The American Journal of Human Genetics, Volume *101*

Supplemental Data

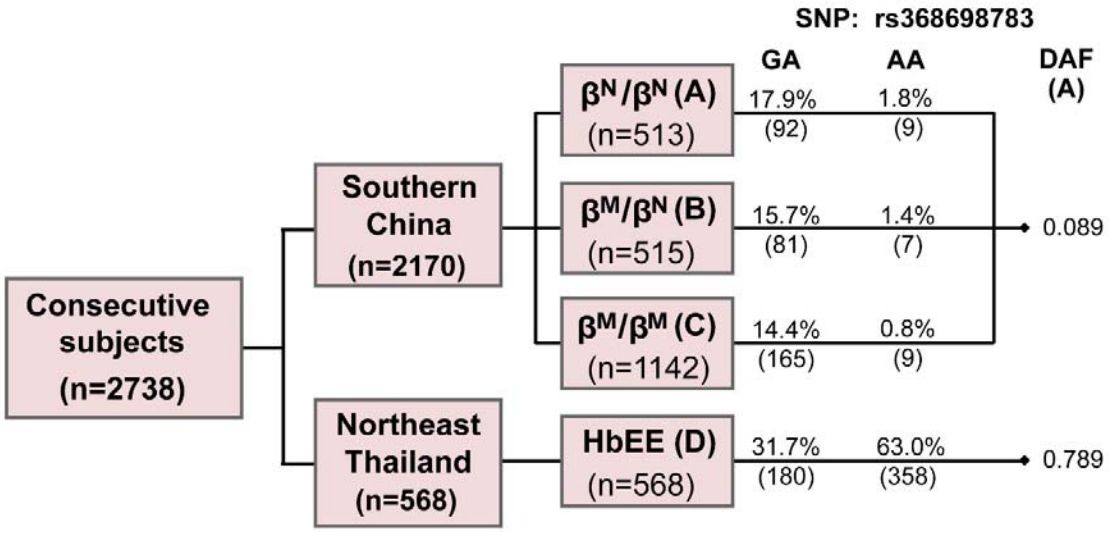
A Genetic Variant Ameliorates β-Thalassemia Severity by Epigenetic-Mediated Elevation of Human Fetal Hemoglobin Expression

Diyu Chen, Yangjin Zuo, Xinhua Zhang, Yuhua Ye, Xiuqin Bao, Haiyan Huang, Wanicha Tepakhan, Lijuan Wang, Junyi Ju, Guangfu Chen, Mincui Zheng, Dun Liu, Shuodan Huang, Lu Zong, Changgang Li, Yajun Chen, Chenguang Zheng, Lihong Shi, Quan Zhao, Qiang Wu, Supan Fucharoen, Cunyou Zhao, and Xiangmin Xu

Figure S1. LD Blocks identified in the β-globin cluster.

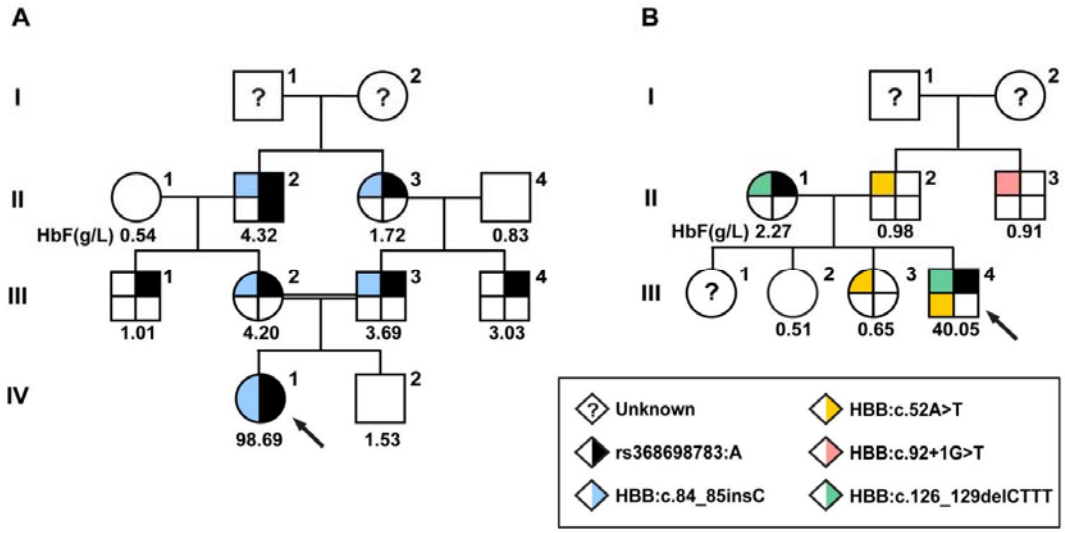
Seven LD blocks containing 163 out of 271 common SNPs were identified in the 80-kb region based on Haploview. Independent LD blocks with highly LD were indicated by black triangle frames. Small box represents D' with strong LD (red), no LD (white) or lack of statistical evidence (blue). See the attached JPG file for Figure S1.

Figure S2. Sample design of the present study.



