Epigenetic inactivation of *ERF* reactivates γ-globin expression in β-thalassemia

Xiuqin Bao,[1,4,5,22] Xinhua Zhang,[2,22] Liren Wang,[3,22] Zhongju Wang,[1,4,5] Jin Huang,[1,4,5] Qianqian Zhang,[1,4,5] Yuhua Ye,[1,4,5] Yongqiong Liu,[1,4,5] Diyu Chen,[1,4,5] Yangjin Zuo,[1,4,5] Qifa Liu,[6,7] Peng Xu,[8] Binbin Huang,[9] Jianpei Fang,[10] Jinquan Lao,[11] Xiaoqin Feng,[12] Yafeng Li,[13,14] Ryo Kurita,[15] Yukio Nakamura,[16,17] Weiwei Yu,[18] Cunxiang Ju,[18] Chunbo Huang,[19,20] Narla Mohandas,[21] Dali Li,[3,23] Cunyou Zhao,[1,4,5,*] and Xiangmin Xu[1,4,5,*]

Summary

The fetal-to-adult hemoglobin switch is regulated in a developmental stage-specific manner and reactivation of fetal hemoglobin (HbF) has therapeutic implications for treatment of β-thalassemia and sickle cell anemia, two major global health problems. Although significant progress has been made in our understanding of the molecular mechanism of the fetal-to-adult hemoglobin switch, the mechanism of epigenetic regulation of HbF silencing remains to be fully defined. Here, we performed whole-genome bisulfite sequencing and RNA sequencing analysis of the bone marrow-derived $GYPA^+$ erythroid cells from β-thalassemia-affected individuals with widely varying levels of HbF groups (HbF ≥ 95th percentile or HbF ≤ 5th percentile) to screen epigenetic modulators of HbF and phenotypic diversity of β-thalassemia. We identified an ETS2 repressor factor encoded by *ERF*, whose promoter hypermethylation and mRNA downregulation are associated with high HbF levels in β-thalassemia. We further observed that hypermethylation of the *ERF* promoter mediated by enrichment of DNMT3A leads to demethylation of γ-globin genes and attenuation of binding of ERF on the *HBG* promoter and eventually re-activation of HbF in β-thalassemia. We demonstrated that *ERF* depletion markedly increased HbF production in human $CD34^+$ erythroid progenitor cells, HUDEP-2 cell lines, and transplanted NCG-Kit-V831M mice. ERF represses γ-globin expression by directly binding to two consensus motifs regulating γ-globin gene expression. Importantly, *ERF* depletion did not affect maturation of erythroid cells. Identification of alterations in DNA methylation of *ERF* as a modulator of HbF synthesis opens up therapeutic targets for β-hemoglobinopathies.

Introduction

The developmental switch from fetal to adult hemoglobin (Hb) in humans is a stage-specific regulatory process.[1] Abnormally high levels of fetal hemoglobin (HbF) in adult life, known as hereditary persistence of fetal hemoglobin (HPFH), described 60 years ago, ameliorate the debilitating clinical manifestation of β-globin mutations in β-thalassemia (MIM: 613985) and sickle cell disease.[1,2] The molecular machinery involved in human Hb switching has been extensively studied, leading to identification of genetic modifier genes and the defining of the contribution of modifier variants in modulating the phenotypic diversity of β-thalassemia and sickle cell disease by regulating HbF levels.[3–5]

Two transcription factors (TFs), BCL11A and ZBTB7A/LRF, have been identified as major erythroid-specific HbF repressors,[6–9] and both have recently been shown to bind *cis*-acting elements in the γ-globin promoter.[10,11] A domain-focused CRISPR screen has identified heme-regulated inhibitor[12] as a potentially druggable target involved in HbF silencing through the regulation of BCL11A.[13] In addition, a series of non-deleted HPFH mutations have been characterized in association with elevated levels of HbF through disruption of BCL11A or ZBTB7A binding[2,10,11] or creation of a *de novo* binding site for the erythroid master regulators

[1]Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, 510515 Guangdong, China; [2]Department of Hematology, 923rd Hospital of the People's Liberation Army, Nanning, Guangxi 530021, China; [3]Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai 200241, China; [4]Guangdong Engineering and Technology Research Center for Molecular Diagnostics of Human Genetic Diseases, Guangzhou, 510515 Guangdong, China; [5]Guangdong Engineering and Technology Research Center for Genetic Testing, Guangzhou, 510515 Guangdong, China; [6]Department of Hematology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China; [7]Guangdong Provincial Key Laboratory of Construction and Detection in Tissue Engineering, Guangzhou 510515, China; [8]Hematology Center of Cyrus Tang Medical Institute, Soochow University, Suzhou 215123, China; [9]Department 1 of Internal Medicine, Sixth People's Hospital of Nanning, Nanning, Guangxi 530021, China; [10]Department of Pediatrics, Sun-Yat-Sen Memorial Hospital, Sun-Yat-Sen University, Guangzhou, 510120 Guangdong, China; [11]Department of Pediatrics, Liuzhou Worker's Hospital, Liuzhou, Guangxi 545005, China; [12]Department of Pediatrics, Nanfang Hospital, Southern Medical University, Guangzhou, 510515 Guangdong, China; [13]Department of Nephrology, The Shanxi Provincial People's Hospital, Taiyuan, Shanxi 030012, China; [14]Precision Medicine Center, The Shanxi Provincial People's Hospital, Taiyuan, Shanxi 030012, China; [15]Department of Research and Development, Central Blood Institute, Japanese Red Cross Society, Shibadaimon, Minato-ku, Tokyo 105-8521, Japan; [16]Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki, Japan; [17]Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan; [18]GemPharmatech Co., Ltd., Nanjing 210000, Jiangsu, China; [19]Guangzhou Huayin Medical Laboratory Center Co., Ltd., Guangzhou, 510663 Guangdong, China; [20]Guangzhou Jiexu Gene Technology Co., Ltd., Guangzhou, 510530 Guangdong, China; [21]Red Cell Physiology Laboratory, New York Blood Center, New York, NY 10065, USA
[22]These authors contributed equally
[23]Senior author
*Correspondence: cyzhao@smu.edu.cn (C.Z.), gzxuxm@pub.guangzhou.gd.cn (X.X.)
https://doi.org/10.1016/j.ajhg.2021.03.005.

TAL1, KLF1, and GATA1.[2,14–16] The identification of these naturally occurring mutations has provided potential therapeutic targets for genetic treatment of β-hemoglobinopathies, and some of these are currently being explored. Substantial efforts have also been made to introduce these mutations in the human hematopoietic stem and progenitor cells (HSPCs) with high efficiency through a genome-editing strategy to allow for stable and autologous reactivation of HbF.[1,2,17–20]

Silencing of the γ-globin genes (*HBG1* [MIM: 142200]/*HBG2* [MIM: 142250]) is regulated by the combined action of numerous factors, among which BCL11A and ZBTB7A are direct repressors essential for this regulation.[1,2,17,18] Epigenetic regulators such as the NuRD chromatin complex, EHMT, DNMT3A, and KDM1A, along with lineage-defining factors such as KLF1 and MYB, play crucial roles in coordinating the expression of *BCL11A* (MIM: 606557) or *ZBTB7A* (MIM: 605878).[1,21–23] In addition, a recent study unraveled *LIN28B*-mediated genetic silencing of HbF through a direct suppression of *BCL11A* mRNA expression,[24] indicating additional post-transcriptional or epigenetic mechanisms in regulation of HbF expression. In this context, we recently performed a next-generation sequencing (NGS)-based screening of variants in globin gene clusters in a large cohort of β-thalassemia-affected individuals and identified that a common regulatory single nucleotide polymorphism (rSNP), rs368698783, in the *HBG* promoter is a major genetic modifier capable of ameliorating the severity of thalassemia major through the epigenetic-mediated elevation of HbF expression.[25] More recently, we identified a missense mutation, c.2633G>A (p.Ser878Phe), in DNA methyltransferase 1 (DNMT1) in β-thalassemia-affected individuals as a modulator of HbF synthesis by attenuating the interactions of DNMT1 with BCL11A, GATA1, and HDAC1/2 and reducing the recruitment of DNMT1 to the *HBG* promoters.[26] These findings suggest that more genetic and epigenetic factors contributing to HbF reactivation involved in ameliorating β-thalassemia severity remain unidentified.

Alternatively, the present study focused on exploring the protective nature of HPFH resulting from intrinsic epigenetically related regulators in addition to genetic alterations.[1,27] Clinically, natural variations in HbF levels produce a wide range of expression from no expression to nearly 100% in β-thalassemia-affected individuals,[28] thereby providing a valuable tool to identify candidate epigenetic modulators of HbF by genome-wide analysis of samples from β-thalassemia-affected individuals. We designed a case-based strategy by combining whole-genome bisulfite sequencing (WGBS) and RNA sequencing (RNA-seq) analysis of bone marrow (BM)-derived $GYPA^+$ erythroid cells from β^0-thalassemia-affected individuals stratified into low- and high-HbF expression groups to screen for potential epigenetic modulators of HbF expression. This strategy led to the identification of ERF as an HbF repressor in the discovery cohort of individuals with hypermethylation of the *ERF* (MIM: 611888) promoter as a distinguishing feature of individuals with high HbF levels. This finding was further validated in the $GYPA^+$ cells from an independent cohort of 47 β-thalassemia-affected individuals. Furthermore, using a series of *in vivo* and *in vitro* assays, we demonstrated that ERF acted as a transcription repressor of γ-globin genes through direct binding on their distant regulatory elements. Our study provided an epigenetic mechanism for the reactivation of fetal γ-globin expression.

Material and methods

Study design

BM-derived $GYPA^+$ erythroblasts from six unrelated Chinese β^0/β^0-thalassemia-affected individuals with widely varying levels of HbF were stratified into low-HbF (HbF_L: 0.1–0.4 g/dL, n = 3) and high-HbF (HbF_H: 8.9–9.2 g/dL, n = 3) groups and screened for candidate epigenetic modulators of HbF expression via WGBS and RNA-seq analysis. The clinical features and genetic analysis of the six affected β-thalassemia individuals and their pedigrees are presented in the supplemental information (Tables 1 and S1, Figure S1). All six subjects genotyped as β^0/β^0 and αα/αα and tested negative for deletional HPFH and δβ-thalassemia. Peripheral blood (PB)-derived $GYPA^+$ erythroblasts from 47 unrelated individuals with β-thalassemia comprising of low- and high-HbF groups (34 versus 13) were used as a validation set for the identified target gene through bisulfite pyrosequencing. To uncover the role of epigenetic mechanisms in regulating the reactivation of γ-globin expression, we analyzed the function of a target gene by using various *in vitro* assays with primary human $CD34^+$ HSPCs or HUDEP-2 cells and *in vivo* assays by engrafting of edited human $CD34^+$ HSPCs into immunodeficient NCG-Kit-V831M mice. To investigate the mechanistic role of ERF in repressing γ-globin, we performed chromatin immunoprecipitation sequencing (ChIP-seq), ChIP-quantitative real-time PCR, and dual luciferase reporter assays following genome editing. All studies included at least two to three biological replicates. Written informed consent was obtained from all the participants and/or their parents. Approval for the extended cohort study was obtained as outlined by the protocol #NFEC-2019-039 approved by Medical Ethics Committee of Nanfang Hospital of Southern Medical University. The study was conducted in accordance with the Declaration of Helsinki.

Genome editing by CRISPR/Cas9

To generate the *ERF*, *ZBTB7A*, *BCL11A*, UEBS, and DEBS knockout (KO) HUDEP-2 cells, we designed sgRNAs as previously described.[31] SgRNA oligonucleotides were cloned into the lenti-CRISPR vector (Addgene plasmid ID 52961) through restriction site *BsmB*I. Lentiviral transduction was performed as previously described.[32] To generate the *ERF*, UEBS, and DEBS KO $CD34^+$ HSPCs, we chemically modified sgRNAs as previously described[33] to optimize knockout efficiency. $CD34^+$ HSPCs were cultured in StemSpan SFEM medium supplemented with 100 ng/mL hTPO, 100 ng/mL Flt3l, and 100 ng/mL hSCF for 24 h prior to electroporation. Cells were incubated with 20 μg Cas9 (A36496, Thermo, USA) and 0.2 nmol sgRNA complex for 10 min at room temperature and electroporated via Lonza 4D Nucleofector (V4XP3032) according to the manufacturer's instructions. After the electroporation, $CD34^+$ HSPCs were subjected to differentiation as previously described.[34] On day 8, we harvested one-quarter of the cells

Table 1. The phenotypic and genotypic data in six subjects

	HbF_L group[a]			HbF_H group[a]			
Characteristics	**YH**	**DZ**	**JW[c]**	**JS**	**FN**	**JQ**	**p value[b]**
Age (years)/sex	7/M	7/M	9/M	10/F	15/M	26/M	0.158/–
Age of onset (months)	8	6	7	108	72	84	0.016
No. of transfusions/year	12	17	18	2[d]	1[d]	0	–
Hb (g/dL)[e]	7.6	4.5	4.4	9.2	9.3	9.7	0.063
HbF (%)[f]	1.3	8.9	2.3	98.9	95.7	94.8	0.000
HbA2 (%)	2.6	2.8	3.4	1.7	5.1	5.7	0.959
MCV (fL)	67.3	63.0	67.9	78.2	74.0	69.0	0.468
MCH (pg)	20.4	20.1	16.3	25.9	25.6	22.0	0.589
MCHC (g/dL)	33.3	33.7	34.6	31.3	34.7	31.8	0.341
HBB genotype[g]	c.[126_129 delCTTT]; [52A>T]	c.[126_129 delCTTT]; [92+1G>T]	c.[126_129 delCTTT]; [52A>T]	c.[84_85 insC]; [84_85insC]	c.[216_217 insA]; [52A>T]	c.[126_129 delCTTT]; [92+1G>T]	–
HBA genotype	αα/αα	αα/αα	αα/αα	αα/αα	αα/αα	αα/αα	–
Clinical diagnosis[h]	TM	TM	TM	TI	TI	TI	–
Methylation level (%)[i]	35.9	40.6	44.2	53.3	55.8	61.7	0.002
Modifier genes genotype							
BCL11A rs766432	AA	AA	AA	AC	AC	AC	–
HBS1L-MYB rs9399137	TT	TT	TT	TT	TT	TC	–
KLF1 mutation	WT	WT	WT	WT	WT	WT	–
HBG1 rs368698783	GG	GA	GG	AA	GG	GA	–
HBG2 rs7482144	CC	CT	CC	TT	CC	CT	–
Deletional HPFH	N	N	N	N	N	N	–

M, male; F, female; No., number; TM, β-thalassemia major; TI, β-thalassemia intermedia; Hb, hemoglobin; HbA2, hemoglobin A2; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; WT, normal; N, no deletional HPFH mutation. Dashes in the p value column denote no value. Diagnostic data and clinical details of these six individuals are included in the supplemental information.

[a]Physical examination showed hepatosplenomegaly in all six individuals.

[b]p values were determined by Student's t test.

[c]The individual died of severe anemia in 2015 and did not regularly receive blood transfusion.

[d]After the illness, they received blood transfusions one or two times.

[e]To more comprehensively reflect the endogenous situation of hematological parameters for our individuals, hematological data were collected before blood transfusion.

[f]HbF expressed in g/dL for individuals YH, DZ, JW, JS, FN, and JQ was as follow: 0.1, 0.4, 0.1, 9.1, 8.9, and 9.2 g/dL, respectively.

[g]*HBB*: c.126_129delCTTT, *HBB*: c.52A>T, *HBB*: c.216_217insA, *HBB*: c.92+1G>T, and *HBB*: c.84_85insC genotype categories are defined as β^0/β^0.

[h]The definition of TM or TI is based on the clinical indications, as described previously.[29,30]

[i]The mean methylation levels of the eight CpG sites on the *ERF* promoter calculated by bisulfite sequencing of BM-derived $GYPA^+$ cells from individuals.

for genomic DNA extraction to test the KO efficiency. We used the edited cells to perform quantitative real-time PCR and high-performance liquid chromatography (HPLC), as well as flow cytometry to measure γ-globin expression levels. ChIP assay was performed with *ERF*, UEBS, or DEBS KO HUDEP-2 cells.

We performed targeted CpG site methylation alteration in HUDEP-2 and $CD34^+$ HSPCs with a dCas9/sgRNA system to induce hypermethylation of the target CpG site in the *ERF* promoter, as previously described.[35] More details are described in the supplemental information.

Erythroid differentiation

Granulocyte colony-stimulating factor (G-CSF)-mobilized adult human $CD34^+$ HSPCs were isolated from the PB of unaffected individuals and separated with a CD34 microbead kit (Miltenyi Biotec, Germany). $CD34^+$ HSPCs were cultured according to a two-phase erythroid differential protocol.[34] HUDEP-2 cells were cultured as previously described.[36] To induce differentiation, we maintained cells in StemSpan SFEM supplemented with FBS (10%), Dox (1 μg/mL), EPO (3 IU/mL), and hSCF (100 ng/mL) for 5 days and then transferred them into StemSpan SFEM supplemented with FBS (30%), Dox (1 μg/mL), EPO (3 IU/mL), and hSCF (100 ng/mL) for 6 days.

Statistical analysis

The data analysis of WGBS and RNA-seq were performed with the BSseq and the NOIseq package, respectively. Details are provided in the supplemental information. A two-tailed Student's t test and ANOVA from SPSS v.20 software were used for comparisons between the indicated groups studied. Data are shown as mean ± standard deviation (SD). p values of less than 0.05 were considered to be statistically significant.

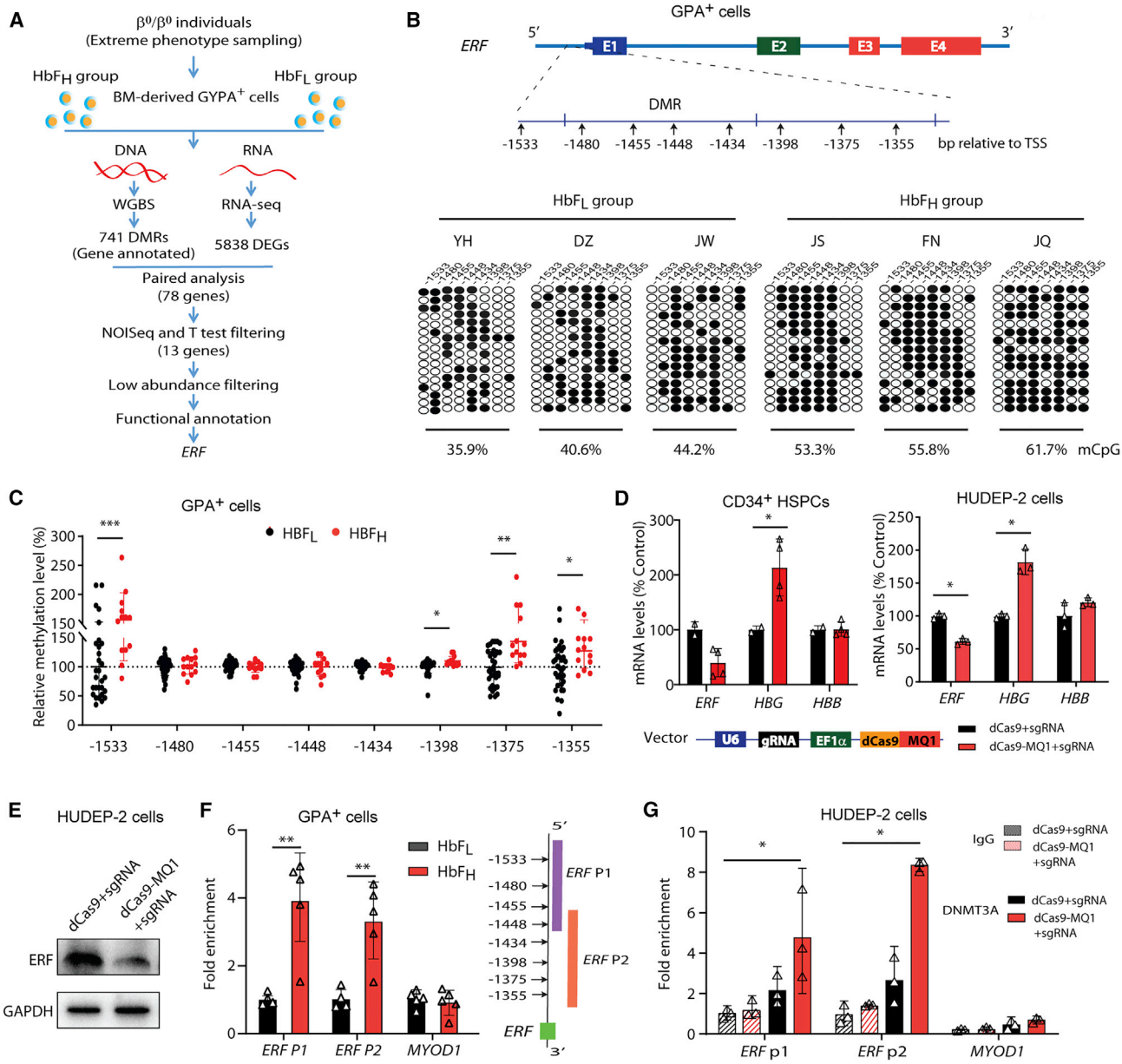


Figure 1. Identification of ERF as an HbF repressor through transcriptomic and methylation studies in β-thalassemia-affected individuals

(A) A flowchart for the screening of candidate modulators of HbF. The threshold for a positive DMR is defined as differential methylation level ≥ 10% (see the supplemental materials). We screened 741 out of the top 3,000 DMRs (top 3,000 hyper-DMR or hypo-DMR), which are annotated to be located in the promoter regions. 5,835 DEGs were identified according to the following criteria: $|\log_2 FC| > 0.8$ and divergence probability > 0.8 via NOISeq. 78 genes were enrolled in the candidate list as they presented significantly differential expression in both mRNA and methylation level between HbF_L and HbF_H groups. We further screened 13 out of the 78 DE-mRNAs ($|\log_2 FC| > 0.8$ and divergence probability > 0.8 in the NOISeq analysis and $p < 0.1$ in the further Student's t test) containing DMRs in their promoter regions, from which we prioritized six candidate genes by filtering out the low abundance with FPKM < 1.0 in either the HbF_H or HbF_L group. We finally performed functional annotation and identified *ERF* as the candidate gene with top priority for functional validation.

