao, Z., Glass, K., Bauer, D.E., Pinello, L., Van Handel, B., Hou, S., Stamatoyannopoulos, J.A., Mikkola, H.K.A., Yuan, G.-C., and Orkin, S.H. (2012). Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. Dev. Cell *23*, 796–811.

Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T.B., Cheng, Y., Yu, D., Hardison, R., Weiss, M.J., Orkin, S.H., et al. (2009). Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Mol. Cell *36*, 682–695.

iScience, Volume *23*

Supplemental Information

GATA Factor-Mediated Gene Regulation in Human Erythropoiesis

Oriana Romano, Luca Petiti, Tristan Felix, Vasco Meneghini, Michel Portafax, Chiara Antoniani, Mario Amendola, Silvio Bicciato, Clelia Peano, and Annarita Miccio

Figure S1

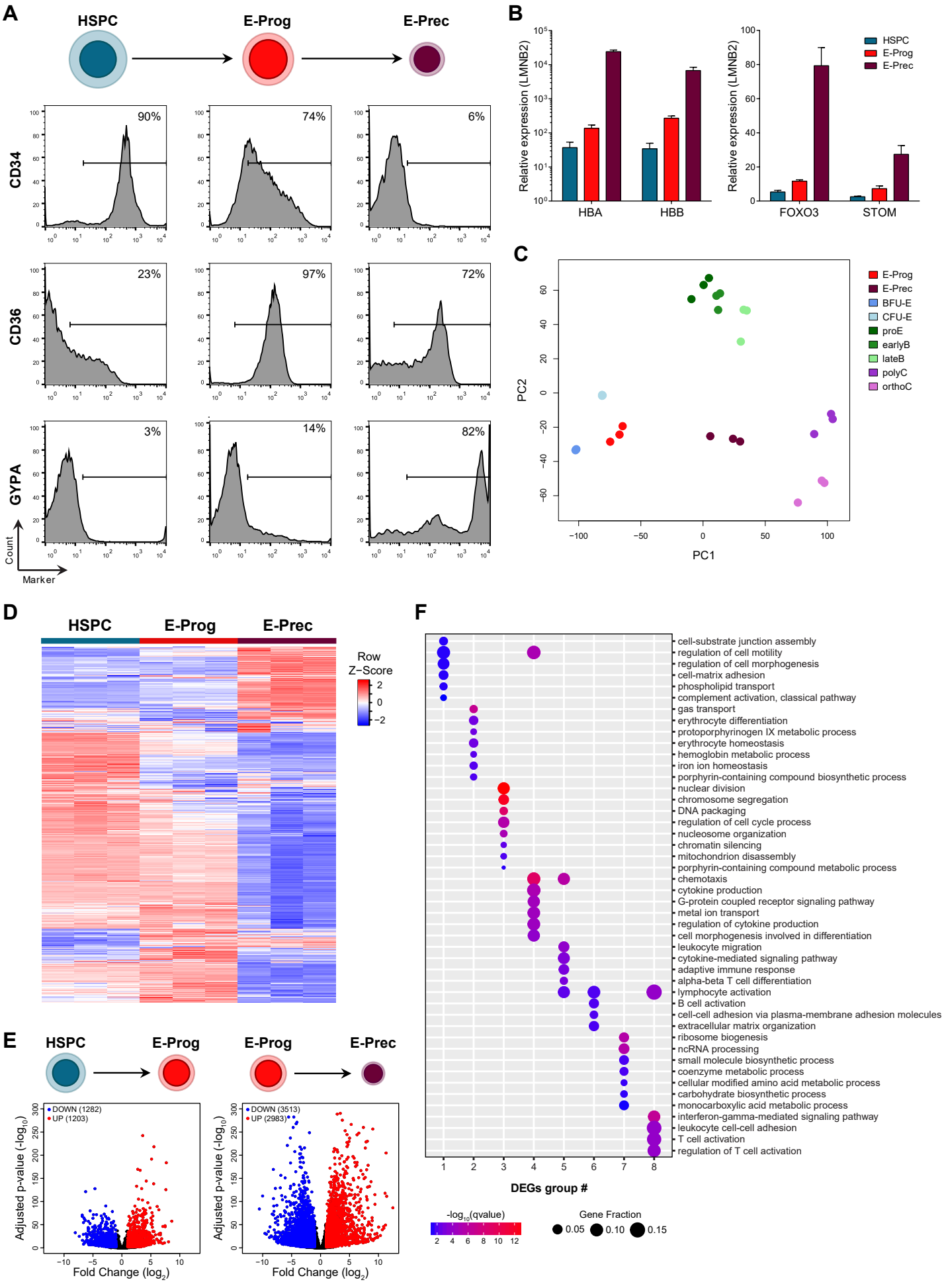

A
HSPC
E-Prog
E-Prec
CD34
CD36
GYPA
90%
74%
6%
23%
97%
72%
3%
14%
82%
Count
Marker
B
Relative expression (LMNB2)
HBA
HBB
FOXO3
STOM
HSPC
E-Prog
E-Prec
C
PC1
PC2
E-Prog
E-Prec
BFU-E
CFU-E
proE
earlyB
lateB
polyC
orthoC
D
HSPC
E-Prog
E-Prec
Row Z-Score
E
HSPC
E-Prog
E-Prog
E-Prec
DOWN (1282)
UP (1203)
DOWN (3513)
UP (2983)
Adjusted p-value (-log10)
Fold Change (log2)
F
cell-substrate junction assembly
regulation of cell motility
regulation of cell morphogenesis
cell-matrix adhesion
phospholipid transport
complement activation, classical pathway
gas transport
erythrocyte differentiation
protoporphyrinogen IX metabolic process
erythrocyte homeostasis
hemoglobin metabolic process
iron ion homeostasis
porphyrin-containing compound biosynthetic process
nuclear division
chromosome segregation
DNA packaging
regulation of cell cycle process
nucleosome organization
chromatin silencing
mitochondrion disassembly
porphyrin-containing compound metabolic process
chemotaxis
cytokine production
G-protein coupled receptor signaling pathway
metal ion transport
regulation of cytokine production
cell morphogenesis involved in differentiation
leukocyte migration
cytokine-mediated signaling pathway
adaptive immune response
alpha-beta T cell differentiation
lymphocyte activation
B cell activation
cell-cell adhesion via plasma-membrane adhesion molecules
extracellular matrix organization
ribosome biogenesis
ncRNA processing
small molecule biosynthetic process
coenzyme metabolic process
cellular modified amino acid metabolic process
carbohydrate biosynthetic process
monocarboxylic acid metabolic process
interferon-gamma-mediated signaling pathway
leukocyte cell-cell adhesion
T cell activation
regulation of T cell activation
DEGs group #
-log10(qvalue)
Gene Fraction

Figure S2

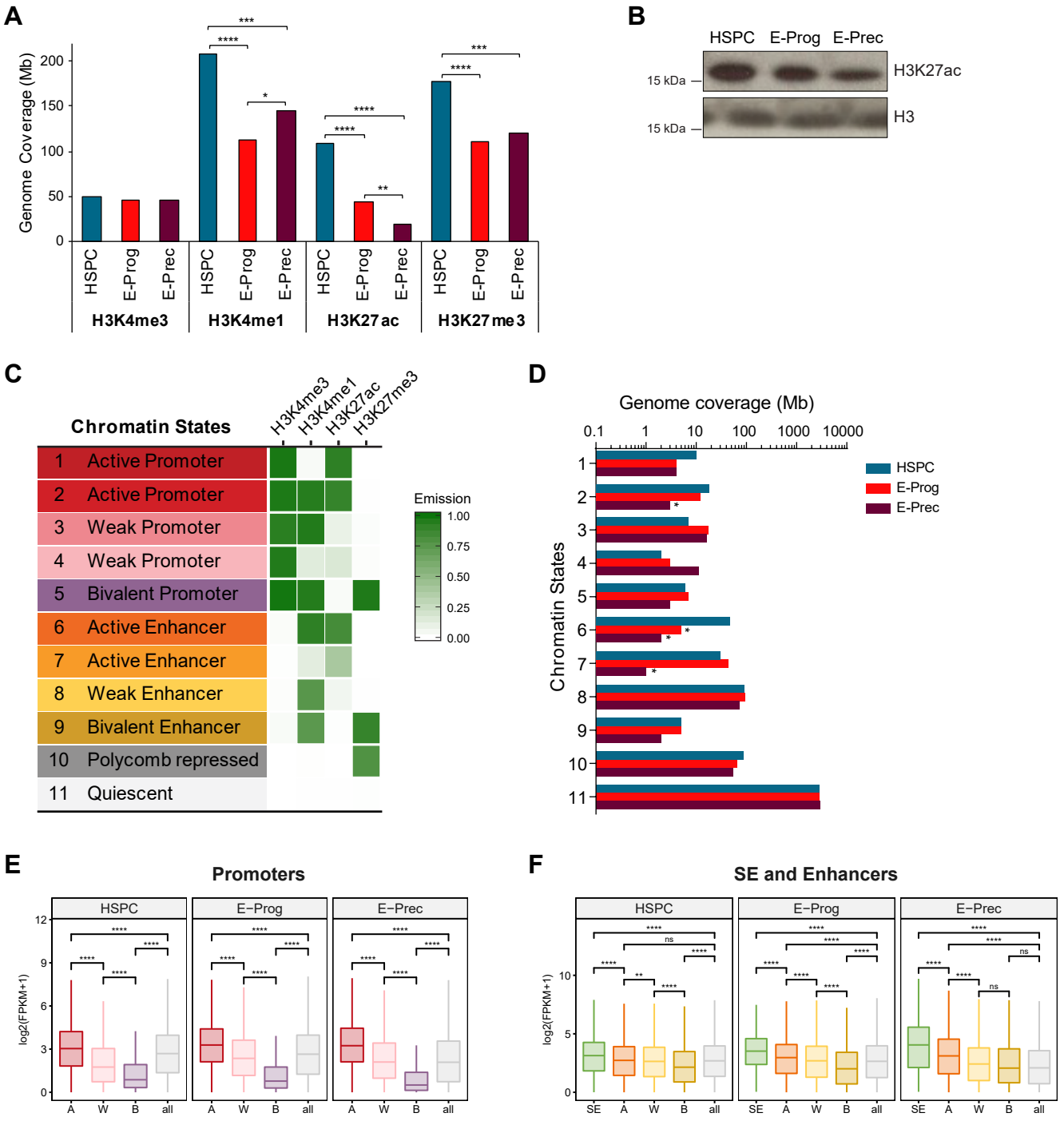

A
Genome Coverage (Mb)
HSPC
E-Prog
E-Prec
H3K4me3
H3K4me1
H3K27ac
H3K27me3
B
HSPC E-Prog E-Prec
15 kDa
H3K27ac
H3
C
Chromatin States
H3K4me3
H3K4me1
H3K27ac
H3K27me3
1 Active Promoter
2 Active Promoter
3 Weak Promoter
4 Weak Promoter
5 Bivalent Promoter
6 Active Enhancer
7 Active Enhancer
8 Weak Enhancer
9 Bivalent Enhancer
10 Polycomb repressed
11 Quiescent
Emission
D
Genome coverage (Mb)
Chromatin States
HSPC
E-Prog
E-Prec
E
Promoters
log2(FPKM+1)
A W B all
F
SE and Enhancers
SE A W B all