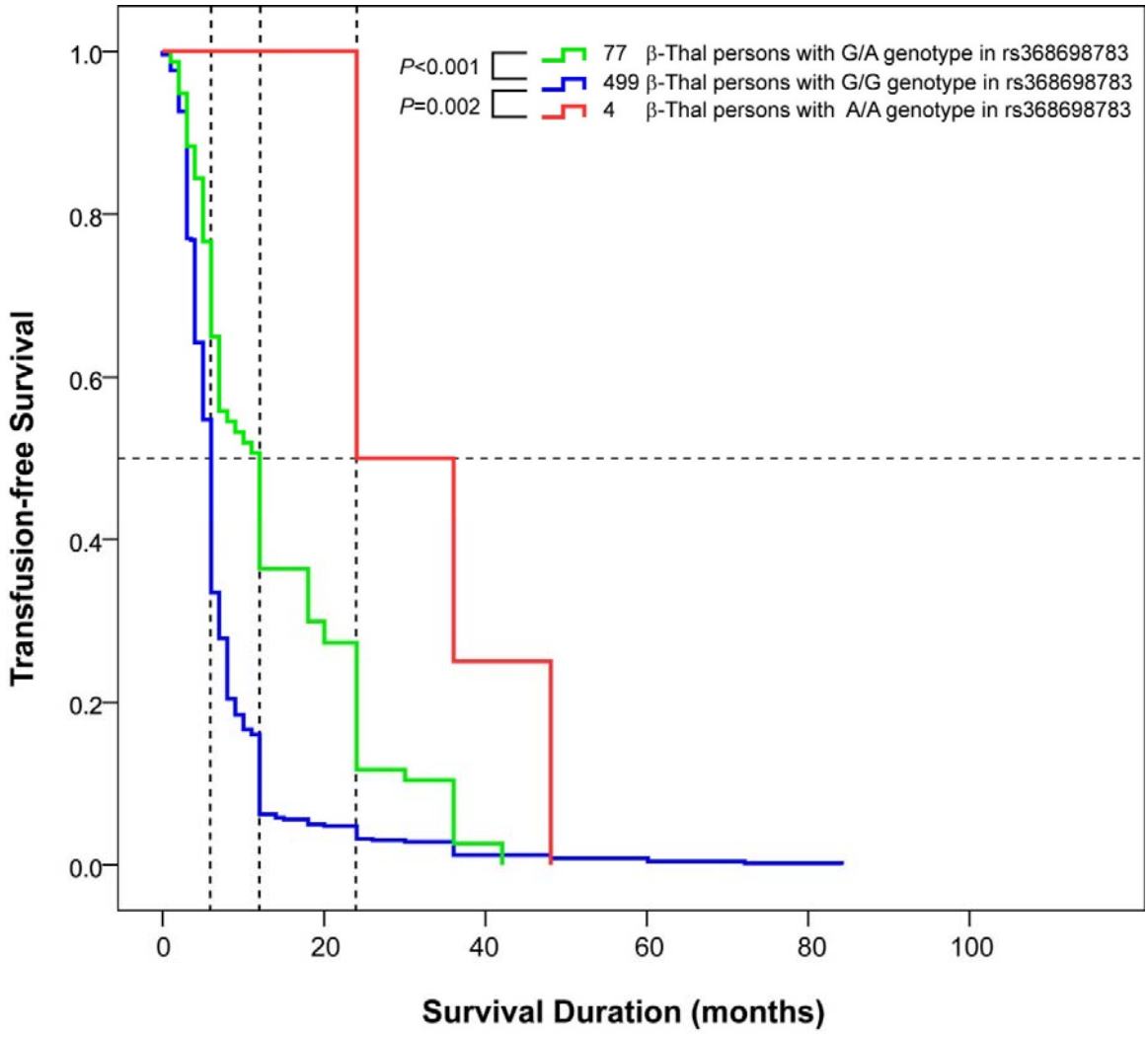
Four consecutive cohorts (n=2738), including 2170 participants from southern China and 568 Thai individuals with HbEE disease, were recruited for this study. The frequency and number of GA and AA genotypes of rs36869873 in each of four cohorts, as well as the derived allele (A) frequency (DAF) in southern Chinese and northeastern Thai subpopulations, are shown. All HbEE individuals had the homozygous AA genotype for *HBB* (c.79 G>A). Definition of thalassemia major (TM) or TI in this study is based on the following 4 clinical indications as described:[1,2] (1) onset of anemia: < 6 months, 6-24 months (TM), or >24 months (TI); (2) transfusion before 4 years of age: symptomatic anemia requiring more than 8 transfusions/year before 4 years of age (TM) or none/occasional transfusion before 4 years of age (TI); (3) steady-state hemoglobin levels: <60 g/L (TM) or 60-100 g/L (TI); (4) liver/spleen enlargement: severe (>4 cm, TM) or moderate (0-4 cm, TI); (5) growth and development: delayed (TM) or normal (TI).

Figure S3. Pedigree analysis.



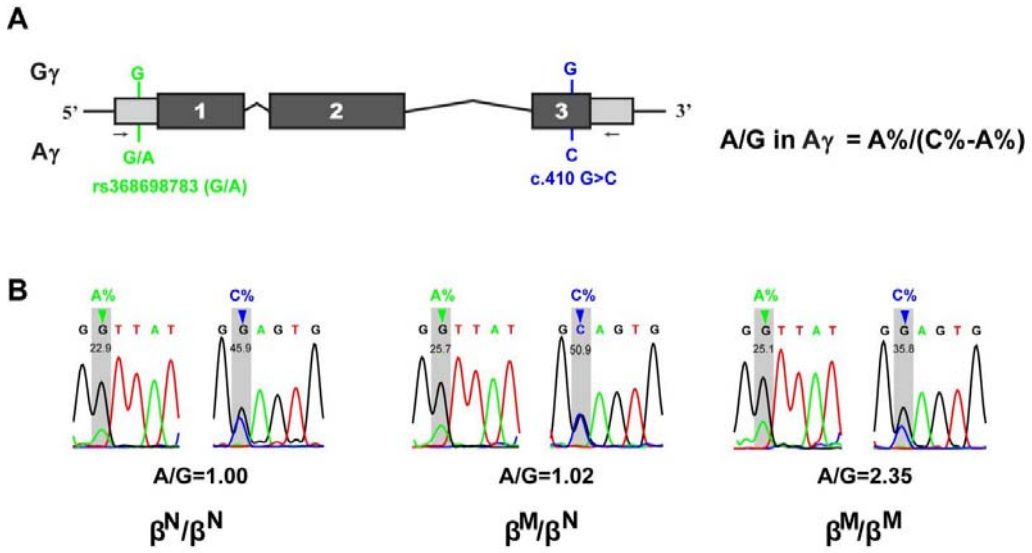
Pedigrees for families A **(A)** and B **(B),** with *HBB* gene mutations and HbF levels indicated for each family member. HbF (g/L) was calculated based on the total Hb level and HbF (%). Squares, males; circles, females; arrows, the proband in each family. To obtain data reflecting the endogenous hemoglobin levels as much as possible, the Hb levels (g/L) were untransfused or pre-transfusion data and HbF (g/L) was calculated from total Hb level and HbF (%) using our previous methods.[3] Hematological parameters were assessed with an automated hematology analyzer (Sysmex F-820; Sysmex, Japan), and hemoglobin analysis was performed using high-performance liquid chromatography (Variant II, Bio-Rad Laboratories, USA). Determination of human γ-globin peptide level was performed as previously described.[3,4] Briefly, 50μl of venous blood were diluted in 1 ml deionized water. Samples were centrifuged at 3000r/min for 10 min to remove cells debris. An equal volume of plasma was prepared in the same way for a blank control. We used a Shimadzu LC-20AT chromatographic system (Shimadzu, Kyoto, Japan), chromatographic separation with a Jupiter C18 HPLC column (4.6 mm×250 mm,5μm, 300A, Phenomenex, Torrance, CA, USA) and a SecurityGuard C18 column (4.0 mm×30 mm, 5μm, 300A, Phenomenex). Relative quantification was carried out by measuring the percentage of the peak area of the heme and globin chains with UV detection at 280 nm. To obtain DNA/RNA, peripheral blood was mixed with an equal volume of RBC lysis buffer (30 mM Tris-HCl, 5 mM EDTA, 50 mM NaCl) and centrifuged at 3000 rpm for 10 min at 4°C. Supernatant was discard and pellet containing leukocytes and reticulocytes were used for DNA extraction using standard phenol and chloroform method and RNA extraction using Trizol reagent (Life Tech, USA).

Figure S4. Kaplan-Meier survival curve analysis.



