REPORT

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

Diyu Chen,[1,13] Yangjin Zuo,[1,13] Xinhua Zhang,[2] Yuhua Ye,[1] Xiuqin Bao,[1] Haiyan Huang,[3] Wanicha Tepakhan,[4] Lijuan Wang,[1] Junyi Ju,[5] Guangfu Chen,[6] Mincui Zheng,[7] Dun Liu,[1] Shuodan Huang,[8] Lu Zong,[1] Changgang Li,[9] Yajun Chen,[10] Chenguang Zheng,[11] Lihong Shi,[12] Quan Zhao,[5] Qiang Wu,[3] Supan Fucharoen,[4] Cunyou Zhao,[1,*] and Xiangmin Xu[1,*]

A delayed fetal-to-adult hemoglobin (Hb) switch ameliorates the severity of β-thalassemia and sickle cell disease. The molecular mechanism underlying the epigenetic dysregulation of the switch is unclear. To explore the potential *cis*-variants responsible for the Hb switching, we systematically analyzed an 80-kb region spanning the β-globin cluster using capture-based next-generation sequencing of 1142 Chinese β-thalassemia persons and identified 31 fetal hemoglobin (HbF)-associated haplotypes of the selected 28 tag regulatory single-nucleotide polymorphisms (rSNPs) in seven linkage disequilibrium (LD) blocks. A Ly1 antibody reactive (LYAR)-binding motif disruptive rSNP rs368698783 (G/A) from LD block 5 in the proximal promoter of hemoglobin subunit gamma 1 (*HBG1*) was found to be a significant predictor for β-thalassemia clinical severity by epigenetic-mediated variant-dependent HbF elevation. We found this rSNP accounted for 41.6% of β-hemoglobinopathy individuals as an ameliorating factor in a total of 2,738 individuals from southern China and Thailand. We uncovered that the minor allele of the rSNP triggers the attenuation of LYAR and two repressive epigenetic regulators DNA methyltransferase 3 alpha (DNMT3A) and protein arginine methyltransferase 5 (PRMT5) from the *HBG* promoters, mediating allele-biased γ-globin elevation by facilitating demethylation of *HBG* core promoter CpG sites in erythroid progenitor cells from β-thalassemia persons. The present study demonstrates that this common rSNP in the proximal ^{A}γ-promoter is a major genetic modifier capable of ameliorating the severity of thalassemia major through the epigenetic-mediated regulation of the delayed fetal-to-adult Hb switch and provides potential targets for the treatment of β-hemoglobinopathy.

The human β-globin cluster, spanning a 70-kb region, is composed of five genes (5′-ε-γ^{G}-γ^{A}-δ-β-3′; *HBE* [MIM 142100]-*HBG2* [MIM 142250]-*HBG1* [MIM 142200]-*HBD* [MIM 142000]-*HBB* [MIM 141900]) and a distal regulatory element known as the locus control region (LCR).[1] The clinical manifestations of β-thalassemia (MIM 613985) mainly depend on the mutation in the β-globin gene.[2] However, individuals with identical β-thalassemia genotypes can exhibit variable clinical severities.[3] Several genetic modulators[4-8] and *cis*-regulatory elements[8-12] involved in the regulation of human fetal hemoglobin (HbF), and concomitant α-thalassemia (MIM 604131) have been identified as ameliorators of β-thalassemia.[13] These modifiers included *HBG2*: −158C>T (NC_000011.9: g.5276169G>A, rs7482144, or *Xmn*I polymorphism) in the β-globin cluster identified to be linked to *HBG1*: +25G>A polymorphism (NC_000011.9: g.5271063C>T or rs368698783),[12,14] and the master genes involved in the regulation of fetal-to-adult hemoglobin (Hb) switching, including B cell CLL/lymphoma 11A (*BCL11A* [MIM 606557]),[5,8,15] Kruppel-like factor 1 (*KLF1* [MIM 600599]),[6,15] and *MYB* (MIM 189990).[7,15] Identifying regulators of Hb switching including these genetic variants could provide promising predictors of β-thalassemia severity and therapeutic targets for re-activating HbF production.[15,16] The proposed modes of Hb switching based on current discoveries rely on various epigenetic and transcriptional regulatory factors that could have interacting roles in this developmental event.[17-20] Recently, leukemia/lymphoma-related factor encoded by zinc finger and BTB domain containing 7A (*ZBTB7A* [MIM 605878]) was identified as a regulator that represents an alternative mechanism independent of the γ-globin repressor BCL11A, as it can occupy γ-globin genes directly.[21] Ly1 antibody reactive (LYAR, HGNC:26021) is a zinc finger transcription factor (TF) that modulates Hb switching by

[1]Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, and Guangdong Technology and Engineering Research Center for Molecular Diagnostics of Human Genetic Diseases, Guangzhou, Guangdong, 510515, China; [2]Department of Hematology, 303rd Hospital of the People's Liberation Army, Nanning, Guangxi, 530021, China; [3]Key Laboratory of Systems Biomedicine (Ministry of Education), Center for Comparative Biomedicine, Institute of Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, 200240, China; [4]Centre for Research and Development of Medical Diagnostic Laboratories, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, 40002, Thailand; [5]The State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, Jiangsu, 210046, China; [6]Department of Pediatrics, Shenzhen Second People's Hospital, Shenzhen, Guangdong, 518035, China; [7]Department of Hematology, Hunan Children's Hospital, Changsha, Hunan, 410007, China; [8]Genetic Disease Prevention Center, Meizhou Maternal and Child Health Hospital, Meizhou, Guangdong, 514021, China; [9]Department of Hematology & Oncology, Shenzhen Children's Hospital, Shenzhen, Guangdong, 518026, China; [10]Genetic and Prenatal Diagnostic Center, Shaoguan Maternity and Children Healthcare Hospital, Shaoguan, Guangdong, 512026, China; [11]Prenatal Diagnostic Center, Guangxi Zhuang Autonomous Region Women and Children Care Hospital, Nanning, Guangxi, 530003, China; [12]State Key Laboratory of Experimental Hematology, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin, 300041, China
[13]These authors contributed equally to this work
*Correspondence: cyzhao@smu.edu.cn (C.Z.), gzxuxm@pub.guangzhou.gd.cn (X.X.)
http://dx.doi.org/10.1016/j.ajhg.2017.05.012.

Table 1. Stepwise Cox Proportional Hazards Analysis for 1142 β-Thalassemia Individuals in Cohort C

Ameliorating alleles	p	Hazard Ratio	95% CI
KLF1 mutations	2.294×10^{-7}	0.219	0.123-0.389
HBB mutations (β^+)	3.286×10^{-48}	0.379	0.333-0.432
HBG1: rs368698783 (A)	3.029×10^{-14}	0.552	0.473-0.643
HBS1L-MYB: rs9399137 (C)	1.175×10^{-10}	0.710	0.640-0.788
HBA mutations	1.875×10^{-10}	0.713	0.643-0.791
BCL11A: rs4671393 (A)	1.830×10^{-5}	0.806	0.730-0.889

A backward stepwise Cox proportional hazards model in 1,142 β-thalassemia individuals in cohort C was conducted to evaluate the associations between putative ameliorating alleles and the age at first transfusion. The covariates were classified based on the number of copies of putative modifying allele. Motif-disrupting SNP rs368698783 alone with ten known modifiers (*HBB* mutations, *HBA* mutations, *KLF1* mutations, *HBS1L-MYB*: rs9399137, rs4895441, rs9402686, rs1427407; *BCL11A*: rs4671393, rs11886868, and rs766432) were included in the analysis. The discriminative ability of the model was high (Harrell's concordance index = 0.708, $R^2 = 0.274$). The performance of the model was measured by Harrell's concordance index by using the *Hmisc* and *rms* package in R version 3.3.1.
HBB genotype categories are defined as (β^0): NM_000518(*HBB*_v001): c.126_129delCTTT, NM_000518(*HBB*_v001): c.52A>T, NM_000518(*HBB*_v001): c.316-197C>T, NM_000518(*HBB*_v001): c.216_217insA, NM_000518(*HBB*_v001): c.92+1G>T, NM_000518(*HBB*_v001): c.130G>T, NM_000518(*HBB*_v001):c.84_85insC, NM_000518(*HBB*_v001): c.91A>G, NM_000518(*HBB*_v001):c.45_46insC, NM_000518(*HBB*_v001): c.165_177delTATGGGCAACCCT, NM_000518(*HBB*_v001):c.315+1G>A, NM_000518(*HBB*_v001):c.287_288insA, NM_000518(*HBB*_v001):c.113G>A, NM_000518(*HBB*_v001):c.93-1G>C;
(β^+): NM_000518(*HBB*_v001): c.-78A>G, NM_000518(*HBB*_v001): c.79G>A, NM_000518(*HBB*_v001): c.-79A>G, NM_000518(*HBB*_v001): c.315+5G>C, NM_000518(*HBB*_v001): c.-140C>T, NM_000518(*HBB*_v001): c.-81A>C, NM_000518(*HBB*_v001): c.92+5G>C.
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.

binding the γ-globin gene and epigenetically silencing HbF.[22] However, the underlying mechanisms driving epigenetic regulation of Hb switching involved in ameliorating β-thalassemia severity are largely unclear.