(B) Top: schematic of *ERF* gene body. Bottom: analysis of methylation levels at CpG sites (indicated by the distance relative to the transcription start site, TSS) within the DMR in the *ERF* promoter, evaluated by sequencing. Data were generated with BM-derived $GYPA^+$ cells from individuals with β^0-thalassemia (HbF_H: n = 3 or HbF_L: n = 3) and used in WGBS assays. Each row of eight CpG sites within a group represents a single bisulfite-treated clone with methylated CpGs (●) or unmethylated CpGs (○).

(C) Quantitative measurement of methylation levels of the *ERF* promoter by bisulfite pyrosequencing in an independent cohort of 47 samples with β-thalassemia (HbF_H: n = 13 or HbF_L: n = 34). Methylation percentage of each subject relative to the mean level for the subjects with HbF_L (the horizontal dashed line) is shown for each CpG site.

(D) Effects of *ERF* promoter hypermethylation by dCas9-MQ1/sgRNA on *ERF*, *HBG*, and *HBB* expression levels in $CD34^+$ HSPCs (left) and HUDEP-2 cells (right). The schematic of dCas9-MQ1-sgRNA assay is shown in the bottom. MQ1 was methyltransferase that mediated

(legend continued on next page)

Results

Epigenetic dysregulation of *ERF* is associated with HbF levels in β-thalassemia

To identify potential epigenetic modulators of HbF in β-thalassemia, we integrated the WGBS and RNA-seq data of the BM-derived GYPA$^+$ erythroid cells from six β-thalassemia-affected individuals with widely varying levels of HbF, designated as HbF_H (HbF ≥ 95th percentile) and HbF_L (HbF ≤ 5th percentile) groups (three subjects in each group; Tables 1 and S1, Figures 1A and S1). We identified 13 candidate genes among 78 differentially expressed genes (DEGs) with differentially methylated regions (DMRs) in their promoter regions that displayed either hypermethylation and downregulation or hypomethylation and upregulation in the HbF_H group compared with the HbF_L group (Figures 1A and S2A–S2C). After ruling out seven genes because of their low abundance (FPKM < 1), we focused on ETS2 repressor factor (ERF), which was a transcription factor among the six remaining genes (Table S2). ERF has recently been shown to be required for effective primitive and definitive hematopoiesis in mice.[37] Moreover, we noted that most of the eight CpG sites within the *ERF* promoter were hypermethylated in all three individuals in the HbF_H group from WGBS data (Figure S2D). We subsequently validated this result with a bisulfite cloning method (Figure 1B), suggesting a possible association of *ERF* promoter hypermethylation with an elevated HbF level.

Furthermore, we performed bisulfite pyrosequencing analysis of DNA from PB GYPA$^+$ cells in an independent cohort of 47 randomly chosen samples of β-thalassemia (Table S3) and noted that four out of the eight CpG sites in the *ERF* promoter displayed hypermethylation in the HbF_H groups compared with the HbF_L groups (Figure 1C). Regression analysis showed that methylation levels at three (−1,355, −1,375, and −1,533) of these four CpG sites of the *ERF* promoter are significantly associated with HbF levels (Figure S3). Moreover, *in vitro* study confirmed that the *ERF* promoter hypermethylation is responsible for its low expressed mRNA (Figure 1D) and protein level (Figure 1E) as determined by site-specific methylation through the dCas9-MQ1-sgRNA system,[35] which targeted the eight CpG sites within the *ERF* promoter in CD34$^+$ HSPCs and HUDEP-2 cells (Figures S2E and S2F), leading to elevated γ-globin production (Figure 1D). These findings imply that hypermethylation-mediated *ERF* downregulation is associated with higher levels of HbF expression in β-thalassemia-affected individuals. Furthermore, enrichment of DNA methyltransferase 3A (DNMT3A) at the *ERF* promoter was significantly enhanced in the HbF_H group compared with the HbF_L group (Figure 1F), which was also noted in the *ERF* promoter site-specific methylated HUDEP-2 cells (Figure 1G). These results imply that the epigenetic silencing of ERF in the HbF_H group is associated with DNMT3A-mediated hypermethylation of the *ERF* promoter.

ERF depletion elevates γ-globin expression *in vitro* in human CD34$^+$ HSPCs and HUDEP-2 cells

We then used CRISPR/Cas9 editing to generate *ERF* KO human CD34$^+$ HSPCs for functional assays. Electroporation Cas9/sgRNA RNP into CD34$^+$ HSPCs from a healthy donor[19] resulted in a 90% KO efficiency on *ERF* (Figure S4A). We noted that the proportion of γ-globin mRNA as a percentage of total globin transcripts as assessed by quantitative real-time PCR was significantly increased (increasing from mean of 1.8% to 28.9%) in the *ERF*-depleted CD34$^+$ HSPCs compared with the controls (Figure 2A), while HbF level was increased from mean 1.2% to 13.1% as assessed by HPLC (Figure 2B) and from mean 0.529% to 25.0% as assessed by flow cytometry assay for measurement of the percentage of HbF-positive cells (Figure 2C) as a consequence of *ERF* depletion. We further validated the ERF function in *ERF* KO HUDEP-2 cells (Figures S4B and S4C) and observed that five independent single *ERF* KO HUDEP-2 clones (Figures 2D and S5) displayed consistent increased HbF levels as assessed by HPLC (Figure 2E) and by flow cytometry (Figure 2F). Intriguingly, HbA level was observed to be decreased in *ERF* KO CD34$^+$ HSPCs (Figure 2B) and HUDEP-2 cells (Figure 2E) as assessed by HPLC. Moreover, *ERF* knockdown (KD) and overexpression (OE) in HUDEP-2 cells and CD34$^+$ HSPCs also confirmed the role of ERF as an HbF repressor (Figure S6).

Validation of ERF as an HbF repressor via engraftment assays in immunodeficient mice

To test the impact of *ERF* editing on CD34$^+$ HSPCs function, we performed *in vivo* transplantation experiments by engrafting edited *ERF*-deleted CD34$^+$ HSPCs into immunodeficient NCG-Kit-V831M mice (Figure 3A), an experimental system extensively used to study stem cell function.[38] Compared to unedited controls, *ERF* deletion did not affect either lymphoid (mean 42.3% for unedited versus mean 42.55% for edited) or myeloid (mean 50.37% for unedited versus mean 49.6% for edited) differentiation of the CD34$^+$ in the BM 16 weeks following

hypermethylation in the target DNA region according to the guide of sgRNA. Two sgRNAs were designed to cover the region of eight CpG sites in the *ERF* promoter.
(E) Immunoblotting analysis in control and *ERF* promoter-targeted methylated HUDEP-2 cells. GAPDH served as a loading control.
(F) ChIP-quantitative real-time PCR assays performed with DNMT3A antibody in GYPA$^+$ cells from the HbF_H group (n = 5) and HbF_L group (n = 5). *ERF* P1 and P2 covered the region of eight CpG sites in the *ERF* promoter, as shown in the right panel.
(G) ChIP-quantitative real-time PCR assays of DNMT3A performed in control (dCas9-sgRNA) and methylated (dCas9-MQ1-sgRNA) HUDEP-2 cells. *MYOD1* served as the negative control. Data from ≥3 independent experiments are presented as means ± SD (*$p < 0.05$; **$p < 0.01$).

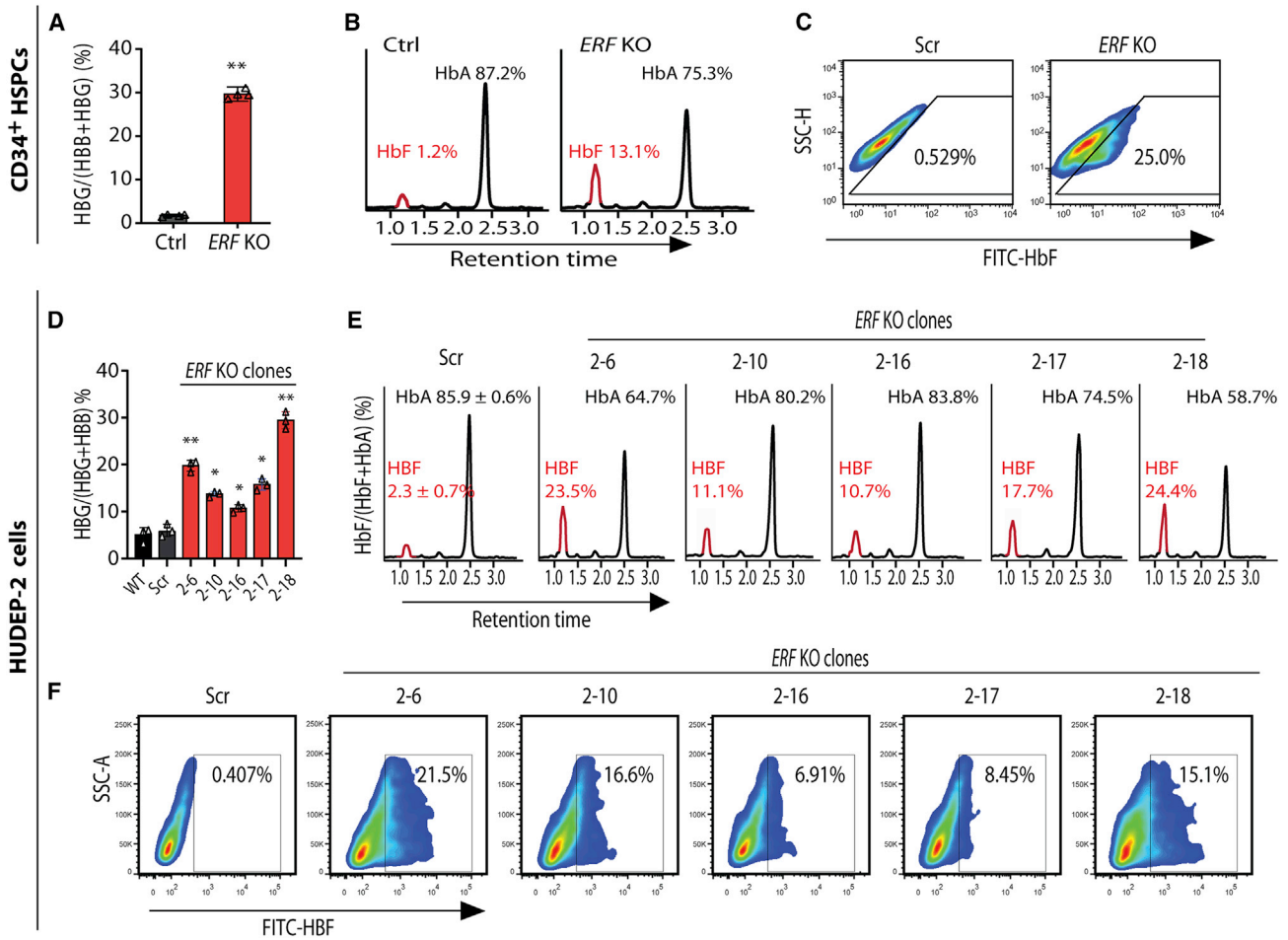


Figure 2. ***ERF* depletion elevates γ-globin expression *in vitro***
(A–C) Quantitative measurement of *HBG* mRNA expression by quantitative real-time PCR (A) and HbF or HbA production by high performance liquid chromatography (HPLC) (B) and by flow cytometry analysis with FITC-conjugated anti-HbF antibody (C) in control (Ctrl or Scr, the non-edited control that has been subject to the same processes as the experimental lines without editing) and *ERF* KO CD34$^+$ HSPCs.
(D–F) Quantitative measurement of *HBG* mRNA expression by quantitative real-time PCR (D) and HbF or HbA production by HPLC (E) and by flow cytometry analysis (F) in wild type (WT), Scr, and five independent single *ERF* KO HUDEP-2 clones. Data from ≥3 independent experiments are presented as means ± SD (*p < 0.05; **p < 0.01).

HSPCs transplantation (Figure 3B). However, a modest decline of erythroid cells was noted for the edited group compared with the control group (Figure 3B). These findings are consistent with previous reports of impaired repopulation ability and quantitative differences in the differentiation of hematopoietic stem cells in the *Erf*-deficient mice.[37]

We further noted that the variation of indel frequencies in the engrafted cells from edited mice were modest, 92% to 96% in BM cells and 85% to 90% in PB cells (Figure 3C), and that the indel frequencies in edited cells were consistent with that of input cells (88%), suggesting a high editing efficiency (Figure S7). Most importantly, in BM cells, we observed significant induction of γ-globin in edited cells (mean of 22.4% of total β-like globin in edited cells compared with mean of 6.9% in unedited cells; Figure 3D). These findings provided critical *in vivo* evidence for the role of *ERF* in repressing γ-globin synthesis.

ERF represses γ-globin synthesis by direct binding to the motifs regulating *HBG* expression

To define the mechanistic basis for ERF in repressing γ-globin, we performed RNA-seq analysis of *ERF* KO HUDEP-2 cells to evaluate genome-wide gene expression changes induced by *ERF* depletion. Gene expression profiling and quantitative real-time PCR analysis demonstrated that depletion or overexpression of *ERF* did not affect the expression of *BCL11A*, *ZBTB7A*, *KLF1*, and *MYB* (Figures S8A–S8D). Gene Ontology (GO) analysis of differentially expressed genes induced by *ERF* KO in HUDEP-2 cells was not evident in gene expression signatures relevant to known erythroid transcription factors (Figure S8E). These results suggest that HbF silencing mediated by ERF is independent of previously defined erythroid repressors or complexes. Subsequent, *in vitro* study indicated that *ERF/ZBTB7A* or *ERF/BCL11A* double-knockout (DKO) cells exhibited significantly higher γ-globin

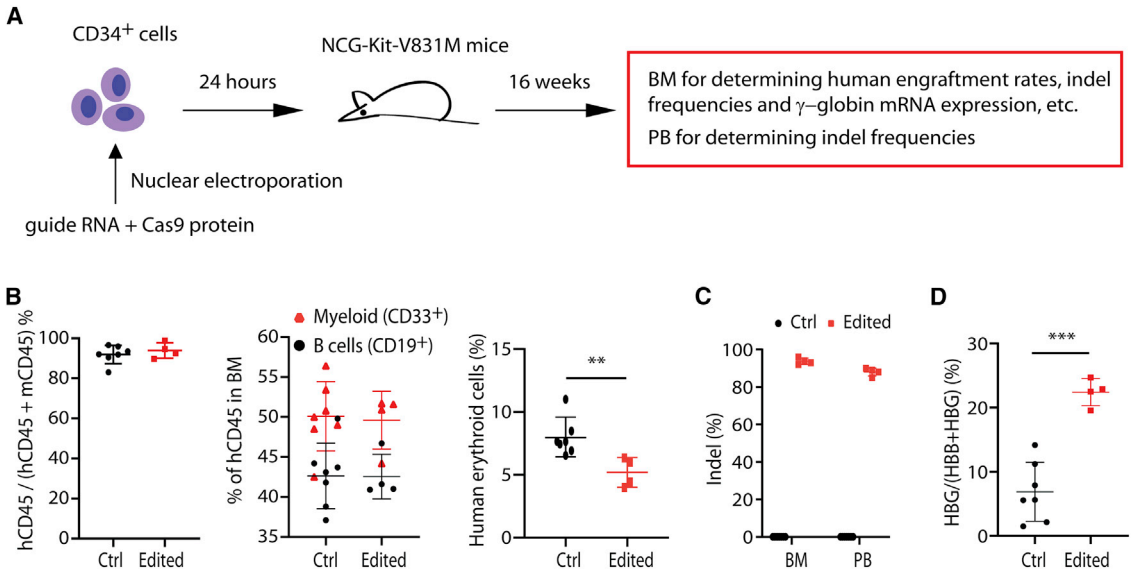


Figure 3. ***ERF* depletion elevates γ-globin expression *in vivo***

(A) The experimental design for *in vivo* functional validations of *ERF*. *ERF*-targeted gRNA and Cas9 protein were electroporated into CD34$^+$ HSPCs and after 24 h engrafted into immunodeficient mice by intravenous tail injection. Bone marrow (BM) and peripheral blood (PB) cells were harvested at week 16 after engraftment for further analysis.

(B) Flow cytometry analysis in mouse BM 16 weeks after transplantation for determination of human engraftment rates (left), human cell multilineage reconstitution (myeloid and B cells, middle), and human erythroid cells (right).

(C) Determination of the indel frequencies by Synthego analysis after sequencing of PCR products in PB and BM from engrafted mice.

(D) Measurement of *HBG* mRNA expression by quantitative real-time PCR in mouse BM 16 weeks after engraftment (n = 4 for edited mice and n = 7 for unedited mouse controls). Data are presented as means ± SD (**p < 0.01; ***p < 0.001).

expression than *ZBTB7A*, *BCL11A*, or *ERF* single-KO cells (Figures S8F and S8G). This result supports that ERF does not regulate HbF through known HbF modifiers.

We then performed ChIP-seq by using anti-ERF antibody to identify potential ERF-binding sites within the human β-globin gene cluster in HUDEP-2 cells. Intriguingly, we observed two significant ERF-binding signals at 3.6 kb upstream of *HBG2* (termed as upstream ERF-binding site, UEBS) and at 1.5 kb downstream (termed as downstream ERF-binding site, DEBS) of *HBG1* (Figure 4A, track 1). Intriguingly, UEBS and DEBS are accessible for several known transcription factors and histone modification markers based on public ENCODE ChIP-seq datasets (Figure S9).[39,40] Moreover, ChIP-seq results obtained from *ERF* KO HUDEP-2 cells showed a consistent absence of peaks in UEBS and DEBS (Figure 4A, track 3), which was further confirmed by ChIP-quantitative real-time PCR with anti-ERF antibody in CD34$^+$ HSPCs (Figure 4B) and in *ERF* KO HUDEP-2 cells (Figure 4C). We also explored whether LRF bound to these two *cis*-elements to coordinately mediate the silencing of γ-globin genes. ChIP-quantitative real-time PCR results demonstrated LRF showed an independent binding pattern (Figure 4D), consistent with previous studies.[11] We further demonstrated that overexpression of *ERF* inhibited the promoter activity of the luciferase reporter containing UEBS or DEBS in HUDEP-2 cells (Figure 4E).

To evaluate the effects of these two *cis*-elements on regulation of HbF expression, we disrupted the UEBS or DEBS motifs by using Cas9 system in HUDEP-2 cells and in CD34$^+$ HSPCs (Figure S10A). We carried out copy number analysis on β-globin clusters to rule out the off-target effects that could potentially be caused by the highly duplicated nature of *HBG1* and *HBG2* and found both *HBG*s and their intergenic regions to be intact except for the ERF-binding sites (Figures S10A–S10C). We established independent UEBS or DEBS KO HUDEP-2 clones (Figures S10D and S10E) and determined the *HBG* mRNA level and HbF levels in these cells. We observed significant increases of expressed *HBG* mRNA (UEBS: mean 20.1% of total β-like globin from four clones; DEBS: mean 16.2% from two clones) and of HbF protein levels (UEBS: mean 21.0% of total hemoglobin; DEBS: mean 19.5%) in edited HUDEP-2 cells (Figures 5A–5C). Similar increases in *HBG*-expressed mRNA and HbF level were also observed in edited CD34$^+$ HSPCs (expressed mRNA: mean 14.4% and 16.5% in UEBS KO and DEBS KO cells, respectively, and HbF level: mean 17.7% and 16.9% in UEBS KO and DEBS KO cells, respectively) (Figures 5D, S10F, and S10G). The observed changes in *HBG* mRNA expression in various gene-edited HUDEP-2 and CD34$^+$ HSPCs by disruption of *ERF* and/or of the two binding sites (Figure 5E), as well as absence of ERF-binding activity in UEBS and DEBS KO HUDEP-2 cells (Figure 5F), confirm an essential role of both ERF and its binding motifs as *cis*-elements in regulation of HbF expression.

