The American Journal of Human Genetics, Volume *108*

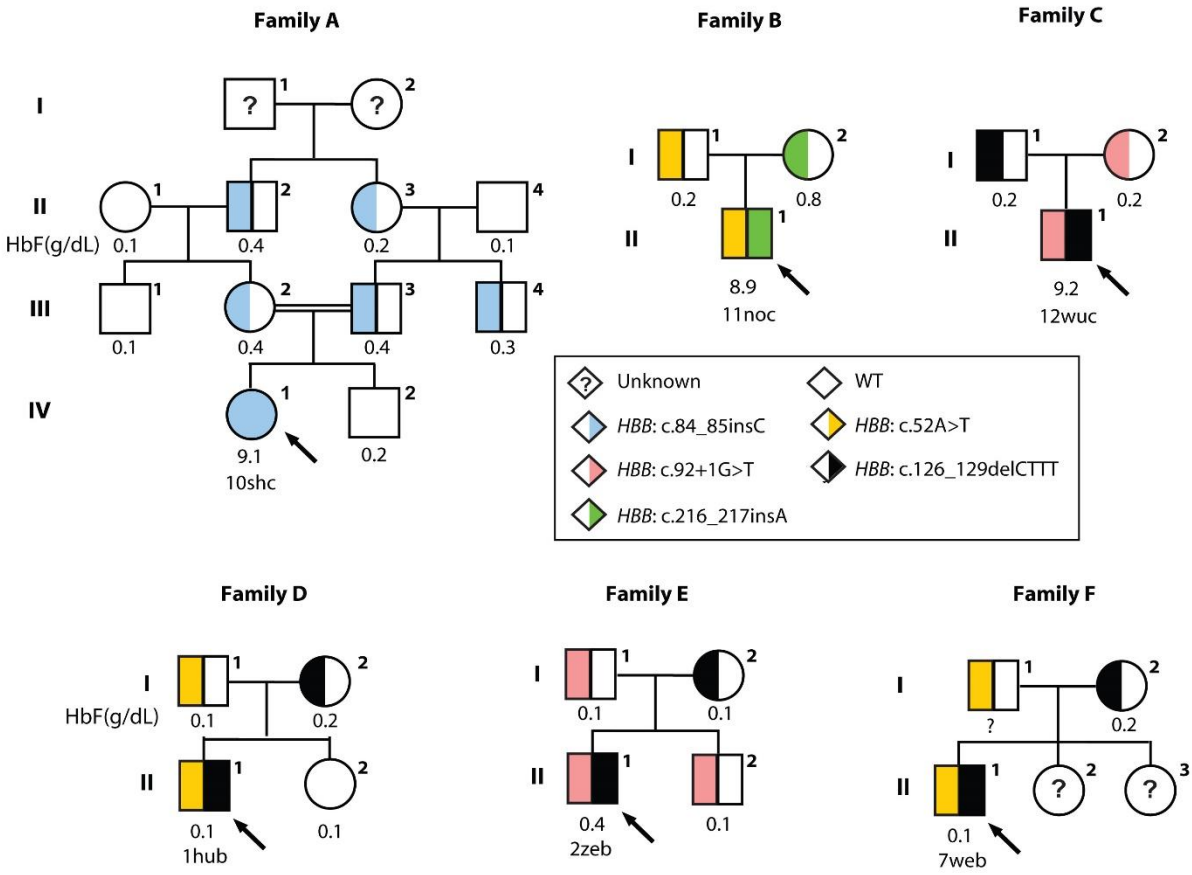
Supplemental information

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

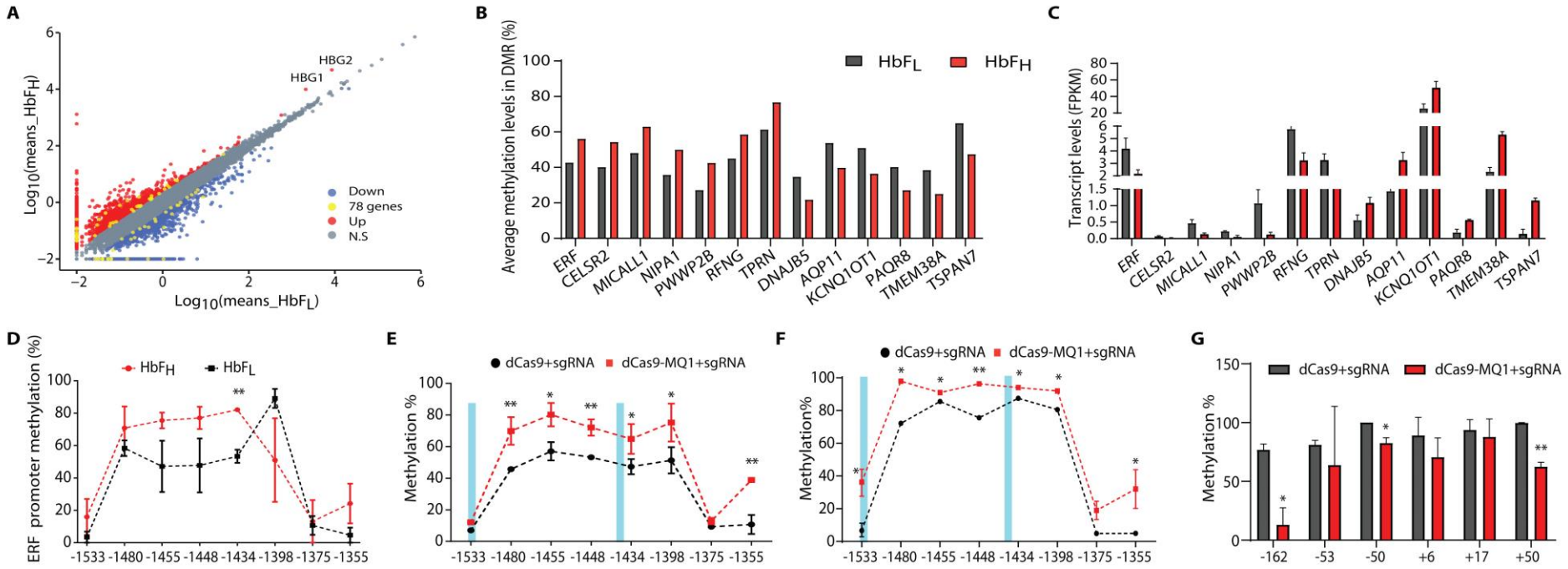
Supplemental Data

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.