To explore the potential genetic cis-variants that ameliorate β-thalassemia severity, we performed capture-based next generation sequencing (NGS) analysis of an 80-kb region spanning the β-globin cluster (Table S1) from the genomic DNA of 1142 Chinese β-thalassemia individuals (The procedures followed were accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and proper informed consent was obtained), and discovered 271 common single-nucleotide polymorphisms (SNPs, Minor allele frequency [MAF] > 0.01) in the cluster. Taking the HbF *Z* score as dependent variable in single SNP association study by Plink, we identified 107 out of 271 SNPs associated with the HbF levels with Bonferroni correction ($p < 0.0002$; Table S2). Furthermore, seven linkage disequilibrium (LD) blocks containing 163 out of 271 common SNPs were identified in the 80-kb region (Figure S1) and 28 from 163 SNPs were selected as tag SNPs ($r^2 = 0.5$ and $p = 0.05$) based on Haploview and Plink program. Haplotype-based association analysis showed that 31 haplotypes of the selected 28 tag SNPs in 7 LD blocks significantly associated with HbF levels ($p < 0.05$; Table S3). Among these tag SNPs, rs368698783 ($p = 5.03E\text{-}18$) from LD block 5 was found embedded within a highly conserved hexanucleotide LYAR-binding motif in the proximal promoter of *HBG1*.[14,22] To further validate the effect of this SNP on clinical severity of β-thalassemia persons, age at first blood transfusion was introduced as a dependent variable in Stepwise Cox proportional hazards analysis including SNP rs368698783, as well as β^+-thalassemia mutations, hemoglobin subunit alpha (*HBA* [MIM 141800-141850]) defects and the known functional variants in *KLF1, HBS1L-MYB* (MIM 612450-189990) intergenic region and *BCL11A*, which had been verified previously in our cohort as ameliorating factors.[6] The analysis showed that SNPs rs368698783 was identified as a critical variant of ameliorating factors ($HR = 0.552$, $p = 3.029 \times 10^{-14}$) after well-known mutations at *KLF1* and *HBB* in a ranking of Hazard ratio (Table 1). These results together with rs368698783 embedded within a highly conserved hexanucleotide LYAR-binding motif required for methylation-related γ-globin gene silencing as previously described,[22] suggested that rs368698783-mediated epigenetic regulation of HbF elevation might be involved in the amelioration of β-thalassemia severity.

To evaluate the prevalence of rs368698783, the genotypes of this rSNP from NGS-analyzed 1,142 individuals, as well as another 1,596 consecutive individuals from two subpopulations determined by Sanger sequencing showed that the derived allele frequencies (DAFs) to be 0.089 and 0.789 for the Chinese and Thai participants, respectively (Figure S2), which are similar to those in previous reports (0.114-0.289) based on other ethnic subpopulations and confirmed the presence of a common allele.[14,23] In 2,170 individuals from southern China, *HBG1*-rs368698783 (G/A) was highly linked with *HBG2*-rs7482144 (C/T), except for 14 individuals with CT (*HBG2*)-GG (*HBG1*) combination genotypes of these two SNPs, which were almost completely linked in 568 homozygous hemoglobin E (HbEE) individuals from the Thai subpopulation with 31.7% for the heterozygous and 63.0% for the homozygous individuals (Figure S2). We also identified a total of 712 of 1,710 individuals with β-hemoglobinopathy who carried the A allele of rSNP rs368698783 and accounted for 41.6% of the individuals from the two subpopulations (Figure S2).

Effects of rs368698783 A allele on HbF levels examined from above two ethnic subpopulations showed that the genotypes GA and AA of *HBG1*-rs368698783 exhibited a significantly elevated level of HbF ($p < 0.05$) compared with the genotype GG in each of three cohorts from China (Figure 1A). The *HBG1*-rs368698783 AA genotype exhibited significantly elevated HbF levels compared with GA in thalassemia (cohorts B and C) and HbEE (cohort D) individuals, as well as in two unrelated Chinese

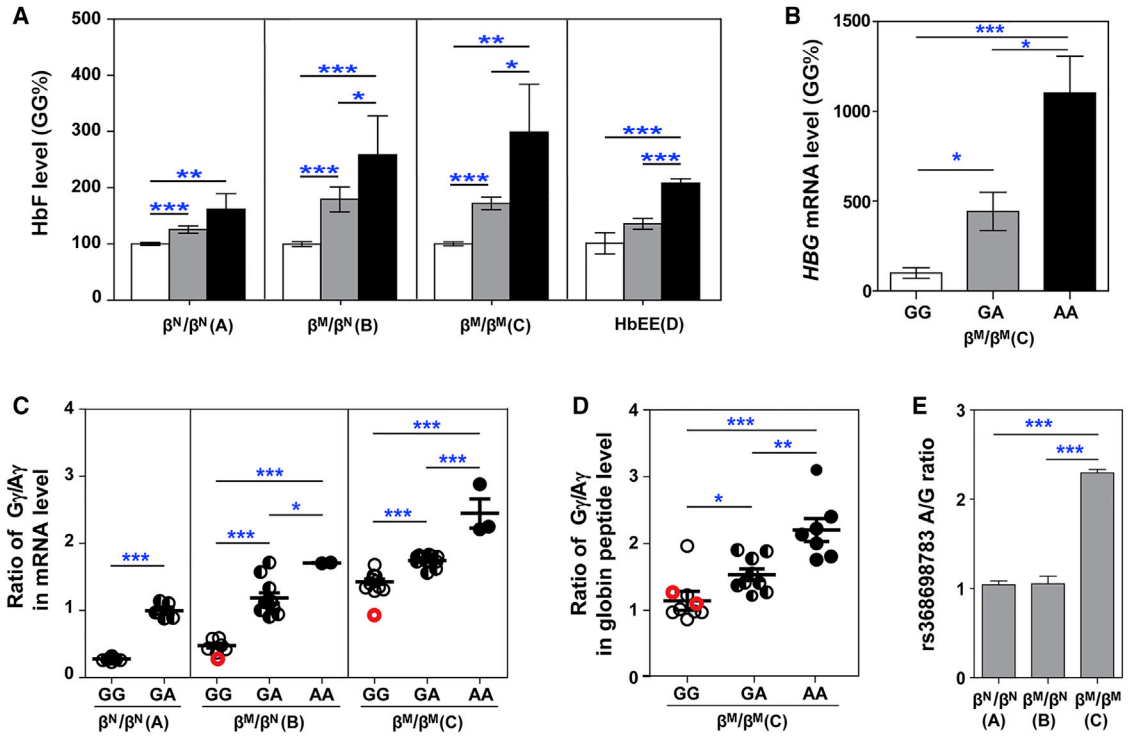


Figure 1. Ameliorating Effects of rs368698783 on β-Thalassemia Severity

(A) The effects of rs368698783 genotypes on the levels of HbF in peripheral red blood cells. Relative HbF level (GG%, g/L) determined as described[6] from each of four cohorts (the numbers of samples for cohorts A = 513, B = 515, C = 1142 and D = 568, successively) are shown in columns (white, GG; gray, GA; black, AA) with standard errors indicated by bars.

(B) Relative *HBG* mRNA level in BM-derived CD235a$^+$ erythroblasts from thalassemia individuals with the GG (n = 3), GA (n = 4), or AA (n = 2) genotypes at rs368698783 are quantified by qPCR ([$2^{-\Delta\Delta Ct}$, SYBR Premix Ex Taq [Takara, China]) with β-actin as a reference gene and shown in columns with standard errors indicated by bars.

(C and D) The effects of rs368698783 on the ratio of Gγ/Aγ mRNA (C) as described in Figure S5 and globin protein (D) as described in Figure S3 expression in peripheral blood. Open circles, GG; half-filled circles, GA; full-filled circles, AA. The red circles indicate individuals with the *HBG1*-rs368698783 GG and *HBG2*-rs7482144 CT genotype. The numbers of samples for each genotype are as follows: in B: GG = 9 and GA = 5 in cohort A; GG = 7, GA = 12 and AA = 2 in cohort B; GG = 11, GA = 10 and AA = 3 in cohort C; in C: GG = 9, GA = 9 and AA = 7 in cohort C. Solid lines represent as the mean ± SEM.