To uncover the underlying mechanism of the inhibition of γ-globin genes induced by the epigenetic silencing of *ERF*, we evaluated the methylation levels of *HBB* and *HBG* promoters in the GYPA$^+$ cells from the same cohort

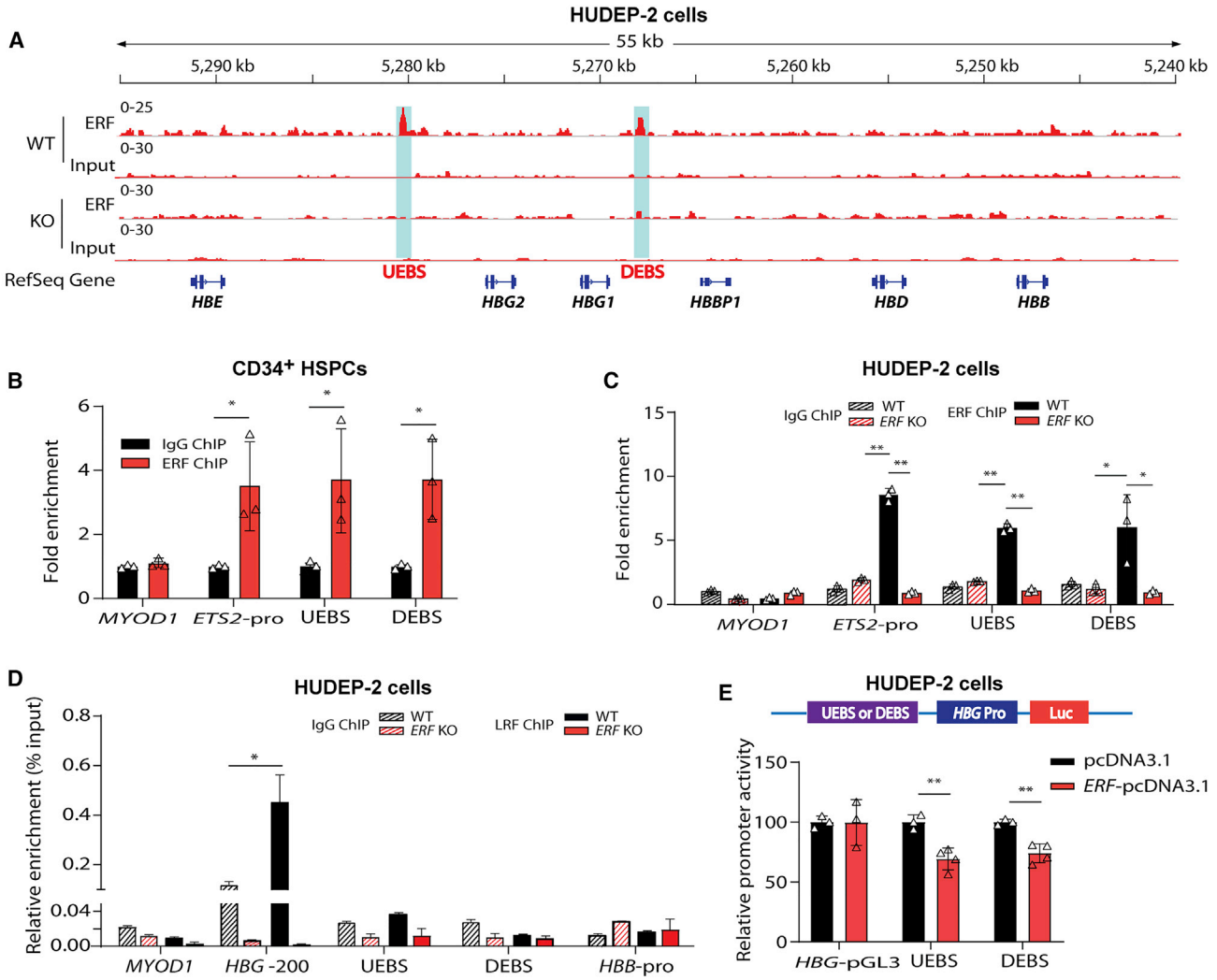


Figure 4. Identification of ERF-binding sites in β-globin gene clusters

(A) ChIP-seq binding patterns from HUDEP-2 cells with anti-ERF antibody at the β-globin cluster. Two ERF-binding positive signals 3.6 kb upstream of *HBG2* (named UEBS) and 1.5 kb downstream of *HBG1* (named DEBS) are marked by the light blue shadows.

(B) Detection of ERF-binding sites by ChIP-quantitative real-time PCR in human CD34$^+$ HSPCs.

(C and D) Detection of ERF-binding (C) or LRF-binding (D) activity in *ERF* KO HUDEP-2 cells (clone #2–6). IgG served as the negative control for ChIP. *MYOD1* and *HBB*-pro served as the negative control for quantitative real-time PCR. *ETS2*-pro served as the positive control for quantitative real-time PCR. *HBG*-200, −200 bp upstream of the TSS in the *HBG* promoter, served as the positive control for LRF binding.

(E) Top: the schematic of dual luciferase reporter assay. Bottom: effects of UEBS or DEBS as regulatory elements on *HBG* promoter activity in a pGL3 luciferase reporter in HUDEP-2 cells without (black) or with (bright red) ERF overexpression (OE). The data of each group were normalized to the control (pcDNA3.1). Data are presented as means ± SD (*$p < 0.05$; **$p < 0.01$).

of 47 β-thalassemia-affected individuals in the above pyrosequencing analysis and found that epigenetic silencing of *ERF* is significantly associated with general hypomethylation of the *HBG* promoter (Figure 5G). We validated this finding by knocking out *ERF* or targeting hypermethylation of *ERF* in CD34$^+$ HSPCs and found that the *ERF* depletion and promoter hypermethylation led to reactivation of HbF through hypomethylation of γ-globin genes promoter, typically in −162, −53, −50, +17, and +50 CpG sites (Figures 5H and S2G). In addition, ChIP-quantitative real-time PCR assays showed that the hypomethylation of γ-globin genes promoter induced by the depletion of *ERF* is likely to be mediated by impaired recruitment of DNMT3A on this CpG island (Figure 5I).

Knockout of *ERF* has no significant impact on the erythroid cell differentiation

Finally, we examined the effect of ERF on erythroid differentiation. We observed that *ERF* depletion exerts no significant impact on either erythroid maturation or terminal differentiation of CD34$^+$ HSPCs, as assessed by CD235a, CD49d, and CD233 surface expression (Figure 6A). Furthermore, the morphology, enucleation rate, and growth of the erythroid cells following *ERF* depletion was comparable to that of wild-type cells (Figures 6B, 6C, and S11A). Similar findings were noted following KO of UEBS or DEBS (Figure S11B). We also observed that depletion of *ERF*, DEBS, or UEBS exerts no significant impact on erythroid differentiation

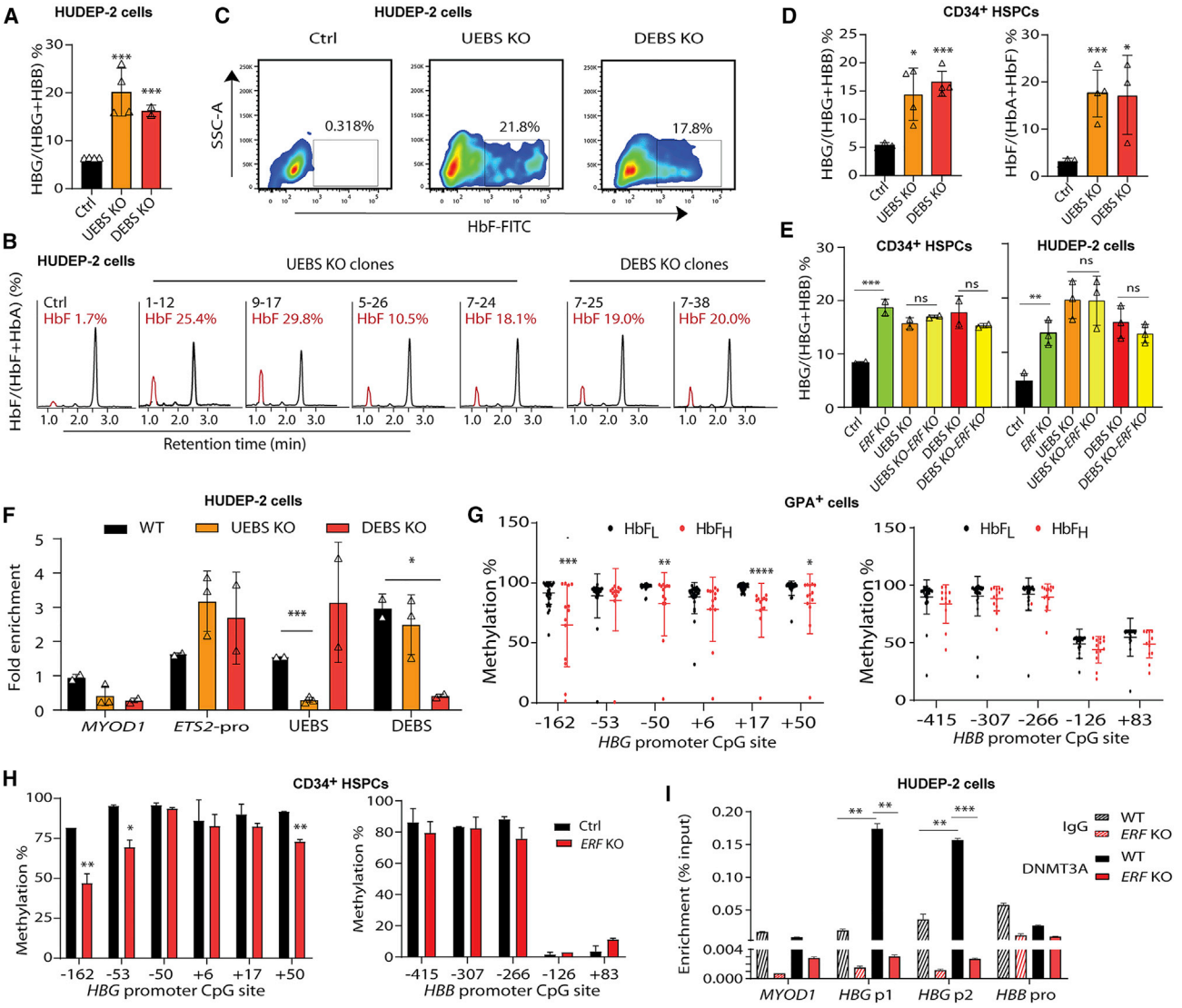


Figure 5. Knockout of ERF-binding sites led to elevation of HbF in HUDEP-2 and CD34$^+$ HSPCs

(A–C) Quantitative measurement of *ERF* mRNA expression by quantitative real-time PCR (A) and HbF production by HPLC (B) and by flow cytometry analysis (C) in UEBS/DEBS KO HUDEP-2 clones (n = 4 for UEBS and n = 2 for DEBS, each point indicates the mean value for each clone). The HPLC profiles of HbF from each of the six independent single UEBS/DEBS KO HUDEP-2 clones are shown.

(D) Quantitative measurement of *HBG* mRNA expression by quantitative real-time PCR (left) and HbF production by HPLC (right) in UEBS/DEBS KO CD34$^+$ HSPCs (editing efficiency: 30% for UEBS, 23% for DEBS).

(E) Quantitative measurement of *HBG* mRNA expression by quantitative real-time PCR in single or double KO of ERF and/or UEBS (UEBS KO-*ERF* KO) or DEBS (DEBS KO-*ERF* KO) CD34$^+$ HSPCs (left) and HUDEP-2 cells (right). The γ-globin expression levels were determined as a percentage of the total β-like globin level (*HBG*+*HBB*). One-way ANOVA was used for comparison of the indicated groups. *$p < 0.05$; ***$p < 0.01$. ns, non-significant ($p > 0.05$).

(F) Detection of ERF-binding sites by ChIP-quantitative real-time PCR in UEBS or DEBS KO HUEDP-2 clones (n = 3 and n = 2, respectively). Data are presented as mean ± SD for each clone.

(G) Quantitative measurement of methylation levels of the *HBG* (left) and *HBB* (right) promoter by bisulfite sequencing in the independent cohort of 47 samples with β-thalassemia (HbF_H: n = 13 or HbF_L: n = 34).

(H) Quantitative measurement of methylation levels of the *HBG* promoter (left) and *HBB* promoter (right) by bisulfite sequencing in the *ERF* KO CD34$^+$ HSPCs.

(I) ChIP-quantitative real-time PCR assay performed with DNMT3A antibody in *HBG* promoter in WT and *ERF* KO HUDEP-2 cells. HBG p1 and p2 covered the region of CpG sites from −162 to +50. *MYOD1* and *HBB* pro served as controls.

in HUDEP-2 cells (Figures S11C–S11F). Taken together, these findings suggest that ERF might be a promising therapeutic target for the reactivation of HbF without significantly altering the terminal erythroid differentiation.

Discussion

It is important to identity genome-wide epigenetic modulators that account for phenotypic diversities in a Mendelian disorder. In the present study, we elucidated the epigenetic

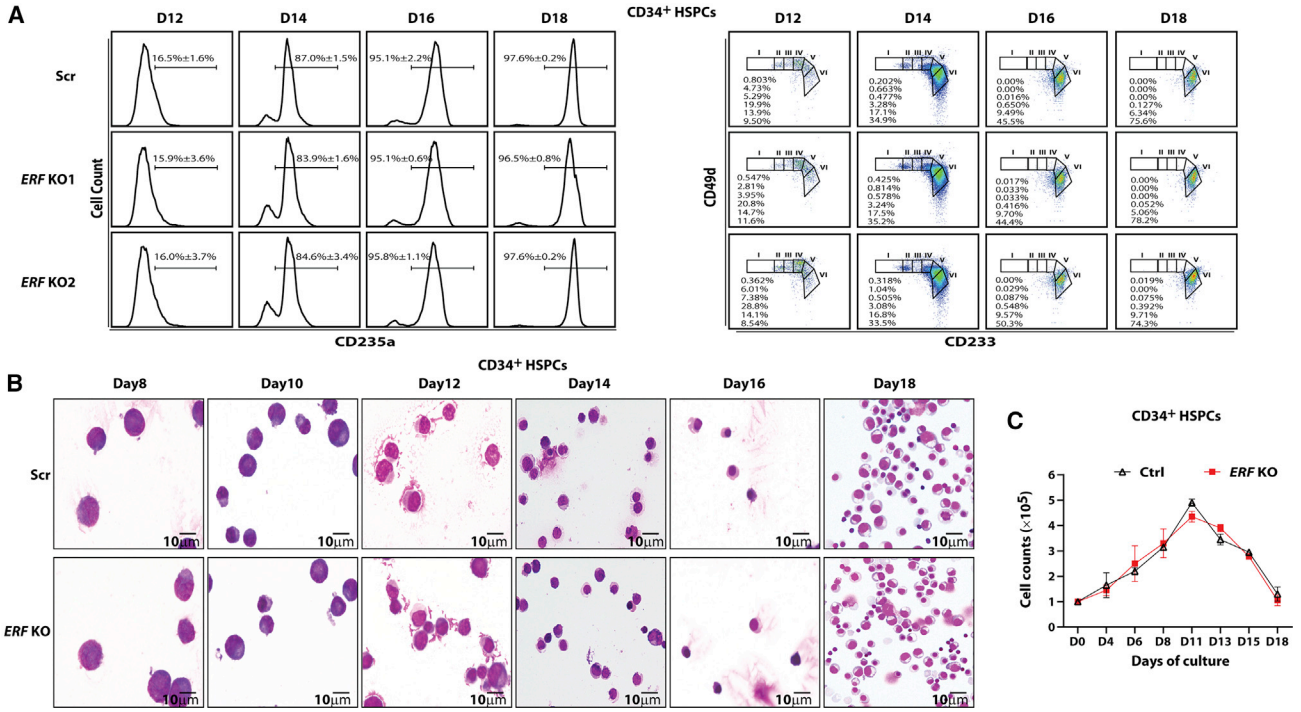


Figure 6. The impact of ERF on the differentiation of CD34$^+$ HSPCs

(A) Left: the expression of CD235a on different time points of culture of CD34$^+$ HSPCs in percentage (mean ± SD, n = 3). Right: terminal erythroid differentiation was examined on indicated days by flow cytometric analysis based on the expression of CD233 and CD49d. Representative plots of CD233 versus CD49d of CD235a$^+$ cells are shown, and the erythroblasts are separated into six populations: pro-erythroblasts (I), early baso erythroblasts (II), late baso erythroblasts (III), polychromatic (IV), early orthochromatic (V), and late orthochromatic (VI).

(B) Representative images of Wright-Giemsa staining of cytospins in different time points of differentiated CD34$^+$ cells (objective lens, 100×).

(C) Cell growth curves of control and *ERF* KO CD34$^+$ cells (mean ± SD, n = 2).

dysregulation of fetal-to-adult Hb switch in β-thalassemia and discovered epigenetic modulators associated with HbF levels. This strategy based on methylomic and transcriptomic profiling enabled us to identify regulators that cannot be readily detected by conventional GWASs because of the high conservation of their genomic sequences or alterations by epigenetic mechanisms.

Integrative studies of RNA-seq and WGBS led to the identification of candidate genes that were significant in both DEG and DMR analysis. Of the 13 candidate genes identified, ERF is a transcription factor that was able to regulate its target genes through a direct binding on its distant regulatory elements.[41,42] ERF was previously reported to be expressed in hematopoietic stem cells and progenitors and displayed a dynamic expression pattern during erythroid differentiation; its expression increased from HSC to proerythroblast stage with subsequent decreased expression at late stages of terminal erythroid differentiation (Figure S12).[43,44] Interestingly, previous studies showed that ERF acted as a transcription repressor indispensable for erythroid differentiation.[42] Importantly, comprehensive *in vitro* and *in vivo* studies enabled us to document that downregulation of *ERF* in erythroid cells led to significant elevation of HbF, validating the hypothesis that ERF is a transcription repressor of γ-globin genes and its epigenetic downregulation results in reactivation of HbF. By systematically screening methylation states in the promoters of *ERF*, *HBG1*, and *HBG2* in an independent and extended cohort of 47 β-thalassemia-affected individuals, we defined that hypermethylation of the *ERF* promoter had led to the haploinsufficiency of ERF, which eventually resulted in the hypomethylation of *HBG* promoters and re-activation of HbF (Figure 1D). This process of dysregulation in *ERF* and *HBG* promoters are mediated by altered recruitment of DNMT3A (Figures 1F, 1G, and 5I). The highly consistent findings from study of two independent cohorts imply that regulating hypermethylation of the *ERF* promoter represents a protective pathway to decrease ineffective erythropoiesis in β-thalassemia by increased expression of HbF.

In detailed analysis of the methylation states of the DMR, we noted that three of the four CpG sites (−1,533, −1,375, and −1,355) in the cohort of 47 β-thalassemia-affected individuals had statistically significant effects on HbF level. Interestingly, we observed that the differentially methylated sites detected in the extended cohort showed a different pattern compared with that of the six samples subjected to WGBS. This difference is probably due to the considerable dynamics of CpG islands at different developmental stages and different tissues and among different individuals. A single CpG site is also regarded as

a methylation variable position (MVP), in which a CpG site contributing to phenotypic alterations is functionally equivalent to a rSNP.[45] Therefore, the dysregulated methylation of diverse CpG sites within one single DMR can result in the downregulation of ERF among different individuals with high HbF levels. Future studies using a larger cohort of β-thalassemia-affected individuals should enable more systematic assessment of the heterogeneity of DNA methylation states among these CpG sites within the *ERF* promoter.

In screening for the known HbF modifier variants in the six β-thalassemia-affected individuals studied, we identified the three individuals in the HbF_H group to be carriers of rs766432 in BCL11A, a functional SNP favorable for elevated HbF levels. Our previous studies indicated that the contribution of this variant to the elevation of HbF is relatively mild compared with *XmnI* (GenBank: NC_000011.9, g.5276169G>A) and *KLF1* mutations in the Chinese population,[4,25] suggesting the existence of other epigenetic factors responsible for high levels of HbF in this group. To further exclude the potentially confounding effects caused by this variant, we used the same panel for genetic screening relevant to HbF-regulating variants in the extended cohort of 47 β-thalassemia-affected individuals and did not find significant differential distributions in the genotypes of *BCL11A*, *MYB-HBS1L*, *XmnI*, and *KLF1* between the HbF_L and HbF_H group, indicating that DMRs in the *ERF* promoter are independently associated with the elevation of HbF levels in β-thalassemia-affected individuals.

In terms of mechanistic understanding, we identified two ERF-binding sites, UEBS and DEBS, specifically located upstream of *HBG2* and downstream of *HBG1*, respectively, that are functionally relevant (Figure 1D). A series of *in vitro* and *in vivo* studies demonstrate that depletion of *ERF* led to a considerable increase in γ-globin expression, while the expression of β-globin appeared to decrease proportionally (Figures 2). In addition, RNA-seq data of *ERF* KO HUDEP-2 cell lines indicated that *ERF* depletion does not lead to dramatic changes in the known key erythroid genes (Figures S8B–S8E). Moreover, we demonstrated that depletion of *ERF* attenuated recruitment of DNMT3A to the *HBG* promoter (Figure 5I), which resulted in demethylation-mediated *HBG* transcriptional activation. Intriguingly, we also observed that UEBS is associated with recruitments of the active chromatin marker H3K27ac or its deacetylase HDAC based on ENCODE ChIP-seq dataset (Figure S9),[39,40] suggesting that ERF binding might interfere with H3K27 acetylation-mediated positive regulation through recruitment of HDAC or DNMT3A. Taken together, ERF acts as a specific repressor of γ-globin genes through an independent pathway of epigenetic regulation.