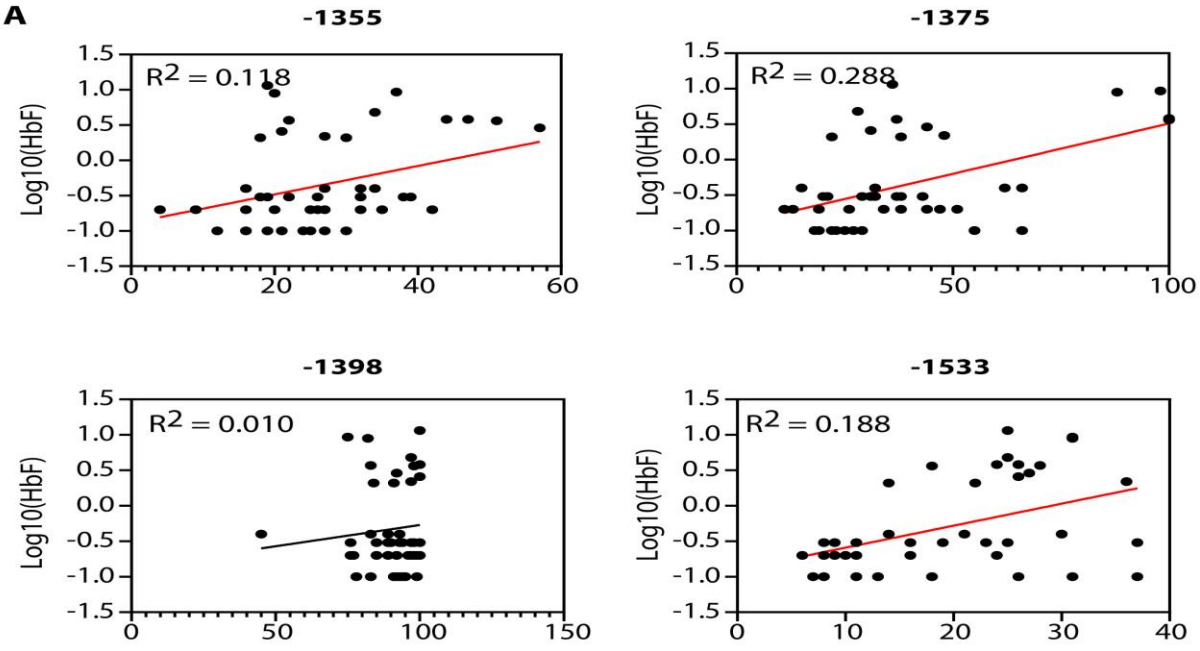
Figure S2. Methylation and expression profiles of 13 candidate genes.