Figure S3

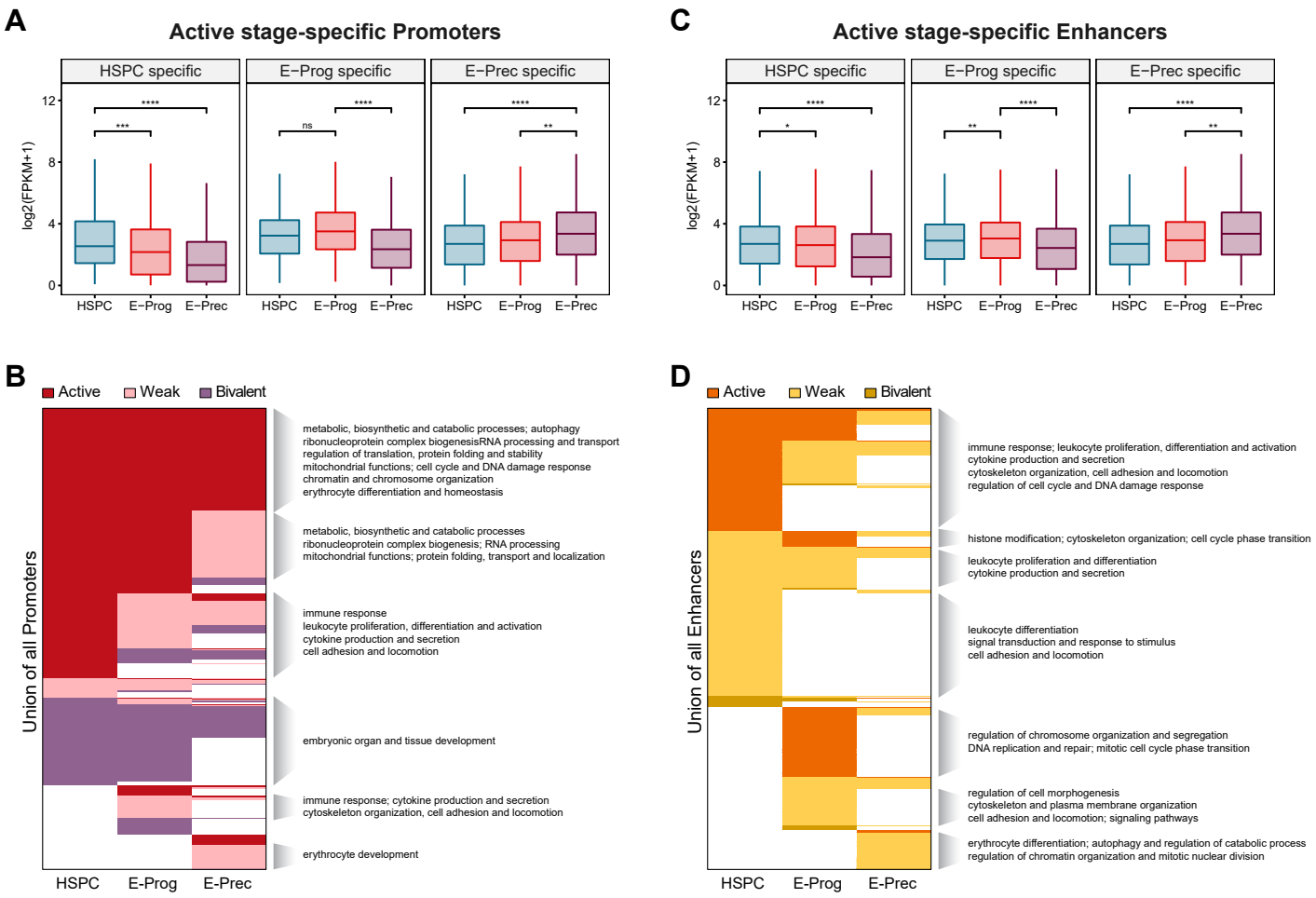

A
Active stage-specific Promoters
HSPC specific
E−Prog specific
E−Prec specific
log2(FPKM+1)
HSPC
E−Prog
E−Prec
C
Active stage-specific Enhancers
B
Active
Weak
Bivalent
Union of all Promoters
metabolic, biosynthetic and catabolic processes; autophagy
ribonucleoprotein complex biogenesisRNA processing and transport
regulation of translation, protein folding and stability
mitochondrial functions; cell cycle and DNA damage response
chromatin and chromosome organization
erythrocyte differentiation and homeostasis
metabolic, biosynthetic and catabolic processes
ribonucleoprotein complex biogenesis; RNA processing
mitochondrial functions; protein folding, transport and localization
immune response
leukocyte proliferation, differentiation and activation
cytokine production and secretion
cell adhesion and locomotion
embryonic organ and tissue development
immune response; cytokine production and secretion
cytoskeleton organization, cell adhesion and locomotion
erythrocyte development
D
Union of all Enhancers
immune response; leukocyte proliferation, differentiation and activation
cytokine production and secretion
cytoskeleton organization, cell adhesion and locomotion
regulation of cell cycle and DNA damage response
histone modification; cytoskeleton organization; cell cycle phase transition
leukocyte proliferation and differentiation
cytokine production and secretion
leukocyte differentiation
signal transduction and response to stimulus
cell adhesion and locomotion
regulation of chromosome organization and segregation
DNA replication and repair; mitotic cell cycle phase transition
regulation of cell morphogenesis
cytoskeleton and plasma membrane organization
cell adhesion and locomotion; signaling pathways
erythrocyte differentiation; autophagy and regulation of catabolic process
regulation of chromatin organization and mitotic nuclear division

Figure S4

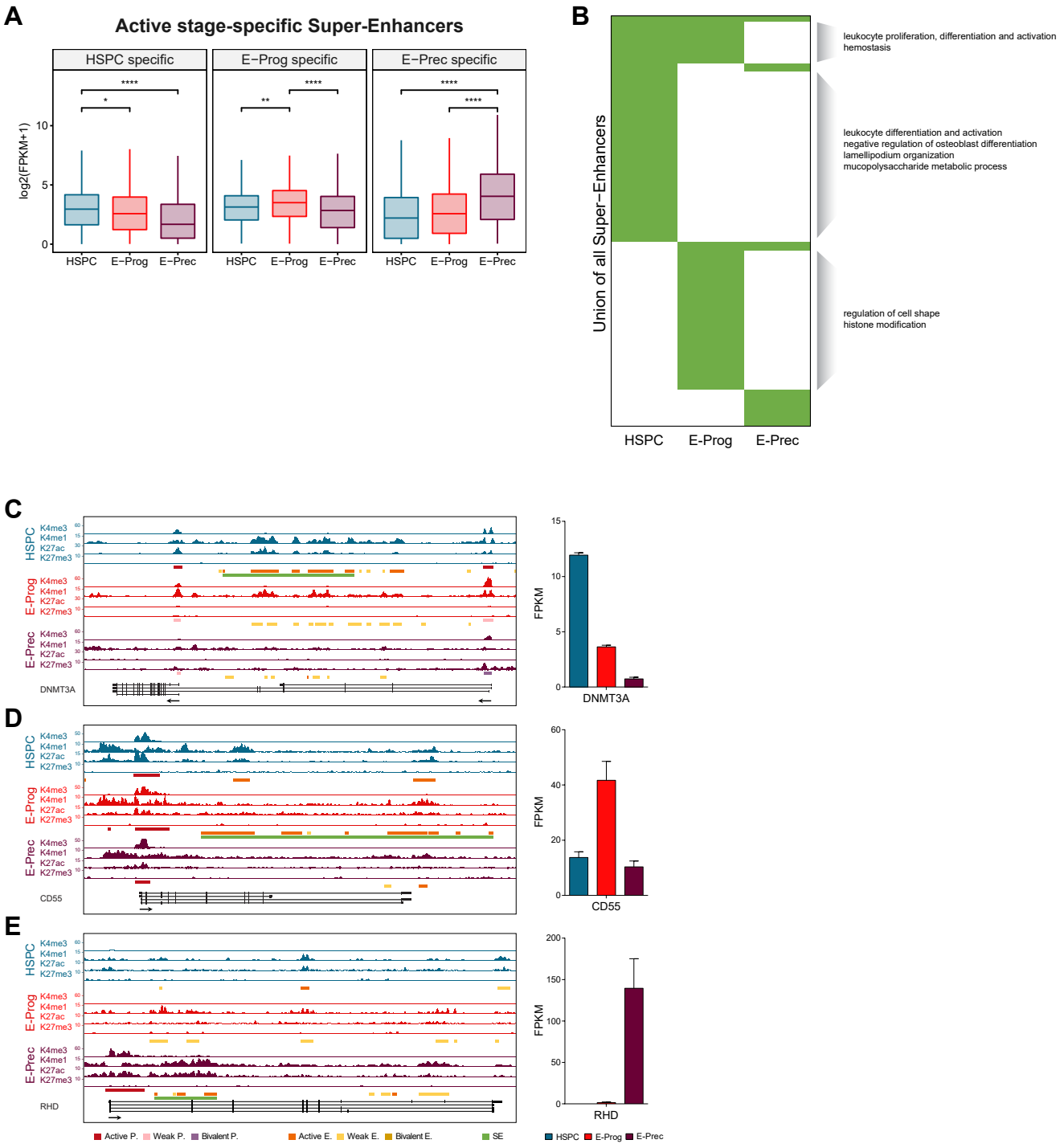
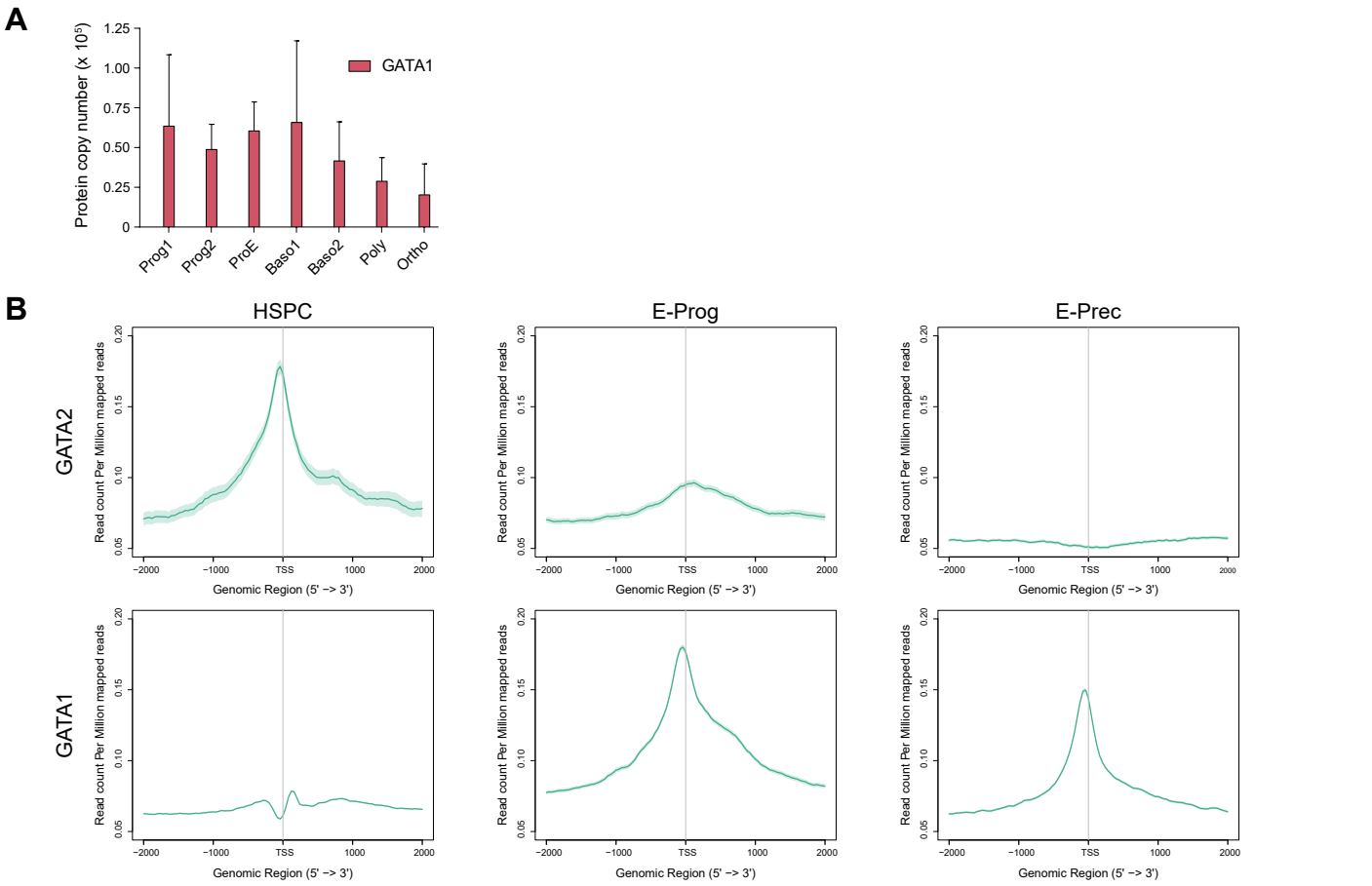