In summary, we developed a strategy for the identification of epigenetic modulators of HbF expression through extensive phenotype sampling (EPS) followed by transcriptomic and methylomic profiling. Through a series of functional validation studies, we identified an epigenetic pathway for silencing of γ-globin gene expression. This pathway initiated by hypermethylation of the *ERF* promoter mediated by enrichment of DNMT3A leads to hypomethylation of γ-globin genes and impaired binding of ERF on the *HBG* promoter and eventually re-activation of HbF in β-thalassemia-affected individuals. Moreover, we found that *ERF* knockout does not lead to observable changes in either erythroid differentiation or enucleation rate of human primary $CD34^+$ HSPCs and HUDEP-2 cells. Given the high efficiency of site-specific methylation via the current dCas9-MQ1-sgRNA system, the *ERF* promoter-specific methylation might be a promising target for genetic therapies of β-hemoglobinopathies.

Data and code availability

All the data generated in this study were shown in the main text and supplemental information. The data of RNA-seq in *ERF* KO and ChIP-seq are available in GEO with the number GEO: GSE160603. The source data underlying Figures 1A and S2A–S2D are provided in Table S6. Additional information is also available upon reasonable request to the corresponding authors.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.03.005.

Acknowledgments

We thank the individuals for their willingness to participate in this study; Xinping Yang, Shiqi Jia, Xichen Bao, Erwei Song, Yuxuan Wu, and Bingtao Hao for valuable advice and comments regarding this work; and Yi Wu, Feijin Chen, and Dun Liu and colleagues for assistance in collecting samples from individuals and controls. We thank the Central Laboratory of Southern Medical University for providing pyrosequencing reagents and platform. Financial supports from the National Key R&D Program of China (2018YFA0507800 and 2018YFA0507803), the Guangdong Science and Technology Foundation (2019B030316032), and NIH (DK32094) (N.M.) are gratefully acknowledged.

Declaration of interests

The authors declare no competing interests.

Received: December 27, 2020
Accepted: March 1, 2021
Published: March 17, 2021

Web resources

ENCODE, https://www.encodeproject.org/
HUGO Gene Nomenclature Committee, https://www.genenames.org/
OMIM, https://omim.org/

References

1. Vinjamur, D.S., Bauer, D.E., and Orkin, S.H. (2018). Recent progress in understanding and manipulating haemoglobin

switching for the haemoglobinopathies. Br. J. Haematol. *180*, 630–643.
2. Wienert, B., Martyn, G.E., Funnell, A.P.W., Quinlan, K.G.R., and Crossley, M. (2018). Wake-up Sleepy Gene: Reactivating Fetal Globin for β-Hemoglobinopathies. Trends Genet. *34*, 927–940.
3. Mettananda, S., and Higgs, D.R. (2018). Molecular Basis and Genetic Modifiers of Thalassemia. Hematol. Oncol. Clin. North Am. *32*, 177–191.
4. Liu, D., Zhang, X., Yu, L., Cai, R., Ma, X., Zheng, C., Zhou, Y., Liu, Q., Wei, X., Lin, L., et al. (2014). KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of β-thalassemia. Blood *124*, 803–811.
5. Chondrou, V., Kolovos, P., Sgourou, A., Kourakli, A., Pavlidaki, A., Kastrinou, V., John, A., Symeonidis, A., Ali, B.R., Papachatzopoulou, A., et al. (2017). Whole transcriptome analysis of human erythropoietic cells during ontogenesis suggests a role of VEGFA gene as modulator of fetal hemoglobin and pharmacogenomic biomarker of treatment response to hydroxyurea in β-type hemoglobinopathy patients. Hum. Genomics *11*, 24.
6. Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H., et al. (2007). A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat. Genet. *39*, 1197–1199.
7. Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science *322*, 1839–1842.
8. Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V.G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G., et al. (2008). Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc. Natl. Acad. Sci. USA *105*, 1620–1625.
9. Masuda, T., Wang, X., Maeda, M., Canver, M.C., Sher, F., Funnell, A.P., 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.
10. Liu, N., Hargreaves, V.V., Zhu, Q., Kurland, J.V., Hong, J., Kim, W., Sher, F., Macias-Trevino, C., Rogers, J.M., Kurita, R., et al. (2018). Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. Cell *173*, 430–442.e17.
11. Martyn, G.E., Wienert, B., Yang, L., Shah, M., Norton, L.J., Burdach, J., Kurita, R., Nakamura, Y., Pearson, R.C.M., Funnell, A.P.W., et al. (2018). Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. Nat. Genet. *50*, 498–503.
12. Chen, J.J. (2007). Regulation of protein synthesis by the heme-regulated eIF2alpha kinase: relevance to anemias. Blood *109*, 2693–2699.
13. Grevet, J.D., Lan, X., Hamagami, N., Edwards, C.R., Sankaranarayanan, L., Ji, X., Bhardwaj, S.K., Face, C.J., Posocco, D.F., Abdulmalik, O., et al. (2018). Domain-focused CRISPR screen identifies HRI as a fetal hemoglobin regulator in human erythroid cells. Science *361*, 285–290.
14. Wienert, B., Funnell, A.P.W., Norton, L.J., Pearson, R.C.M., Wilkinson-White, L.E., Lester, K., Vadolas, J., Porteus, M.H., Matthews, J.M., Quinlan, K.G.R., and Crossley, M. (2015). Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. Nat. Commun. *6*, 7085.
15. Wienert, B., Martyn, G.E., Kurita, R., Nakamura, Y., Quinlan, K.G.R., and Crossley, M. (2017). KLF1 drives the expression of fetal hemoglobin in British HPFH. Blood *130*, 803–807.
16. Martyn, G.E., Wienert, B., Kurita, R., Nakamura, Y., Quinlan, K.G.R., and Crossley, M. (2019). A natural regulatory mutation in the proximal promoter elevates fetal *globin* expression by creating a de novo GATA1 site. Blood *133*, 852–856.
17. Traxler, E.A., Yao, Y., Wang, Y.D., Woodard, K.J., Kurita, R., Nakamura, Y., Hughes, J.R., Hardison, R.C., Blobel, G.A., Li, C., and Weiss, M.J. (2016). A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. Nat. Med. *22*, 987–990.
18. Wu, Y., Zeng, J., Roscoe, B.P., Liu, P., Yao, Q., Lazzarotto, C.R., Clement, K., Cole, M.A., Luk, K., Baricordi, C., et al. (2019). Highly efficient therapeutic gene editing of human hematopoietic stem cells. Nat. Med. *25*, 776–783.
19. Wang, L., Li, L., Ma, Y., Hu, H., Li, Q., Yang, Y., Liu, W., Yin, S., Li, W., Fu, B., et al. (2020). Reactivation of γ-globin expression through Cas9 or base editor to treat β-hemoglobinopathies. Cell Res. *30*, 276–278.
20. Humbert, O., Radtke, S., Samuelson, C., Carrillo, R.R., Perez, A.M., Reddy, S.S., Lux, C., Pattabhi, S., Schefter, L.E., Negre, O., et al. (2019). Therapeutically relevant engraftment of a CRISPR-Cas9-edited HSC-enriched population with HbF reactivation in nonhuman primates. Sci. Transl. Med. *11*, eaaw3768.
21. Gnanapragasam, M.N., Scarsdale, J.N., Amaya, M.L., Webb, H.D., Desai, M.A., Walavalkar, N.M., Wang, S.Z., Zu Zhu, S., Ginder, G.D., and Williams, D.C., Jr. (2011). p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. Proc. Natl. Acad. Sci. USA *108*, 7487–7492.
22. Roosjen, M., McColl, B., Kao, B., Gearing, L.J., Blewitt, M.E., and Vadolas, J. (2014). Transcriptional regulators Myb and BCL11A interplay with DNA methyltransferase 1 in developmental silencing of embryonic and fetal β-like globin genes. FASEB J. *28*, 1610–1620.
23. Renneville, A., Van Galen, P., Canver, M.C., McConkey, M., Krill-Burger, J.M., Dorfman, D.M., Holson, E.B., Bernstein, B.E., Orkin, S.H., Bauer, D.E., and Ebert, B.L. (2015). EHMT1 and EHMT2 inhibition induces fetal hemoglobin expression. Blood *126*, 1930–1939.
24. Basak, A., Munschauer, M., Lareau, C.A., Montbleau, K.E., Ulirsch, J.C., Hartigan, C.R., Schenone, M., Lian, J., Wang, Y., Huang, Y., et al. (2020). Control of human hemoglobin switching by LIN28B-mediated regulation of BCL11A translation. Nat. Genet. *52*, 138–145.
25. Chen, D., Zuo, Y., Zhang, X., Ye, Y., Bao, X., Huang, H., Tepakhan, W., Wang, L., Ju, J., Chen, G., et al. (2017). A Genetic Variant Ameliorates β-Thalassemia Severity by Epigenetic-Mediated Elevation of Human Fetal Hemoglobin Expression. Am. J. Hum. Genet. *101*, 130–138.
26. Gong, Y., Zhang, X., Zhang, Q., Zhang, Y., Ye, Y., Yu, W., Shao, C., Yan, T., Huang, J., Zhong, J., et al. (2020). A natural DNMT1 mutation elevates the fetal hemoglobin via epigenetic de-repression of gamma-globin gene in beta-thalassemia. Blood. Published online November 23, 2020. https://doi.org/10.1182/blood.2020006425.

27. Adelvand, P., Hamid, M., and Sardari, S. (2018). The intrinsic genetic and epigenetic regulator factors as therapeutic targets, and the effect on fetal globin gene expression. Expert Rev. Hematol. *11*, 71–81.
28. Sripichai, O., and Fucharoen, S. (2016). Fetal hemoglobin regulation in β-thalassemia: heterogeneity, modifiers and therapeutic approaches. Expert Rev. Hematol. *9*, 1129–1137.
29. Karimi, M., Cohan, N., De Sanctis, V., Mallat, N.S., and Taher, A. (2014). Guidelines for diagnosis and management of Beta-thalassemia intermedia. Pediatr. Hematol. Oncol. *31*, 583–596.
30. Galanello, R., and Origa, R. (2010). Beta-thalassemia. Orphanet J. Rare Dis. *5*, 11.
31. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.
32. Ishikawa, Y., Maeda, M., Pasham, M., Aguet, F., Tacheva-Grigorova, S.K., Masuda, T., Yi, H., Lee, S.U., Xu, J., Teruya-Feldstein, J., et al. (2015). Role of the clathrin adaptor PICALM in normal hematopoiesis and polycythemia vera pathophysiology. Haematologica *100*, 439–451.
33. Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., Wilkens, A.B., et al. (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nat. Biotechnol. *33*, 985–989.
34. Sun, Z., Wang, Y., Han, X., Zhao, X., Peng, Y., Li, Y., Peng, M., Song, J., Wu, K., Sun, S., et al. (2015). miR-150 inhibits terminal erythroid proliferation and differentiation. Oncotarget *6*, 43033–43047.
35. Lei, Y., Zhang, X., Su, J., Jeong, M., Gundry, M.C., Huang, Y.H., Zhou, Y., Li, W., and Goodell, M.A. (2017). Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. Nat. Commun. *8*, 16026.
36. 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.
37. Peraki, I., Palis, J., and Mavrothalassitis, G. (2017). The Ets2 Repressor Factor (Erf) Is Required for Effective Primitive and Definitive Hematopoiesis. Mol. Cell. Biol. *37*, e00183–17.
38. McIntosh, B.E., Brown, M.E., Duffin, B.M., Maufort, J.P., Vereide, D.T., Slukvin, I.I., and Thomson, J.A. (2015). Nonirradiated NOD,B6.SCID Il2rγ-/- Kit(W41/W41) (NBSGW) mice support multilineage engraftment of human hematopoietic cells. Stem Cell Reports *4*, 171–180.
39. Consortium, E.P.; and ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
40. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. *46* (D1), D794–D801.
41. Bain, M., Mendelson, M., and Sinclair, J. (2003). Ets-2 Repressor Factor (ERF) mediates repression of the human cytomegalovirus major immediate-early promoter in undifferentiated non-permissive cells. J. Gen. Virol. *84*, 41–49.
42. Balasubramanian, M., Lord, H., Levesque, S., Guturu, H., Thuriot, F., Sillon, G., Wenger, A.M., Sureka, D.L., Lester, T., Johnson, D.S., et al.; DDD Study (2017). Chitayat syndrome: hyperphalangism, characteristic facies, hallux valgus and bronchomalacia results from a recurrent c.266A>G p.(Tyr89Cys) variant in the *ERF* gene. J. Med. Genet. *54*, 157–165.
43. 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.
44. Bagger, F.O., Kinalis, S., and Rapin, N. (2019). BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. Nucleic Acids Res. *47* (D1), D881–D885.
45. Rakyan, V.K., Down, T.A., Balding, D.J., and Beck, S. (2011). Epigenome-wide association studies for common human diseases. Nat. Rev. Genet. *12*, 529–541.

The American Journal of Human Genetics, Volume *108*

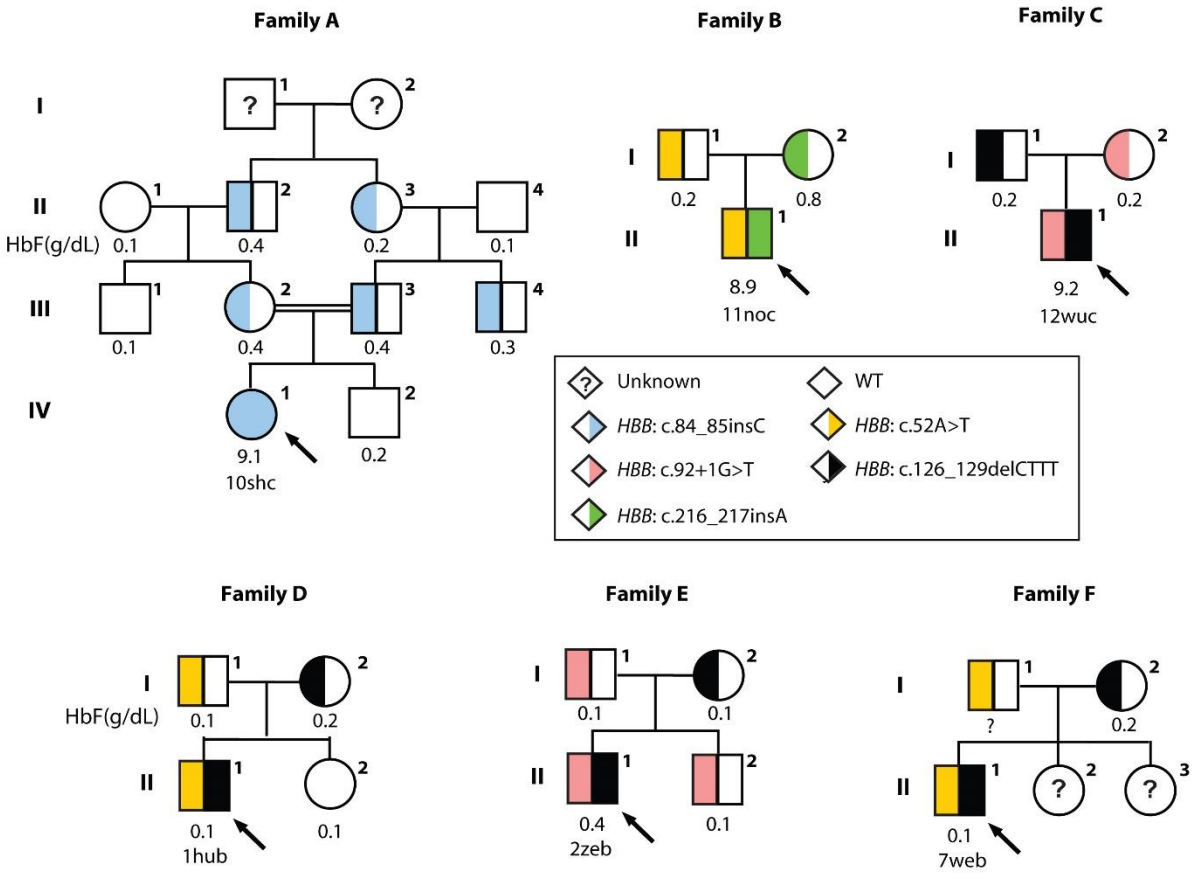
Supplemental information

Epigenetic inactivation of *ERF* reactivates γ-globin expression in β-thalassemia

Xiuqin Bao, Xinhua Zhang, Liren Wang, Zhongju Wang, Jin Huang, Qianqian Zhang, Yuhua Ye, Yongqiong Liu, Diyu Chen, Yangjin Zuo, Qifa Liu, Peng Xu, Binbin Huang, Jianpei Fang, Jinquan Lao, Xiaoqin Feng, Yafeng Li, Ryo Kurita, Yukio Nakamura, Weiwei Yu, Cunxiang Ju, Chunbo Huang, Narla Mohandas, Dali Li, Cunyou Zhao, and Xiangmin Xu

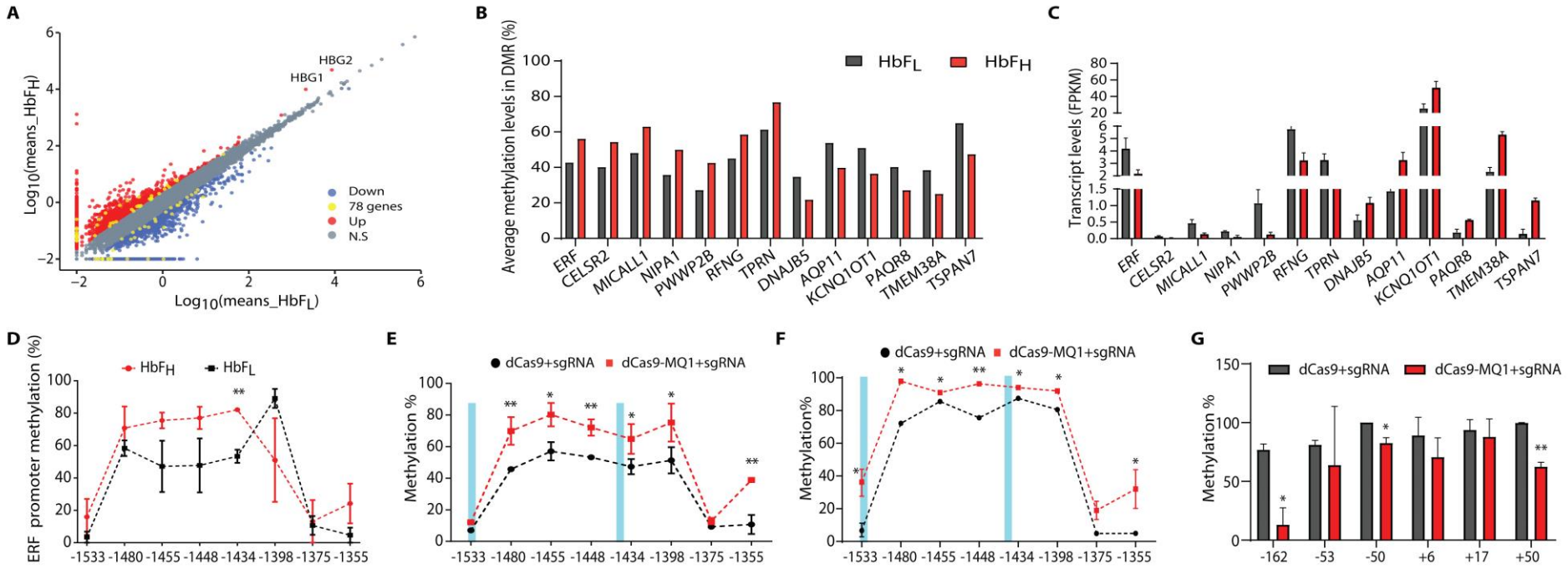
Supplemental Data

Figure S1. Pedigrees of the six β-thalassemia families.



The arrows indicate the probands. The *HBA* genotypes of all the members were αα/αα. The *HBB* mutations in all six probands were inherited from the parents.