(**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

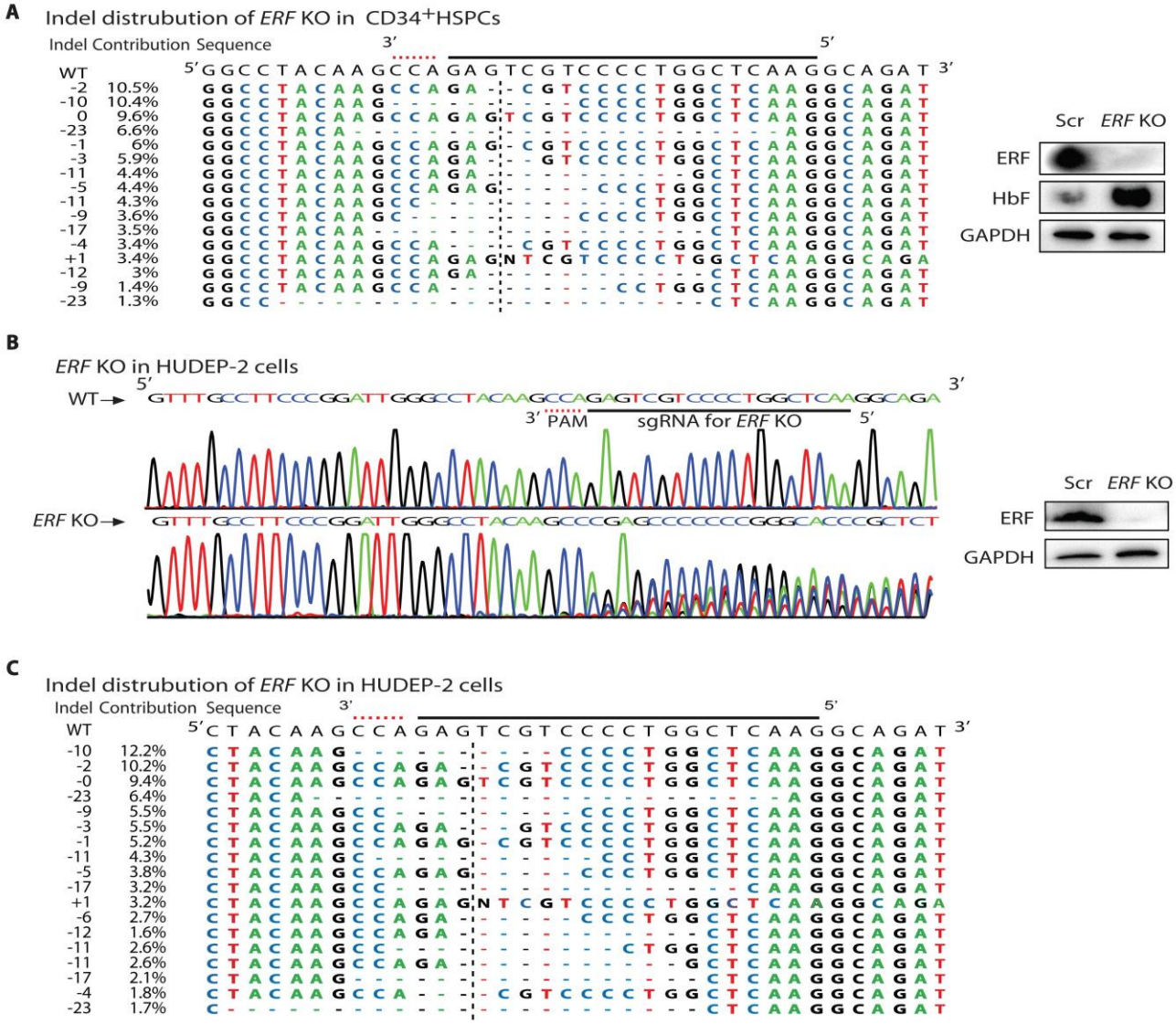


B

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

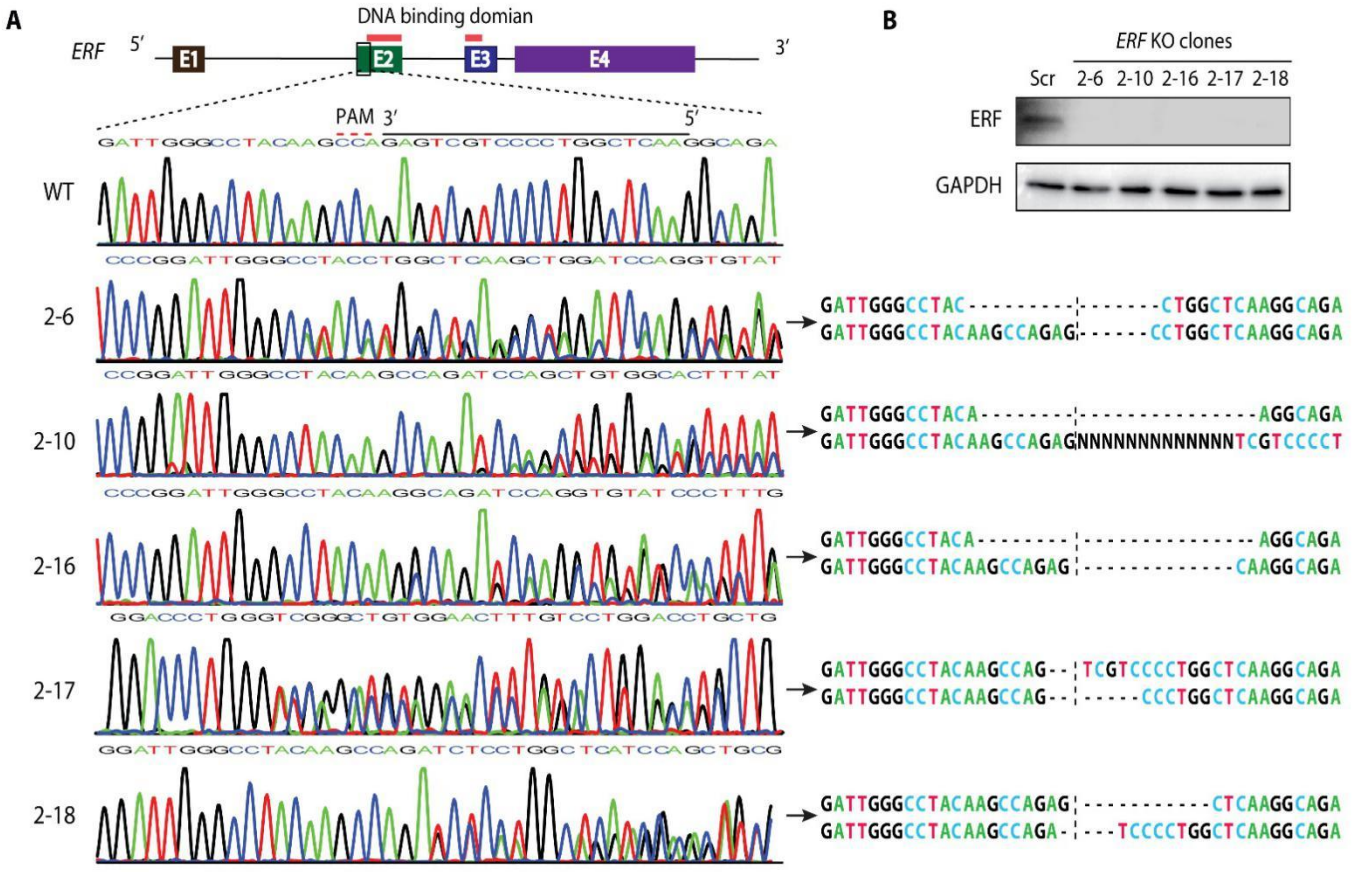
(A) The scatter diagram shows the correlation between HbF level and *ERF* methylation of the four CpG sites (-1355, -1375, -1398, -1533) in *ERF* promoter. **(B)** The statistics of the regression analysis. B, beta, indicated partial regression coefficient. R^2, coefficient of determination. CI, Confidence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

Figure S4. *ERF* KO efficiency.