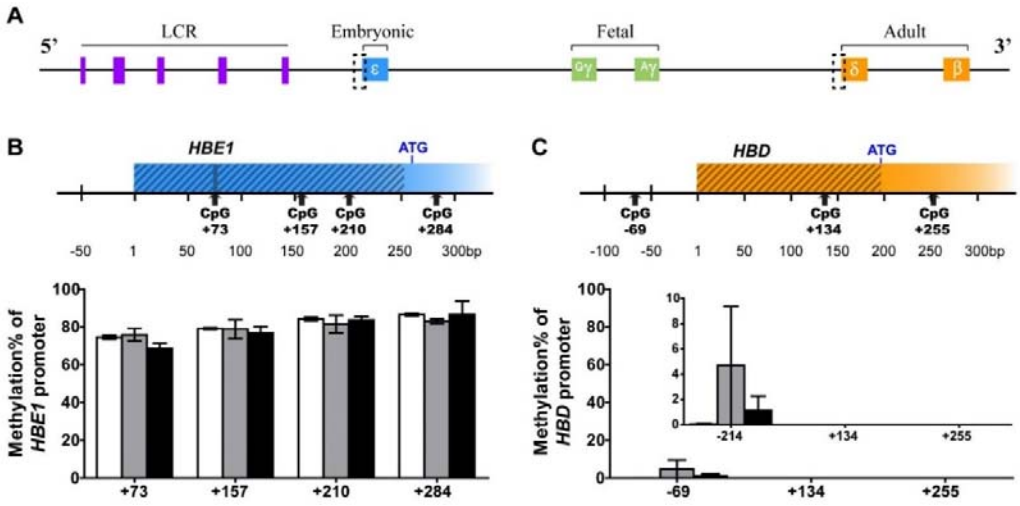
Chinese β-thalassemia individuals with similar genetic variations at the genotypes β^0/β^0, αα/αα, *KLF1* (WT), *BCL11A*-rs4671393 (GG or GA), and *HBS1L-MYB*-rs9399137 (TT or TC) analyzed by Sanger sequencing or high-resolution melting (HRM) as described [3] were recruited into our study as described in **Table S2**. Survival curve analysis was generated using the Kaplan-Meier log rank test in SPSS v20.0 to compare the median age at first transfusion (Survival duration) between individuals with GG genotypes (n=500) and those with GA genotypes (n=77) or AA genotypes (n=4) for the rs368698783 polymorphism.

Figure S5. Analysis of genotype-dependent RNA expression by RT sequencing.



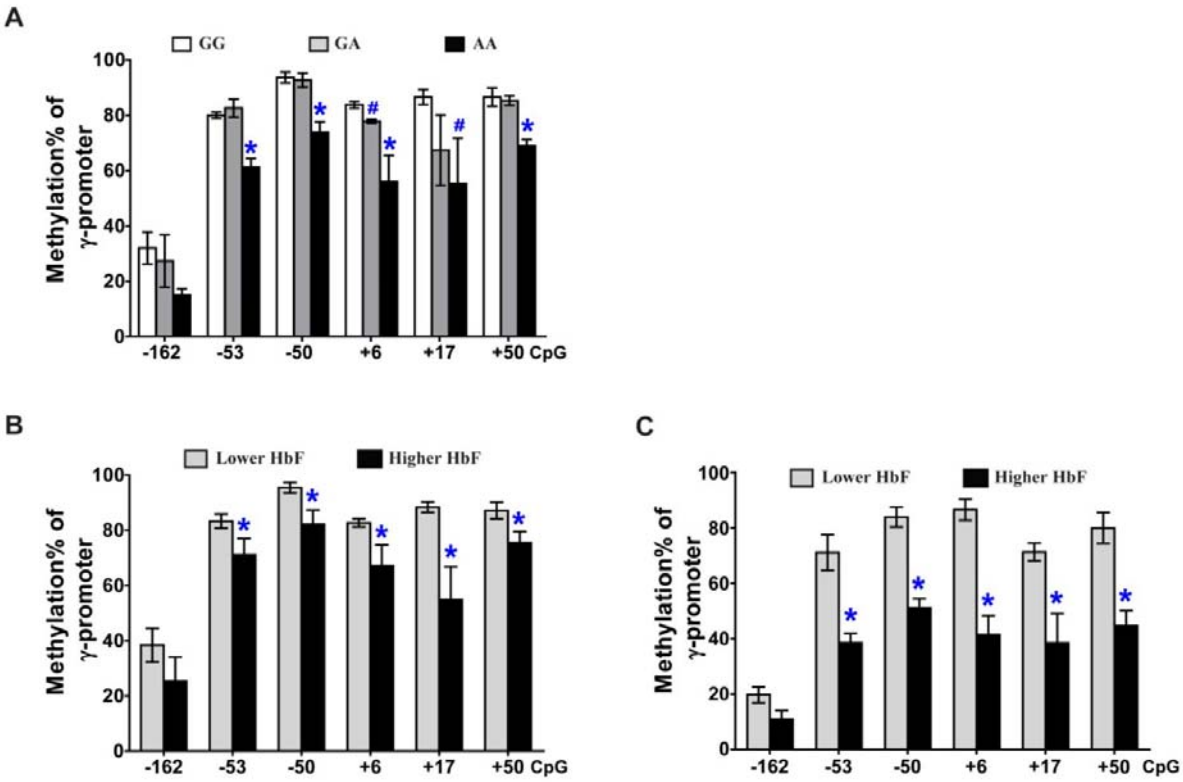
(A) The structure of the γ-globin gene and the locations of markers employed for quantification of the $^{G}\gamma/^{A}\gamma$ ratio or the *HBG1*-rs368698783 allelic A/G ratio. The grey boxes represent non-coding exons, black boxes represent coding exons, and solid lines represent introns. To quantify the ratio of $^{G}\gamma/^{A}\gamma$-globin mRNA expression shown in Figure 1C, the *HBG*: c.410G>C polymorphism in exon 3 of *HBG* mRNA was used as a marker and the allelic expression of $^{A}\gamma$-rs368698783 (**Figure 1E**) was determined based on the G (in $^{G}\gamma$ mRNA only) and C (in $^{A}\gamma$ mRNA only) allele peaks observed on the sequencing chromatographs from the reverse-transcript PCR products obtained using the BioEdit Sequence Alignment Editor.[5] Reverse transcription from total RNA was performed to generate cDNA template using the PrimeScript RT Reagent Kit (Takara, Dalian, China). Reaction was carried out with 2 μg of total RNA, random hexamers and PrimeScript RT Reagent Kit (Takara) for 10 min at 25°C, 30 min at 48°C, 5 min at 95°C, and stopped by the addition of 10 nM ethylene diaminetetra acetic acid. The $^{G}\gamma$ mRNA expression was determined by quantification of the G allele of c.410 in the *HBG* gene. The $^{A}\gamma$ mRNA expression was determined by quantification of the C allele of c.410 in the *HBG* gene. The rs368698783 allelic A/G ratio in *HBG1* mRNA was calculated by A%/(C%-A%), where A% or C% represents the allele peak frequency on the sequencing chromatograph. **(B)** The representative sequencing chromatographs of RT-PCR products from non-thalassemia controls (β^{N}/β^{N}), β-thalassemia carriers (β^{M}/β^{N}), or β-thalassemia individuals (β^{M}/β^{M}) with heterozygous GA genotypes for SNP rs368698783. The *HBG1*-rs368698783 allelic A/G ratio as determined by mRNA analysis for each of three representative samples is indicated.