(E) Measurements of allelic ratios in the peripheral blood of non-thalassemia controls ($\beta^N/\beta^N = 5$), thalassemia carriers ($\beta^M/\beta^N = 12$) and thalassemia persons ($\beta^M/\beta^M = 10$) as described in Figure S5. Data are represented as the mean ± SEM from three independent experiments with triplication. *p < 0.05, **p < 0.01, ***p < 0.001 based on the Mann-Whitney *U* test (A) or the *T*-test (B–E). The genotype of β^N/β^N represents the normal *HBB* genotype for non-thalassemia controls, β^N/β^M (β^M representing β^0 or β^+ genotype in *HBB* as illustrated in Table 1 legend) for β-thalassemia carriers, and β^M/β^M for β-thalassemia persons.

families with β-thalassemia intermedia (Figure S3 and Table S4). Moreover, the genotypes AA and GA of *HBG1*-rs368698783 exhibited a significantly elevated *HBG* mRNA level ($p < 0.05$) compared with the genotype GG in bone marrow (BM)-derived CD235a$^+$ erythroblasts from the individuals with β-thalassemia (Figure 1B). Univariate analysis after correction for *HBB*, *HBA*, *KLF1*, *BCL11A*, and *HBS1L-MYB* genotypes demonstrated that variant *HBG1*-rs368698783 could act as a modifier associated with elevated levels of HbF in β-thalassemia ($p = 7.79 \times 10^{-10}$) and HbEE ($p = 1.34 \times 10^{-7}$) and could ameliorate the severity of β-thalassemia in age at onset ($p = 3.82 \times 10^{-10}$) and transfusion requirements ($p = 1.44 \times 10^{-9}$) and reduce the risk of developing β-thalassemia major ($p = 9.17 \times 10^{-18}$; Table 2, S5–S8). Moreover, AA ($p = 0.002$) or GA ($p < 0.001$) genotypes had significantly increased age at first transfusion with β-thalassemia according to the Kaplan-Meier log-rank test (Figure S4). We also observed that the A allele significantly increased the $^G\gamma/^A\gamma$ ratio through an increase in the production of γ-globin mRNA (Figure 1C) and peptide (Figure 1D), as well as allele-biased RNA expression in β-thalassemia individuals (β^M/β^M) with an A/G allelic ratio of 2.4 compared with non-thalassemic controls (β^N/β^N) or β-thalassemic traits (β^M/β^N) with an A/G allelic ratio of 1.1 (Figures 1E and S5). This evidence supports that the HbF level in β-thalassemia major is elevated by a variant-dependent activation of the γ-globin gene with an increasing $^G\gamma/^A\gamma$ ratio expression.

To explore whether the presence of rs368698783-A in β-thalassemia disrupts the LYAR-binding motif (GGTTAT) in the *HBG1* promoter and can induce demethylation-mediated γ-globin elevation, we examined the relationship of the methylation levels of *HBE1*, *HBG*, *HBD*, and *HBB* loci in BM-derived CD235a$^+$ erythroblasts from ten β^0/β^0-thalassemic subjects (GG = 4, GA = 3 and AA = 3)

Table 2. Univariate Analysis of *HBG1*-rs368698783 in 581 β-Thalassemia and 386 HbEE Individuals

Characteristics	rs368698783			p
β^M/β^M Cohort[a]	**GG (n = 500)**	**GA (n = 77)**	**AA (n = 4)**	
Gender (Male: Female)	326:174	48:29	2:2	0.733
Age of onset (months), median (5th-95th percentile)	6 (2.0-19.9)	12 (2.0-36.0)	30 (24.0-42.0)	3.82×10^{-10}
Hb (g/L)[c]	73.35 ± 22.82	73.74 ± 21.62	69.50 ± 21.42[e]	0.891
HbF (g/L)[d]	9.14 ± 11.11	17.65 ± 16.57	37.52 ± 24.03	7.79×10^{-10}
Requirement for systematic transfusion (No.)	479 (95.8%)	62 (80.5%)	2 (50.0%)	1.44×10^{-9}
Category of anemia (No.) TI: TM	60:440	38:39	3:1	9.17×10^{-18}
HbEE Cohort[b]	**GG (n = 24)**	**GA (n = 128)**	**AA (n = 234)**	
Gender (Male: Female)	8:16	49:79	114:120	0.086
Hb (g/L)	118.83 ± 16.30	113.87 ± 13.80	115.82 ± 13.98	0.075
HbE (g/L)	107.35 ± 16.86	103.44 ± 16.24	104.01 ± 16.05	0.499
HbF (g/L)	4.12 ± 4.45	5.14 ± 4.99	8.41 ± 6.24	1.34×10^{-7}

Univariate analysis was conducted according to our previous operation.[6]
[a]The individuals with the similar genetic variants of β^0/β^0, αα/αα, *KLF1* (WT), *BCL11A*-rs766432 (AA or AC), and *HBS1L-MYB*- rs9399137 (TT or CT) in β^M/β^M thalassemia cohort.
[b]The individuals with the similar genetic variants of αα/αα, *KLF1* (WT), *BCL11A*-rs4671393 (GG or GA), *HBS1L-MYB*-rs4895441 (AA or AG), and *HBS1L-MYB*-rs9399137 (TT or TC) in HbEE cohort.
[c]Hemoglobin levels were untransfused or pre-transfusion data.
[d]HbF (g/L) was calculated from total Hb level and HbF (%).
[e]Lower hemoglobin levels in the individuals with the genotype of AA were most likely to be correlated with low frequency of transfusion. The odds ratio (95% CI) of requirement for systematic transfusion was 0.181 (0.089-0.370, $p = 1.16 \times 10^{-5}$) between the GA group and the GG group, 0.044 (0.006-0.327, p = 0.011) between the AA group and the GG group. The odds ratio (95% CI) of TI diagnosis was 0.140 (0.083-0.236, $p = 4.47 \times 10^{-16}$) between the GA group and the GG group, 0.045 (0.005-0.444, p = 0.007) between the AA and the GG group.

with respect to their genotypes (Figures 2A, S6, and S7). Participants with the AA genotype of *HBG1*-rs368698783 exhibited significant hypomethylation at CpG site positions −53, −50, +6, +17, and +50 in the *HBG* promoter regions compared with the GG genotypes (Figure 2A). Additionally, individuals with higher levels of HbF (> 90 g/L) exhibited significant hypomethylation at the *HBG* promoters than individuals with lower levels of HbF (< 5 g/L, Figure S7). This result demonstrated a motif-disrupting variant in *HBG1* involved in the demethylation-mediated elevation of γ-globin expression.

To further examine the fine CpG methylation patterns in the proximal promoters of *HBG2* and *HBG1*, we performed DNA methylation analysis of the two probands from two Chinese families (Figure S3 and Table S4) using $CD235a^+$ BM-derived erythroblasts, and the AA genotype displayed different degrees of hypomethylation at the six CpG sites flanking the TSS in both *HBG2* and *HBG1* (−162 to +50) compared with the two distinct GG genotype controls (Figure 2B). This finding was further validated using the BS-seq method (Figure S8). These results suggested that the T allele of the *HBG2*-rs7482144 polymorphism alone could not reduce the levels of methylation in both *HBG* promoters and that the *HBG1* promoter exhibited a higher level of methylation in this case. Thus, these results further indicated rs368698783-mediated hypomethylation responsible for efficient HbF elevation.

To test the de-repression of LYAR due to a motif-disrupting variant, we constructed a dual luciferase reporter gene containing DNase hypersensitive site 4 (HS4) and the full length of *HBG1* and *HBG2* promoters with wild-type or mutant allele at the two rSNP sites to evaluate the role of variant *HBG1*-rs368698783 alone and its combined effect with *HBG2*-rs7482144 in regulation of *HBG* transcriptional activity (Figure 3A). When transfected into K562 cells, the *HBG1*-rs368698783 mutant (Aγ+25MT) combined either with the *HBG2*-rs7482144 wild-type (Gγ-158WT) reporter or with the *HBG2*-rs7482144 mutant reporter had significantly increased promoter activity compared to the wild-type of both promoters or the *HBG2*-rs7482144 mutant-type alone (Figure 3B). When induced by transiently expressed LYAR, decreased endogenous *HBG* mRNA expression in K562 cells was also observed (Figure 3C), supporting the hypothesis that the motif-disrupting variant *HBG1*-rs368698783 led to the de-repression of γ-promoter activity via diminished binding activity of LYAR. Moreover, the results shown in Figure 3B indicated that both *HBG* promoters were obviously upregulated in the presence of this motif-disrupting variant, suggesting effect of *HBG1*-rs368698783A on the activation of γ-globin expression. This is consistent with the observation that the *HBG2*-rs7482144 polymorphism alone exhibited the lowest level of *HBG* expression in both RNA and protein production (Figures 1C and 1D), as well as hypomethylation can be dominated by rs368698783 rather than rs7482144 (Figure 2B).