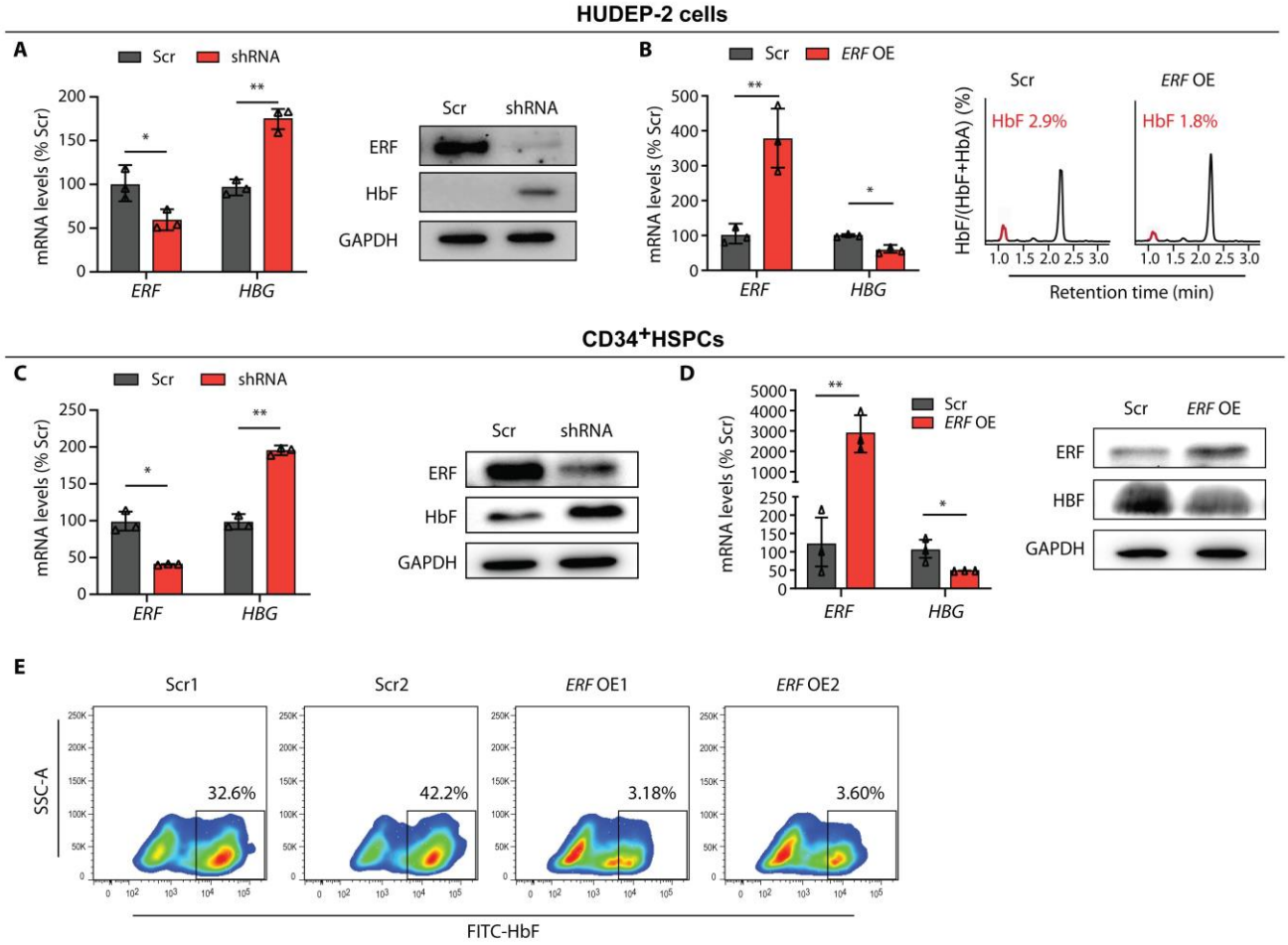
(A) Left: The indel distribution pattern of *ERF* KO in CD34^{+} HSPCs. The black line represents sgRNA targeting *ERF*. The red dashed line indicates PAM. Right: Western blotting analysis of the KO efficiency in CD34^{+} HSPCs. **(B)** Left: Representative sequencing chromatographs of PCR products of *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.