Figure S6. Determination of DNA methylation in *HBE1* and *HBD* loci.



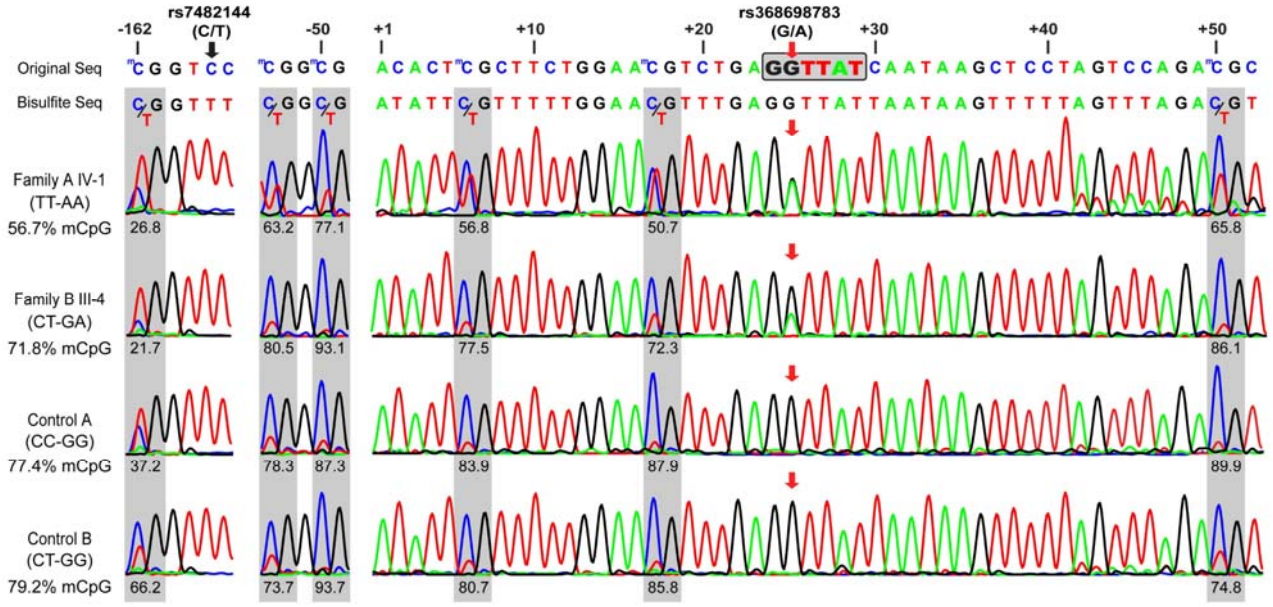
(A) A diagram of the human β-globin cluster. The panels of **(B)** and **(C)** show the locations of CpGs at *HBE1* and *HBD*, respectively. The gray hatched region represents the 5'UTR of *HBE1* **(B)** and *HBD* **(C)**. The effects of the rs368698783 genotypes on the methylation of the core and flank regions of the proximal promoters of *HBE1* and *HBD* in CD235a+ erythroblasts from ten β^0/β^0 thalassemia individuals (GG=4, GA=3, AA=3) are shown. The mean methylation percentage from BS-seq method is shown in the columns (white, GG; gray, GA; black, AA), with the standard error indicated by bars. There were no significant differences between the CpGs of these loci. Bisulfite modification of genomic DNA was performed using sodium metabisulfite (2.0 M) and hydroquinone (0.5 mM) as described.[5] Briefly, DNA was performed by denaturing 1 mg genomic DNA with 0.3M NaOH at 42°C for 20 min, followed by 95°C for 3min and 0°C for 1 min, and incubating at pH 5.0 with sodium metabisulfite (2.0M) and hydroquinone (0.5mM) at 55°C for 16h in the dark overlaid with mineral oil. Modified DNA was purified with Promega Wizard DNA Clean Up System (Madison, WI, USA). The eluted DNA was incubated with NaOH (0.3M) at 37°C for 15 min and neutralized by 3M NH4-acetate to pH 7.0. The neutralized DNA was precipitated by 75% ethanol and recovered in 20 ml of TE buffer. The target DNA segments in the bisulfite-modified DNA were amplified with nested-PCR primers (**Table S8**). The methylation levels of the target CpG sites in the β-globin gene cluster were determined by cloning the PCR products into the pEASY-T5 Zero cloning vector (TransGene, China) and sequencing 10 to 12 clones (BS-clone method; **Figure 2B**). Alternatively, methylation levels were determined by direct sequencing (BS-seq method, **Figure S8**) and validated based on the C and T allele peaks observed on the sequencing chromatographs obtained using the BioEdit Sequence Alignment Editor v7.0.9.0 (Carlsbad, USA).[5]

Figure S7. Determination of DNA methylation of the *HBG* promoter.



(A) The effects of the *HBG1*-rs368698783 genotypes on *HBG* promoter methylation were determined by the BS-seq method in the β^M/β^M thalassemia individuals. The mean methylation frequency for six CpG sites around the γ-globin promoter is shown in the columns (white=GG; gray=GA; black=AA), with the standard error indicated by bars. Each group contained three subjects. * denotes the AA genotypes as significantly different (P<0.05) from both the GG and GA genotypes. # denotes the GA or AA genotypes as significantly different (P<0.05) from the GG genotypes. The P value was determined using a two-tailed Student's t-test. The *HBG* promoter **(B)** methylation levels were determined using the BS-clone method and were further validated using the BS-seq method **(C)** in β^M/β^M thalassemia individuals with lower HbF (<5 g/L, n=5, grey) and higher HbF (>90 g/L, n=4, black) levels. The mean methylation levels (%) for each of the two groups are shown in the columns, with the standard error indicated by bars. * denotes significant differences in methylation levels between the lower and higher HbF groups. **n.s.** denotes no significant difference.

Figure S8. Analysis of *HBG* promoter methylation levels using the BS-seq method.