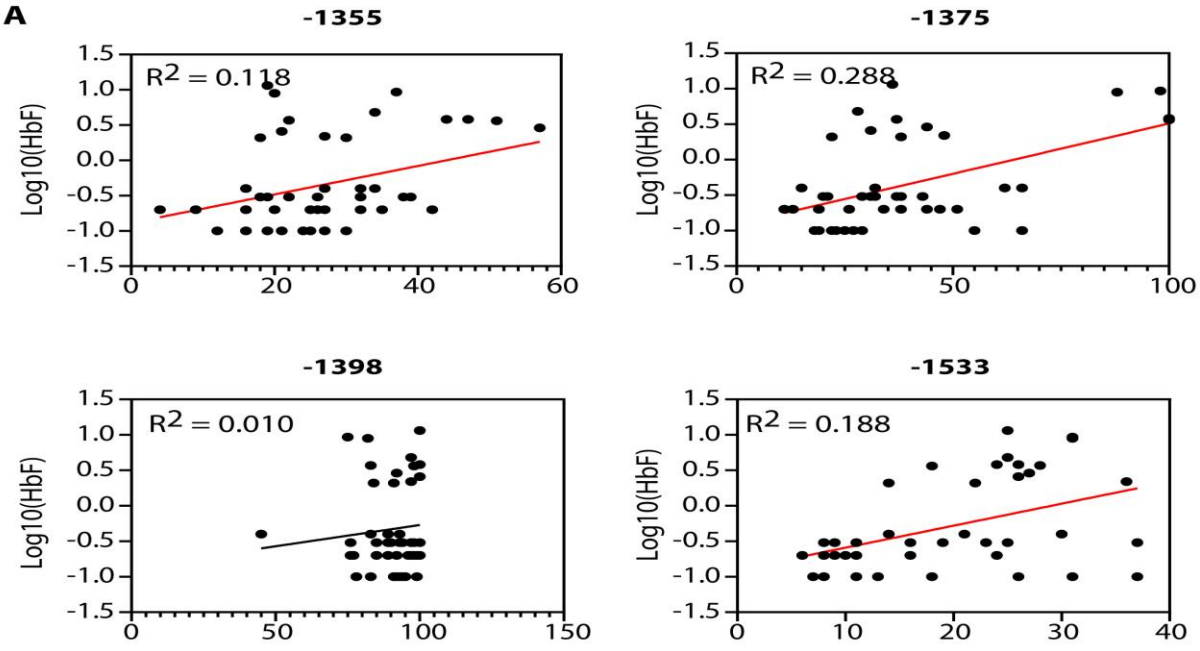
Figure S2. Methylation and expression profiles of 13 candidate genes.



(**A**) RNA-seq analysis between HbF_H and HbF_L group. Each dot represents an individual gene; differentially expressed genes are indicated according to the mean of FPKM (Fragments Per Kilobase Per Million Mapped Fragments) values. The yellow dots highlighted the 78 candidate genes. Red dots indicated up-expression genes in HbF_H group, while blue dots represented down-expression genes. The gray dots showed the none significant genes. (**B**) The average methylation levels of the promoter DMR in the 13 selected genes identified by WGBS from six subjects with β^0/β^0-thalassemia. **(C)** Transcript levels of 13 selected genes identified by integrative analysis of WGBS and RNA-seq from six subjects with β^0/β^0-thalassemia. The FPKM values were obtained from RNA-seq. The black bars represent individuals in the HbF_L group, and the red bars represent those in the HbF_H group. **(D)** The methylation level of CpG sites within the *ERF* promoter DMR, quantified by WGBS from six subjects with β^0/β^0-thalassemia. The abscissa denotes the base pairs relative to the *ERF* transcription start site (TSS). **(E and F)** Altered methylation levels in the *ERF* promoter after treatment with targeted *ERF* DNA methylation in HUDEP-2 cells **(E)** and $CD34^+$ HSPCs **(F)**. The light blue bars represent the locations of *ERF* sgRNA-binding sites. **(G)** Altered methylation levels in the *HBG* promoter after treatment with targeted *ERF* DNA methylation in $CD34^+$ HSPCs**.** Data are shown as the mean ± SD from at least three independent experiments performed in triplicate. $^*p < 0.05$; $^{**}p < 0.01$.

Figure S3. Regression analysis of the *ERF* methylation level and HbF level in 47 β-thalassemia individuals.

A

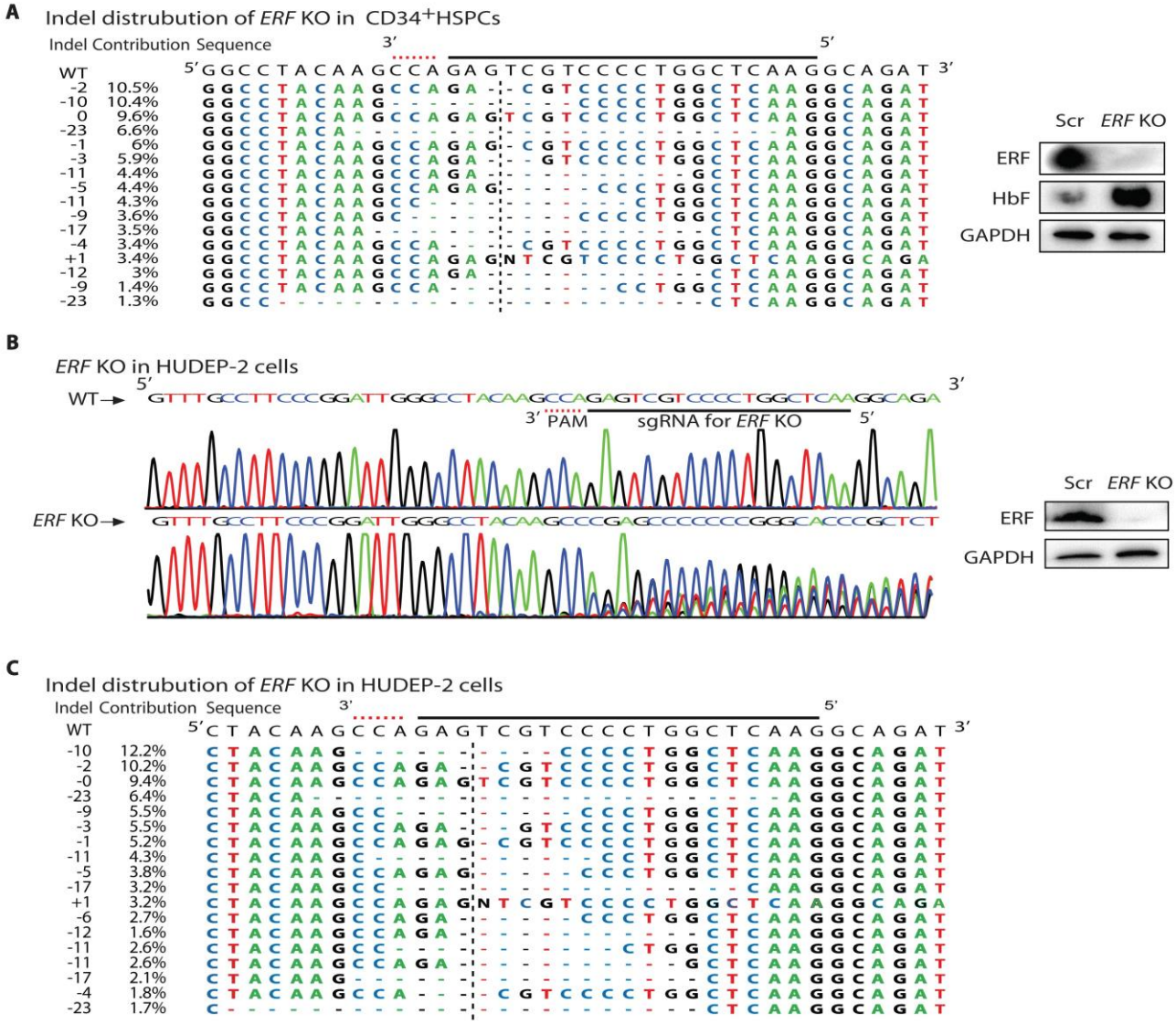


B

CpG sites	B	R^2	p value	95% CI
-1355	0.020	0.118	0.018	0.004-0.037
-1375	0.014	0.288	0.000	0.007-0.021
-1398	0.006	0.010	0.510	-0.012-0.024
-1434	-0.032	0.146	0.130	-0.055--0.008
-1448	0.007	0.016	0.404	-0.001-0.023
-1455	-0.019	0.095	0.455	-0.037--0.001
-1480	0.002	0.002	0.774	-0.014-0.018
-1533	0.031	0.188	0.004	0.010-0.052

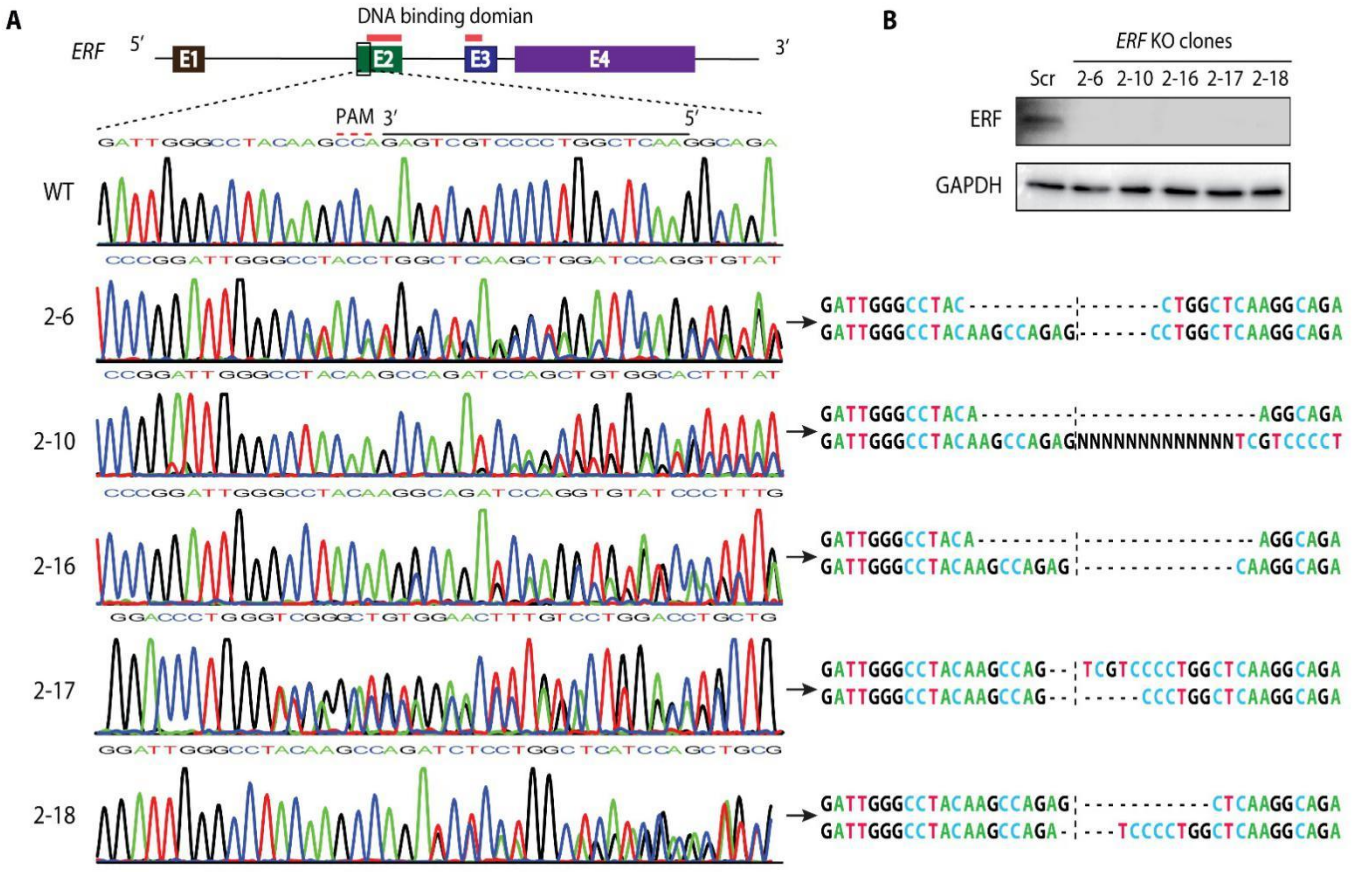
(A) The scatter diagram shows the correlation between HbF level and *ERF* methylation of the four CpG sites (-1355, -1375, -1398, -1533) in *ERF* promoter. **(B)** The statistics of the regression analysis. B, beta, indicated partial regression coefficient. R^2, coefficient of determination. CI, Confidence interval limits.

Figure S4. *ERF* KO efficiency.



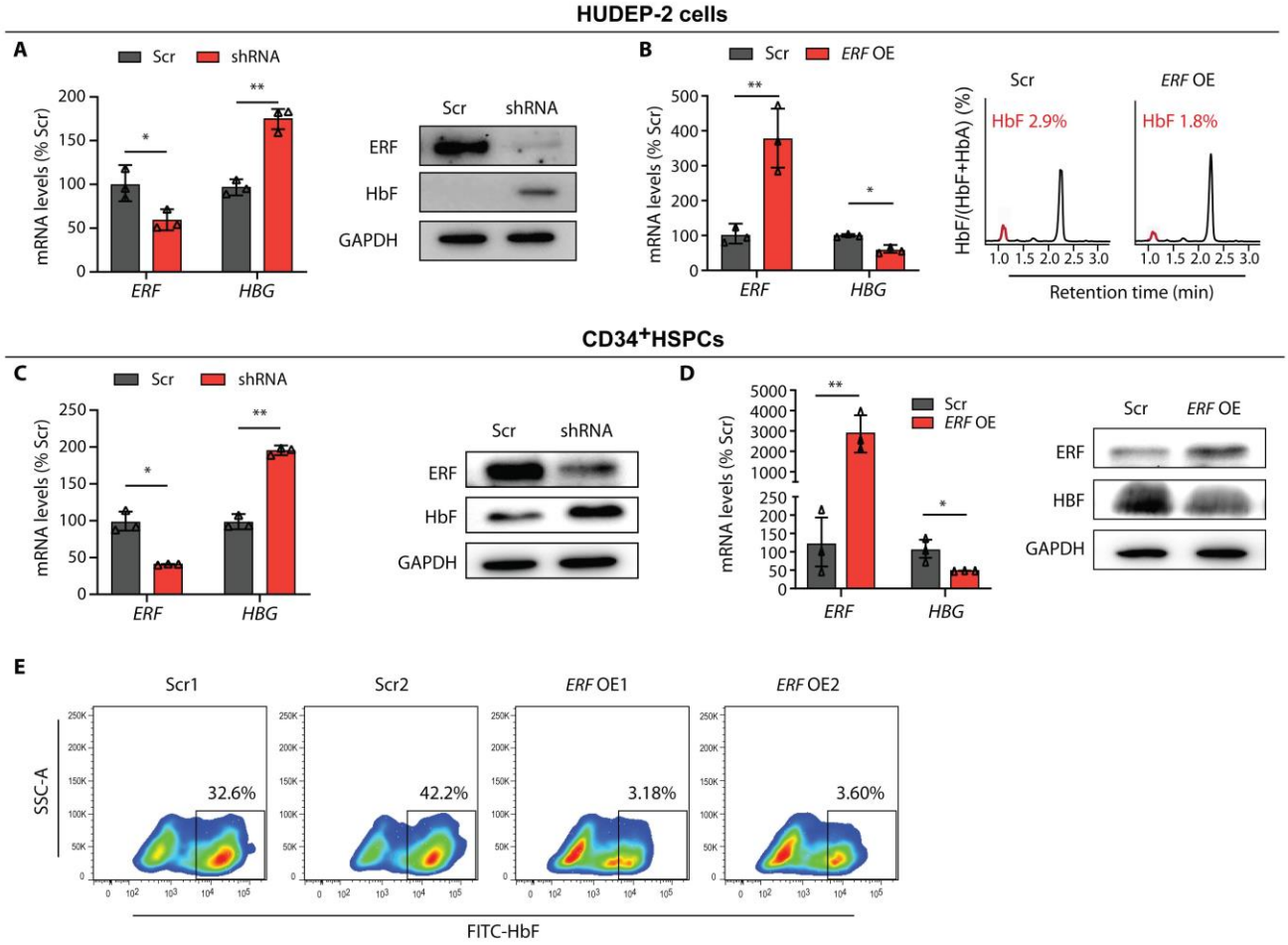
(A) Left: The indel distribution pattern of *ERF* KO in CD34^{+} HSPCs. The black line represents sgRNA targeting *ERF*. The red dashed line indicates PAM. Right: Western blotting analysis of the KO efficiency in CD34^{+} HSPCs. **(B)** Left: Representative sequencing chromatographs of PCR products of *ERF* WT and KO HUDEP-2 cells. Right: Western blot analysis of the KO efficiency. GAPDH served as a loading control. **(C)** The indel distribution pattern of *ERF* KO in HUDEP-2 cells.

Figure S5. *ERF* KO clones of HUDEP-2 cells.



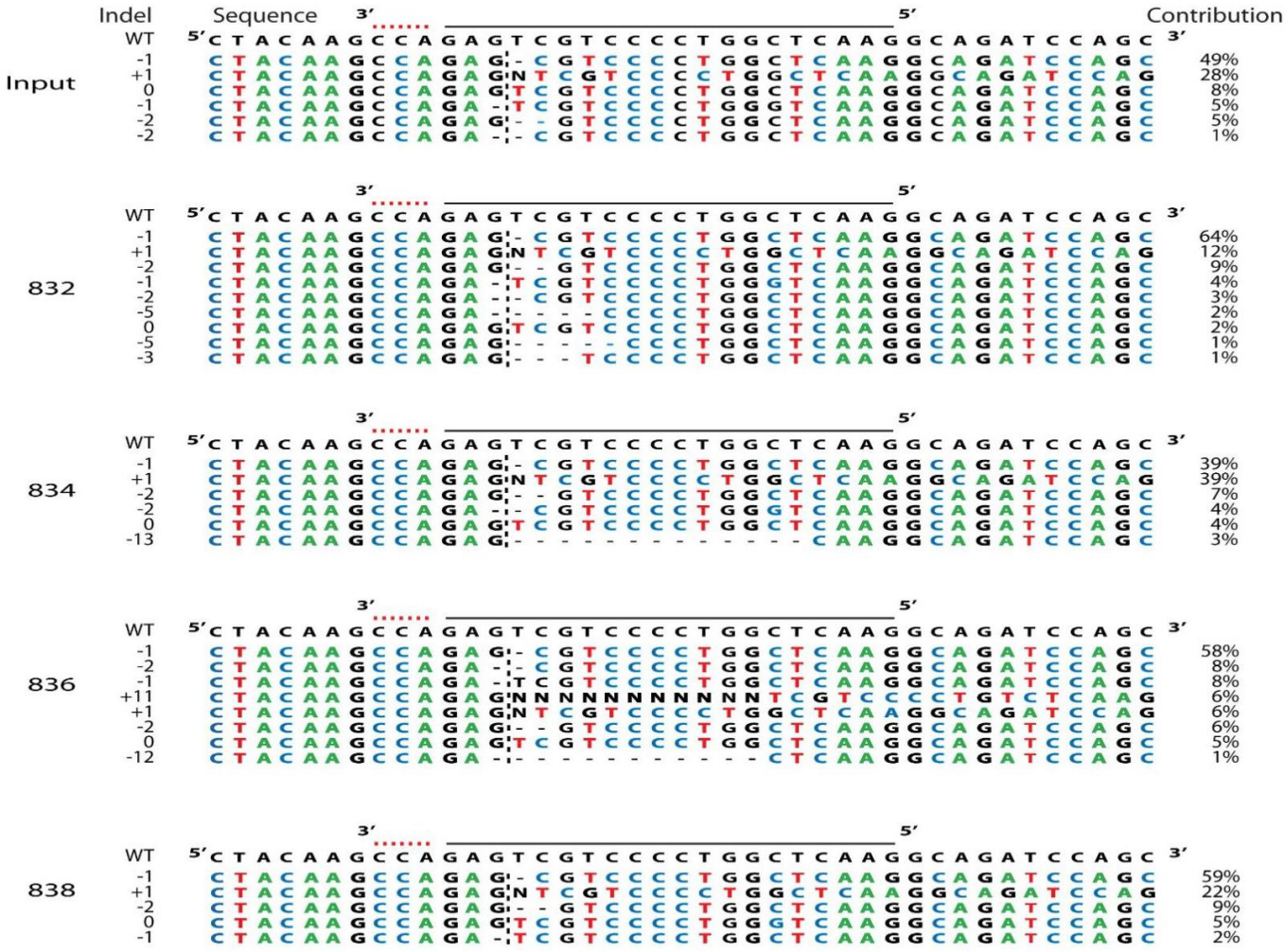
(A) Top: schematic of the *ERF* gene structure. E1, Exon 1. Bottom: sequencing chromatographs of PCR products of *ERF* WT and KO HUDEP-2 clones. The black line represents sgRNA targeting *ERF*. The red dashed line indicates PAM. Right: the indel pattern of each *ERF* KO clone; the black dashed line indicates the cutting site. **(B)** Western blotting analysis of *ERF* KO HUDEP-2 clones. GAPDH served as a loading control.