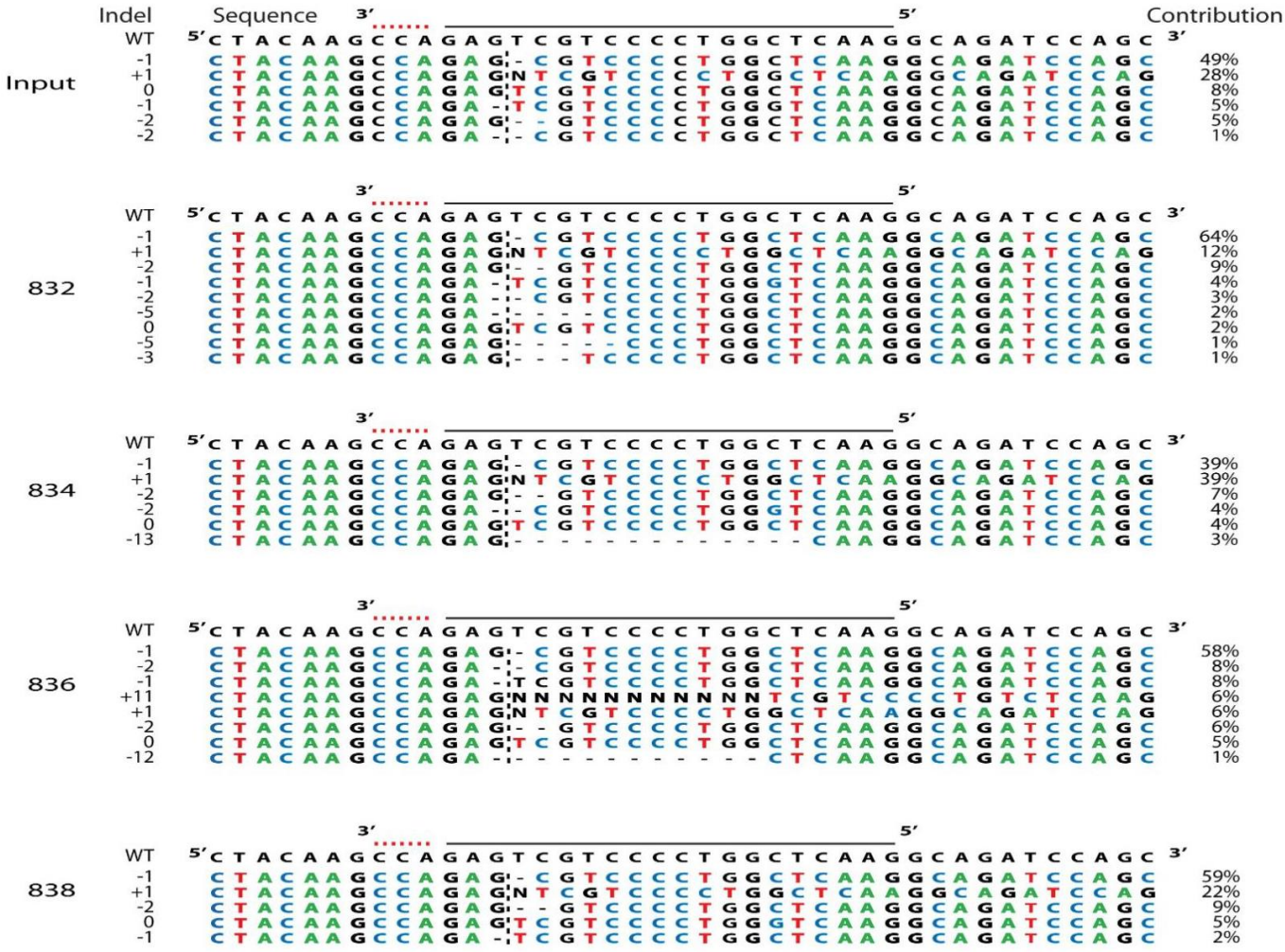
(A) Top: schematic of the *ERF* gene structure. E1, Exon 1. Bottom: sequencing chromatographs of PCR products of *ERF* WT and KO HUDEP-2 clones. The black line represents sgRNA targeting *ERF*. The red dashed line indicates PAM. Right: the indel pattern of each *ERF* KO clone; the black dashed line indicates 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.**



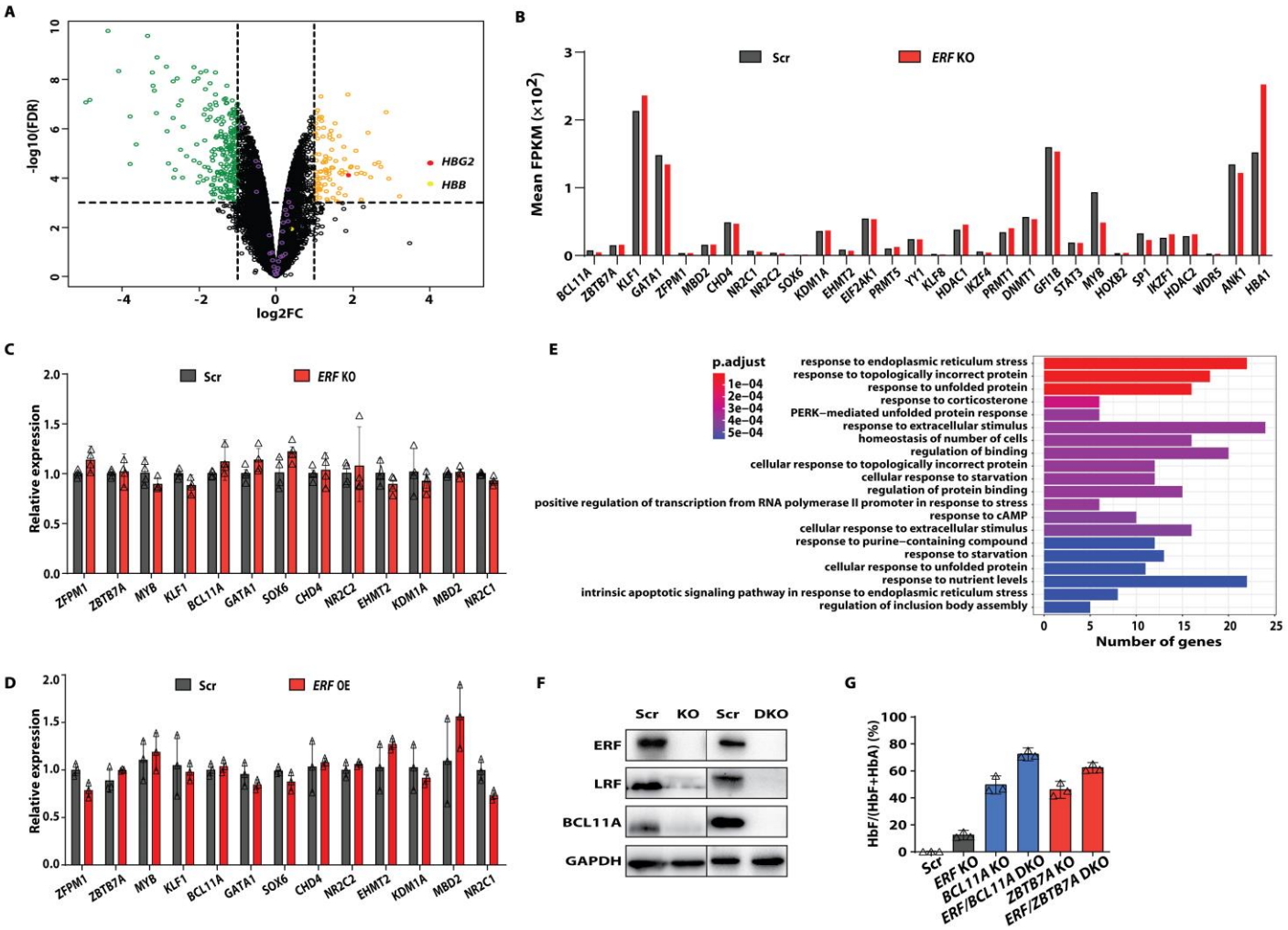
(A) Analysis was performed with qPCR analysis (left) and western blotting (right) in ERF KD HUDEP-2 cells. **(B)** Left: qPCR analysis of γ-globin and *ERF* expression in HUDEP-2 cells with *ERF* OE. Right: representative HPLC profiles of HUDEP-2 cells without (Scr) or with *ERF* OE. **(C)** qPCR analysis (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.