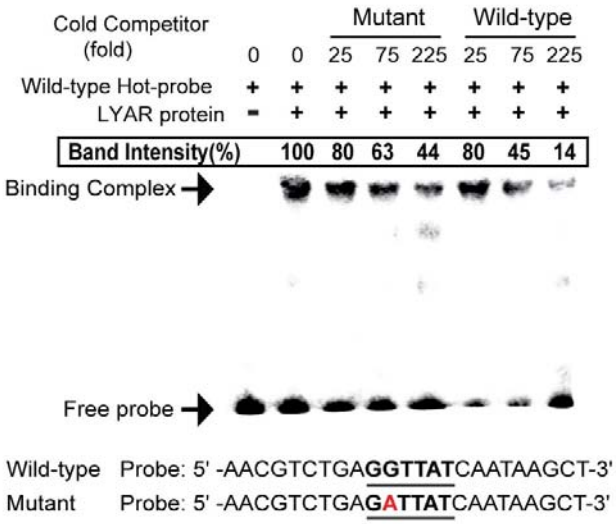


Methylation levels of the γ-globin promoter were determined by the BS-seq method for the CD235a$^+$ cells from human BM **(Figures 2A, S6A&C, S7 and S8**) or the cultured CD34$^+$ cells transduced by the LV3-LYARsiRNA lentivirus for the *LYAR* RNA interference (**Figure 4D**). Four subjects with TT-AA, CT-GA, CC-GG, and CT-GG genotype combinations of the rs7482144 and rs36869783 SNPs. The CpG sites, rs7482144 (C/T) and rs36869783 (G/A) SNPs, and LYAR binding motif (gray box) in the γ-globin gene are indicated in the original sequence. Because an unmethylated C is replaced with a T, the presence of any C in six CpG dinucleotides on the bisulfite sequence represents a methylated C allele (highlighted in grey) according to the BS-seq method. The methylation levels of each CpG dinucleotide (C/(C+T)%) are indicated under the chromatograph, and the average methylation levels of these six CpGs are shown on the left. The rs7482144 (C/T) genotype cannot be displayed in a bisulfite sequence where the unmethylated C is converted to a T. The rs368698783 SNP (G/A) exists only in the ^{A}γ-globin promoter and not in the ^{G}γ-globin promoter. Bisulfite-PCR products are mixtures of the ^{A}γ- and ^{G}γ-globin promoters such that the A allele peak is equal to the G allele peak in the rs368698783 SNP (G/A) site in the chromatograph. In Family A, the proband with the AA genotype and the A alleles has one-third as many G alleles as that in Family B with the GA genotype, and the peaks are only GG alleles for controls with the GG genotype.

The mononuclear cells of peripheral blood or bone marrow (BM) from β-thalassemia individuals were isolated by a Ficoll-Hypaque density gradient method and was then enriched for CD34$^+$ or CD235a$^+$ cells by immunomagnetic separation Microbeads Kit (Miltenyi Biotec, Germany). CD34$^+$ cells from isolated

mononuclear cells were sorted using magnetic beads that bind human CD34 (Multisort CD34 Microbeads) to obtain CD34+ cells through positive selection. $CD235a^+$ cells from isolated mononuclear cells were sorted using magnetic beads that bind human CD235a (Multisort CD235a Microbeads) to obtain $CD235a^+$ cells through positive selection. $CD34^+$ cells were cultured as previously described.[6] Briefly, isolated $CD34^+$ cells were cultured in erythroid differentiation StemSpan serum free expansion medium (Stemcell technologies, USA) supplemented with 10% FBS (Life Technologies, USA), 50 ng/mL SCF (stem cell factor, R&D systems, USA), 1 IU/mL erythropoietin (EPO, KIRIN, Japan) and 10 ng/mL Interleukin-3 (IL-3, Sigma, USA) for 6 days. From day 6, only 30% FBS and 1 IU/mL erythropoietin were used as supplement until day 14 on which the cells were harvested for analysis. For lentivirus infection, the cultured $CD34^+$ cells were infected with viral supernatants on days 4. Transduced cells were selected for GFP expression by fluorescence-activated cell sorting on day 14. Cell surface marker analysis with CD71 and Glycophorin A (GPA) indicated that more than 90% of cultured cells were at erythroid lineage. The siRNA target sequences of RNA interference for LYAR were inserted into the BamHI/EcoRI sites in the LV3 (H1/GFP&Puro) lentiviral, and packing and purification of LV3-LYARsiRNA lentivirus were custom-made in GenePharma Company (Shanghai, China). The oligonucleotides were: Human *LYAR* siRNA: CCTGGTCATCTTTAACAAG.

Figure S9. EMSA competition analysis.



EMSA competition analysis with the indicated amounts (25-, 75- or 225-fold molar excess) of cold wild-type or mutant competitors using LYAR expressed by TNT®Quik Coupled Transcription/Translation Systems. DNA binding complex bands and a free-probe band are indicated by arrows. Band intensity for each binding complex was shown. EMSAs were performed using the LightShift EMSA optimization and control kit (Pierce, USA) to evaluate the affinity of LYAR to the wild-type or mutant *HBG1* promoter. Briefly, nuclear extracts prepared from K562 cells (**Figure 2D**) or the in vitro expressed LYAR by TNT®Quik Coupled Transcription/Translation Systems (Promega; **Figure S9**) were incubated in binding buffer for 20 min at room temperature with 2 pM of double-strand biotin-labeled hot wild-type probe and a serial dilution of unlabeled cold mutant or wild-type probe as competitors. The reaction samples were run on 6% native polyacrylamide gels in 0.5× TBE (Tris/Boric acid/EDTA buffer) buffer. The binding reactions were transferred to nylon membranes using a Bio-Rad wet transfer apparatus, and DNA was cross-linked to membrane with an ultraviolet illuminator. The biotin-labeled DNA complexes were visualized and quantified using a chemiluminescent imaging system (Tanon 4200; Tanon, China).

Table S1. Target fragments of β-globin cluster region and probe coverage in SureSelect

Target gene region	Interval (hg19)	Length (kbp)	Coverage (%)
HBB	chr11:5240000-5252000	12.0	61.88
HBD	chr11:5252000-5260000	8.0	97.14
HBBP1	chr11:5260000-5266000	6.0	97.83
HBG1	chr11:5266000-5273000	7.0	88.42
HBG2	chr11:5273000-5284000	11.0	61.39
HBE1	chr11:5284000-5295000	11.0	59.52
LCR	chr11:5295000-5320000	25.0	90.66
β-globin cluster	chr11:5240000-5320000	80.0	79.03

Target fragments from the β-globin cluster region of genomic DNA from 1142 β-thalassemia individuals in cohort C were enriched using the SureSelect DNA Standard Design Wizard (https://earray.chem.agilent.com/suredesign; Agilent Tech, USA) and then sequenced on a HiSeq2000 instrument platform (Illumina, USA). After obtaining the raw data in fastq format, those reads with mean Phred quality lower than 20 or contaminated by adapter sequences were removed. BWA-0.7.8 was applied for alignment with "aln –L –I –k 2 –I 31 –t 4 –i 10" and "sampe –a 500". Aligned files with bam format were produced. Then the bam files were sorted and duplication reads were removed by samtools 0.1.19. Finally, the variants were detected by GATK 2.1.8 and frequencies for each variant were annotated using the retrieved dbSNP data from Hapmap and 1000 genome databases. High quality common variants were obtained by filtering those with score <99 examined by SOAPsnp 1.03.[7] Common variants were identified by setting MAF > 0.01 as the threshold. Due to the sequence similarities between *HBG2* and *HBG1*, the proximal promoter regions spanning from 610 bp upstream to 120 bp downstream of the transcription start site (TSS) of each gene were further sequenced using the ABI 3500Dx Genetic Analyzer with specific primers (**Table S8**). According to the gender, the HbF z-score, the relationship (all individuals are unrelated) and the genotypes of all common variants within the 80kb β-globin gene cluster for each of the sequenced samples, files in map and ped formats were prepared for the association study in Plink v1.07. Taking the HbF z-score as dependent variable, the set-based tests were conducted on seven blocks by Plink v1.07.