We then used electrophoretic mobility shift assay (EMSA) with a 24-bp probe corresponding to the sequence

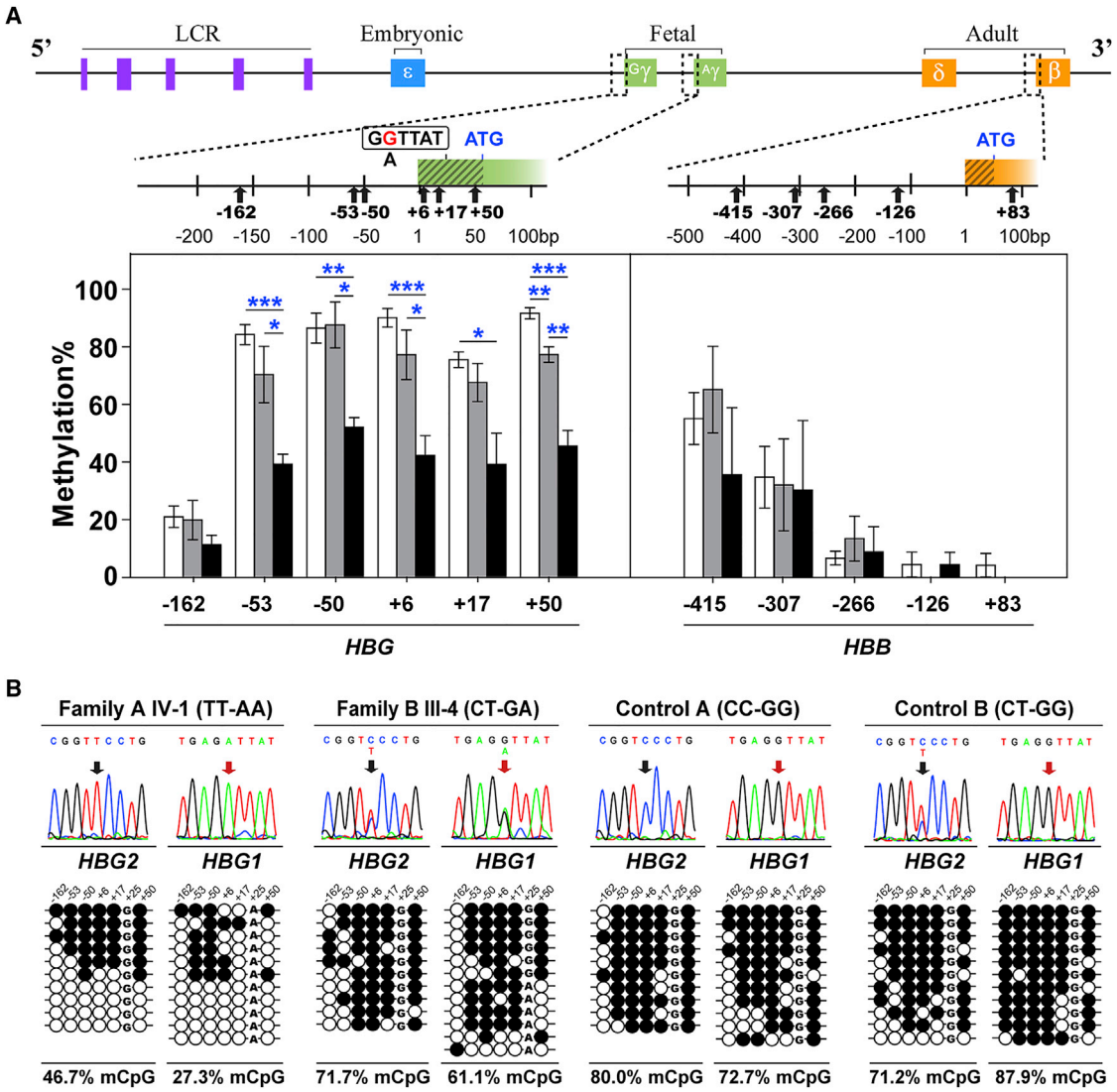


Figure 2. Genotype-Dependent Demethylation of *HBG* Promoters in β-Thalassemia

(A) The effects of the rs368698783 genotype on CpG methylation at *HBG* and *HBB* loci using CD235a$^+$ BM-derived erythroblasts from ten β^0/β^0 thalassemia individuals (GG = 4, GA = 3, and AA = 3). The broken-line boxes represent the analyzed region compassing CpG methylation sites on the top diagram and their enlarged graphs depict the distribution of these sites, with indicated by black arrows. The boxed GGTTAT sequence indicates the LYAR-binding motif. The mean methylation frequencies are shown in column (white, GG; gray, GA; black, AA) with standard error indicated by bars. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ based on the *T*-test.

(B) The effects of two combined SNPs, *HBG2*-rs7482144 (C/T, black arrow), and *HBG1*-rs368698783 (G/A, red arrow), on *HBG* promoter methylation. Sequence variations of four individuals with the genotypes of TT-AA, CT-GA, CC-GG, or CT-GG at the two SNPs are shown in the middle of the sequencing chromatograph, and their relationships with *HBG* methylation levels obtained from BM-derived CD235a$^+$ erythroblasts of individuals are shown at the bottom. Each row within a group represents a single bisulfite-treated clone with methylated CpGs (●) or unmethylated CpGs (○) from the G or A alleles of *HBG1*-rs368698783. See the Figures S6–S8 for the detailed procedures of methylation analysis.

flanking rSNP to test whether variant *HBG1*-rs368698783 could affect the binding activity of LYAR on the *HBG* promoters. We observed a significant LYAR-binding band, which could be easily competed with a cold wild-type probe but only weakly with a cold mutant probe using K562 cell nuclear extracts (Figure 3D) and LYAR expressed by TNT®Quik Coupled Transcription/Translation Systems (Figure S9), confirming that allele A in the mutant probe weakened the LYAR binding activity. Moreover, the indicated binding band was super-shifted by the addition of LYAR antibody in the gel-shift assay (Figure 3E). These results indicated that alterations in the DNA sequence containing the GGTTAT motif affect the binding of the LYAR-containing complex to the *HBG* promoters, which might directly influence the interactions between LYAR and epigenetic regulators that are required for γ-globin gene silencing.

To ascertain whether epigenetic regulators displayed rs368698783 allelic-biased enrichment at the *HBG* promoters, we used chromatin immunoprecipitation (ChIP) assays to analyze the interactions between the reported epigenetic regulators[24,25] and specific *HBG* promoter

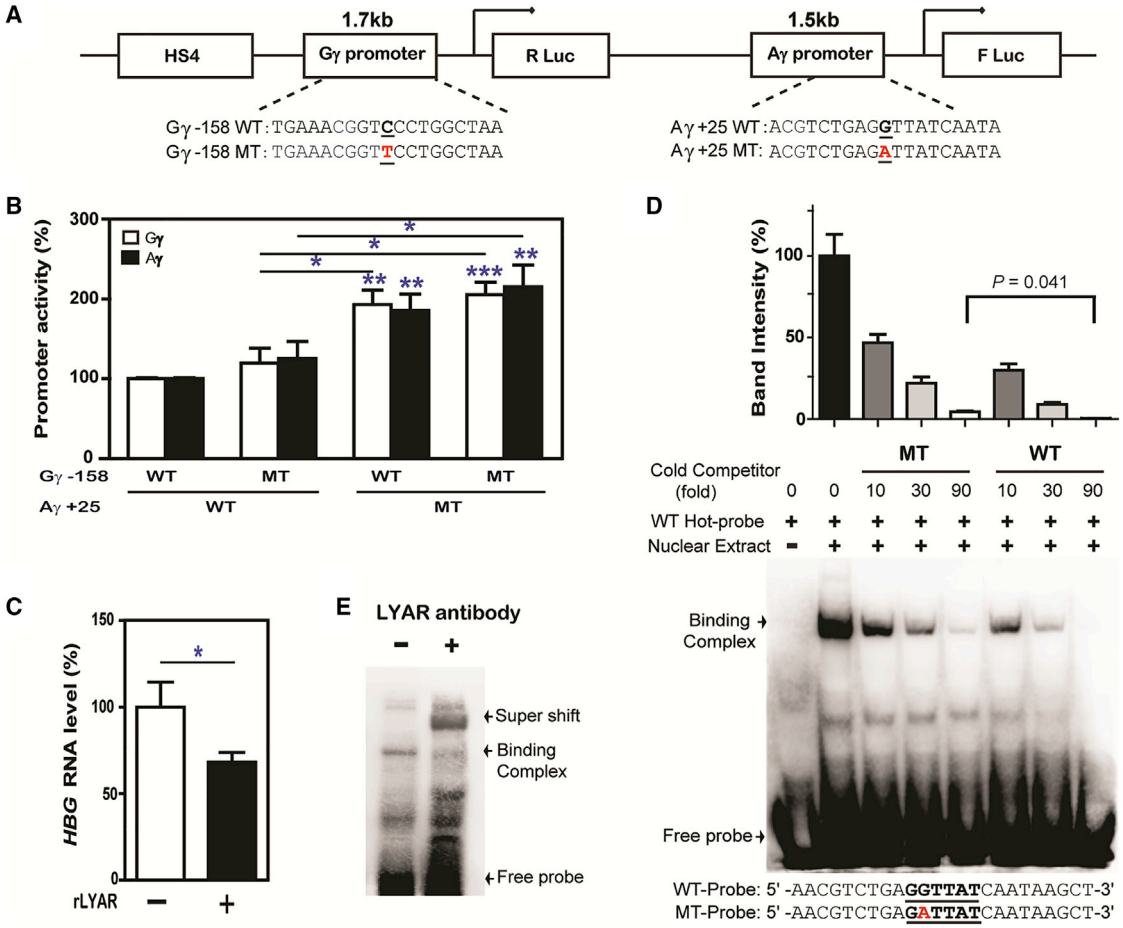


Figure 3. rs368698783-Mediated *HBG1* Transcriptional Activation in K562 Cells

(A) Schematic representation of a Renilla luciferase (R Luc) reporter driven by the *HBG2* promoter containing either the C allele (wild-type, Gγ-158WT) or the T allele (mutant, Gγ-158MT) of rs7482144 and a firefly luciferase (F Luc) reporter driven by the *HBG1* promoter containing either the G allele (wild-type, Aγ+25WT) or the A allele (mutant, Aγ+25MT) of rs36869783. HS4 represents DNase hypersensitive site 4.