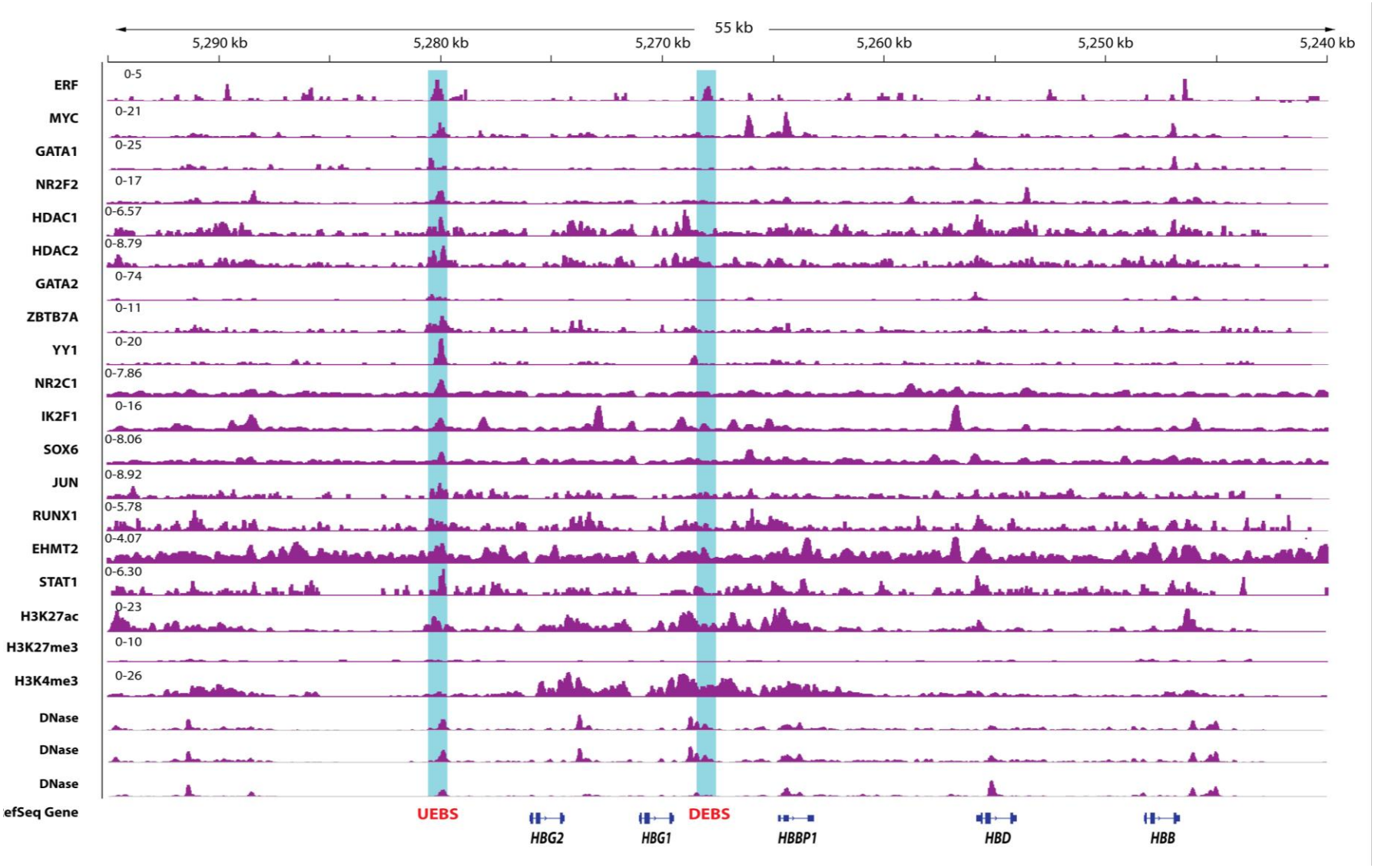
The black line shows the position of sgRNA. The red dashed line shows the PAM. The black dashed line shows the cutting site of sgRNA. The numbers 832, 834, 836 and 838 indicate mouse IDs.