Table S2. Association of 271 common SNPs and seven calculated LD blocks in β-globin cluster with the HbF levels.

See the attached excel file Table_S2.xlsx.

Table S3. Association of haplotypes of the selected tag-SNPs in 7 LD blocks with HbF levels.

See the attached excel file Table_S3.xlsx

Table S4. The phenotypes and genotypes of two Chinese family members

ID.	Hb	HbF	HbA_2	MCV	MCH	MCHC	*HBG1*	*HBG2*	*BCL11A*	*HBS1L -MYB*	*KLF1*	*HBA*	*HBB*
	g/L	g/L	%	fl	pg	g/L	rs368698783	rs7482144	rs4671393	rs9399137	mutations	genotype	genotype
Family A													
II-1	134	0.54	2.8	85.1	29.3	344	GG	CC	GA	TC	[=]+[=]	αα/αα	N/N
II-2	149	4.32	5.5	65.1	21.2	325	AA	TT	AA	TT	[=]+[=]	αα/αα	c.84_85insC/N
II-3	101	1.72	5.1	68.0	20.7	303	GA	CT	GG	TT	[=]+[=]	αα/αα	c.84_85insC/N
II-4	118	0.83	2.6	99.7	31.8	319	GG	CC	GA	CC	[=]+[=]	αα/αα	N/N
III-1	144	1.01	2.8	87.9	29.9	340	GA	CT	GA	TC	[=]+[=]	αα/αα	N/N
III-2	100	4.20	5.5	73.0	22.5	309	GA	CT	GA	TC	[=]+[=]	αα/αα	c.84_85insC/N
III-3	119	3.69	5.2	68.7	20.4	297	GA	CT	GA	TC	[=]+[=]	αα/αα	c.84_85insC/N
III-4	121	3.03	5.1	71.4	21.0	294	GA	CT	GA	TC	[=]+[=]	αα/αα	c.84_85insC/N
IV-1*	106	98.69	1.7	89.3	27.0	303	AA	TT	GA	CC	[=]+[=]	αα/αα	c.84_85insC/ c.84_85insC
IV-2	109	1.53	3.0	91.0	29.0	319	GG	CC	GA	TT	[=]+[=]	αα/αα	N/N
Family B													
II-1	108	2.27	4.9	67.6	20.3	301	GA	CT	GA	TC	[=]+[=]	αα/αα	c.126_129delCTTT/N
II-2	122	0.98	5.1	70.4	21.9	310	GG	CC	GA	TC	[=]+[=]	αα/αα	c.52A>T/N
II-3	130	0.91	5.1	66.7	20.2	304	GG	CC	AA	TT	[=]+[=]	αα/αα	c.92+1G>T/N
III-2	127	0.51	2.9	88.0	29.2	332	GG	CC	GA	TC	[=]+[=]	αα/αα	N/N
III-3	108	0.65	4.8	63.8	19.1	299	GG	CC	GG	TC	[=]+[=]	αα/αα	c.52A>T/N
III-4*	90	40.05	2.2	86.1	27.8	323	GA	CT	GA	CC	[=]+[=]	αα/αα	c.52A>T/ c.126_129delCTTT

*Probands

Hb, hemoglobin;

MCV, mean corpuscular volume;

MCH, mean corpuscular haemoglobin;

N, normal *HBB* allele.

Table S5. Phenotypic and genotypic data from 513 non-thalassemic individuals in cohort A

Characteristics	Genotype (rs368698783)			*P*†	HWE-*P*‡
	GG(n=412)	GA (n=92)	AA (n=9)		0.152
Gender (n)					
Males : Females	199:213	47:45	4:5	0.787	
Hematological data*					
Hb (g/L)	137.39 ± 17.38	140.65 ± 19.27	143.11 ± 16.48	0.240	
MCV (fl)	94.25 ± 4.36	94.39 ± 4.30	97.31 ± 5.96	0.301	
MCH (pg)	30.45 ± 1.41	30.43 ± 1.29	30.77 ± 1.26	0.778	
HbF (%)	0.43 ± 0.22	0.54 ± 0.25	0.66 ± 0.40	<0.001	
HbF(g/L)	0.59± 0.32	0.76± 0.36	0.92± 0.51	<0.001	
HbA_2 (%)	2.90 ± 0.29	2.88 ± 0.21	3.02 ± 0.51	0.709	
HBG2: rs7482144 (n, %)					0.282
CC	406 (98.5%)	0 (0.0%)	0 (0.0%)	<0.001	
CT	6 (1.5%)	92 (100.0%)	0 (0.0%)	<0.001	
TT	0 (0.0%)	0 (0.0%)	9 (100.0%)	<0.001	

*Hematological data were shown in mean ± standard deviation.

†*P* value was determined by either a Kruskal-Wallis test or the x^2 test among 3 genotypes of rs368698783.

‡HWE-P for the *P* value from the Hardy-Weinberg equilibrium (HWE) test.

Table S6. Phenotypic and genotypic data from 515 β-thalassemia heterozygotes in cohort B

Characteristics	Genotype (rs368698783)			*P*†	HWE-*P*‡
	GG (n=427)	GA (n=81)	AA (n=7)		0.168
Gender (n)					
Males : Females	192:235	36:45	5:2	0.368	
Hematological data*					
Hb (g/L)	112.86 ± 18.34	113.77 ± 20.05	123.86 ± 18.41	0.301	
MCV (fl)	67.84 ± 5.89	69.32 ± 6.35	64.84 ± 3.31	0.060	
MCH (pg)	20.60 ± 1.97	21.30 ± 2.40	20.29 ± 1.24	0.053	
HbF (%)	1.63 ± 1.37	2.94 ± 3.37	3.83 ± 2.76	<0.001	
HbF(g/L)	1.81 ± 1.54	3.30 ± 3.54	4.65 ± 3.30	<0.001	
HbA_2 (%)	5.28 ± 0.62	5.08 ± 0.71	5.06 ± 0.42	0.045	
***HBB* genotype (n, %)**					
β^+/β^N	60 (14.1%)	18 (22.2%)	0 (0.0%)	0.091	
β^0/β^N	367 (85.9%)	63 (77.8%)	7 (100.0%)	0.091	
***HBA* genotype (n, %)**					
αα/αα	427 (100.0%)	81 (100.0%)	7 (100.0%)	1.000	
***HBG2*: rs7482144 (n, %)**					0.308
CC	421 (98.6%)	0 (0.0%)	0 (0.0%)	<0.001	
CT	6 (1.4%)	81 (100.0%)	0 (0.0%)	<0.001	
TT	0 (0.0%)	0 (0.0%)	7 (100.0%)	<0.001	

*Hematological data were shown in mean ± standard deviation.