A
Active stage-specific Super-Enhancers
HSPC specific
E−Prog specific
E−Prec specific
log2(FPKM+1)
HSPC
E−Prog
E−Prec
B
Union of all Super−Enhancers
leukocyte proliferation, differentiation and activation
hemostasis
leukocyte differentiation and activation
negative regulation of osteoblast differentiation
lamellipodium organization
mucopolysaccharide metabolic process
regulation of cell shape
histone modification
HSPC
E-Prog
E-Prec
C
D
E
K4me3
K4me1
K27ac
K27me3
DNMT3A
CD55
RHD
FPKM
Active P.
Weak P.
Bivalent P.
Active E.
Weak E.
Bivalent E.
SE

Figure S5



D

	Motif Name	Consensus	GATA2 only in HSPC	GATA switch	GATA1 only in E-Prog	GATA1 common	GATA1 only in E-Prec
GATA	GATA(Zf)	SAGATAAGRV	2.01	3.21	2.89	3.31	2.91
	GATA(Zf):TAL1(E-box)	CRGCTGBNGNSNNSAGATAA	2.56	4.15	2.80	3.84	3.87
ETS	ERG(ETS)	ACAGGAAGTG	1.68	1.26	0.19	-0.01	0.19
	FLI1(ETS)	NRYTTCCGGH	2.01	1.59	0.25	0.12	0.51
	GABPA(ETS)	RACCGGAAGT	1.90	1.55	0.29	0.11	0.57
	PU.1(ETS)	AGAGGAAGTG	2.19	1.87	0.41	0.02	0.12
	ELF1(ETS)	AVCCGGAAGT	2.06	1.54	0.37	0.30	0.83
	ELK1(ETS)	HACTTCCGGY	1.97	1.58	0.34	0.29	0.88
	ELK4(ETS)	NRYTTCCGGY	2.06	1.55	0.33	0.35	0.92
	ETS(ETS) - Promoter	AACCGGAAGT	2.22	1.73	0.50	0.50	1.00
	ETS:E-box	AGGAARCAGCTG	3.47	2.06	0.06	-0.20	0.04
Others	RUNX1(Runt)	AAACCACARM	1.47	1.38	0.88	0.37	-0.04
	RUNX(Runt)	SAAACCACAG	1.71	1.68	1.14	0.28	-0.16
	Sp1(Zf) - Promoter	GGCCCCGCCCCC	0.27	0.65	0.14	1.04	1.31
	NFY(CCAAT) - Promoter	RGCCAATSRG	-0.06	0.12	0.22	0.43	1.10
	KLF1(Zf)	NWGGGTGTGGCY	-0.68	0.13	0.69	1.08	1.70

Log2FC
Enriched
Depleted

Figure S6

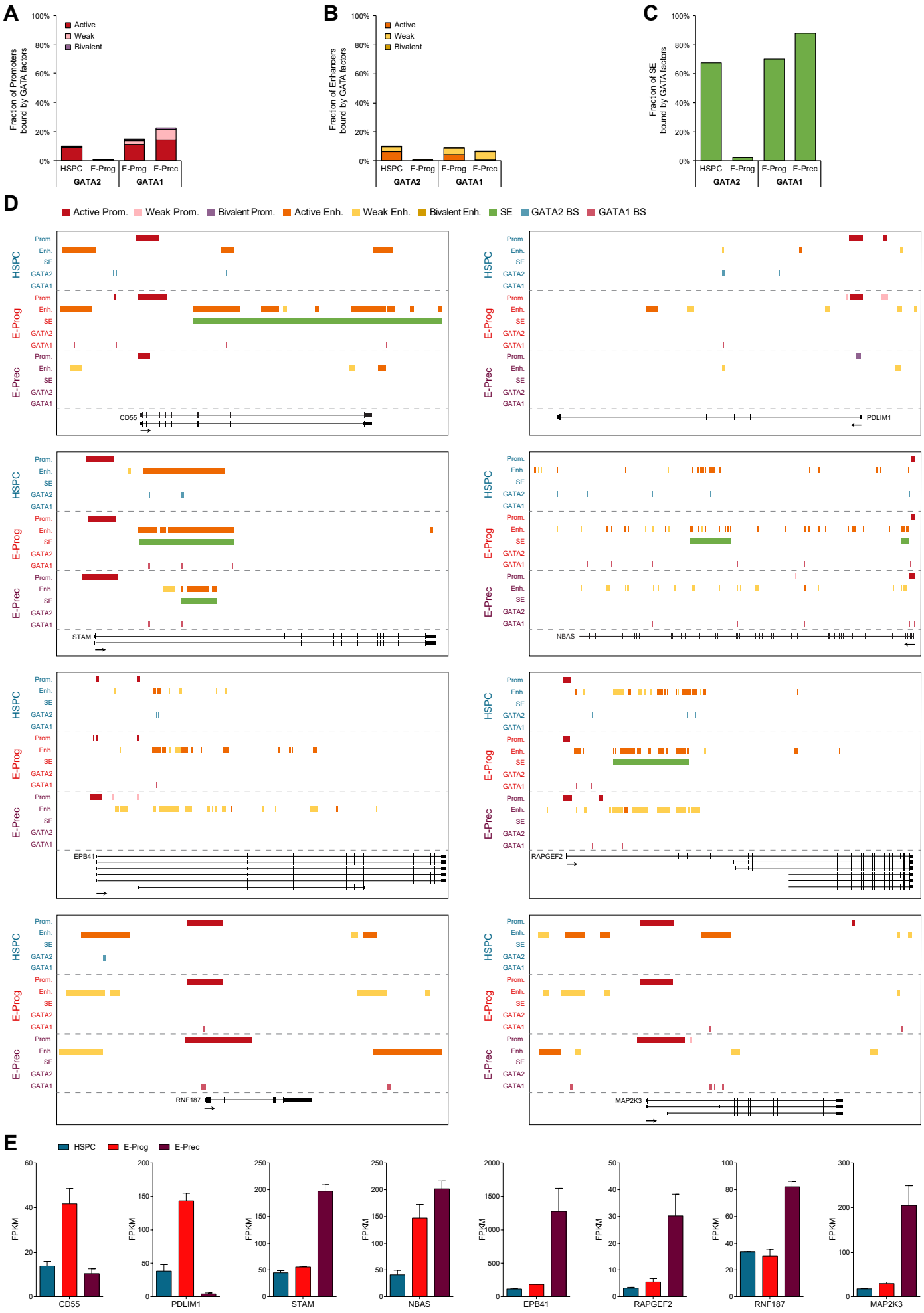

A
Fraction of Promoters bound by GATA factors
Active
Weak
Bivalent
HSPC
E-Prog
E-Prec
GATA2
GATA1
B
Fraction of Enhancers bound by GATA factors
C
Fraction of SE bound by GATA factors
D
Active Prom.
Weak Prom.
Bivalent Prom.
Active Enh.
Weak Enh.
Bivalent Enh.
SE
GATA2 BS
GATA1 BS
CD55
PDLIM1
STAM
NBAS
EPB41
RAPGEF2
RNF187
MAP2K3
E
FPKM

Figure S7

	Motif Name	Consensus	E-Prog vs HSPC		E-Prec vs E-Prog	
			UP-regulated gene	DOWN-regulated genes	UP-regulated gene	DOWN-regulated genes
GATA	GATA(Zf)	SAGATAAGRV	3.17	3.27	3.20	3.21
	GATA(Zf):TAL1(E-box)	CRGCTGBNGNSNNSAGATAA	3.78	3.98	4.37	3.98
ETS	ERG(ETS)	ACAGGAAGTG	0.45	0.88	0.10	0.66
	Fli1(ETS)	NRYTTCCGGH	0.64	1.11	0.35	0.82
	GABPA(ETS)	RACCGGAAGT	0.53	0.84	0.31	0.86
	PU.1(ETS)	AGAGGAAGTG	0.82	1.45	0.02	0.83
	ELF1(ETS)	AVCCGGAAGT	0.32	0.85	0.53	0.98
	ETS(ETS) - Promoter	AACCGGAAGT	0.46	1.12	0.42	1.45
Others	KLF1(Zf)	NWGGGTGTGGCY	1.65	0.69	1.54	1.20
	E2F4(E2F)	GGCGGGAAAH	0.72	0.15	0.68	0.40