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



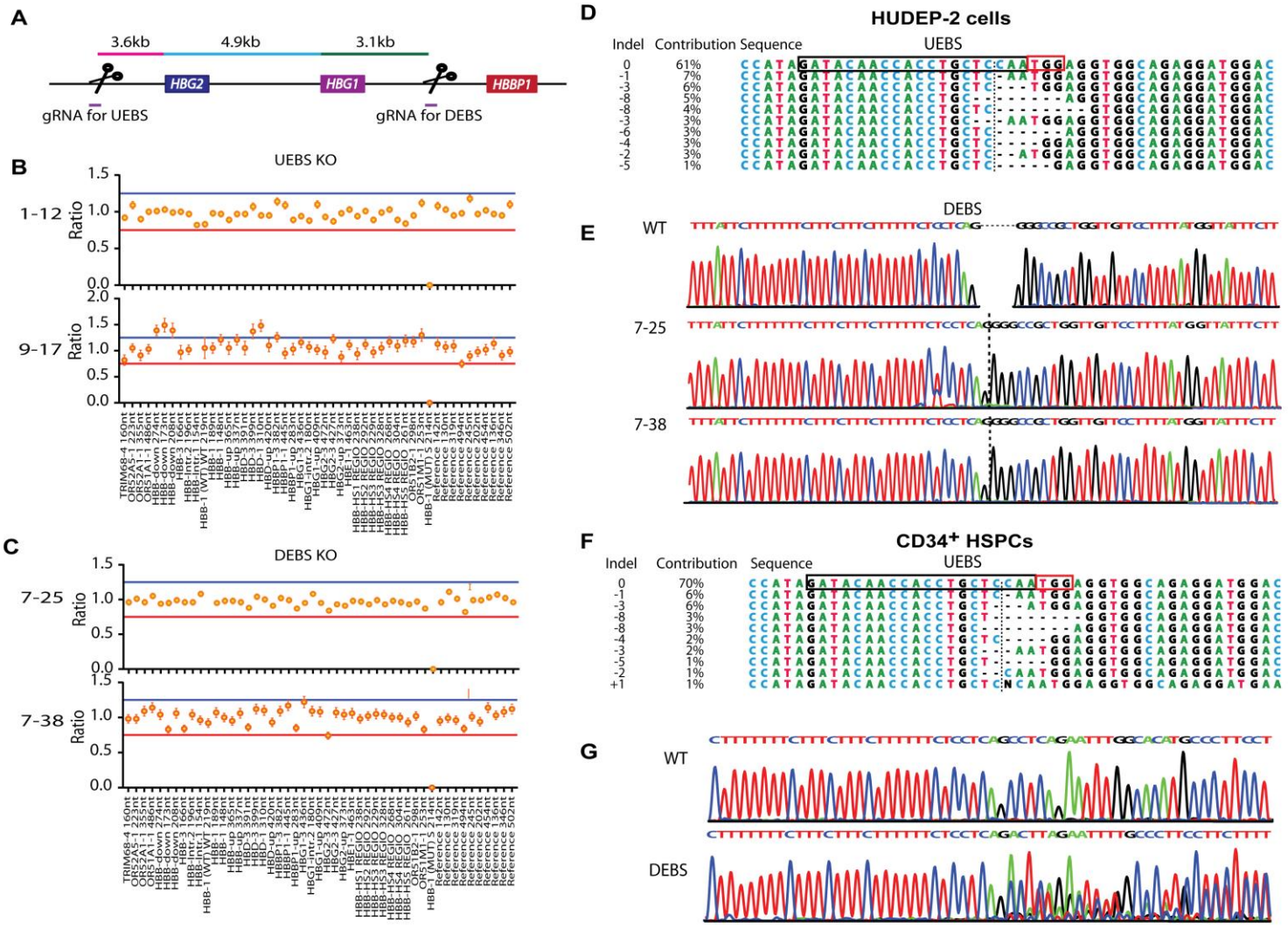
(A) RNA-seq was performed in *ERF* KO HUDEP-2 cells. The green and orange circles denote significant transcripts according to the screening standards: fold change (FC) > 2, FDR < 10^{-3}. The purple circles represent the known γ-globin repressors shown in **(B)**. Red solid circle: *HBG2*. Blue solid 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.