†*P* value was determined by either a Kruskal-Wallis test or the x^2 test among 3 genotypes of rs368698783.

‡HWE-P for the *P* value from the Hardy-Weinberg equilibrium (HWE) test.

Table S7. Phenotypic and genotypic data from 1142 β-thalassemia individuals in cohort C

Characteristics	Genotype (rs368698783)			*P*†	HWE-*P*‡
	GG (n=968)	GA (n=165)	AA (n=9)		0.503
Gender (n)					
Males : Females	626:342	102:63	4:5	0.246	
Hematological data*					
Hb (g/L)	72.24 ± 22.51	71.99 ± 20.03	81.00 ± 22.15	0.532	
MCV (fl)	80.17 ± 7.55	77.21 ± 8.39	82.49 ± 6.19	<0.001	
MCH (pg)	25.75 ± 3.35	24.23 ± 3.61	26.09 ± 3.33	<0.001	
HbF (g/L)	13.06 ± 13.52	22.28 ± 18.40	39.06 ± 33.34	<0.001	
Systematic transfusion (n, %)§	825 (85.2%)	95 (57.6%)	2 (22.2%)	<0.001	
Age at first transfusion (months), median (5^{th}-95^{th} percentile)	7 (3-60)	12 (3-172)	36 (3-360)	<0.001	
Diagnosed as thalassemia intermedia (n, %)	324 (33.5%)	108 (65.5%)	7 (77.8%)	<0.001	
***HBB* genotype** (n, %)¶					
β^+/β^+	10 (1.0%)	0 (0.0%)	1 (11.1%)	0.848	
β^+/β^0	290 (30.0%)	67 (40.6%)	0 (0.0%)	0.107	
β^0/β^0	668 (69.0%)	98 (59.4%)	8 (88.9%)	0.102	
***HBA* genotype** (n, %)#					
αα/αα	832 (85.9%)	144 (87.2%)	7 (77.8%)	0.916	
-α/αα	53 (5.5%)	9 (5.5%)	2 (22.2%)	0.363	
$\alpha\alpha^T/\alpha\alpha$	11 (1.2%)	3 (1.8%)	0 (0.0%)	0.600	
--/αα	68 (7.0%)	9 (5.5%)	0 (0.0%)	0.310	
$-\alpha/\alpha^T\alpha$	1 (0.1%)	0 (0.0%)	0 (0.0%)	0.679	
--/-α	1 (0.1%)	0 (0.0%)	0 (0.0%)	0.679	
$--/\alpha^T\alpha$	2 (0.2%)	0 (0.0%)	0 (0.0%)	0.559	
HBG2: rs7482144 (n, %)					0.549
CC	966 (99.8%)	0 (0.0%)	0 (0.0%)	<0.001	
CT	2 (0.2%)	165 (100.0%)	0 (0.0%)	<0.001	
TT	0 (0.0%)	0 (0.0%)	9 (100.0%)	<0.001	
***BCL11A*:** rs4671393 (n, %)					0.608
GG	571 (59.0%)	78 (47.3%)	3 (33.3%)	0.002	
GA	348 (36.0%)	75 (45.4%)	3 (33.3%)	0.045	
AA	49 (5.0%)	12 (7.3%)	3 (33.3%)	0.010	
***HBS1L-MYB*:** rs9399137 (n, %)					0.328
TT	660 (68.2%)	103 (62.4%)	5 (55.6%)	0.102	
TC	268 (27.7%)	60 (36.4%)	3 (33.3%)	0.029	
CC	40 (4.1%)	2 (1.2%)	1 (11.1%)	0.247	
***KLF1* mutations** (n, %)	12 (1.2%)	0 (0.0%)	0 (0.0%)	0.134	

*Hematological data are shown as the mean ± standard deviation.

†*P* value was determined by either the Kruskal-Wallis test or the x^2 test between the 3 genotypes of rs368698783.

‡HWE-*P* for the *P* value from the Hardy-Weinberg equilibrium (HWE) test.

§Systematic transfusion was defined as requiring more than 8 transfusions/year.

¶The *HBB* [NM_000518 (*HBB*_v001)] genotype categories are defined as **(β^0):** *HBB*:c.126_129delCTTT (39.4%), *HBB*:c.52A>T (25.3%), *HBB*:c.316-197C>T (10.1%), *HBB*:c.216_217insA (4.1%), *HBB*:c.92+1G>T (2.2%), *HBB*:c.130G>T (1.0%), *HBB*: c.84_85insC (0.5%), *HBB*:c.91A>G (0.1%), *HBB*:c.45_46insC (0.1%), *HBB*:c.165_177delTATGGGCAACCCT (0.1%), *HBB*:c.315+1G>A (0.1%), *HBB*:c.287_288insA (0.1%), *HBB*:c.113G>A (0.1%), *HBB*:c.93-1G>C (0.1%); **(β^+):** *HBB*:c.-78A>G (9.4%), *HBB*:c.79G>A (5.2%), *HBB*:c.-79A>G (1.2%), *HBB*:c.315+5G>C (0.7%), *HBB*:c.-140C>T (0.1%), *HBB*:c.-81A>C (0.1%), and *HBB*:c.92+5G>C (0.1%).

#*HBA* genotype categories are defined as **(-α):** NG_000006.1: g.34247_38050del, NC_000016.9: g.219817_(223755_224074)del; **(--):** NG_000006.1:g.26264_45564del19301, NG_000006.1: g.10664_44164del33501; **($\alpha^T\alpha$):** NM_000517.4(*HBA2*_v001):c.427T>C, NM_000517.4(*HBA2*_v001): c.369C>G, NM_000517.4(*HBA2*_v001):c.377T>C.

A total of 884 of 1142 participants in this table were employed in the previous study.