(B) Relative dual luciferase activities were measured using a Wallac Victor V 1420 Multilabel Counter (PerkinElmer, USA) 30 hr after K562 cells were transfected with four genotypes of the *HBG2* and *HBG1* promoters using the 4D-Nucleofector System (Lonza, Switzerland). The data from three independent experiments with triplications are shown in columns (white, Gγ; black, Aγ) and standard errors indicated by bars. The Aγ+25 MT promoter combined either with the Gγ-158 WT reporter or with the Gγ-158 MT reporter had significantly increased promoter activity compared to the WT of both promoters shown with two or three asterisks (*). The promoter activity for the Aγ+25 WT promoter combined with the Gγ-158 MT reporter significant different from the Aγ+25 MT promoter was shown by asterisk above the indicated comparison. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ based on the *T*-test.

(C) Endogenous *HBG* mRNA expression in K562 cells co-transfected with (+) or without (−) the recombinant LYAR (rLYAR) cloned in the pcDNA 3.1 vector at the *Xho*I and *Bam*HI sites were quantified by SYBR Premix Ex Taq (Takara) with β-actin as a reference gene. The data from three independent experiments with triplications are shown in columns and standard errors indicated by bars. *$p < 0.05$ based on the *T*-test.

(D) EMSA competition analysis with the indicated amounts (10-, 30-, or 90-fold molar excess) of cold wild-type or mutant competitors in K562 nuclear extract. Relative intensity of binding complex from three-independent experiments are shown in columns with standard errors indicated by bars and significant differences are marked by p value from the *T*-test above the indicated comparison (upper panel). DNA binding complex bands and a free-probe band are indicated by arrows and sequences of probes are shown at the bottom (lower panel).

(E) The gel super shift assay of K562 nuclear extracts with (+) or without (–) the anti-LYAR antibody. The super-shifted band is indicated.

regions containing rSNP using CD235a$^+$ erythroblasts from β-thalassemia individuals. When the input DNA was normalized to contain equivalent amounts of both alleles for the individuals with the AA genotype (50% A-allele from *HBG1* and 50% G-allele from *HBG2*, horizontal dashed line in Figure 4A), LYAR (44% A-allele, $p = 0.008$), DNA methyltransferases 3A (DNMT3A [MIM 602769]), 39%, $p = 0.005$), and protein arginine methyltransferase 5 (PRMT5 [MIM 604045]), 33%, $p = 0.022$) enrichments were less frequent, binding to the A- than G-alleles of rs368698783. Additionally, LYAR ($p = 0.022$) and DNMT3A ($p = 0.035$) displayed different recruitments in erythroblasts of the *HBG1*-rs368698783 AA individuals from the GG individuals (Figure 4B). Furthermore, we observed an elevated *HBG* mRNA level (Figure 4C) accompanied by the hypomethylation at CpG site positions

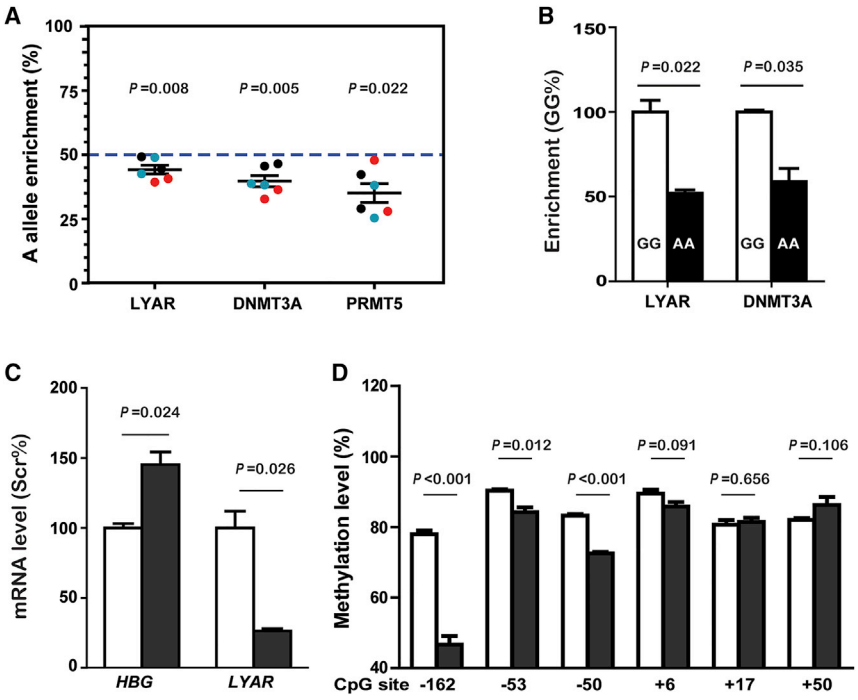


Figure 4. Epigenetic Regulation of *HBG* Transcription in β-Thalassemia Primary Cells

(A) ChIP analysis of the CD235a^{+} erythroblasts of three β^{M}/β^{M} thalassemia individuals (red-, black-, and blue-filled circles) with the AA genotype for the A-allele in the *HBG1* and the G-allele in the *HBG2* at rs368698783 using EZ-ChIP kit (Upstate, USA). Three antibodies against LYAR (home-made),[22] DNMT3A (Abcam, USA) or PRMT5 (Sigma-Aldrich, USA) used in this assay are indicated at the bottom. Because of highly similarity in promoter sequences between *HBG1* and *HBG2* in the PCR amplification region, qPCR products from ChIP-DNAs are the mixtures of the *HBG1* and *HBG2* promoter, and the A/G ratio of Input DNA is 1 for the AA *HBG1*-rs368698783 subject with 50% A-allele from *HBG1* and 50% G-allele from *HBG2*. Abundance of the rs368698783 A-allele in the qPCR products containing a heterozygous A/G mixture from both *HBG1* and *HBG2* DNA was quantified by two TaqMan probes, one for G allele labeled with FAM-dye and another for A allele labeled with HEX-dye, in each reaction after the input DNA was normalized to contain equivalent amounts of both alleles (50% A-allele, horizontal dashed line). The data are shown as the mean ± SEM from two independent experiments (same color circles) with duplication. The p values compared to the input DNA were obtained from the *T*-test.

(B) ChIP analysis of the CD235a^{+} erythroblasts of two β^{M}/β^{M} thalassemia individuals with the GG genotype (white column) or the AA genotype (black column) of the *HBG1*-rs368698783. Enrichment of *HBG* promoter in the PCR products containing *HBG1* and *HBG2* from ChIP assay using anti-LYAR or anti-DNMT3A antibodies was quantified by qPCR with a TaqMan probe and a pair of primers targeting the common region of *HBG1* and *HBG2* promoter and *GAPDH* promoter as a reference gene. The data are shown as the mean ± SEM from two independent experiments with duplication. The p values between the indicated group were obtained from *T*-test.

(C) qPCR analysis of *HBG* and *LYAR* mRNA levels normalized to β-actin mRNA from LYAR-knockdown (LYAR-KD, black column) or negative scrambled control (Scr, white column) erythroid progenitor cells from β-thalassemia individuals. The data are shown as the mean ± SEM from two independent experiments with duplication. The p values between the indicated group were obtained from *T*-test.

(D) Methylation levels of *HBG* promoter in the LYAR-KD (black column) relative to the Scr (white column) erythroid progenitor cells from β-thalassemia individuals. The mean methylation levels of each of six CpG sites obtained by BS-seq method are shown in column with standard error indicated by bars. The p values between the indicated group were obtained from *T*-test. The procedure for LYAR-KD in the cultured erythroid progenitor cells was shown in Figure S8 legend.

−162, −53, and −50 in the *HBG* promoter regions (Figure 4D) in the LYAR knockdown erythroid progenitor cultured cells from BM of β-thalassemia individuals. These results demonstrate that the presence of the rs368698783 A-allele triggers the attenuation of LYAR and two repressive epigenetic regulators DNMT3A and PRMT5 from the *HBG* promoters, thereby resulting in demethylation-mediated elevation of γ-globin expression.