Figure S6. ***ERF*** **KD and OE in HUDEP-2 cells and CD34$^+$ HSPCs.**



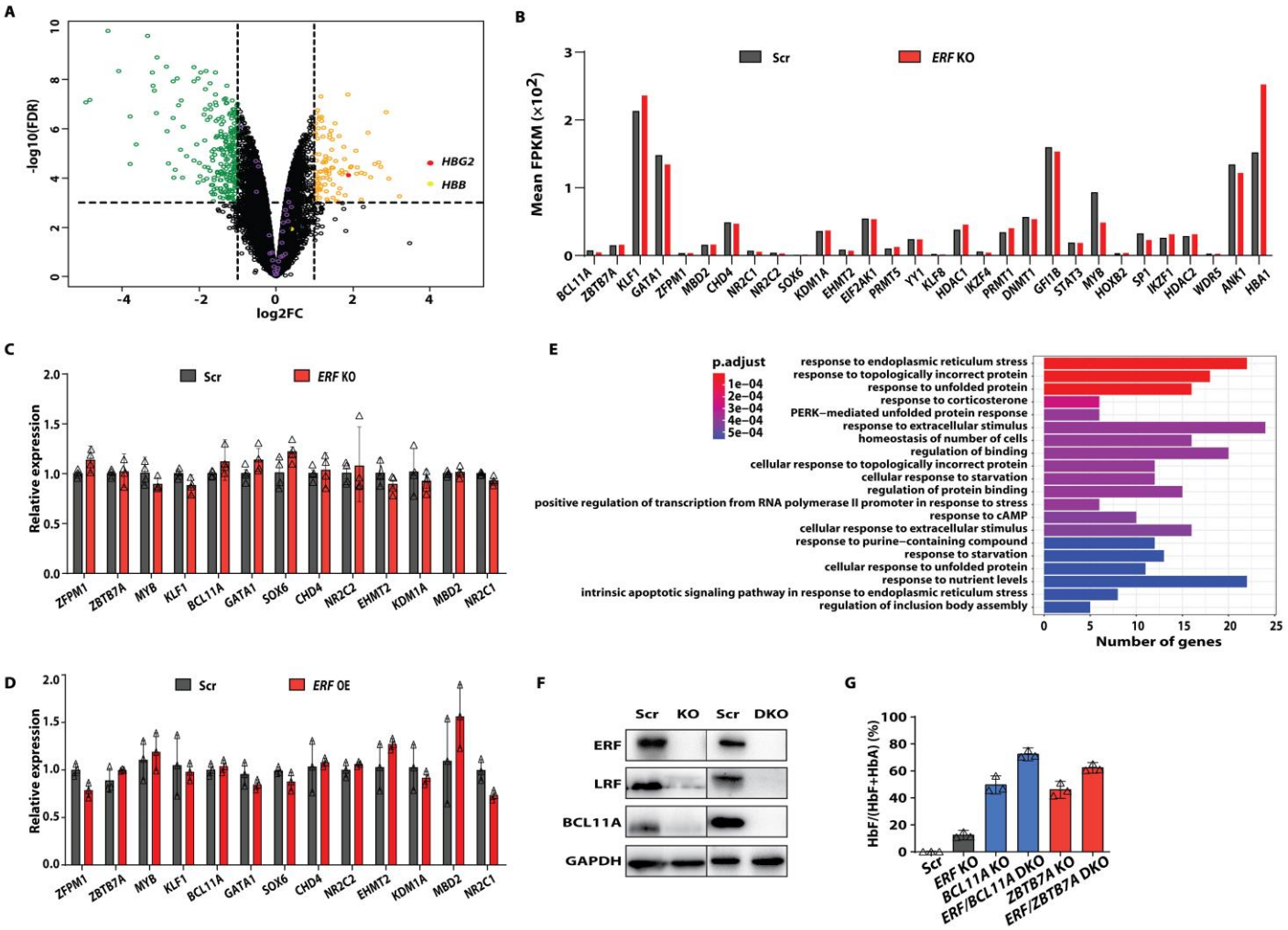
(A) Analysis was performed with qPCR analysis (left) and western blotting (right) in ERF KD HUDEP-2 cells. **(B)** Left: qPCR analysis of γ-globin and *ERF* expression in HUDEP-2 cells with *ERF* OE. Right: representative HPLC profiles of HUDEP-2 cells without (Scr) or with *ERF* OE. **(C)** qPCR analysis (left) and western blotting (right) in ERF KD CD34$^+$ HSPCs. **(D)** qPCR (left) and western blotting (right) analysis of γ-globin and *ERF* expression in CD34$^+$ HSPCs with *ERF* OE. **(E)** Representative FACS profiles of HbF levels in CD34$^+$ HSPCs with *ERF* OE. Data are shown as the mean ± SD from at least three independent experiments performed in triplicate. $^*p < 0.05$; $^{**}p < 0.01$.

Figure S7. Representative indel distribution pattern of *ERF* KO in BM cells from engrafted mice.



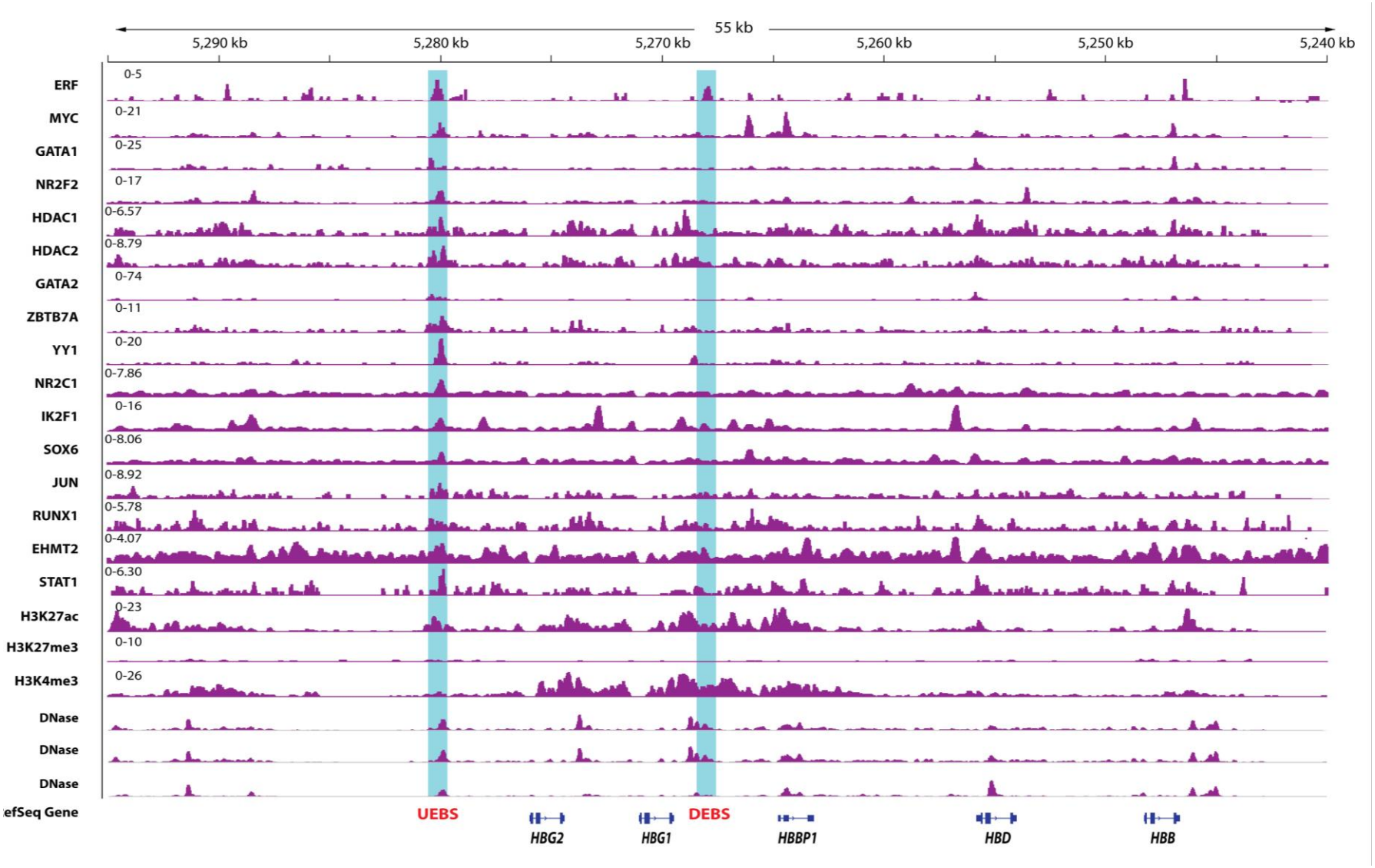
The black line shows the position of sgRNA. The red dashed line shows the PAM. The black dashed line shows the cutting site of sgRNA. The numbers 832, 834, 836 and 838 indicate mouse IDs.

Figure S8. Effects of *ERF* KO on expression of the known γ-globin repressors in HUDEP-2 cells.



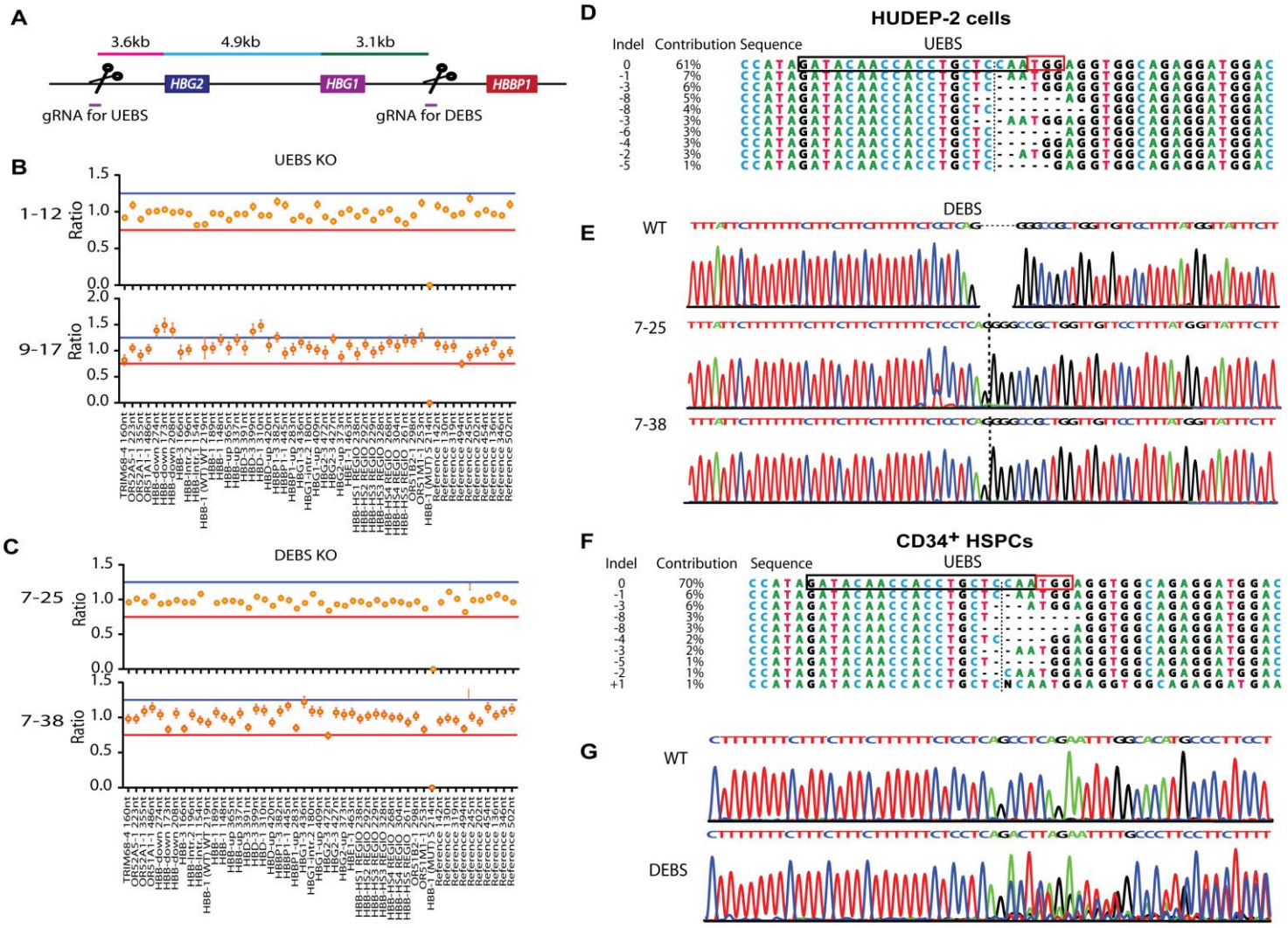
(A) RNA-seq was performed in *ERF* KO HUDEP-2 cells. The green and orange circles denote significant transcripts according to the screening standards: fold change (FC) > 2, FDR < 10^{-3}. The purple circles represent the known γ-globin repressors shown in **(B)**. Red solid circle: *HBG2*. Blue solid circle: *HBA2*. Yellow solid circle: *HBB*. The horizontal dashed line represents FDR=10^{-3}. The vertical dashed lines represent $|\log_2 FC| = 1$. **(B)** Bar graphs show transcript levels of know γ-globin regulator in control and *ERF* KO HUDEP-2 cells. Mean FPKM values from two to three independent samples per genotype are shown. **(C, D)** qPCR analysis of the major known γ-globin repressors in *ERF* KO (C) and over expression (D) HUDEP-2 cells. **(E)** Gene Ontology (GO) analysis was performed based on FDR < 10^{-3} and |Log2FC| ≥ 1 in expression magnitude between ERF KO and WT HUDEP-2 cells. **(F)** Knockout of *ERF*, *ZBTB7A*, *BCL11A*, *ERF*/*ZBTB7A* double KO and *ERF*/*BCL11A* double KO (DKO) was confirmed by western blotting with anti-ERF, anti-LRF or anti-BCL11A antibodies. GAPDH was served as a loading control. **(G)** Quantitative measurement of hemoglobin by HPLC. The percentage of HbF relative to the sum of HbF and adult globin (HbA) is shown for controls (Scr) and *ERF* KO, *BCL11A* KO, *ZBTB7A* KO, *ERF*/*BCL11A* double KO or *ERF*/*ZBTB7A* double KO HUDEP-2 cells. Two-way ANOVA was used to evaluate the interaction between *ERF* KO and the DKO HUDEP-2 cells and no statistically significance was found (p = 0.1879 for the interaction between ERF and BCL11A, and p = 0.5357 for the interaction between ERF and ZBTB7A).

Figure S9. The signals of ChIP-seq showed that known γ-globin regulators bound to UEBS and DEBS.



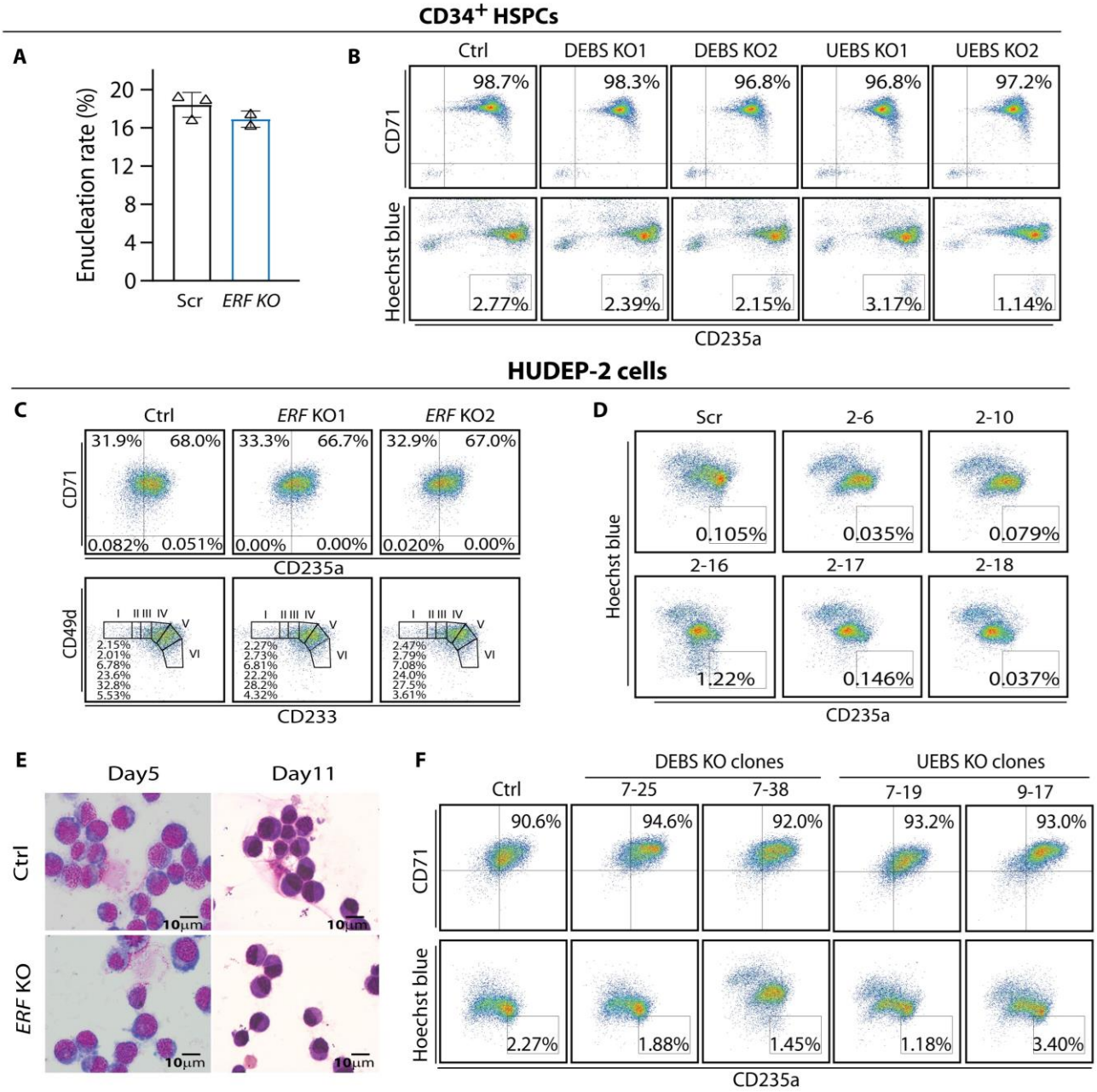
The ChIP-seq data were attained from the ENCODE. The data ID of each transcription factor are as followed: MYC (ENCFF112BJR), GATA1 (ENCFF226FPS), NR2F2 (ENCFF335LTU), HDAC1 (ENCFF355ZSE), HDAC2 (ENCFF532SIS), ZBTB7A (ENCFF457PSO), YY1 (ENCFF480ZET), NR2C1(ENCFF490IPR), IK2F1 (ENCFF715OCK), SOX6 (ENCFF757HAQ), JUN (ENCFF825SKH), RUNX1 (ENCFF962DJQ), EHMT2 (ENCFF982RMW), STAT1 (ENCFF985QWF), H3K27ac (ENCFF010PHG), H3K27me3 (ENCFF312LYO), H3K4me3 (ENCFF715DGL) and DNase (ENCFF113ZZB, ENCFF286KTH, ENCFF352SET). The light blue shadows indicate the regions of UEBS (left) and DEBS (right).

Figure S10. UEBS and DEBS KO in HUDEP-2 cells and CD34⁺ HSPCs.



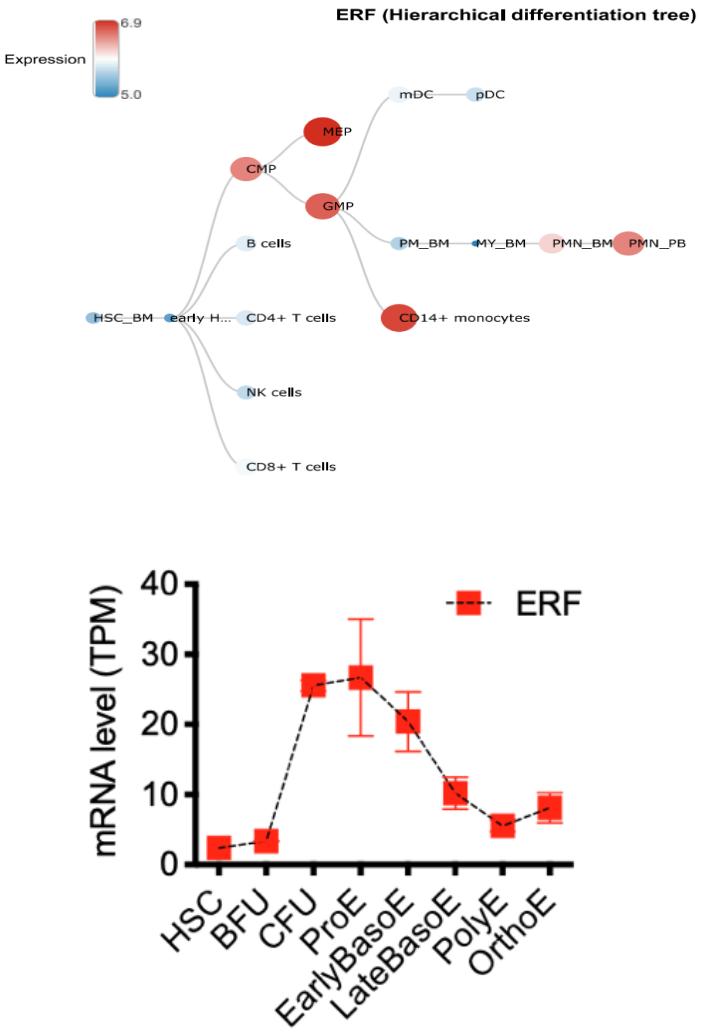
(A) Map of *HBG2*, *HBG1* and *HBBP1* showing cleavage sites for Cas9/gRNA (scissors). **(B, C)** Representative MLPA profiles of several UEBS **(B)** or DEBS KO **(C)** HUDEP-2 clones. The ratios of all probes ranged from 0.75 (red line) to 1.25 (blue line), which indicates normal copy number. **(D, E)** Indel distribution pattern of UEBS KO **(D)** and the sequencing chromatographs of PCR products of DEBS KO clones **(E)** in HUDEP-2 cells. Due to the restriction of PAM (NGG), we designed only one sgRNA to target the DEBS fragment. The black rectangular box shows the position of sgRNA. The red rectangular box shows the PAM. The dashed line shows the cutting site of sgRNA. in H UDEP-2. We used two sgRNAs to target the DEBS fragment. The black dashed lines indicated the ligation site. **(F, G)** Indel distribution pattern of UEBS KO **(F)** and the sequencing chromatographs of PCR products of DEBS KO **(G)** in CD34⁺ HSPCs.