Table S8. Phenotypic and genotypic data of 568 Thai HbEE individuals in cohort D

Characteristics	Genotype (rs368698783)			*P*†	HWE-*P*‡
	GG (n=30)	**GA (n=180)**	**AA (n=358)**		0.242
Gender (n) Males : Females	14:16	76:104	169:189	0.545	
Hematological data*					
Hb (g/L)	121.15 ± 19.07	114.52 ± 19.02	116.18 ± 19.17	0.359	
MCV(fl)	60.54 ± 6.33	62.68 ± 5.12	63.07 ± 6.00	0.277	
MCH(pg)	20.00 ± 1.48	21.07 ± 1.80	21.27 ± 2.48	0.010	
HbE (g/L)	108.77 ± 16.11	103.93 ± 16.37	103.51 ± 15.78	0.134	
HbF (g/L)	4.81 ± 4.93	6.51 ± 6.08	9.97 ± 7.21	<0.001	
***HBA* genotype (n, %)**					
αα/αα	26 (86.7%)	159 (88.3%)	298 (83.2%)	0.285	
-α/αα	2 (6.7%)	11 (6.1%)	30 (8.4%)	0.632	
$\alpha\alpha^{T}$/αα	0 (0.0%)	2 (1.1%)	7 (2.05%)	0.463	
--/αα	1 (3.3%)	8 (4.5%)	22 (6.1%)	0.622	
-α/α^{T}α	1 (3.3%)	0 (0.0%)	1 (0.3%)	0.131	
HBG2: rs7482144 (n, %)					0.220
CC	30 (100.0%)	0 (0.0%)	0 (0.0%)	<0.001	
CT	0 (0.0%)	180 (100.0%)	1 (0.35)	<0.001	
TT	0 (0.0%)	0 (0.0%)	357 (99.7%)	<0.001	
***BCL11A*: rs4671393 (n, %)**					0.246
GG	19 (63.3%)	94 (52.2%)	208 (58.1%)	0.319	
GA	11 (36.7%)	68 (37.8%)	126 (35.2%)	0.839	
AA	0 (0.0%)	18 (10.0%)	24 (6.7%)	0.109	
***HBS1L-MYB*:**					
rs4895441 (n, %)					0.012
AA	24 (80.0%)	115 (63.9%)	213 (59.5%)	0.069	
AG	6 (20.0%)	54 (30.0%)	117 (32.7%)	0.326	
GG	0 (0.0%)	11 (6.1%)	28 (7.8%)	0.237	
rs9399137 (n, %)					<0.001
TT	25 (83.3%)	117 (65.0%)	222 (62.0%)	0.062	
TC	5 (16.7%)	51 (28.3%)	103 (28.85)	0.363	
CC	0 (0.0%)	12 (6.7%)	33 (9.2%)	0.150	
***KLF1* mutations (n, %)**	3 (10.0%)	11 (6.15)	34 (9.5%)	0.392	

*Hematological data were shown in mean ± standard deviation.

†*P* value was determined by either a Kruskal-Wallis test or the x^2 test among three genotypes of rs368698783.

‡HWE-P for the *P* value from the Hardy-Weinberg equilibrium (HWE) test.

Table S9. Information of primers and probes used in this study (based on hg19)

Purpose	Gene/Loci		5' primer/WT probe	3' primer/MT probe	Product
			Primer/Probe sequence (5'-3')	Primer/Probe sequence (5'-3')	Length (bp
Genotyping	**rs368698783**		TACTGCGCTGAAACTGTGG	TACCTTCCCAGGGTTTCTCC	777
	rs7482144		CCTGCACTGAAACTGTTGC	TACCTTCCCAGGGTTTCTCC	774
Bisulfite sequencing	***HBE1***	1st	GAAGATGATGAAGAGGGTAAAAAAG	TCTATAAAATAACACCATATCAAATACA	532
		2nd	GAAATTTGTGTTGTAGATAGATGAG	TCTTAAAAACTTTCCCAATCAACTTAC	450
	HBG	1st	TTAAAAATTTTGGATTTATGTTTA	CAAATTACCAAAACTATCAAAAAACC	793
		2nd	TTAAATTATAGGTTTTATTGGAGTT	AATCAAAAAATACCACAAATCC	635
	HBD	1st	AGAGGTAAAGAAGAATTTTATATTGAGT	CTCTATCTACACATACCCAATTTCC	609
		2nd	AGTATAAAGTGATAGAAATAAATAAGTT	CTCTTATAACCTTAATACCAACCTAC	500
	HBB	1st	TAAGAAAAATAATAATAAATGAATGTA	TCTCCACATACCCAATTTCTATTAATC	804
		2nd	ATTAGAAGGTTTTAATTTAAATAAGGA	ACCTTAATACCAACCTACCCAAAAC	669
mRNA-seq	***HBG***		ACTCGCTTCTGGAACGTCT	TAAAGCCTATCCTTGAAAGCTCT	536
EMSA	**rs368698783**		AACGTCTGAGGTTATCAATAAGCT	AACGTCTGAGATTATCAATAAGCT	/
Luciferase construct	**HS4**		CACAGCAAACACAACGACCC	TGAATGAGAGCCTCTGGGGA	983
	HBG2		AGCCGCCTAACACTTTGAGCA	TACCTTCCCAGGGTTTCTCC	1524
	HBG1		GGCTACTTCATAGGCAGAGT	TACCTTCCCAGGGTTTCTCC	1771
cDNA cloning	***LYAR***		ATGGTATTTTTTACATGCAATG	TCATTTCACAAGCTTGACTTTG	1140
Real-time qPCR	***HBG***		TGGGTCATTTCACAGAGGAG	AGAGGCAGAGGACAGGTTG	158
	LYAR		TCCAACAGCGAACCAGTC	ACGGCGTCTTTCACTTTG	113
	GAPDH		GTGAAGGTCGGAGTCAACG	TGAGGTCAATGAAGGGGTC	112
	β-actin		GGGAAATCGTGCGTGACATT	GGAGTTGAAGGTAGTTTCGTG	227
ChIP-PCR	***HBG*-rs368398783**		CCCTTCAGCAGTTCCACACA	GGCGTCTGGACTAGGAGCTTATT	70
	rs368698783-TaqMan		FAM-TGGAACGTCTGAGGTT-BHQ-X	HEX-TGGAACGTCTGAGATT-BHQ-X	/
	***HBG*-1/2**-pro		CGGTCCCTGGCTAAACTCCA	GAAATGACCCATGGCGTCTG	227
	HBG-1/2-pro-TaqMan		FAM-CATGGGTTGGCCAGCCTTGCCT-TAMRA		/
	GAPDH		TACTAGCGGTTTTACGGGCG	TCGAACAGGAGGAGCAGAGAGCGA	166

Supplemental References

1. Thuret, I., Pondarre, C., Loundou, A., Steschenko, D., Girot, R., Bachir, D., Rose, C., Barlogis, V., Donadieu, J., de Montalembert, M., et al. (2010). Complications and treatment of patients with beta-thalassemia in France: results of the National Registry. Haematologica 95, 724-729.

2. Weatherall, D.J., and Clegg, J.B. (2008). Human Haemoglobin. In The Thalassaemia Syndromes. (Blackwell Science Ltd), pp 63-120.

3. 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.

4. Wan, J.H., Tian, P.L., Luo, W.H., Wu, B.Y., Xiong, F., Zhou, W.J., Wei, X.C., and Xu, X.M. (2012). Rapid determination of human globin chains using reversed-phase high-performance liquid chromatography. Journal of chromatography B, Analytical technologies in the biomedical and life sciences 901, 53-58.

5. 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.

6. 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.

7. Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., and Wang, J. (2009). SNP detection for massively parallel whole-genome resequencing. Genome Res 19, 1124-1132.