In summary, we found a batch of HbF-associated genetic variants including haplotypes and SNPs from 1,142 Chinese β-thalassemia individuals by systematical analysis of an 80-kb region spanning the β-globin cluster based on NGS method. Then a common rSNP rs368698783, a LYAR-binding motif-disruptive SNP located in the *HBG1* proximal promoter, was found to be a significant predictor of clinical severity by elevating HbF levels in β-thalassemia. Furthermore, univariate analysis using a matched case-controls and transfusion-free survival curve analysis, supported rs368698783 accounting for 41.6% of β-hemoglobinopathy individuals as an ameliorating factor for the clinical severity of β-thalassemia phenotype. Finally, we uncovered that the minor allele of the rSNP impairs the LYAR-binding activity and triggers the attenuation of repressive epigenetic regulators DNMT3A and PRMT5 from the *HBG* core promoter resulting in the demethylation of the promoter CpG sites and the elevation of γ-globin gene expression in erythroid progenitor cells from β-thalassemia individuals. In conclusion, this finding delineates the mechanism insights gained from the epigenetic regulation of the fetal-to-adult hemoglobin switch, which is driven by the rSNP rs368698783 in the *HBG1* proximal promoter highly linked with another mechanism-unidentified well-known rSNP rs7482144 in the *HBG2* proximal promoter, expands our knowledge of Hb switch regulation from theoretical and practical points of views and provides potential targets for the treatment of β-hemoglobinopathy.

Supplemental Data

Supplemental Data include nine figures and nine tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2017.05.012.

Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China (NSFC)-Guangdong Joint Fund (U1201222), the NSFC (31671314 and 31471291), the National Key Technology Research & Development Program of China (2012BAI09B01), the Doctoral Fund of Ministry of Education of China-Key Program of Priority Fields (20134433130001), and the Talents Program in Higher Education of Guangdong (2050205). We thank the individuals for their willingness to participate in this study; Drs. Feng Zhang, Hongyan Wang, Xin Xu, and Mingding Li for valuable advices and comments on this work; Yi Cheng, Qiang Zhang, Kui Hong, Qifa Liu, Juan Tang, Xiaowei Wu, Junneng Zhang, Yuan Yang, and others for assistance in collecting samples from β-thalassemia individuals and controls.

Received: December 28, 2016
Accepted: May 18, 2017
Published: June 29, 2017

Web Resources

The URLs for data presented herein are as follows:
HUGO Gene Nomenclature Committee, http://www.genenames.org/
Mutalyzer, https://mutalyzer.nl/index
NCBI SNP, https://www.ncbi.nlm.nih.gov/snp/
OMIM, http://www.omim.org/
R statistical software, the *Hmisc* and *rms* package, http://www.r-project.org/
Sequence Variant Nomenclature, http://varnomen.hgvs.org/

References

1. Levings, P.P., and Bungert, J. (2002). The human beta-globin locus control region. Eur. J. Biochem. *269*, 1589–1599.
2. Higgs, D.R., Engel, J.D., and Stamatoyannopoulos, G. (2012). Thalassaemia. Lancet *379*, 373–383.
3. Thein, S.L. (2013). Genetic association studies in β-hemoglobinopathies. Hematology (Am Soc Hematol Educ Program) *2013*, 354–361.
4. 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.
5. Zhou, D., Liu, K., Sun, C.W., Pawlik, K.M., and Townes, T.M. (2010). KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. Nat. Genet. *42*, 742–744.
6. 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.
7. Stadhouders, R., Aktuna, S., Thongjuea, S., Aghajanirefah, A., Pourfarzad, F., van Ijcken, W., Lenhard, B., Rooks, H., Best, S., Menzel, S., et al. (2014). HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. J. Clin. Invest. *124*, 1699–1710.
8. Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., et al. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science *342*, 253–257.
9. Lettre, G., Sankaran, V.G., Bezerra, M.A.C., Araújo, A.S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F.F., Hirschhorn, J.N., and Orkin, S.H. (2008). DNA polymorphisms at the BCL11A, HBS1L-MYB, and β-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc. Natl. Acad. Sci. USA *105*, 11869–11874.
10. Sankaran, V.G., Xu, J., Byron, R., Greisman, H.A., Fisher, C., Weatherall, D.J., Sabath, D.E., Groudine, M., Orkin, S.H., Premawardhena, A., and Bender, M.A. (2011). A functional element necessary for fetal hemoglobin silencing. N. Engl. J. Med. *365*, 807–814.
11. Farrell, J.J., Sherva, R.M., Chen, Z.Y., Luo, H.Y., Chu, B.F., Ha, S.Y., Li, C.K., Lee, A.C., Li, R.C., Li, C.K., et al. (2011). A 3-bp deletion in the HBS1L-MYB intergenic region on chromosome 6q23 is associated with HbF expression. Blood *117*, 4935–4945.
12. Gilman, J.G., and Huisman, T.H. (1985). DNA sequence variation associated with elevated fetal G gamma globin production. Blood *66*, 783–787.
13. Mettananda, S., Gibbons, R.J., and Higgs, D.R. (2015). α-Globin as a molecular target in the treatment of β-thalassemia. Blood *125*, 3694–3701.
14. Bianchi, N., Cosenza, L.C., Lampronti, I., Finotti, A., Breveglieri, G., Zuccato, C., Fabbri, E., Marzaro, G., Chilin, A., De Angelis, G., et al. (2016). Structural and Functional Insights on an Uncharacterized Aγ-Globin-Gene Polymorphism Present in Four β0-Thalassemia Families with High Fetal Hemoglobin Levels. Mol. Diagn. Ther. *20*, 161–173.
15. Sankaran, V.G., and Weiss, M.J. (2015). Anemia: progress in molecular mechanisms and therapies. Nat. Med. *21*, 221–230.
16. Dulmovits, B.M., Appiah-Kubi, A.O., Papoin, J., Hale, J., He, M., Al-Abed, Y., Didier, S., Gould, M., Husain-Krautter, S., Singh, S.A., et al. (2016). Pomalidomide reverses γ-globin silencing through the transcriptional reprogramming of adult hematopoietic progenitors. Blood *127*, 1481–1492.
17. Amaya, M., Desai, M., Gnanapragasam, M.N., Wang, S.Z., Zu Zhu, S., Williams, D.C., Jr., and Ginder, G.D. (2013). Mi2β-mediated silencing of the fetal γ-globin gene in adult erythroid cells. Blood *121*, 3493–3501.
18. Mabaera, R., Richardson, C.A., Johnson, K., Hsu, M., Fiering, S., and Lowrey, C.H. (2007). Developmental- and differentiation-specific patterns of human γ- and β-globin promoter DNA methylation. Blood *110*, 1343–1352.
19. Lessard, S., Beaudoin, M., Benkirane, K., and Lettre, G. (2015). Comparison of DNA methylation profiles in human fetal and adult red blood cell progenitors. Genome Med. *7*, 1.
20. Forster, L., McCooke, J., Bellgard, M., Joske, D., Finlayson, J., and Ghassemifar, R. (2015). Differential gene expression analysis in early and late erythroid progenitor cells in β-thalassaemia. Br. J. Haematol. *170*, 257–267.
21. 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.
22. Ju, J., Wang, Y., Liu, R., Zhang, Y., Xu, Z., Wang, Y., Wu, Y., Liu, M., Cerruti, L., Zou, F., et al. (2014). Human fetal globin gene expression is regulated by LYAR. Nucleic Acids Res. *42*, 9740–9752.
23. Kersey, P.J., Allen, J.E., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., Hughes, D.S., Humphrey, J., Kerhornou, A.,

Khobova, J., et al. (2014). Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic Acids Res. *42*, D546–D552.

24. Zhao, Q., Rank, G., Tan, Y.T., Li, H., Moritz, R.L., Simpson, R.J., Cerruti, L., Curtis, D.J., Patel, D.J., Allis, C.D., et al. (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. Nat. Struct. Mol. Biol. *16*, 304–311.
25. Rank, G., Cerruti, L., Simpson, R.J., Moritz, R.L., Jane, S.M., and Zhao, Q. (2010). Identification of a PRMT5-dependent repressor complex linked to silencing of human fetal globin gene expression. Blood *116*, 1585–1592.