Figure S11. The impact of ERF on erythroid differentiation.



(A) Enucleation rate in day 18 *ERF* KO CD34$^+$ HSPCs. The enucleation rate was displayed as the percentage of enucleated cells in the cytospins. 2-3 cytospins in each group were calculated. **(B)** FACS analysis of the surface markers CD235a and CD71 and enucleation rate in UEBS or DEBS KO CD34$^+$ HSPCs. **(C)** Representative fluorescence-activated cell sorting (FACS) analysis of the surface markers CD71 and CD235a (top), and CD49d and CD233 (bottom) at day 11 control (Scr) and *ERF* KO HUDEP-2 clones. **(D)** Hoechst blue and CD235a were used to monitor the enucleation rate of Scr and ERF KO (2-6, 2-10, 2-16, 2-17 and 2-18 single clones) HUDEP-2 cells. The enucleation rate was gated on Hoechst blue negative and CD235a positive. **(E)** Representative images of Wright-Giemsa staining of cytospins in day 5 (the last day of the first culture stage with 10% FBS) and day 11 (the last day of the second culture stage in the differentiation culture media with 30% FBS) Scr and *ERF* KO HUDEP-2 cells. **(F)** FACS analysis of the surface markers CD235a and CD71 and enucleation rate in UEBS or DEBS KO HUDEP-2 cells.

Figure S12. ***ERF*** **expression pattern during erythropoiesis from primary CD34⁺ HSPCs**



ERF expression profile in blood spot dataset (**upper**)[1] or transcriptome analyses of human erythroid progenitors (**lower**)[2].

Table S1. The phenotypes and genotypes of family members from the six probands included in the discovery cohort.

ID	HB	HbF	HbA2	MCV	MCH	MCHC	*HBG1*	*HBG2*	*HBS1L-MYB*	*BCL11A*	*KLF1*	*HBA*	*HBB*
	g/dL	%	%	fL	pg	g/dL	rs368698783	rs7482144	rs9399137	rs766432	mutations	genotype	genotype
Family A													
II-1	13.4	0.7	2.8	85.1	29.3	34.4	GG	CC	TC	AC	WT	αα/αα	N/N
II-2	14.9	2.7	5.5	65.1	21.2	32.5	AA	TT	TT	CC	WT	αα/αα	c.84_85insC/N
II-3	10.1	2.0	5.1	68.0	20.7	30.3	GA	CT	TT	AA	WT	αα/αα	c.84_85insC/N
II-4	11.8	0.8	2.6	99.7	31.8	31.9	GG	CC	CC	AC	WT	αα/αα	N/N
III-1	14.4	0.7	2.8	87.9	29.9	34.0	GA	CT	TC	AC	WT	αα/αα	N/N
III-2	10.0	4.0	5.5	73.0	22.5	30.9	GA	CT	TC	AC	WT	αα/αα	c.84_85insC/N
III-3	11.9	3.4	5.2	68.7	20.4	29.7	GA	CT	TC	AC	WT	αα/αα	c.84_85insC/N
III-4	12.1	2.5	5.1	71.4	21.0	29.4	GA	CT	TC	AC	WT	αα/αα	c.84_85insC/N
IV-1[a]	9.2	98.9	1.7	78.2	25.9	31.3	AA	TT	CC	AC	WT	αα/αα	c.84_85insC/c.84_85insC
IV-2	10.9	1.8	3.0	91.0	29.0	31.9	GG	CC	TT	AA	WT	αα/αα	N/N
Family B													
I-1	12.0	1.7	5.2	64.1	19.8	30.9	GG	CC	TT	AC	WT	αα/αα	c.52A>T/N
I-2	11.4	7.0	6.0	63.5	19.8	31.2	GG	CC	TT	AA	WT	αα/αα	c.216_217insA/N
II-1[a]	9.3	95.7	5.1	74.0	25.6	34.7	GG	CC	TT	AC	WT	αα/αα	c.216_217insA/c.52A>T
Family C													
I-1	10.6	1.9	5.6	77.0	25.2	32.7	GG	CC	TT	CC	WT	αα/αα	c.126_129delCTTT/N
I-2	10.9	1.8	5.6	70.9	22.3	31.5	GA	CT	TC	AC	WT	$-\alpha^{4.2}$/αα	c.92+1G>T/N
II-1[a]	9.7	94.8	5.7	69.0	22.0	31.8	GA	CT	TC	AC	WT	αα/αα	c.126_129delCTTT/c.92+1G>T
Family D													
I-1	13.2	0.8	5.9	69.0	27.8	34.9	GG	CC	TT	AA	WT	$-\alpha^{4.2}$/αα	c.52A>T/N
I-2	11.1	1.8	5.4	62.1	19.4	31.3	GG	CC	TT	AA	WT	αα/αα	c.126_129delCTTT/N
II-1[a]	7.6	1.3	2.6	67.3	20.4	33.3	GG	CC	TT	AA	WT	αα/αα	c.126_129delCTTT/c.52A>T
II-2	11.4	0.9	2.8	61.1	18.6	30.5	GG	CC	TT	AA	WT	αα/αα	N/N
Family E													
I-1	12.8	0.8	5.6	63.7	19.8	31.0	GA	CT	TT	AC	WT	αα/αα	c.92+1G>T/N
I-2	12.0	0.8	4.6	62.7	19.7	31.5	GG	CC	TC	AC	WT	αα/αα	c.126_129delCTTT/N
II-1[a]	4.5	8.9	2.8	63.0	20.1	33.7	GA	CT	TT	AA	WT	αα/αα	c.126_129delCTTT/c.92+1G>T
II-2	10.3	1.0	5.7	59.8	18.3	30.7	GA	CT	TC	CC	WT	αα/αα	c.92+1G>T/N
Family F													
I-2	11.5	0.0	5.6	60.0	19.8	33.1	GG	CC	TC	AC	WT	αα/αα	c.126_129delCTTT/N
II-1[a]	4.4	2.3	3.4	67.9	16.3	34.6	GG	CC	TT	AA	WT	αα/αα	c.126_129delCTTT/c.52A>T

[a]Probands; WT, wild type; N, normal *HBB* allele.

Table S2. The methylation and expression of 6 candidate genes.

		WGBS						RNA-seq				
Gene.id	**Gene annotation**	**Chrom position**	**HbF_H methy Level**	**HbF_L methy Level**	**Ratio[a]**	**Difference level**	**Status [b]**	**Means HbF_L**	**Means HbF_H**	**log2Ratio**	**Status[c]**	**Probability**
ERF	Ets2 repressor factor	chr19:42760641-42760789	0.56	0.43	1.31	0.13	hyper	4.18	2.17	-0.95	Down	0.83
RFNG	RFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyl transferase	chr17:80010596-80010820	0.59	0.45	1.30	0.14	hyper	5.77	3.24	-0.83	Down	0.83
TPRN	taperin	chr9:140096043-140096276	0.77	0.61	1.25	0.15	hyper	3.27	1.76	-0.89	Down	0.83
AQP11	aquaporin 11	chr11:77299848-77300374	0.40	0.54	0.74	-0.14	hypo	1.43	3.27	1.19	Up	0.83
KCNQ1OT1	KCNQ1 opposite strand/antisense transcript 1 (non-protein coding)	chr11:2721681-2722358	0.37	0.51	0.72	-0.15	hypo	25.65	50.62	0.98	Up	0.83
TMEM38A	transmembrane protein 38A	chr19:16770368-16770793	0.25	0.39	0.65	-0.13	hypo	2.33	5.33	1.19	Up	0.83

[a]The fold change of mean methylation level of DMR in HbF_H group compared to HbF_L group.

[b]The mean methylation level of HbF_H group in the DMR region compared to that of HbF_L group.

[c]The gene expression level of HbF_H group compared to that of HbF_L group.

Table S3. The phenotypes and genotypes of the validation cohort including 47 individuals with β^0/β^0-thalassemia.

Characteristics	HbF_L (N = 34)	HbF_H (N = 13)	P value[a]
Median age (range) - yr	9.5(4-17)	10.0(5-22)	0.711
Sex - no. (%)			
Female	18(52.9)	7(53.8)	0.956
Male	16(47.1)	6(46.2)	
Median absolute amount of fetal hemoglobin (range) - %	2.1(0.7-5.3)	42.7.(24.3-90.8)	<0.001
Median hemoglobin (range) - g/dL	9.5(5.5-13.3)	8.8(7.2-12.6)	0.372
Median mean corpuscular volume (range) - fL	84.7(73.1-91.2)	74.6(64.8-85.7)	<0.001
Median mean corpuscular hemoglobin (range) - pg	28.1(23.5-30.7)	24.6(20.9-27.1)	<0.001
Median mean corpuscular hemoglobin concentration (range) - g/dL	33.4(30.9-35.2)	32.9(29.0-34.2)	0.049
Median time since previous transfusion (range) - days	15(11-43)	19.5(16-47)	0.014
Median age at first transfusion (range) - months	4.5(2-14)	14.5(8-72)	<0.001
Median transfusion frequencies (range) - times/yr	18(12-26)	12(6-18)	<0.001
Median RBC transfusion burden (range) - units/yr	42.5(24-80)	24.0(12-54)	<0.001
Splenectomy - no. (%)	5(14.7)	3(23.1)	0.803
Non-transfusion-dependent persons - no. (%)	0(0.0)	1(7.7)	NA
***HBB* genotype categories - no. (%)**			
c.126_129del/c.126_129del	13(38.3)	3(23.1)	
c.126_129del/c.52A>T	10(29.4)	3(23.1)	
c.126_129del/c.85dup	0(0.0)	2(15.4)	
c.126_129del/c.217dup	3(8.8)	0(0.0)	
c.126_129del/c.92+1G>A	2(5.9)	2(15.4)	
c.126_129del/c.316-197C>T	0(0.0)	1(7.7)	
c.126_129del/c.130G>T	0(0.0)	1(7.7)	
c.52A>T/c.52A>T	5(14.7)	1(7.7)	
c.52A>T/c.217dup	1(2.9)	0(0.0)	
Modifier genes genotype categories - no. (%)			
***BCL11A* rs766432**			0.218
C/C	3(8.8)	0(0.0)	
C/A	8(23.5)	6(46.2)	
A/A	23(67.6)	7(53.8)	
***HBS1L-MYB* intergenic region rs9399137**			0.162
T/T	31(91.2)	9(69.2)	
T/C	2(5.9)	3(23.1)	
C/C	1(2.9)	1(7.7)	
***HBG2* rs7482144 (XmnI)**			0.059
C/C	31(91.2)	9(69.2)	
C/T	3(8.8)	4(30.8)	
T/T	0(0.0)	0(0.0)	
***HBG1* rs368698783**			0.059
G/G	31(91.2)	9(69.2)	
G/A	3(8.8)	4(30.8)	
A/A	0(0.0)	0(0.0)	
***KLF1* mutants - no. (%)**	0(0.0)	1(16.7)[b]	

yr, year; NA, not available; RBC, red blood cell.

[a]Individuals carrying deletional HPFH mutations or α-thalassemia mutations were excluded from the cohort. Characteristics of the individuals were compared between the HbF_H group and HbF_L group with the Pearson chi-square test for categorical variables and the Wilcoxon rank-sum test for continuous variables.

[b]One individual carried a heterozygous modifying mutation in *KLF1* (c.519_525dup).

Table S4. Primers used in this study.

Purpose	Gene/loci		5' primer/WT probe	3' primer/MT probe
			Primer/Probe sequence (5'-3')	Primer/Probe sequence (5'-3')
Bisulfite sequencing	*ERF*	1st	GTATAATGGTAAAGGGTAAATGTGG	CTCAAACCTCCTCCTTAAAAAAAC
		2nd	TTTTGGTTGTAAGATTTAGTGGA	CAAACCTCCTCCTTAAAAAAAC
Pyrosequencing	PCR	BS1	GTATAATGGTAAAGGGTAAATGTGGT	Bio-CCAAACATTATCAAAAACATCTTCTCAT
		BS2	GGGATTTTGTGAGGATTAAATGAGAA	Bio-ACCAAAAACTCCAAAAAAACTTTC
	Sequence	S1	GTGGTTTGGAATTGGTTA	
		S2	AGAAGATGTTTTTGATAATGT	
		S3	ATGATTATTATTGTTGGATTAT	
Dual luciferase assay	UEBS		CTCACCCTGGTAGCTGAAGACA	CAATGGAGGTGGCAGAGGATG
	DEBS		GCTCTGCCTATGCAGTAGTCATTC	AACTTCTCCAACATCTCTGCCTGG
ChIP-qPCR	*MYOD1*		CAGACTGTCATCCCCACCACA	CCCAACTGTGTGATTTTGTGGA
	UEBS		CCCACCTGATGGTCCTTCCT	GCCAGGAGGTGGCACTTTCTA
	DEBS		AGTGCATATTCTGAAACGGTAGTG	GCACATGCCCTTCCTTCTTTC
	ETS pro		TTACTTCCTCCAGAGACTGACGA	CGCCGGCCAGAGACGAT
Validation for KO	*ERF*		CTCCTGTCTAGCCCCTCACA	ACGCACTTAGGTGTGGTTCC
	UEBS		TAGAGGAAGCAGCGGGAAAAGC	TTCCTGAGAGCCGAACTGTAGTG
	DEBS		CCATGGCAATGTCAGAAAGTTACCC	GACCCAAATAATAAGCCTGCGCC
	BCL11A		CCGAGCCTCTTGAAGCCATT	TCATCCTCTGGCGTGACCT
	ZBTB7A		GGGGCTTTGGCTGTACTGT	TACACGTTCTGCTGGTCCAC
RT-qPCR	*HBG*		GGTCATTTCACAGAGGAGGACAAG	CCAAAGCTGTCAAAGAACCTCTG
	ERF		GGCCCTGCGCTATTACTATA	CCAGCCAACCCCACATCAA
	ACTB		GGGAAATCGTGCGTGACATT	GGAGTTGAAGGTAGTTTCGTG
	HBB		GCACGTGGATCCTGAGAACT	CACTGGTGGGGTGAATTCTT
	CHD4		AAATATGCGGCCGCGAGCAGCTGG GCCAATTCC	GCTCTAGATCACTGCTGCTGGGCTACCT
	NR2C1		AACACCTGCAGCTCCTAACAGA	TACTGCTCCATAATGACGTCCTG
	NR2C2		TCTTTAGCCCCGATCATCCAG	CCAATCGGTAGGTGTCTTCTG
	EHMT2		TGGTATGACAAGGATGGGCGAT	TGTAGCCGCACCTTGATGCC
	KDMIA		CAGGTGCCCCACAGCCGATT	TGCAATTCTTCCCGCTTCTCGCA
	ZFPM1		GCAGATCAAGCGTTCCCTC	TGAGTTAACATCTGCGCTGGG
	MBD2		CAGTCCAAGTGGTAAGAAGTTC	ATGAGTGTTCCAGCGCAGC
	GATA1		CACTGAGCTTGCCACATCC	ATGGAGCCTCTGGGGATTA
	KLF1		ACACCAAGAGCTCCCACCT	GTAGTGGCGGGTCAGCTC
	BCL11A		CCCAAACAGGAACACATAGCA	GAGCTCCATGTGCAGAACG
	ZBTB7A		TCACCAGGCAGGACAAGCT	AGCAGCTGTCGCACTGGTA

	SOX6		GCTTCTGGACTCAGCCCTTT	GGAGTTGATGGCATCTTTGC
	MYB		AGCAAGGTGCATGATCGTC	GATCACACCATGATGAAGAATCAG
Genotyping	***BCL11A***	**rs766432**	CGCTTCTCAGACCCAAATGCTC	GGCTTTCTAGACTGGTGGACG
	HBS1L-MYB	**rs9399137**	CGGTTCCCTCAGAAGACACTTAC	TTCACTGCCAGAAGCACTTTGG
	KLF1	**promoter**	TACCCAGCACCTGGACCCTC	GAGGCTGTGATAGCCCCTTCG
		exon1	CTAAGGACAGAGAGGAGCCC	CAGCCAGCCCACCTAGAC
		exon2-1	CCAGTGTCCACCGAACCTC	ATCCTCCGAACCCAAAAGCC
		exon2-2	CGAGACTCTGGGCGCATA	GGAAGTGCCCTTGGTACTGA
		exon2-3	GTACCCCGCGATGTACCC	CGGTCTCGGCTATCACACC
		exon2-4	GGGGACTGCAGAGGATCCA	GCGCCCTTTCTCATGTCC
		exon3	CAGACAGTGGCGCTTATGG	CCCCAGTCACTAGGAGAGTCC
		3'UTR1	CATGAAGCGCCACCTTTGAGC	TCTCACTGGGTTTGCACGACA
		3'UTR2	GAGCCACACAGAGATGTCCAAAC	TTACAGCCTCCTGCCATCTTCC
	HBG1	**rs368698783**	TACTGCGCTGAAACTGTGG	TACCTTCCCAGGGTTTCTCC
	HBG2	**rs7482144**	AACTGTTGCTTTATAGGATTTT	AGGAGCTTATTGATAACCTCAGAC

The primers of *BCL11A*、*HBS1L-MYB*、*KLF1*、*HBG1*、*HBG2* were used to genotyping the 6 six probands and their families.

Table S5. SgRNA oligonucleotides used in this study.

Purpose	Gene	Sequence (5'-3')
Knock out	***ERF***	CTTGAGCCAGGGGACGACTC
	BCL11A	TGAACCAGACCACGGCCCGT
	ZBTB7A	GTCGGGGAACGGGATCCCGA
Targeted methylation	***ERF1***	CCCCTCGTTGGGATTTTGTG
	ERF2	ATGAAGATGATTATTATTGC
Binding site KO	**UEBS KO**	GATACAACCACCTGCTCCAA GGAACAACCAGCGGCCCTCG
	DEBS KO	TGTGCCAAATTCTGAGGCTG

Supplemental Materials and Methods

Human subjects

Six subjects with β^0/β^0-thalassemia were recruited from a cohort of 1,142 individuals with β-thalassemia that were previously enrolled by our laboratory to explore the potential cis-variants responsible for the Hb switching.[3] The subjects were from Guangxi province in Southern China, where thalassemia is highly prevalent.[4-6] These six unrelated Chinese β-thalassemia individuals with a large range of Hb F expression levels were stratified into low- and high-HbF groups for this study. Each group comprised of three subjects who were clinically diagnosed with thalassemia major (TM) in the low-HbF group (HbF_L: 0.1–0.4 g/dL) or thalassemia intermedia (TI) in the high-HbF group (HbF_H: 8.9–9.2 g/dL), respectively (**Table 1**). A brief summary of the available transfusion-related records for the six study subjects is as follows: YH: The individual first presented with anemia with pallor at 8 months of age and received regular blood transfusion every 30 days after admission. This study was performed in 2014 prior to transfusion when he was 7 years old. DZ: The individual first presented with anemia at 6 months of age, with a cold and fever, and received regular blood transfusion every 21 days starting at 1 year of age. This study was performed in 2014 prior to transfusion when he was 7 years old. Subsequently, both YH and DZ are maintained at stable mean levels of Hb ranging from 7.5 to 8.5 g/dL through regular blood transfusions. JW: The individual first presented with anemia at 7 months of age and received regular blood transfusion every 20 days until he was 7 years old. He did not receive subsequent regular blood transfusions due to financial constraints. This study was performed in 2014 prior to blood transfusion, when he was 9 years old. JS: The individual first presented with anemia and pallor at 9 years of age. She received blood transfusions only twice following bouts with cold and fever. This study was performed in 2014 when she was 10 years old. FN: The individual first presented with anemia and pallor at 6 years old

and did not receive prior blood transfusions. He received a single blood transfusion subsequently during a bout with cold and fever and study was performed in 2014 when he was 15 years old. JQ: The individual first presented with anemia at 7 years of age, with pallor and nosebleed. The individual never received blood transfusions and the study was performed in 2014 when he was 26 years old. The fetal hemoglobin levels of the 6 individuals in this study were determined when they presented first time with anemia. Family members of the study subjects were recruited for validation studies (**Figure S1 and Table S1**).

We further recruited an extended cohort of β^0/β^0-thalassemia to validate the functions of candidate genes identified from analysis of WGBS and RNA-seq data in the 6 β-thalassemia individuals described above. Prior to this, a cutoff was set according to the highest and lowest quartile for absolute amount of HbF observed in our previously studies.[4] Thus, a total of 47 β^0/β^0-thalassemia individuals with extreme HbF levels were entered into our validation cohort that consisted of 34 subjects categorized in HbF_L group (HbF $\leq$ 0.4 g/dL) and 13 subjects in HbF_H group (HbF $\geq$ 2.0 g/dL) (**Table S3**). To minimize the bias of the hematological parameters caused by blood transfusion, especially for the HbF_L individuals who are clinically more severe, hematological data were collected just prior to blood transfusion. $GYPA^+$ erythroid cells were isolated using microbeads and subjected to RNA-seq and WGBS assays to exclude contribution from non-erythroid cells.