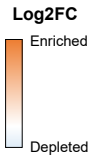


Figure S8

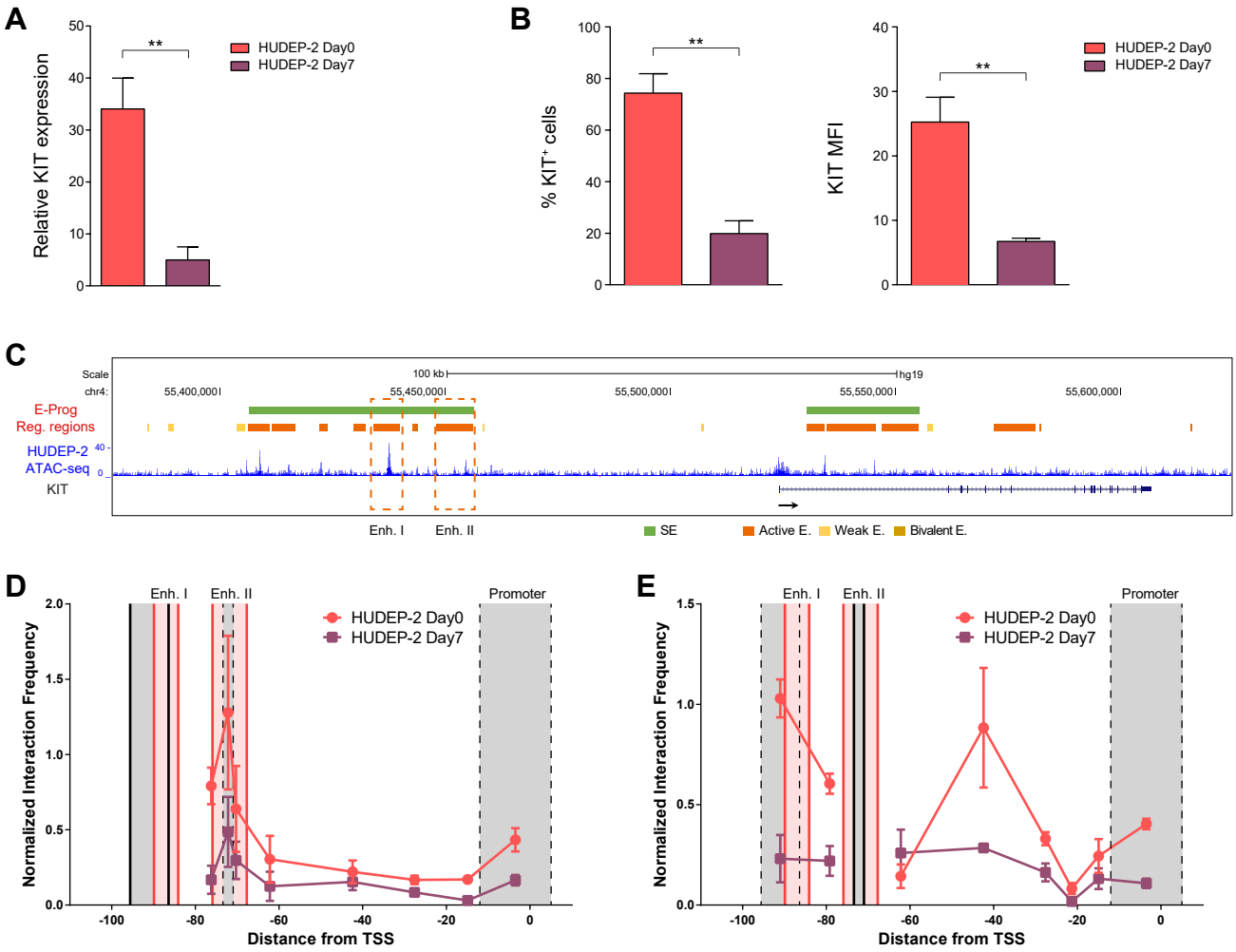

A
Relative KIT expression
HUDEP-2 Day0
HUDEP-2 Day7
**
B
% KIT+ cells
KIT MFI
C
Scale
100 kb
hg19
chr4:
55,400,000
55,450,000
55,500,000
55,550,000
55,600,000
E-Prog
Reg. regions
HUDEP-2
ATAC-seq
KIT
Enh. I
Enh. II
SE
Active E.
Weak E.
Bivalent E.
D
E
Promoter
Normalized Interaction Frequency
Distance from TSS

Figure S9

A

Sequence	InDel Type	InDel Lenght (nt)	InDel Frequency	GATA motif disruption
CCTCCGCAGGCCTCTGCTCCCAGCA|GATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	WT	0	52.2%	no
CCTCCGCAGGCCTCTGCTCCCAGCA|--TAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-2	8.3%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|-------GACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-7	5.7%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|-ATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-1	4.3%	yes
CCTCCGCAGGCCTCTGCTCCCAGC-|GATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-1	3.5%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|---AATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-3	3.2%	yes
CCTCCGCAGGCCTCTGCTCCCAG--|GATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-2	2.9%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|nGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACA	INS	1	2.8%	yes
CCTCCGCAGGCCTCTGCTCCC----|-ATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-5	2.4%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAG	INS	4	1.6%	possible
CCTCCGCAGGCCTCTGCTCCCAGCA|nnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGAC	INS	2	1.5%	possible
CCTCCGCAGGCCTCTGCTCCCAGCA|------AGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-6	1.3%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|----ATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-4	1.0%	yes
CCTCCGCAGGCCTCTGCTCCCA---|GATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-3	1.0%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGA	INS	3	0.9%	possible
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnnnnnnnnnnnnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGT	INS	15	0.8%	possible
CCTCCGCAGGCCTCTGCTCCC----|------AGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-10	0.8%	yes
CCTCCGCAGGCCTCTGCTCCC----|----------GTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-16	0.5%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnnnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAA	INS	6	0.5%	possible
CCTCCGCAGGCCTCTGCTCCCAGCA|----------GGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-11	0.4%	yes
CCTCCGCAGGCCTCTGCTCCC----|GATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-4	0.4%	yes
CCTCCGCAGGCCTCTGCTCCCA---|----------TGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-13	0.4%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnnnnnnnnnnnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTT	INS	14	0.3%	possible
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnnnnnnnnnnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTG	INS	13	0.2%	possible
CCTCCGCAGGCCTCTGCTCCCAG--|-------GACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-9	0.2%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAA	INS	5	0.2%	possible
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnnnnnnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTA	INS	9	0.2%	possible
CCTCCGCAGGCCTCTGCTCCCAGCA|nnnnnnnnnnnnGATAATAGACTGGTCATGAGGTATTTTTAAGCAGTTGC	INS	12	0.1%	possible
CCTCCGCAGGCCTCTGCTCCCA---|------AGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-9	0.1%	yes
CCTCCGCAGGCCTCTGCTCCCAGCA|--------------CATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-14	0.1%	yes
CCTCCGCAGGCCTCTGCTCCC----|----ATAGACTGGTCATGAGGTATTTTTAAGCAGTTGCTTAAAAAGACAG	DEL	-8	0.1%	yes

TAL1 motif
GATA1 motif

B

Sequence	InDel Type	InDel Lenght (nt)	InDel Frequency	GATA motif disruption
CTAGGACTCCTGTGACTCATTACTG|TTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	WT	0	27.4%	no
CTAGGACTCCTGTGACTCATTACTG|----------AAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-10	6.8%	yes
CTAGGACTCCTGTGACTCATTACTG|--------ATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-8	4.1%	yes
CTAGGACTCCTGTGACTCATTACTG|nTTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAA	INS	1	3.6%	no
CTAGGACTCCTGTGACTCATTAC--|-TGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-3	3.4%	no
CTAGGACTCCTGTGACTCATTACTG|------AGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-6	3.4%	no
CTAGGACTCCTGTGACTCATTACTG|-TGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-1	3.1%	no
CTAGGACTCCTGTGACTCATTACTG|nnTTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGA	INS	2	3.0%	no
CTAGGACTCCTGTGACTCATTACTG|---------TAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-9	2.7%	yes
CTAGGACTCCTGTGACTCATTACTG|--GAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-2	2.7%	no
CTAGGACTCCTGTGACTCATTACTG|-----GAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-5	2.7%	no
CTAGGACTCCTGTGACTCATTACTG|----------GAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-12	2.6%	yes
CTAGGACTCCTGTGACTCATTACTG|----------AGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-11	2.5%	yes
CTAGGACTCC----------|----------TGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-30	2.3%	yes
CTAGGACTCCTGTGACTCATTACTG|----------TGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-15	1.9%	yes
CTAGGACTCCTGTGACTC--------|----------AATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-20	1.7%	yes
CTAGGACTCCTGTGACTC--------|----------ATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-21	1.6%	yes
CTAGGACTCCTGTGACTC--------|----------TGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-22	1.6%	yes
CTAGGACTCCTGTGACTCATTACTG|----GGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-4	1.6%	no
CTAGGACTCCTGTGACTCATTACTG|-------GATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-7	1.4%	yes
CTAGGACTCCTGTGACTCATTACTG|----------ATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-14	1.3%	yes
CTAGGACTCCTGTGACTCATTAC--|----------TGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-17	1.0%	yes
CTAGGACTCC----------|----------ATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-29	1.0%	yes
CTAGGACTCCTGTGACTCATTACTG|nnnTTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGG	INS	3	1.0%	no
CTAGGACTCCTGTGACTCATTACTG|nnnnnTTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTT	INS	5	0.9%	no
CTAGGACTCCTGTGAC---------|----------TGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-24	0.6%	yes
CTAGGACTCCTGTGACTCATTACTG|nnnnnnnnTTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGA	INS	8	0.6%	no
CTAGGACTCCTGT----------|----------AATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-25	0.6%	yes
CTAGGACTCC----------|----------AGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-26	0.5%	yes
CTAGGACTCCTGTGACTCATTACTG|nnnnTTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAGTTG	INS	4	0.4%	no
CTAGGACTCCTGTGACTCATTACTG|----------AATGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-13	0.3%	yes
CTAGGACTCCTGTGACTCATTACT-|----------TGTGGGCTTTGGAGTCATCAAGACAGAGTTGGAAA	DEL	-16	0.3%	yes
CTAGGACTCCTGTGACTCATTACTG|nnnnnnnnnnnnnnnnTTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCA	INS	15	0.1%	no
CTAGGACTCCTGTGACTCATTACTG|nnnnnnnTTGAGGAGATAAGAATGTGGGCTTTGGAGTCATCAAGACAGAG	INS	7	0.1%	no

AP1/HIF1b motif
GATA1 motif
PITX1 motif

Supplemental figure legends

Figure S1. Characterization of erythroid progenitors and precursors. Related to Figure 1.

(A) Representative FACS histograms showing CD34, CD36 and glycophorin A (GYPA) levels in HSPC, E-Prog and E-Prec.

(B) RT-qPCR analysis of *HBA, HBB*, *FOXO3* and *STOM* mRNA expression levels in HSPC, E-Prog and E-Prec. During erythroid commitment and differentiation, *HBA* and *HBB* expression levels progressively increased, while the transcription factor *FOXO3* (essential for terminal maturation)(Liang et al., 2015) and the erythrocyte membrane protein *STOM* were expressed at high levels only in E-Prec. Transcript levels were normalized to *LMNB2* mRNA. Data were plotted as mean with SEM.