The American Journal of Human Genetics, Volume *101*

Supplemental Data

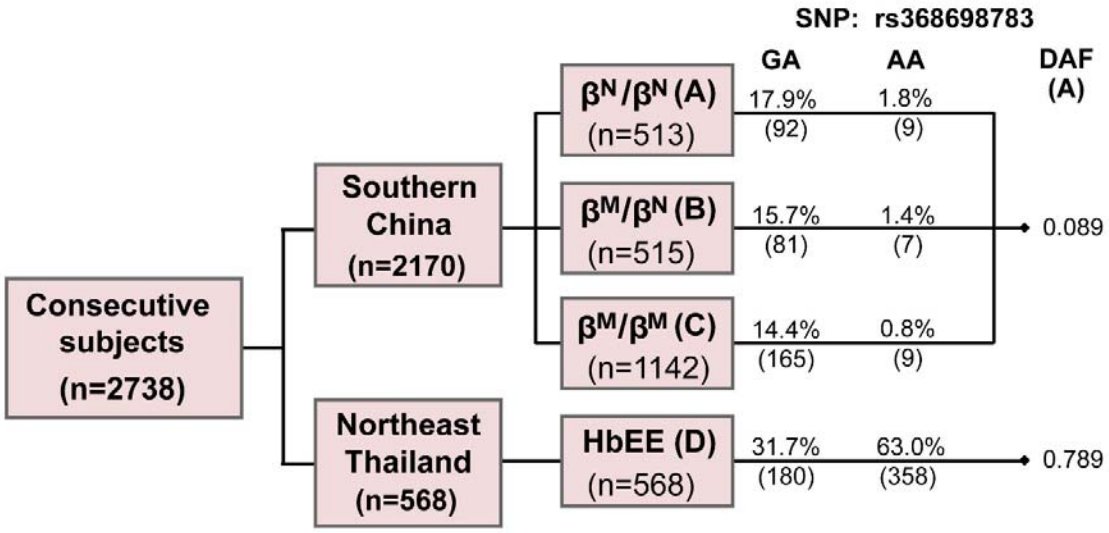
A Genetic Variant Ameliorates β-Thalassemia Severity by Epigenetic-Mediated Elevation of Human Fetal Hemoglobin Expression

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.

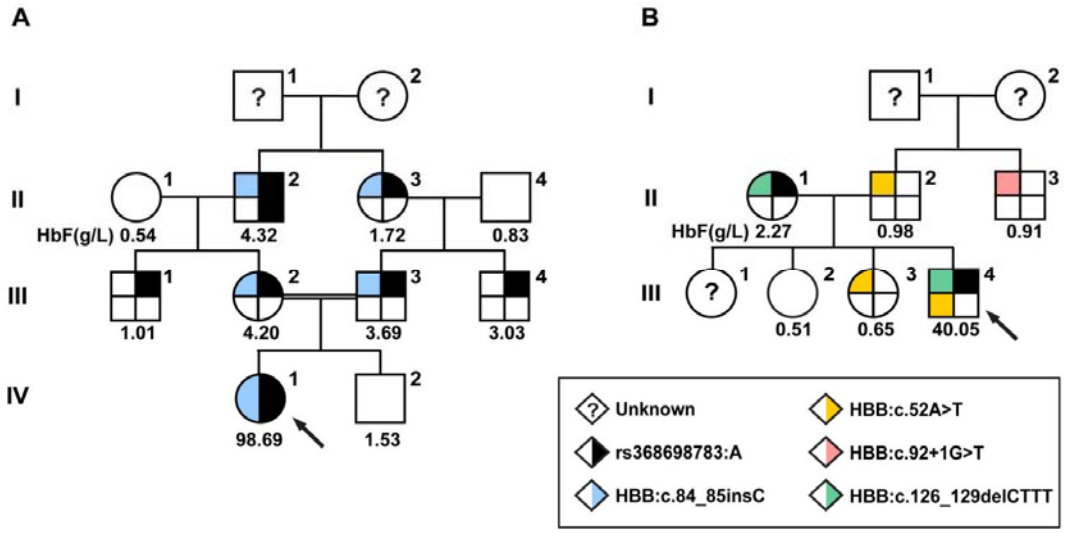
Seven LD blocks containing 163 out of 271 common SNPs were identified in the 80-kb region based on Haploview. Independent LD blocks with highly LD were indicated by black triangle frames. Small box represents D' with strong LD (red), no LD (white) or lack of statistical evidence (blue). See the attached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

Figure S2. Sample design of the present study.



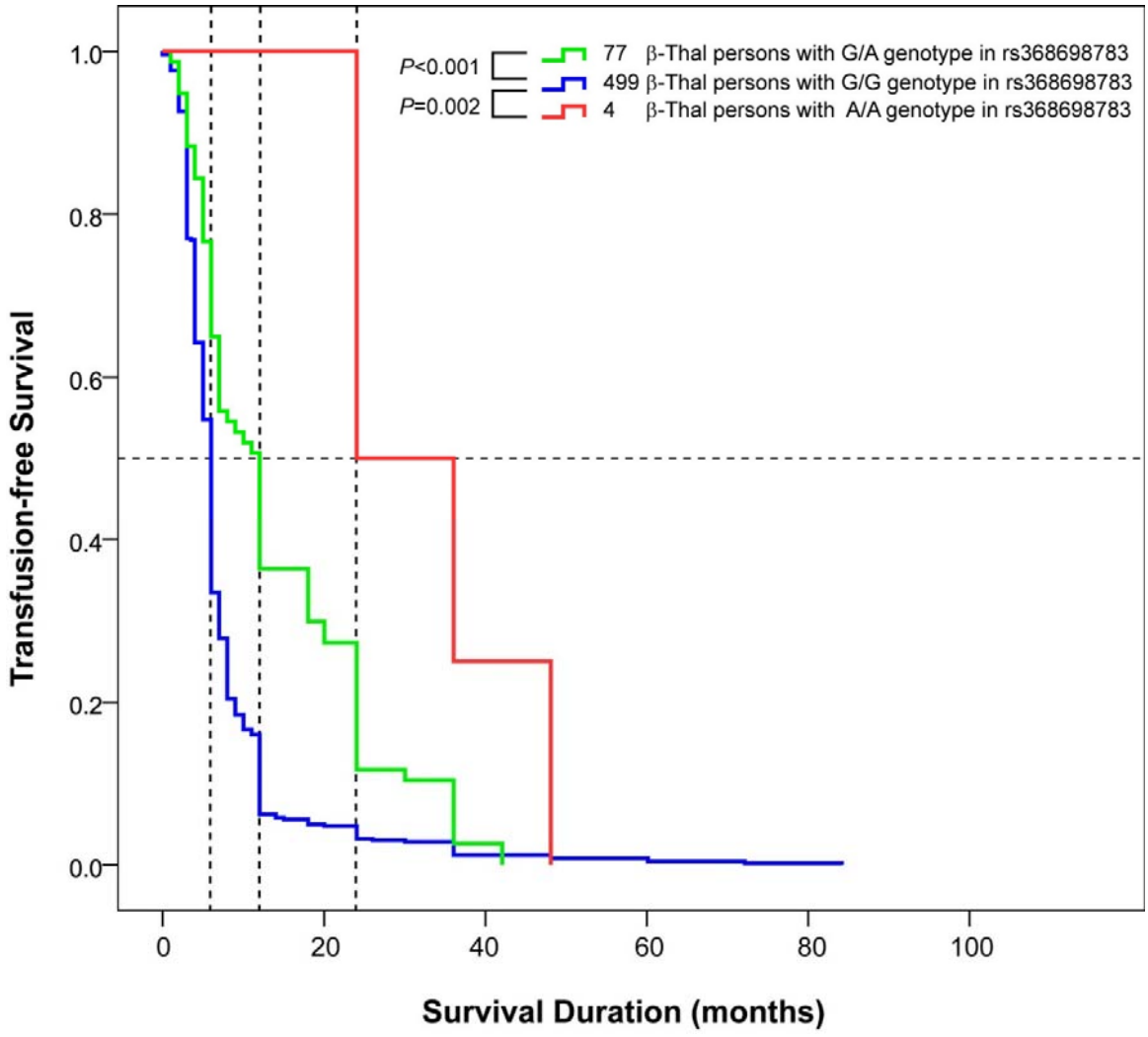
Four consecutive cohorts (n=2738), including 2170 participants from southern China and 568 Thai individuals with HbEE disease, were recruited for this study. The frequency and number of GA and AA genotypes of rs36869873 in each of four cohorts, as well as the derived allele (A) frequency (DAF) in southern 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.



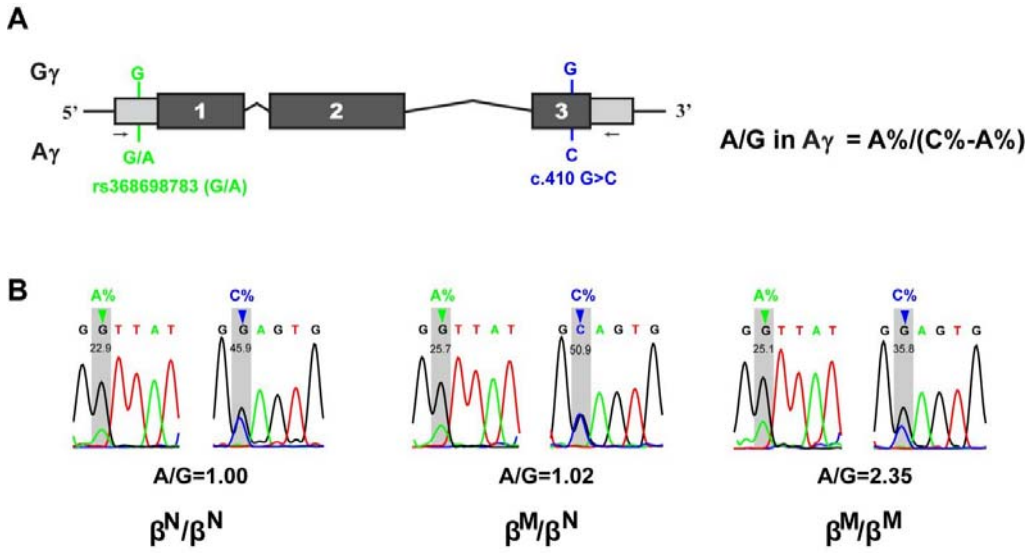
Pedigrees for families A **(A)** and B **(B),** with *HBB* gene mutations and HbF levels indicated for each family member. HbF (g/L) was calculated based on the total Hb level and HbF (%). Squares, males; circles, females; arrows, the proband in each family. To obtain data reflecting the endogenous hemoglobin 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.