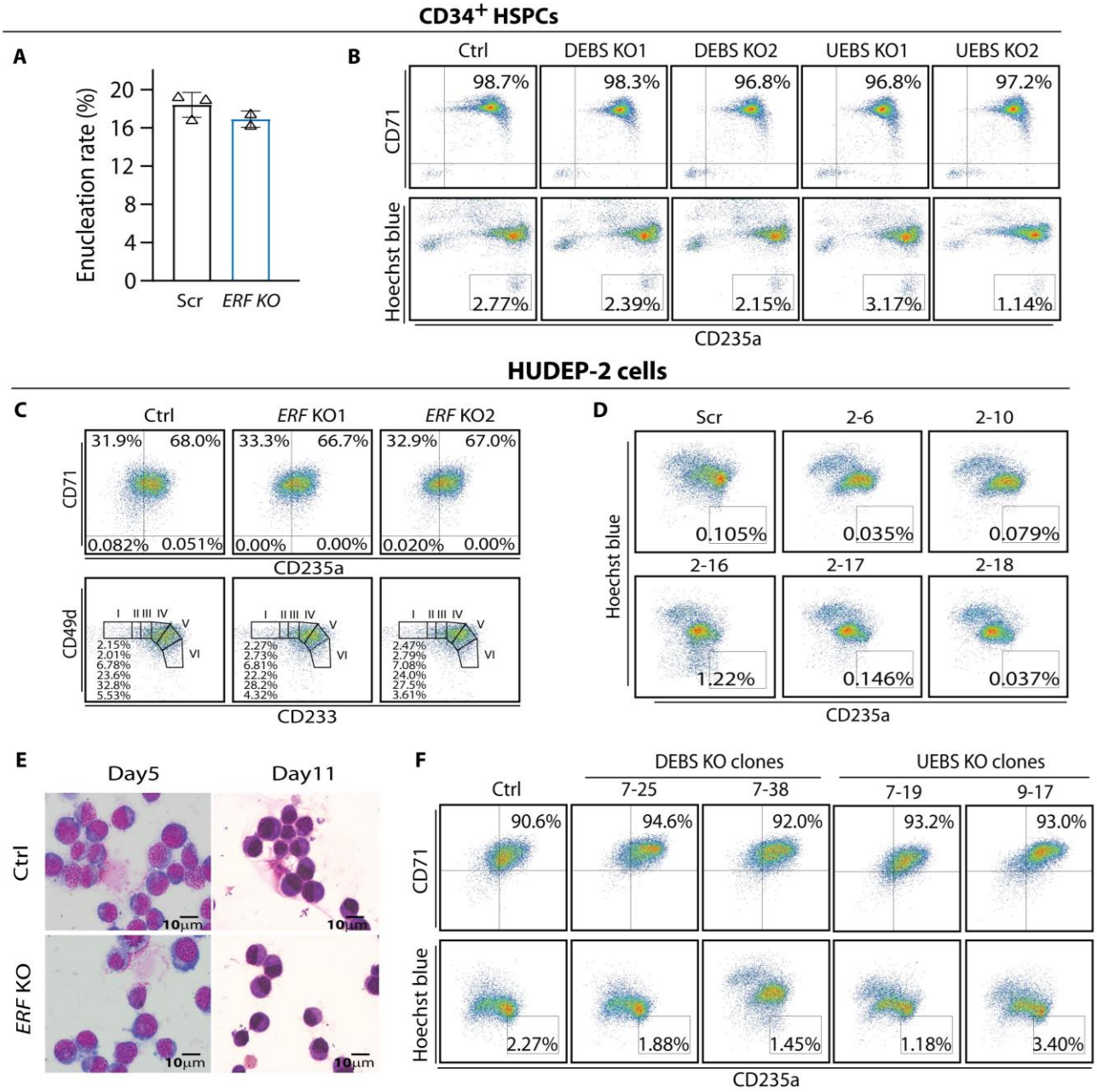
The ChIP-seq data were attained from the ENCODE. The data ID of each transcription factor are as followed: MYC (ENCFF112BJR), GATA1 (ENCFF226FPS), NR2F2 (ENCFF335LTU), HDAC1 (ENCFF355ZSE), HDAC2 (ENCFF532SIS), ZBTB7A (ENCFF457PSO), YY1 (ENCFF480ZET), NR2C1(ENCFF490IPR), IK2F1 (ENCFF715OCK), SOX6 (ENCFF757HAQ), 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.



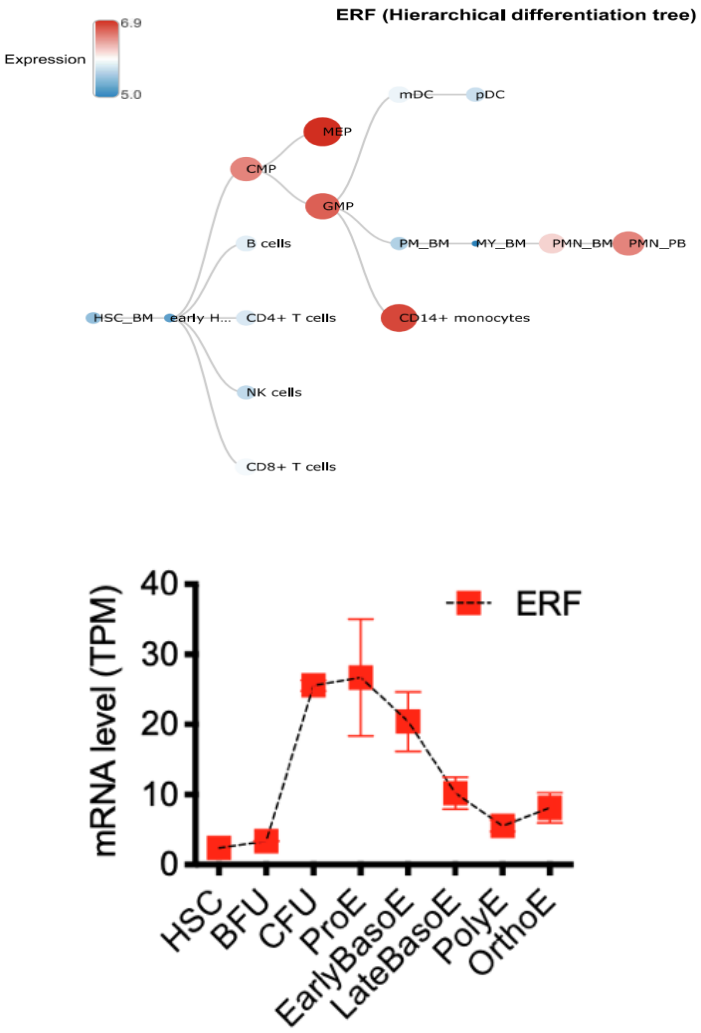
(A) Map of *HBG2*, *HBG1* and *HBBP1* showing cleavage sites for Cas9/gRNA (scissors). **(B, C)** Representative MLPA profiles of several UEBS **(B)** or DEBS KO **(C)** HUDEP-2 clones. The ratios of all probes ranged from 0.75 (red line) to 1.25 (blue line), which indicates normal copy number. **(D, 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.