(C) Comparison of gene expression profiles of E-Prog, E-Prec and erythroid progenitors and precursors from publicly available datasets (An et al., 2014; Li et al., 2014). PCA analysis showed that E-Prog cluster with human erythroid progenitors (BFU-E and CFU-E), while E-Prec exhibit gene expression profiles similar to human polychromatic erythroid precursors. The 2D plot show PCA coordinates for each biological replicate. Principal components (PC1 – PC2) are shown on the x and y axes. Abbreviations: erythroid burst-forming units, BFU-E; erythroid colony-forming units, CFU-E; proerythroblasts, proE; early basophilic erythroblasts, earlyB; late basophilic erythroblasts, lateB; polychromatic erythroblasts, polyC; orthochromatic erythroblasts, orthoC.

(D) Unsupervised clustering of expression levels in HSPC, E-Prog and E-Prec of genes with CPM ≥ 1 in at least 3 samples (expressed genes). Blue and red indicate higher and lower than average expression levels, respectively. As reported by An et colleagues (An et al., 2014), a general reduction of gene expression was observed during differentiation from E-Prog to E-Prec.

(E) Volcano plots showing global transcriptional changes upon erythroid commitment (left panel) and differentiation (right panel). Down-regulated genes are indicated in blue, up-regulated genes in red, genes not differentially expressed in black. Differentially expressed genes (DEGs) were identified with absolute log2 fold change > 1 and adjusted p-value < 0.01.

(F) Gene Ontology enrichment analysis. Enriched Biological Process (BP) terms are shown on the y-axis; DEGs groups, as defined in Figure 1B, are shown on the x-axis. Dots are color-coded based on the enrichment q-values; dot size indicates the fraction of DEGs in each BP term.

Figure S2. Analysis of histone marks, chromatin states and regulatory elements during erythroid development. Related to Figure 1.

(A) Genome coverage of histone marks in HSPC, E-Prog and E-Prec (Chi-square test).

(B) Western blot analysis of H3K27ac levels in the nuclear fraction of HSPC, E-Prog and E-Prec. Total H3 was used for normalization.

(C) Definition of chromatin states using Spectacle. The heatmap shows emission probabilities of each histone mark for the 11-state model.

(D) Genome coverage of chromatin states in HSPC, E-Prog and E-Prec. Active promoters (state 5 and 6) and enhancers (state 1 and 2) displayed reduced genome coverage in E-Prog and E-Prec compared to HSPC. *: p-value < 0.001 (Chi-square test).

(E) Expression levels of genes associated to promoter regions. Promoters were classified as active ($H3K27ac^{+}H3K27me3^{-}$, A), weak ($H3K27ac^{-}H3K27me3^{-}$, W) or bivalent ($H3K27ac^{-}H3K27me3^{+}$, B). For each promoter state, the boxplot shows the distribution of the expression levels of target genes. As control, the distribution of the expression levels of all expressed genes (all) was reported (Wilcoxon test).

(F) Expression levels of genes targeted by enhancers and super-enhancers (SE). Enhancers were classified as active ($H3K27ac^{+}H3K27me3^{-}$, A), weak ($H3K27ac^{-}H3K27me3^{-}$, W) or bivalent ($H3K27ac^{-}H3K27me3^{+}$, B). For super-enhancers and each enhancer state, the boxplot shows the distribution of the expression levels of target genes. As control, the distribution of expression levels of all expressed genes (all) was reported (Wilcoxon test).

(A, E, F) *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001; ****: p-value < 0.0001

Figure S3. Expression levels and functional enrichment analysis of genes associated with stage-specific or shared promoters and enhancers. Related to Figure 1.

(A) Expression levels of genes associated with active stage-specific promoters in HSPC, E-Prog and E-Prec. The expression levels of genes associated with stage-specific promoters were compared to the expression levels of the same genes in the other cell types. These genes showed higher expression levels in the specific stage compared to the others (Wilcoxon test).

(B) Gene Ontology enrichment analysis of promoter subsets. Each row of the heatmap represents a promoter region. Biological Process (BP) terms enriched in each subset of promoters are shown.

(C) Expression levels of genes targeted by active stage-specific enhancers in HSPC, E-Prog and E-Prec. The expression levels of genes associated with stage-specific enhancers were compared to the expression levels of the same genes in the other cell types. These genes showed higher expression levels in the specific stage compared to the others (Wilcoxon test).

(D) Gene Ontology enrichment analysis of enhancer subsets. Each row of the heatmap represents an enhancer region. Biological Process (BP) terms enriched in each subset of enhancers are shown.

(A, C) ns: not significant; *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001; ****: p-value < 0.0001

Figure S4. Expression levels and functional enrichment analysis of genes associated with stage-specific or shared super-enhancers. Related to Figure 1.

(A) Expression levels of genes associated with HSPC (n=453 genes), E-Prog (n=401 genes), and E-Prec (n=110 genes) stage-specific super-enhancers (Wilcoxon test).

(B) Gene Ontology enrichment analysis of super-enhancer subsets. Each row of the heatmap represents a super-enhancer. Biological Process (BP) terms enriched in each subset of super-enhancers are shown.

(C, D, E) Representative stage-specific super-enhancers and expression levels of the target gene in HSPC (C), E-Prog (D) and E-Prec (E). *DNMT3A* is essential in HSPC biology. *CD55* (encoding the antigen of the Cromer blood group system) and *RHD* (encoding the antigen of the Rh blood group) are highly expressed in E-Prog and E-Prec, respectively. Data were plotted as mean with SEM.

Figure S5. GATA1 expression along erythroid development and transcription factor binding site motifs in GATA-targeted genomic regions. Related to Figure 2.

(A) GATA1 protein copy number across erythropoiesis. Data were retrieved from (Gautier et al., 2016). Data were plotted as mean with SEM.

(B) Density plots of GATA2 and GATA1 ChIP-seq.

(C) Examples of GATA2 and GATA1 peaks at known target genes.

(D) Enrichment or depletion of transcription factor binding motifs in genomic regions targeted by GATA factors only in one cell population (GATA2 only in HSPC, GATA1 only in E-Prog and GATA1 only in E-Prec) or by GATA2 in HSPC and GATA1 in E-Prog (GATA switch) or by GATA1 in both E-Prog and E-Prec (GATA1 common). GATA and GATA-TAL1 binding motifs were enriched in all the categories.

Figure S6. Regulatory regions targeted by GATA factors. Related to Figure 2.

(A, B, C) Bar plot showing the percentage of promoters (A), enhancers (B) and super-enhancers (C) bound by GATA factors in HSPC, E-Prog and E-Prec.

(D) Examples of genes regulated by GATA factors (related to Figure 2E-F). CD55 and PDLIM1 belong to group B; STAM, NBAS, EPB41, RAPGEF2 and RNF187 belong to group C; MAP2K3 belongs to group F. RAPGEF2, MAP2K3 and RNF187 are down-regulated during murine erythropoiesis, but are upregulated during human erythropoiesis and associated with active regulatory elements in E-Prog and E-Prec.

(E) Expression levels of genes reported in panel D. Data were plotted as mean with SEM.

Figure S7. Enrichment analysis of transcription factor binding motifs around GATA1 BS within regulatory regions associated with up- or down-regulated genes in E-Prog and E-Prec. Related to Figure 3.

For each motif, enrichment values over genomic background are reported as log2 fold change for each category of GATA1 BS. GATA and GATA-TAL1 binding motifs were enriched in all the categories.

Figure S8. KIT expression and chromatin interactions in HUDEP-2 cells. Related to Figure 4.

(A) *KIT* expression levels in undifferentiated (Day0) and differentiated (Day7) HUDEP-2 cells, as determined by RT-qPCR. Data were plotted as mean with SEM. **: p-value < 0.01 (unpaired t-test).

(B) Flow cytometry analysis of KIT expression in undifferentiated (Day0) and differentiated (Day7) HUDEP-2 cells. The percentage of KIT^+ cells and the MFI (median fluorescence intensity) are shown. Data were plotted as mean with SEM. **: p-value < 0.01 (unpaired t-test).

(C) E-Prog regulatory elements and ATAC-seq signal within the *KIT* locus in undifferentiated HUDEP-2. Orange dashed boxes indicate the E-Prog-specific *KIT* enhancers.

(D-E) Chromatin interactions within the *KIT* locus in undifferentiated (Day0) and differentiated (Day7) HUDEP-2 cells. We used as anchor a genomic fragment containing the enhancer I (D) or II (E) (flanked by solid black lines). HindIII digestion fragments of interest are flanked by dashed black lines. Distances on the x-axis are in kb counting from the TSS of the *KIT* gene. (C) *KIT* enhancer I showed a higher interaction frequency with *KIT* enhancer II and promoter in undifferentiated HUDEP-2 cells. (D) *KIT* enhancer II showed a higher interaction frequency with *KIT* enhancer I and promoter in undifferentiated HUDEP-2 cells.

Figure S9. Disruption of GATA1 binding sites in *KIT* enhancers I and II. Related to Figure 5.

(A) Representative analysis of GATA1 binding site disruption in enhancer I. In Δ BS I samples, a frequency ranging from 36 to 46% of GATA1 binding sites in enhancer I were modified. Virtually all the editing events disrupted the GATA1 binding site. Most of the editing events disrupted also the TAL1 motif that overlaps with the GATA1 motif (composite TAL1:GATA1 motif). PCR products were subjected to Sanger sequencing and ICE CRISPR Analysis. InDel, insertion or deletion.

(B) Representative analysis of GATA1 binding site disruption in enhancer II. To disrupt GATA1 binding site in enhancer II, we employed 2 gRNA flanking the binding site (5'-end and 3'-end gRNAs). ddPCR revealed the deletion of a 67-bp fragment containing the GATA1

binding site in 7 to 11% of the loci. In addition, Sanger sequencing followed by ICE CRISPR Analysis revealed that, at the remaining loci, the 5'-end gRNA disrupted on average ~34% of the GATA1 binding sites. Consistently, targeted NGS sequencing of a PCR amplicon encompassing the GATA1 binding site in enhancer II showed that the GATA1 motif was disrupted in ~59% of the loci, and the AP1/HIF1b and PITX1 motifs in ~14% and ~42% of the loci, respectively (data not shown). InDel, insertion or deletion.

Supplemental Tables

Table S1. *Cis*-regulatory regions identified in HSPC, E-Prog and E-Prec. Related to Figure 1.