Informed consent was obtained from all the participants prior to the study following presentation of the nature of the procedures. Approval for the extended cohort study was obtained as outlined by the protocol #NFEC-2019-039 approved by Medical Ethics Committee of Nanfang Hospital of Southern Medical University. The study was conducted in accordance with the Declaration of Helsinki.

The definition of TM or TI was based on the following four clinical indications : (1) age of onset (anemia): < 6 months, 6–24 months (TM), or > 24 months (TI); (2) transfusion before 4 years of age: symptomatic anemia requiring more than eight transfusions/year before 4 years of age (TM) or no/occasional transfusion before 4 years of age (TI); (3) steady-state hemoglobin levels: < 7.0 g/dL (TM) or 7.0–10.0 g/dL (TI); (4) growth and development: delayed (TM) or near normal (TI).[7; 8]

Hematological and genetic analysis

Hematological parameters were assessed using an automated hematology analyzer (Sysmex F-820; Sysmex, Japan), and hemoglobin analysis was performed with HPLC (Variant II, Bio-Rad Laboratories, USA). Genetic analysis was performed on genomic DNA extracted from the peripheral blood. We used PCR followed by Sanger sequencing to genotype the six subjects and their family members. The primers used for genotyping *KLF1*, *HBA*, *HBB* and *HBG* were previously described.[3; 4; 9] The primers used in genotyping *BCL11A* are listed in **Table S4.**

Genetic analysis was performed on genomic DNA extracted from the peripheral blood. qRT-PCR was performed as previously described. $CD34^+$ HSPCs or HUDEP-2 cells were analyzed by flow cytometry to monitor HbF expression.

High throughput sequencing of 6 bone marrow derived $GYPA^+$ erythroblasts and data analysis

WGBS was performed with genomic DNA extracted with phenol-chloroform from BM $GYPA^+$ erythroblasts of six subjects with β^0/β^0-thalassemia. A total of 5.2 μg of genomic DNA spiked with 26 ng of lambda DNA was fragmented by sonication to sizes of 200-300 bp with a Covaris S220 instrument and followed by end repair and adenylation. Cytosine-methylated barcodes were ligated to sonicated DNA according to the manufacturer's instructions. These DNA fragments were subsequently treated twice with bisulfate with an EZ DNA Methylation-Gold Kit (Zymo Research).

The resulting single-stranded DNA fragments were PCR amplified with KAPA HiFi HotStart Uracil + ReadyMix (2×). Clustering of the index-coded samples was performed on a cBot Cluster Generation System with a TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. Libraries were sequenced by Novogene Solutions (Tianjin, China) to a depth of ~700 million 100-bp clean paired-end reads per sample on the Illumina HiSeq 2500 platform. After quality control with FastQC and mapping onto hg19 in Bismark software (version 0.12.5) [10], the bisulfite-treated read counts for the methylated cytosine (C) and unmethylated C in the CpG/CHG/CHH context was determined with Bismark to indicate the methylation level for the given C in the CpG/CHG/CHH context: methylation level = methylated C/(methylated C + unmethylated C). The methylation levels after correction for the bisulfite non-conversion rate (r = 0.43%) as (methylation level-r)/(1-r) were subjected to statistical analysis.

We used the R package bsseq to identify the differentially methylated regions (DMRs) between the HbF_L and HbF_H groups. To compare CpG levels across multiple samples, the methylation data was analyzed using the BSmooth method to estimate methylation value as previously described.[11] Default parameters were used as the minimum number of CpGs for each window ns = 70, and the minimum window size h = 1000 bp. We then removed CpGs with reads coverage < 5 to avoid false positives, computed the *t*-statistics for each CpG sites by using the function BSmooth.tstat and estimated the standard deviation in the HbF_L group. To identify DMRs, we defined groups of consecutive CpGs for which the t-statistic was [2.5%，97.5%] a cutoff selected on the basis of the marginal empirical distribution of *t*. We filtered the set of DMRs by requiring each DMR to contain at least three CpGs, to have an average methylation difference 0.1 or greater, and to have at least one CpG every 300 bp and show methylation differences in the same direction.

RNA-seq was performed with RNA extracted with TRIzol from BM $GYPA^+$ erythroblasts of six subjects with β^0/β^0-thalassemia. A minimum RNA integrity value of 7 was required to build a strand-specific RNA library following removal of ribosomal RNAs, and sequencing was performed by BGI Solutions to (Shenzhen, China) to a depth of ~100 million 150 bp paired-end reads per sample on the Illumina Hiseq 2000 platform. Sequenced reads were aligned to the UCSC hg19 (human) genome with BWA and Bowtie, and the coding gene and isoform expression levels for mRNA and lncRNA were quantified by RNA-Seq with the Expectation Maximization (RSEM) package. DE mRNA was carried out with the NOISeq package. To identify 5835 DE-mRNAs between the HbF_H and HbF_L groups, we screened DE-mRNAs according to the following criteria: $|\log_2 FC| > 0.8$ and divergence probability > 0.8 by NOISeq analysis.

Integrative analysis of WGBS and RNA-seq data

To identify HbF reactivation-associated genes showing altered DNA methylation and RNA expression, we firstly screened 741 out of the top 3000 DMRs (top 3000 hyper-DMR and top 3000 hypo-DMR) annotated with the promoter regions. We next screened 78 DE-mRNAs that displayed hypermethylation and downregulation or hypomethylation and upregulation in the HbF_H group compared to their methylation and expression levels in the HbF_L group. Of these 78 DE-mRNAs, we focused on the 13 DE-mRNAs based on the criteria: $|\log_2 FC|>0.8$ and divergence probability> 0.8 in the NOISeq analysis and $p< 0.1$ in Student's *T* test (**Figure S2**). Of the 13 genes identified, 7 genes were filtered out due to their low abundance in either HbF_H or HbF_L groups (FPKM < 1.0). Among the remaining 6 genes, ERF is the only transcription factor that was able to regulate its target genes through a direct binding on its promoter. Moreover, previous studies showed that ERF acted as a transcription repressor indispensable for erythroid differentiation. Therefore, *ERF* was of particular interest for follow-up functional analysis for its role in the regulation of HbF reactivation.

Target CpG methylation modification

Targeted CpG site methylation alteration in HUDEP-2 cells and $CD34^+$ HSPCs were performed with a dCas9/sgRNA system to induce hypermethylation of the target CpG site in the *ERF* promoter, as previously described.[12] Since the targeted window restricts within 50 bp downstream and upstream of the binding site of sgRNA, therefore, two sgRNA oligonucleotides were designed to target the eight CpG sites identified in the *ERF* promoter and then cloned into the lentiviral pCDH-U6-dCas9-MQ1 vector. MQ1 protein can mediate methylation on the CpG sites in the targeted regions following the guiding of sgRNA. HUDEP-2 cells and $CD34^+$ HSPCs were harvested 1 week after transduction and 14 days after differentiation respectively, and DNA and RNA were extracted to detect the *ERF* promoter methylation level and RNA expression. Protein from HUDEP-2 cells was used to detect the protein level of ERF. All sgRNA oligonucleotides used in this study are listed in **Table S5.**

Bisulfite sequencing and cloning

The methylation level of the *ERF* promoter region was further determined with bisulfite Sanger sequencing, cloning or pyrosequencing methods. Bisulfite modification of genomic DNA was performed with an EpiArtTM DNA Methylation Bisulfite Kit (EM101, Vzyme, China). Bisulfite-treated DNA was purified according to the manufacturer's protocol and diluted to a final volume of 10 μl. Nested PCR was performed with 2 μl bisulfite-treated DNA as the template. Bisulfite Sanger sequencing was performed by direct sequencing of the nested PCR products and the methylation level of each CpG site was determined on the basis of the C and T allele sequencing chromatography peaks via BioEdit Sequence Alignment Editor V7.0.9.0 (Carlsbad, USA).[13] Nested PCR products were subjected to Bisulfite cloning into the pMD19T cloning vector (D104,

TAKARA, Japan). 10 to 15 of these clones (**Figure 1B**) were subjected to Bisulfite pyrosequencing with a PyroMark Q24 Advanced System (QIAGEN) according to the manufacturer's protocols.

ERF knock down (KD) and overexpression (OE)

KD of *ERF* was performed with lentiviral short hairpin RNA (shRNA) vectors (pLent-U6-ERF shRNA-GFP-Puro) purchased from Virgen (China). *ERF* OE lentivirus (Ubi-ERF-3FLAG-SV40-EGFP-IRES-puromycin, GV208) and control lentivirus were purchased from Genekai (China). Transduction of viral particles into HUDEP-2 and $CD34^+$ HSPCs was performed according to the manufacturer's protocol. After 72 hours of transduction, transduced HUDEP-2 cells were selected with puromycin (1 μg/ml), and $CD34^+$ HSPCs were selected on the basis of EGFP expression via flow cytometry. KD or OE of *ERF* was confirmed by qRT-PCR and western blotting with anti-ERF antibody (Abcam).

Engraftment studies

NCG-Kit-V831M (T003802) mice were obtained from GemPharmatech in China and the related experiments were approved by the Institutional Animal Care and Use Committee of GemPharmatech. Human $CD34^+$ HSPCs were obtained from healthy donors and edited with sgRNA to generate *ERF* KO $CD34^+$ HSPCs. Non-irradiated NCG- Kit-V831M female mice (3–4 weeks of age) were subjected to tail intravenous injection with 0.675 million cells (resuspended in 300 μl DPBS). Equal numbers of non-edited $CD34^+$ HSPCs were injected into NCG-Kit-V831M mice, which served as controls. We performed flow cytometry analysis of BM cells for monitoring of human xenograft efficiency 16 weeks post engraftment. Flow cytometry analyses of BM cells were performed with FITC anti-human CD235a antibody (Biolegend, 349104), PerCP-cy5.5 anti-human CD45 antibody (BD, 564105), APC anti-mouse CD45 antibody (BD, 559864), Brilliant Violet 421 anti-human CD19 (BD, 562440) and Brilliant Violet 650 CD33 (BD, 744353) antibodies, and

SYTOX Blue (Invitrogen, S34857) for dead cell staining at 4 °C for 30 minutes. The engraftment rate was calculated as $hCD45^+$ cells/ ($hCD45^+$ cells+$mCD45^+$ cells) ×100. B cells ($CD19^+$) and myeloid ($CD33^+$) lineages were gated on the $hCD45^+$ population. Human erythroid cells ($CD235a^+$) were gated on the $mCD45^-hCD45^-$ population.

RNAs extracted from BM cells were used for qPCR analysis of *ERF*, *HBB* or *HBG* expression. DNA extracted from blood mononuclear cells were used to detect the indel frequencies by Sanger sequencing followed by PCR.

Flow cytometry

Cells derived from cultures of $CD34^+$ HSPCs or HUDEP-2 cells were analyzed by flow cytometry with a PE-conjugated anti-CD71 antibody (12-0711-82, eBioscience, USA), a FITC-conjugated anti-CD235a antibody (13-9987-82, eBioscience, USA), BV480-conjugated anti-CD49d antibody (566134, BD), PE-conjugated anti-CD233 antibody (130-121-345, Miltenyi) and Hoechst33342. Single stain of each antibody was used for compensation. FMO was performed for gating. 7-AAD was used to gate live cells. CD235a positive cells were gated to further analyze the expression of surface markers CD233 and CD49d to determine the terminal erythroid differentiation. CD235a and Hoechst33342 were used to determine enucleation rate.

To monitor HbF expression flow cytometry was performed with anti-human HbF (2D12, BD) antibodies. Cells were fixed with 4% paraformaldehyde at RT for 20 minutes and permeabilized with 0.1% Triton X-100 for 10 minutes at RT. After washing with PBS, cells were incubated with the FITC-conjugated HbF antibody for 30 minutes at RT. Cells were washed following incubation and measured with a BD flow cytometry system. Prior to subjecting cells to analysis by BD FACS-Melody cytometer, 5 μl of 7-AAD was added to select for live cells.

RNA-sequencing of *ERF* KO HUDEP-2 cells

Total RNA was extracted with TRIzol from *ERF* KO or wild type (WT) HUDEP-2 cells. Libraries were prepared as previously described. Sequencing was performed on the Illumina Hiseq 2000 platform and 150 bp paired-end reads were generated. The reads were aligned to the UCSC Homo sapiens reference genome hg19 with TopHat version 2.0.4, and multi-aligned reads were removed with SAMtools. We then used HTSeq to generate mapping counts for the known annotated genes. We used Cuffdiff to convert the counts into total FPKM values, which represents the gene expression levels. DEGs between *ERF* KO and WT were identified with the edgeR ver. 2.6.12 statistical package.

Chromatin immunoprecipitation (ChIP)

ChIP assays were performed with cell extracts from HUDEP-2 cells or $CD34^+$ HSPCs with anti-ERF (ab153726, Abcam) antibody as recommended by the manufacturer (EZ-ChIP, Merck Millipore, Germany). A total of 0.5-1 × 10^7 cells were fixed with 1% formaldehyde for 10 minutes at 37°C, and the reaction was terminated by the addition of glycine. Sonication was performed with a Covaris S2 instrument (Covaris). The chromatin solution was first incubated with 60 µl of Protein A/G Dynabeads (Merck Millipore) for 1 hour at 4°C on a rotating shaker to prevent nonspecific binding. ChIP was then performed with 2 µg of antibody overnight at 4°C on a rotating shaker. Antibody/protein complexes were collected after incubation with 60 µl of Protein A/G Dynabeads for 2 hours at 4°C. Beads were sequentially washed once with each of the following buffers: low-salt buffer, high-salt buffer, LiCl buffer and Tris-EDTA buffer. Complexes were then eluted from the beads in buffer containing 50 mM Tris-HCl (pH 8.0), 1% SDS and 10 mM EDTA. After reverse crosslinking (overnight at 65°C) and proteinase K treatment, DNA was extracted (Merck Millipore) for further high-throughput sequencing or for qPCR analysis of target genes. ChIP-seq libraries

were sequenced to a depth of ~100 million 50-bp single-end reads or 150-bp paired-end reads per sample by Novogene Solutions (China) with the Hiseq 2500-SE50 or PE150 platform.

All analyzed reads passed quality control with FASTQC and were mapped onto hg19 with bowtie2. After removal of duplicates with picard, the bam data were ranked, and an index was created with samtools. We used MACS2 to generate peak calls with input as the control and set a 200-bp band-width in the shifting model to denote the peak with $p < 0.01$. We identified 20734 and 825 peaks in WT and *ERF* KO HUDEP-2 cells, respectively.

ChIP-qPCR was performed on a Bio-Rad real-time qPCR system with SYBR green. The *ETS2* promoter served as the positive control, and *MYOD1* served as the negative control. The relative enrichment of ERF on the β-globin cluster was determined relative to the enrichment of IgG. The primers used in qPCR are listed in **Table S4**.

To evaluate whether DNMT3A and LRF bound to the *ERF* promoter and/or γ-globin promoter, ChIP were performed by using anti-DNMT3A and anti-LRF antibodies. *ERF* KO and WT HUDEP-2 cells were used to analysis.

qPCR, Western blotting

Total RNA extracted from cell lines or GPA^+ cells using TRIzol reagent (Life Technologies) was reverse-transcribed into cDNA using a PrimeScript RT Reagent Kit with gDNA Eraser (Takara, China). A comparative qPCR assay with SYBR green dye-containing SuperArray PCR master mix (Takara) was performed on an ABI Prism 7900 system (Life Technologies) with *ACTB* as a reference gene. All the primers used in this study were synthesized by Invitrogen (**Table** S**4**).

Western blotting was performed with anti-ERF (ab153726, Abcam), anti-BCL11A (sc-514842, Santa Cruz), anti-LRF (sc-33683, Santa Cruz), anti-GAPDH (129-10312, Ray antibody), or anti-HbF (ab137096, Abcam) antibodies.

Dual luciferase reporter assay

We inserted the ERF binding fragments (UEBS or DEBS) as enhancers into the vector pGL3, in which the luciferase reporter was preceded by *HBG* promoter, using the restriction sites *Kpn*I and *Hind*III. The ERF coding region was inserted into the pcDNA3.1 vector using *Hind*III/*Xho*I restriction sites to generate an ERF expression vector (ERF-pcDNA3.1). The reporter plasmids were co-transfected with ERF-pcDNA3.1 and PRLTK into HUDEP-2 cells with the 4D-Nucleofector System (Lonza, Switzerland). Dual luciferase activity (Promega) was measured with a Wallac Victor V 1420 Multilabel Counter (PerkinElmer, San Jose, CA, USA) 24 hours following transfection of HUDEP-2 cells.

Multiplex Ligation-dependent Probe Amplification (MLPA)

To examine off-target effects of editing due to the duplicated nature of the *HBG1* and *HBG2* genes, we verified the target region cut within the β-globin gene cluster through copy number analysis with SALSA MLPA Probe P102-C1 HBB (MRC-Holland). In brief, DNA was diluted to 28–30 ng/μl, and 4 μl DNA was added to each tube and placed in a thermocycler. The thermocycler program was started at 98°C for 20 minutes, and then paused at 25°C. We ensured that samples were at 25°C before removing tubes from the thermocycler. Then 0.75 μl MLPA buffer and 0.75 μl probe mix were added and incubated at 95°C for 1 minute and 60°C for 16 hours, then held at 54°C. Subsequently, 12.5 μl ultrapure water, 1.5 μl ligase buffer A, 3 μl ligase buffer B and 0.5 μl Ligase-65 enzyme were added, and the following thermocycler program was used: 54°C for 15 minutes, 98°C for 5 minutes and hold at 20°C. PCR was performed as follows: 3.75 μl ultrapure water, 1 μl

SALSA PCR primer mix and 0.25 µl polymerase master mix were added and incubated in the thermocycler for 35 cycles at 95°C for 30 seconds, 60°C for 20 seconds and 72°C for 60 seconds. After 35 cycles, DNA was incubated at 72°C for 20 minutes and held at 15°C. The PCR products were analyzed with the GenomeLab GeXP Genetics Analysis System (BeckMan).

Supplemental References

1. Bagger, F.O., Kinalis, S., and Rapin, N. (2019). BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. Nucleic Acids Res 47, D881-d885.

2. 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.

3. Chen, D.Y., Zuo, Y.J., Zhang, X.H., Ye, Y.H., Bao, X.Q., Huang, H.Y., Tepakhan, W., Wang, L.J., Ju, J.Y., Chen, G.F., et al. (2017). A Genetic Variant Ameliorates beta-Thalassemia Severity by Epigenetic-Mediated Elevation of Human Fetal Hemoglobin Expression. American Journal of Human Genetics 101, 130-138.

4. Shang, X., Peng, Z., Ye, Y., Asan, Zhang, X., Chen, Y., Zhu, B., Cai, W., Chen, S., Cai, R., et al. (2017). Rapid Targeted Next-Generation Sequencing Platform for Molecular Screening and Clinical Genotyping in Subjects with Hemoglobinopathies. EBioMedicine 23, 150-159.

5. Xiong, F., Sun, M., Zhang, X., Cai, R., Zhou, Y., Lou, J., Zeng, L., Sun, Q., Xiao, Q., Shang, X., et al. (2010). Molecular epidemiological survey of haemoglobinopathies in the Guangxi Zhuang Autonomous Region of southern China. Clin Genet 78, 139-148.

6. Xu, X.M., Zhou, Y.Q., Luo, G.X., Liao, C., Zhou, M., Chen, P.Y., Lu, J.P., Jia, S.Q., Xiao, G.F., Shen, X., et al. (2004). The prevalence and spectrum of alpha and beta thalassaemia in Guangdong Province: implications for the future health burden and population screening. J Clin Pathol 57, 517-522.

7. Karimi, M., Cohan, N., De Sanctis, V., Mallat, N.S., and Taher, A. (2014). Guidelines for diagnosis and management of Beta-thalassemia intermedia. Pediatr Hematol Oncol 31, 583-596.

8. Galanello, R., and Origa, R. (2010). Beta-thalassemia. Orphanet J Rare Dis 5, 11.

9. Liu, D., Zhang, X., Yu, L., Cai, R., Ma, X., Zheng, C., Zhou, Y., Liu, Q., Wei, X., Lin, L., et al. (2014). KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of beta-thalassemia. Blood 124, 803-811.

10. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571-1572.

11. Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol 13, R83.

12. Lei, Y., Zhang, X.T., Su, J.Z., Jeong, M., Gundry, M.C., Huang, Y.H., Zhou, Y.B., Li, W., and Goodell, M.A. (2017). Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. Nat Commun 8, 16026.

13. Pun, F.W., Zhao, C., Lo, W.S., Ng, S.K., Tsang, S.Y., Nimgaonkar, V., Chung, W.S., Ungvari, G.S., and Xue, H. (2011). Imprinting in the schizophrenia candidate gene GABRB2 encoding GABA(A) receptor beta(2) subunit. Mol Psychiatry 16, 557-568.

14. McIver, S.C., Hewitt, K.J., Gao, X., Mehta, C., Zhang, J., and Bresnick, E.H. (2018). Dissecting Regulatory Mechanisms Using Mouse Fetal Liver-Derived Erythroid Cells. Methods Mol Biol 1698, 67-89.