Regulatory region	Cell	Number of regions	Average size (bp)	Median size (bp)
Active promoters	HSPC	13,548	2,417.91	2,200
	E-Prog	9,788	2,594.91	2,400
	E-Prec	6,156	3,141.42	2,800
Weak promoters	HSPC	1,217	727.86	600
	E-Prog	4,957	1,250.31	1,000
	E-Prec	6,479	1,589.94	1,400
Bivalent promoters	HSPC	5,074	1,075.68	800
	E-Prog	5,605	1,295.24	1,200
	E-Prec	3,233	1,256.05	1,000
Active enhancers	HSPC	20,202	2,939.47	2,000
	E-Prog	20,083	1,857.50	1,200
	E-Prec	1,725	2,596.06	1,800
Weak enhancers	HSPC	27,251	1,013.06	800
	E-Prog	24,327	1,247.87	800
	E-Prec	19,051	1,478.14	1,000
Bivalent enhancers	HSPC	1,955	1,046.65	800
	E-Prog	1,991	993.47	600
	E-Prec	561	870.94	600
Super-enhancers	HSPC	497	32,928.37	28,800
	E-Prog	436	30,368.81	27,400
	E-Prec	135	14,594.07	12,000

Table S5. GATA factor binding sites identified in HSPC, E-Prog and E-Prec. Related to Figure 2.

TF	Cell	Number of BS	Average size (bp)	Median size (bp)
GATA2	HSPC	15171	276.24	241
	E-Prog	419	456.17	453
	E-Prec	0	-	-
GATA1	HSPC	0	-	-
	E-Prog	23268	186.15	166
	E-Prec	11005	231.37	202

Table S6. Primer for RT-qPCR analysis. Related to Figures S1 and 5.

Primer name	Sequence (5'-3')
HBA Forward primer	CGGTCAACTTCAAGCTCCTAA
HBA Reverse primer	ACAGAAGCCAGGAACTTGTC
HBB Forward primer	GCAAGGTGAACGTGGATGAAGT
HBB Reverse primer	TAACAGCATCAGGAGTGGACAGA
KIT Forward primer	ATGGCACGGTTGAATGTAAGGC
KIT Reverse primer	TCTCCTCAACAACCTTCCACTG
FOXO3 Forward primer	TGTTGGTTTGAACGTGGGGA
FOXO3 Reverse primer	TGTCCACTTGCTGAGAGCAG
STOM Forward primer	GGAGCCAAAGGACCTGGTTT
STOM Reverse primer	GACCACACCATCCACGCTAA
GYPA Forward primer	TCCAGAAGAGGAAACCGGAGA
GYPA Reverse primer	AAAGGCACGTCTGTGTCAGG
GATA1 Forward primer	GCCCAAGAAGCGAATGATTG
GATA1 Reverse primer	GTGGTCGTTTGACAGTTAGTGCAT
GATA2 Forward primer	ACCACAAGATGAATGGACAGAA
GATA2 Reverse primer	GTCGTCTGACAATTTGCACAAC

Table S7. Primer for 3C experiments in *KIT* locus using promoter as anchor. Related to Figure 4.

Primer name	Description	Sequence (5'-3')
-99.2 kb primer		AATTCACCTGCTCAAACCCT
-86.4 kb primer	Enhancer I	AAAGACAGCATTGCGTGACC
-76.2 kb primer		CCCCACACCCAGCCAAATTA
-72.2 kb primer	Enhancer II	AGTGTACATGCTCAAGCCCA
-44.7 kb primer		GCAGCGTGGTTTAAGAATCCC
-27.6 kb primer		TGAACCATCAAGCCTTGCCT
-18.8 kb primer		GGGCCTGGAAGGATGAGTTG
-3.5 kb primer	Promoter (anchor)	CTGGGTGGCTGGAAGGTAAA

Table S8. Primer for 3C experiments in *KIT* locus using Enhancer I as anchor. Related to Figure S8.

Primer name	Description	Sequence (5'-3')
-91.0 kb primer	Enhancer I (anchor)	AAAGACAGCATTGCGTGACC
-76.2 kb primer		AGACTGGACCTTCAAAATGGTG
-72.2 kb primer	Enhancer II	GGGAAAGACTCCGAGTGAGC
-70.3 kb primer		CAATTGGTCACAGCCAGTGC
-62.2 kb primer		TTGGGGATCTGGCCATTCAG
-42.4 kb primer		TCCACAATTGGACTGCCCTC
-27.6 kb primer		GCTCTCTAAGGTGGCACAGT
-14.9 kb primer		GCCATAGCATGGCATTCAAGA
-3.5 kb primer	Promoter	CTGGGTGGCTGGAAGGTAAA

Table S9. Primer for 3C experiments in KIT locus using Enhancer II as anchor. Related to Figure S8.

Primer name	Description	Sequence (5'-3')
-91.0 kb primer	Enhancer I	AAAGACAGCATTGCGTGACC
-76.2 kb primer		TGGAATCCTGGAAAATCGCA
-72.2 kb primer	Enhancer II (anchor)	AGTGTACATGCTCAAGCCCA
-62.2 kb primer		TTGGGGATCTGGCCATTCAG
-42.4 kb primer		AATTGGACTGCCCTCTCCAC
-27.6 kb primer		GCTCTCTAAGGTGGCACAGT
-21.3 kb primer		AGACAAGACCACAAAACATAAGGA
-14.9 kb primer		AGCCTTGTTTCTTGCCAATTCT
-3.5 kb primer	Promoter	CTGGGTGGCTGGAAGGTAAA

Table S10. gRNAs for *KIT* Enhancers deletion or GATA1 BS disruption. Related to Figure 5.

Name	Description	Sequence (5'-3')	Chr	Strand	Start	End
gRNA Ctr	Luciferase	CTTCGAAATGTCCGTTCGGT	na	na	na	na
gRNA1	Exon 2	GCCTAATCTCGTCGCCCACG	4	-	55,561,754	55,561,776
gRNA2	Enhancer I 5'-end	GGAGATCCTAGTTTACAG	4	+	55,436,361	55,436,378
gRNA3	Enhancer I 3'-end	GTTTATGACAATCCCTCA	4	-	55,440,103	55,440,121
gRNA4	Enhancer I GATA1bs	GACCAGTCTATTATCTGC	4	-	55,437,596	55,437,613
gRNA5	Enhancer II 5'-end	GTACACGGTATGTTGCGGGG	4	-	55,450,377	55,450,396
gRNA6	Enhancer II 3'-end	GACTGGCCACAGTCTCACGA	4	+	55,457,052	55,457,071
gRNA7	Enhancer II GATA BS 5'-end	CCTGTGACTCATTACTGTTG	4	+	55,451,941	55,451,960
gRNA8	Enhancer II GATA BS 3'-end	TTCAGCTCCACCACTAAATG	4	+	55,452,008	55,452,027

na, not applicable

Table S11. Primer for detecting GATA1 binding site disruption. Related to Figure S9.

Primer Name	Description	Sequence (5'-3')
F Enh I primer	Enhancer I GATA 1 BS (Sanger sequencing)	GCAAAGCCTTTCGTTTTGCC
R Enh I primer		TCCCTGACGCTAGAAGGAGT
F Enh II primer	Enhancer II GATA 1 BS (Sanger sequencing)	CGAATTTATGCGTGGGTGCC
R Enh II primer		AATCAGCTCCTGGGCTTTGG
F Enh II ddPCR primer	Enhancer II GATA 1 BS (ddPCR)	CCAAGAAACATCACTAGGACTCC
R Enh II ddPCR primer		CGAGGCTTAGAGAGGTGCAG
F KIT prom primer	KIT promoter, Chr4 (ddPCR)	GCAGTTAAGAGCCCTAGCCC
R KIT prom primer		GCTGCCAACCCCAGTAATGA
F Enh II NGS primer	Enhancer II GATA 1 BS (NGS)	ACAACTTTCTACTCTCTCATGCTG
R Enh II NGS primer		AGAATCCAGGTTGCTGCAGA

Transparent methods

Culture and purification of primary cells

Isolation of hematopoietic stem/progenitor cells (HSPC)

Human $CD34^+$ cells were purified from umbilical cord blood (CB) samples obtained from healthy donors. CB samples eligible for research purposes were obtained thanks to a convention with the cord blood bank of Saint Louis Hospital (Paris, France). Experiments were performed in accordance with the Declaration of Helsinki. This study was approved by the regional investigational review board (DC 2014-2272, CPP Ile-de-France II "Hôpital Necker-Enfants malades"). Informed consent was obtained from all subjects. Mononuclear cells derived from different healthy donors were pooled and isolated by gradient separation (Lymphocytes separation medium, Eurobio), washed twice with phosphate-buffered saline (PBS) completed with 2% fetal bovine serum (FBS, Hyclone), and then $CD34^+$ cells were purified by immunomagnetic sorting (Indirect CD34 MicroBead Kit, human Miltenyi Biotec). The purity of $CD34^+$ sorted cell populations was evaluated by flow cytometry.

Culture of HSPC

$CD34^+$ cells were seeded at 0.5 x 10^6 cells/ml and cultured for 36 hours in an IMDM-based expansion medium (Lonza) containing 20% FBS (Hyclone) and supplemented with 100 ng/ml human stem cell factor (hSCF), 100 ng/ml human Flt3-ligand (hFlt3-l), 20 ng/ml human trombopoietin (hTPO) and 20 ng/ml human interleukin-6 (hIL-6) (PeproTech). Cells were harvested after 36 hours.

Culture and isolation of erythroid progenitors (E-Prog) and erythroid precursors (E-Prec)

$CD34^+$ cells were cultured as described by Roselli and colleagues (Roselli et al., 2010). Cells were seeded at 10^5 cells/ml in StemSpan medium (Stem Cell Technologies) containing 20% FBS (Hyclone) and supplemented with 50 ng/ml hSCF (Peprotech), 1 U/ml human erythropoietin (EPO; Janssen), 1 ng/ml hIL-3 (Peprotech), 10^{-6} M dexamethasone (Sigma-Aldrich), and 10^{-6} M ß-estradiol (Sigma-Aldrich). At day 7, cells were grown in StemSpan medium with 10% FBS supplemented with 2 U/ml human EPO, and at day 10 they were cultured in StemSpan containing only 10% FBS.

To isolate $CD36^+$ erythroid progenitors (E-Prog) at day 5, cells were labeled with FITC-conjugated anti-CD36 antibody and selected using magnetic-activated cell sorting (MACS) technology in combination with anti-FITC microbeads (Miltenyi Biotec), following manufacturer's instructions. Erythroid precursors (E-Prec) were collected at day 11.

Flow cytometry analysis

Cells were labeled with antibodies against CD34 (CD34-FITC [345801], BD Biosciences; CD34-PE [345802], BD Biosciences), CD36 (CD36-FITC [555454], BD Pharmingen), CD235a/GYPA (CD235a-APC [551336], BD Pharmingen) and CD117/KIT (CD117-PE [332785], BD Biosciences) surface markers, following manufacturer's instructions. FACS analysis was performed using LSRFortessa X-20 flow cytometer (BD Biosciences).

RT-qPCR

RNA was reverse-transcribed using SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen). RT-qPCR was performed using iTaq Universal SYBR Green Supermix (Bio-Rad). LMNB2 primers were used as an internal control. RT-qPCR data were analyzed using the $2^{-\Delta\Delta Ct}$ method. Primer sequences are listed in **Table S6**.

Western Blot

To extract nuclear proteins, $3x10^6$ cells were resuspended in 100 µl lysis buffer (20 mM HEPES pH 7.8, 5 mM potassium acetate, 0.5 mM $MgCl_2$ in water), then crushed 25 times in a douncer. Lysate was centrifuged for 5 min at 1,500 x *g*. Nuclei containing pellets were resuspended in 100 µl sonication buffer (50 mM Tris-HCl pH 8, 1 mM EDTA pH 8, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, 140 mM NaCl in water) and sonicated using Bioruptor Pico (Diagenode) for 10 cycles of a 30 s ON/30 s OFF program. Protein concentration was measured using BCA protein assay kit (Pierce).

Western Blot analysis was performed using Mini Gel Tank (Thermo) by loading the equivalent of $3x10^5$ cells lysate per well. SDS PAGE was performed using NuPage Bis-Tris 4-12% gel (Thermo) in 1X MOPS buffer (Thermo) for 2 hours at 100 V. Protein transfer was performed

on PVDF membrane (Millipore) pre-activated in ethanol, using a Mini Blot Transfer Module (Thermo). After transfer, membranes were incubated for 1 hour in TBS-Tween buffer (50 mM Tris-HCl pH8, 150 mM NaCl, 0,1% Tween 20 in water) completed with 5% skim milk.

Primary antibody staining was performed by incubating the membranes in TBS-Tween buffer with 5% skim milk at 4°C overnight on a shaker using rabbit anti-GATA1 antibody (Abcam, ab11852, diluted 1:400), rabbit anti-GATA2 antibody (Santa Cruz sc299, diluted 1:200), rabbit anti-H3K27ac antibody (Abcam, ab4729, diluted 1 µg/ml) and rabbit anti-H3 antibody (Abcam, ab1791, diluted 1:1000). Blots were washed several times in TBS-Tween and incubated 1 hour at room temperature (RT) in TBS-Tween buffer with 5% skim milk containing anti-rabbit HRP antibody (Thermo, A27036, diluted 1:3000). After 3 washes in TBS-Tween, membranes were incubated for 1 minute with ECL substrate (Pierce) and exposed to film (Hyperfilm ECL, Amersham). The film was developed following manufacturer's instructions.

RNA-seq

RNA extraction, library preparation and sequencing

Total RNA was extracted from 1-2x10^6 HSPC, E-Prog and E-Prec (3 biological replicates for each stage) using RNeasy Micro kit (QIAGEN). RNA quality was assessed using the Agilent High Sensitivity RNA ScreenTape System (Agilent Technologies): RNA Integrity Number (RIN) was > 8 for all the samples. Libraries were prepared starting with 1 µg of total RNA using TruSeq Stranded mRNA kit (Illumina) and 100 bp paired-end sequences were generated on the HiSeq 2500 instrument (Illumina). About 30 million reads per sample were obtained.

Bioinformatic analysis

Reads quality was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw reads were mapped to the human reference genome (hg19 build) using STAR (Dobin et al., 2013). Raw gene counts were obtained using HTSeq (Anders et al., 2015), with stranded option and Ensembl gene annotation (GRCh37 Release 82), and then normalized according to library size to obtain counts per million (CPM) values. Only genes with a CPM ≥ 1 in at least 3 samples were

retained for subsequent analyses. DESeq2 R package was employed to estimate size factors and dispersion in all samples and then define differentially expressed genes (DEGs) between samples of consecutive stages (Love et al., 2014). Functional enrichment analysis of gene ontology (GO) biological process (BP) categories was performed using clusterProfiler R package (Yu et al., 2012). We used Principal Component Analysis (PCA) to compare E-Prog and E-Prec gene expression profiles with public transcriptomic datasets obtained from human erythroid progenitors (BFU-E and CFU-E) and erythroid precursors (proerythroblasts, proE; early basophilic erythroblasts, earlyB; late basophilic erythroblasts, lateB; polychromatic erythroblasts, polyC; orthochromatic erythroblasts, orthoC) (An et al., 2014; Li et al., 2014).

ChIP-seq

ChIP assay

Chromatin from E-Prog and E-Prec (derived from pools of HSPC obtained from multiple donors) was prepared after cross-linking for 10 minutes at RT with 1% formaldehyde-containing medium. Nuclear extracts were sonicated using the Bioruptor Pico Sonication System (Diagenode) to obtain DNA fragments around 150-200 bp in length. For histone modifications, chromatin obtained from 10^7 cells was immunoprecipitated overnight with 10 µg of antibodies against H3K4me1 (ab8895, Abcam), H3K4me3 (ab8580, Abcam), H3K27ac (ab4729, Abcam), and H3K27me3 (07-449, Millipore). For transcription factors, chromatin obtained from $3x10^7$ cells was immunoprecipitated overnight with 30 µg of antibodies against GATA1 (ab11852, Abcam) and GATA2 (sc-9008, Santa Cruz Biotechnology). ChIP assay was performed as previously described (Cui et al., 2009; Romano et al., 2016).

Library preparation and sequencing

ChIP-seq libraries for E-Prog and E-Prec were prepared from 1 ng of immunoprecipitated (IP) DNA and 5 ng of control DNA (Input: non-immunoprecipitated chromatin fragments) following the Diagenode Microplex Library preparation kit. Libraries quality was checked by capillary electrophoresis (Tape Station, Agilent) with the High sensitivity D1000 assay and quantified by Real Time q-PCR using the kit from KAPA Biosystems (Roche). Each library was sequenced in one MiSeq Illumina RUN and 50 bp single-end reads were generated. At least 20 million reads per sample were obtained.

Bioinformatics data analysis

Read quality was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw reads were mapped against the human reference genome (hg19 build) using Bowtie allowing up to 2 mismatches (Langmead et al., 2009). Then, each BAM file was processed using SAMtools and converted into a bed file using BEDTools (Li et al., 2009; Quinlan and Hall, 2010). Sample quality was evaluated using the cross-correlation analysis implemented in spp R package (Kharchenko et al., 2008). Density plots of GATA2 and GATA1 ChIP-seq (**Figure S5B**) were generated using ngs.plot (Shen et al., 2014) and a 4-kb window centered on gene transcription start site (TSS). ChIP-seq peak calling was performed using MACS2, with a q-value threshold of 0.05, and *--broad* option for histone modifications (Zhang et al., 2008). Input data were used to model the background noise. HSPC raw ChIP-seq data were downloaded from the Gene Expression Omnibus (GEO) GSE70660 and GSE45144 (H3K4me3: GSM1816072, H3K4me1: GSM1816068, H3K27ac: GSM1816075, H3K27me3: GSM1816079, GATA1: GSM1816081, GATA2: GSM1097883, input: GSM1816095) and analyzed as described above for E-Prog and E-Prec. The antibodies used in HSPC were the same used in E-Prog and E-Prec except for H3K4me3 antibody (04-745; Millipore).

Motif finding analysis on GATA factor BS was performed using the findMotifsGenome function of Homer software (Heinz et al., 2010). Motifs were searched in 200-bp windows centered on peak summits.

Definition of chromatin states and regulatory regions

Chromatin states were defined using Spectacle (Song and Chen, 2015), training the hidden Markov model (HMM) with regions enriched in histone modifications in the 3 cell stages, setting the number of internal states to 11, and using bins of 200 bp. States with high emission probabilities for H3K4me3 were defined as promoter-like, while states with high emission probabilities for H3K4me1 and low for H3K4me3 were defined as enhancer-like. Overall, we defined 5 promoter states, 4 enhancer states, a polycomb-repressed state, and a quiescent state devoid of any histone mark (**Figure S2C**). Specifically, based on the emission probabilities of H3K27ac and H3K27me3, promoter and enhancer chromatin states were

labeled as active (H3K27ac^{+}H3K27me3^{-}), weak (H3K27ac^{-}H3K27me3^{-}) or bivalent (H3K27ac^{-} H3K27me3^{+}).

Genomic segments with an “active promoter” state were merged with contiguous genomic segments with “active promoter” and “weak promoter” states (if present) to define active promoters; genomic segments with a “bivalent promoter” state were merged with contiguous genomic segments with a “weak promoter” state (if present) to define bivalent promoters; the remaining genomic segments with a “weak promoter” state were defined weak promoters. The same strategy was applied to define active, weak and bivalent enhancers. Finally, we retained only promoters at a distance lower than 5 kb from Ensembl gene TSSs and enhancers at a distance higher than 5 kb from Ensembl gene TSSs.

Active enhancers were used as constituent enhancers for super-enhancer identification. Enhancers were stitched together and super-enhancers were defined using the ROSE algorithm (Lovén et al., 2013; Whyte et al., 2013), with stitching distance of 10 kb and promoter exclusion zone of 10 kb around Ensembl gene TSSs. H3K27ac signal was used for enhancer ranking. The number of regulatory regions defined at each stage of erythroid development, and their average and median size, are reported in **Table S1**.

Regulatory regions were annotated to the nearest 3 genes (with a maximum distance of 100 kb). Functional enrichment analysis of gene ontology (GO) biological process (BP) categories was performed using clusterProfiler R package (Yu et al., 2012).

Identification of epigenetic dynamics in erythropoiesis

To study the dynamic changes in the usage of regulatory elements (promoters, enhancers and super-enhancers) during erythroid development, we generated a unique reference list containing active, weak and bivalent regulatory elements by merging the genomic regions identified at each stage, using the *merge* and *multiintersect* functions of the BEDTools suite (Quinlan and Hall, 2010) (**Figure 1F-H**).

We then integrated the analysis of the dynamics of regulatory regions with the differential expression analysis (**Figure 3A**). Briefly, for each DEG, we quantified the total extension (in kb) of each type of regulatory elements (active, weak, and bivalent promoters and enhancers) by summing up the size of all gene-associated regulatory regions in each cell type. We

obtained a gene-centered matrix containing the coverage of the 6 types of regulatory elements associated to each DEG for each cell stage. Finally, we used the k-means clustering algorithm to identify clusters of DEGs sharing similar epigenetic landscapes (i.e., coverage of regulatory regions). In k-means, the silhouette method was used to determine the optimal number of 19 clusters (as local maximum). The statistical significance of the association of each cluster with gene expression patterns and transcription factor binding sites was assessed using Fisher's exact test.

ATAC-seq

Raw data of ATAC-seq performed in undifferentiated HUDEP-2 cells were downloaded from the Gene Expression Omnibus (GEO) public database (GSE74977) (Masuda et al., 2016). Raw reads were trimmed to remove adapter sequences and mapped against the human reference genome (hg19 build) using Bowtie2 (Langmead and Salzberg, 2012). Then, the BAM file was processed using SAMtools to remove unmapped, not primary alignments, supplementary alignments, reads with alignment quality below 30 and to keep only properly paired reads (Li et al., 2009). Duplicates were also removed together with reads mapping on mitochondrial chromosome. The resulting BAM file was used to generate a normalized coverage track file (bigwig).

HUDEP-2 cell culture

HUDEP-2 cells were cultured as described in Antoniani et al. (2018) (Antoniani et al., 2018). Briefly, HUDEP-2 cells were expanded in StemSpan medium (Stem Cell Technologies) supplemented with 1 µg/ml doxycycline (Sigma-Aldrich), 100 ng/ml hSCF (Peprotech), 3 U/ml EPO (Janssen) and 10^{-6} M dexamethasone (Sigma-Aldrich). HUDEP-2 cells were differentiated for 7 days in IMDM medium (Lonza) containing 5% AB human serum (Biowest) and supplemented with 1 µg/ml doxycycline (Sigma-Aldrich), 100 ng/ml hSCF (Peprotech), 3 U/ml EPO (Janssen), 10 µg/ml human insulin (Sigma-Aldrich), 330 µg/ml holo-transferrin (R&D SYSTEMS) and 2 U/ml heparin (Sigma-Aldrich). Flow cytometry analysis of erythroid markers were performed to monitor differentiation of HUDEP-2 cells (Antoniani et al., 2018).

Chromosome Conformation Capture

Chromosome conformation capture (3C) assays were performed as described in Hagège et al. (2007) (Hagège et al., 2007). Briefly, HUDEP-2 cells were cross-linked for 10 minutes at RT with 1% formaldehyde-containing medium. Crosslinking was quenched by addition of a 3X glycine solution to reach a final concentration of 0.125 M. Nuclei obtained from $2x10^7$ cells were digested with HindIII restriction enzyme overnight at 37 °C while shaking at 900 rpm. The digestion reaction was stopped by addition of SDS (2% final concentration) and heat inactivation at 65 °C for 30 minutes while shaking at 900 rpm. DNA fragments were diluted and ligated using the Quick Ligase (NEB) for 4 hours at 16 °C, followed by an additional step of 30 minutes at RT, while shaking at 900 rpm. After Protease K and RNase A treatment, and reverse cross-linking, ligation products were purified by phenol-chloroform extraction and ethanol precipitation. Ligation products were amplified by qPCR using SYBR Green PCR Master Mix (Applied Biosystems). A control template of randomly ligated DNA fragments was prepared by digestion and ligation of a bacterial artificial chromosome (BAC) containing the human *KIT* locus (RP11-959G16). This control template was used to generate a standard curve and quantify the interaction frequencies. To reduce variability between samples, interaction frequencies detected in the *KIT* locus were normalized to the interaction frequency between two restriction fragments in the *ERCC3* locus, whose spatial conformation was assumed similar in all analyzed samples (Palstra et al., 2003). Primer sequences used to evaluate the interaction frequencies are listed in **Tables S7-9**.

CRISPR/Cas9 genome editing experiments

Single guide RNAs (gRNAs) were designed using ZIFIT and CRISPOR tools (Haeussler et al., 2016; Sander et al., 2010, 2007). COSMID tool was used to select sequence-specific and efficient 18-20 bp-long gRNAs (Cradick et al., 2014). Plasmids expressing Cas9-GFP (pMJ920) and gRNAs (MLM3636) were purchased from Addgene (plasmid # 42234 and 43860). gRNA spacers were cloned in the MLM3636 plasmid or in the MA128 plasmid (provided by Dr. Mario Amendola, Genethon, France) containing an optimized gRNA scaffold (Dang et al., 2015). Sequences and genomic coordinates (hg19 build) of gRNAs are listed in **Table S10**. HUDEP-2 cells ($1x10^6$) were transfected with 4 µg of a Cas9-GFP expressing plasmid and 0.8-1.6 ug of each gRNA-containing vector using Nucleofector I (L-29 program;

Lonza) and AMAXA Cell Line Nucleofector Kit V (VCA-1003). FACS analyses were performed 24 hours after transfection using LSRFortessa X-20 flow cytometer (BD Biosciences) to evaluate GFP and KIT (CD117-PE [332785], BD Biosciences) expression levels. GFP^+ cells were sorted using SH800 Cell Sorter (Sony Biotechnology) for DNA and RNA extraction. To evaluate editing efficiency at gRNA target sites, we performed PCR followed by Sanger sequencing and ICE CRISPR Analysis Tool (Synthego) (Hsiau et al., 2018). Digital Droplet PCR (ddPCR) was performed using EvaGreen mix (Biorad) to quantify the frequency of a 60-bp deletion encompassing the GATA1 binding site in enhancer II. Control primers annealing to a neighboring genomic region (KIT promoter) were used as DNA loading control. Primers use for PCR amplification and Sanger sequencing and for ddPCR are listed in **Table S11**.

Statistical analyses

Chi-square tests were used to compare the genome coverage of histone modifications and chromatin states (**Figure S2**). Wilcoxon tests were used to compare expression levels distributions (**Figure S2, S3 and S4**). Unpaired *t* tests were used to compare expression levels measured by RT-qPCR and FACS analysis (**Figure 4, 5 and Figure S8**). Wilcoxon tests were performed in R (v3.3.1); Chi-square and unpaired *t* tests were performed in Prism6 software (GraphPad, La Jolla, CA, USA). The threshold for statistical significance was set to $p\text{-value} < 0.05$.

Data and software availability

All RNA-seq and ChIP-seq data have been deposited in Gene Expression Omnibus under accession number GSE124165.

Supplemental References

An, X., Schulz, V.P., Li, J., Wu, K., Liu, J., Xue, F., Hu, J., Mohandas, N., Gallagher, P.G., 2014. Global transcriptome analyses of human and murine terminal erythroid differentiation. Blood 123, 3466–3477. https://doi.org/10.1182/blood-2014-01-548305

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Antoniani, C., Meneghini, V., Lattanzi, A., Felix, T., Romano, O., Magrin, E., Weber, L., Pavani, G., El Hoss, S., Kurita, R., Nakamura, Y., Cradick, T.J., Lundberg, A.S., Porteus, M., Amendola, M., El Nemer, W., Cavazzana, M., Mavilio, F., Miccio, A., 2018. Induction of fetal hemoglobin synthesis by CRISPR/Cas9-mediated editing of the human β-globin locus. Blood 131, 1960–1973. https://doi.org/10.1182/blood-2017-10-811505

Cradick, T.J., Qiu, P., Lee, C.M., Fine, E.J., Bao, G., 2014. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. Mol. Ther. - Nucleic Acids 3, e214. https://doi.org/10.1038/mtna.2014.64

Cui, K., Zang, C., Roh, T.-Y., Schones, D.E., Childs, R.W., Peng, W., Zhao, K., 2009. Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation. Cell Stem Cell 4, 80–93. https://doi.org/10.1016/j.stem.2008.11.011

Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P., Wu, H., 2015. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. Genome Biol. 16, 280. https://doi.org/10.1186/s13059-015-0846-3

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Gautier, E.-F., Ducamp, S., Leduc, M., Salnot, V., Guillonneau, F., Dussiot, M., Hale, J., Giarratana, M.-C., Raimbault, A., Douay, L., Lacombe, C., Mohandas, N., Verdier, F., Zermati, Y., Mayeux, P., 2016. Comprehensive Proteomic Analysis of Human Erythropoiesis. Cell Rep. 1–15. https://doi.org/10.1016/j.celrep.2016.06.085

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J., Concordet, J.-P., 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 17, 148. https://doi.org/10.1186/s13059-016-1012-2

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., Forné, T., 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat. Protoc. 2, 1722–1733. https://doi.org/10.1038/nprot.2007.243

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities.

Mol. Cell 38, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Hsiau, T., Maures, T., Waite, K., Yang, J., Kelso, R., Holden, K., Stoner, R., 2018. Inference of CRISPR Edits from Sanger Trace Data. bioRxiv 251082. https://doi.org/10.1101/251082

Kharchenko, P. V, Tolstorukov, M.Y., Park, P.J., 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat. Biotechnol. 26, 1351–1359. https://doi.org/10.1038/nbt.1508

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Hale, J., Bhagia, P., Xue, F., Chen, L., Jaffray, J., Yan, H., Lane, J., Gallagher, P.G., Mohandas, N., Liu, J., An, X., 2014. Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. Blood 124, 3636–3645. https://doi.org/10.1182/blood-2014-07-588806

Liang, R., Campreciós, G., Kou, Y., McGrath, K., Nowak, R., Catherman, S., Bigarella, C.L., Rimmelé, P., Zhang, X., Gnanapragasam, M.N., Bieker, J.J., Papatsenko, D., Ma'ayan, A., Bresnick, E., Fowler, V., Palis, J., Ghaffari, S., 2015. A Systems Approach Identifies Essential FOXO3 Functions at Key Steps of Terminal Erythropoiesis. PLOS Genet. 11, e1005526. https://doi.org/10.1371/journal.pgen.1005526

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., Young, R.A., 2013. Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. Cell 153, 320–334. https://doi.org/10.1016/j.cell.2013.03.036

Masuda, T., Wang, X., Maeda, M., Canver, M.C., Sher, F., Funnell, A.P.W., Fisher, C., Suciu, M., Martyn, G.E., Norton, L.J., Zhu, C., Kurita, R., Nakamura, Y., Xu, J., Higgs, D.R., Crossley, M., Bauer, D.E., Orkin, S.H., Kharchenko, P. V., Maeda, T., 2016. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. Science 351, 285–289. https://doi.org/10.1126/science.aad3312

Palstra, R.-J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., de Laat, W., 2003. The β-globin nuclear compartment in development and erythroid differentiation. Nat. Genet. 35, 190–194. https://doi.org/10.1038/ng1244

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Romano, O., Peano, C., Tagliazucchi, G.M., Petiti, L., Poletti, V., Cocchiarella, F., Rizzi, E., Severgnini, M., Cavazza, A., Rossi, C., Pagliaro, P., Ambrosi, A., Ferrari, G., Bicciato, S., De Bellis, G., Mavilio, F., Miccio, A., 2016. Transcriptional, epigenetic and retroviral signatures identify regulatory regions involved in hematopoietic lineage commitment. Sci. Rep. 6, 24724. https://doi.org/10.1038/srep24724

Roselli, E.A., Mezzadra, R., Frittoli, M.C., Maruggi, G., Biral, E., Mavilio, F., Mastropietro, F., Amato, A., Tonon, G., Refaldi, C., Cappellini, M.D., Andreani, M., Lucarelli, G., Roncarolo, M.G., Marktel, S., Ferrari, G., 2010. Correction of β-thalassemia major by gene transfer in haematopoietic progenitors of pediatric patients. EMBO Mol. Med. 2, 315–328. https://doi.org/10.1002/emmm.201000083

Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K., Dobbs, D., 2010. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. 38, W462–W468. https://doi.org/10.1093/nar/gkq319

Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F., Dobbs, D., 2007. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. Nucleic Acids Res. 35, W599–W605. https://doi.org/10.1093/nar/gkm349

Shen, L., Shao, N., Liu, X., Nestler, E., 2014. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics 15, 284. https://doi.org/10.1186/1471-2164-15-284

Song, J., Chen, K.C., 2015. Spectacle: fast chromatin state annotation using spectral learning. Genome Biol. 16, 33. https://doi.org/10.1186/s13059-015-0598-0

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., Young, R.A., 2013. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell 153, 307–319. https://doi.org/10.1016/j.cell.2013.03.035

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. Omi. A J. Integr. Biol. 16, 284–287. https://doi.org/10.1089/omi.2011.0118

Zhang, Y., Liu, T., Meyer, C. a, Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137