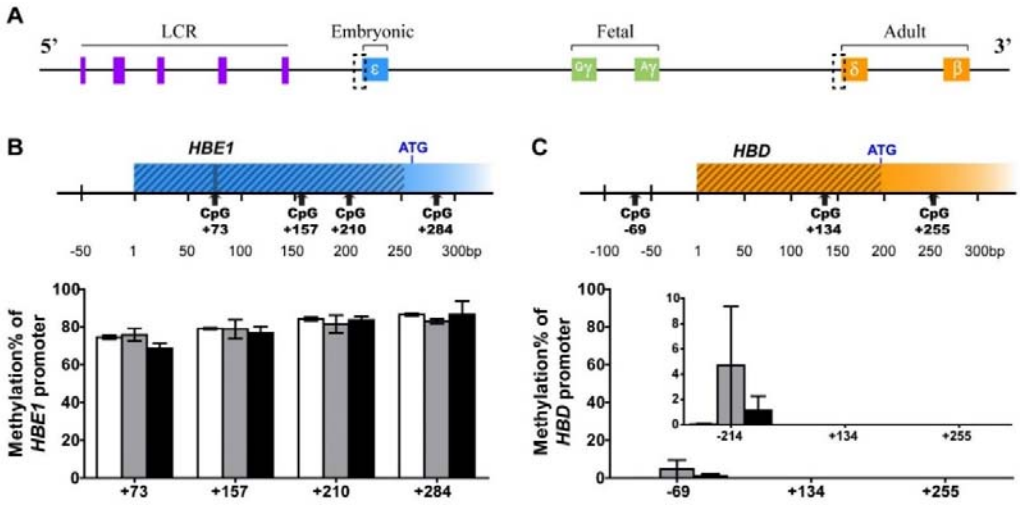
Chinese β-thalassemia individuals with similar genetic variations at the genotypes β^0/β^0, αα/αα, *KLF1* (WT), *BCL11A*-rs4671393 (GG or GA), and *HBS1L-MYB*-rs9399137 (TT or TC) analyzed by Sanger sequencing or high-resolution melting (HRM) as described [3] were recruited into our study as described 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.



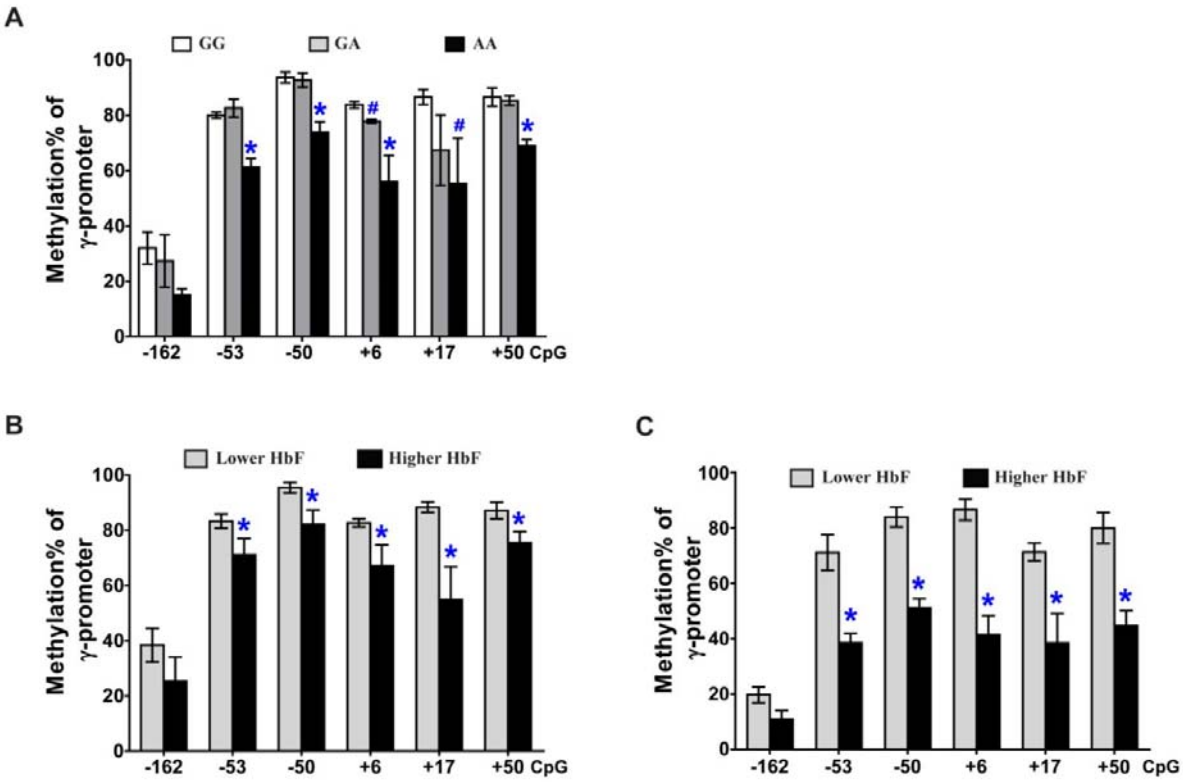
(A) The structure of the γ-globin gene and the locations of markers employed for quantification of the $^{G}\gamma/^{A}\gamma$ ratio or the *HBG1*-rs368698783 allelic A/G ratio. The grey boxes represent non-coding exons, black boxes represent coding exons, and solid lines represent introns. To quantify 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.



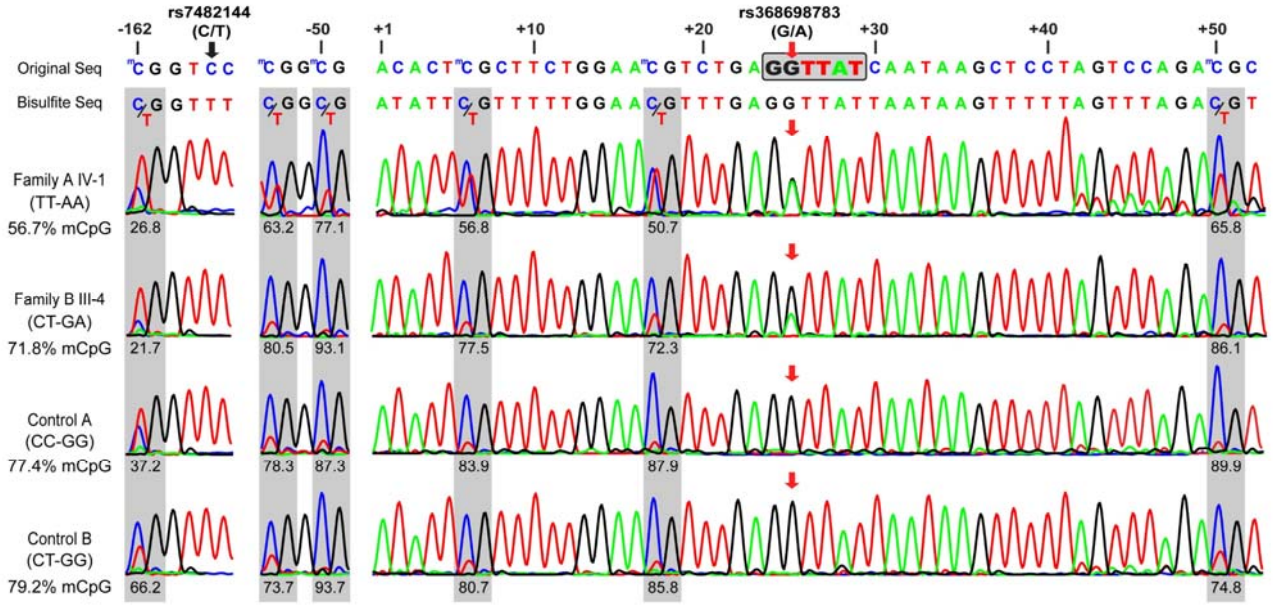
(A) A diagram of the human β-globin cluster. The panels of **(B)** and **(C)** show the locations of CpGs at *HBE1* and *HBD*, respectively. The gray hatched region represents the 5'UTR of *HBE1* **(B)** and *HBD* **(C)**. The effects of the rs368698783 genotypes on the methylation of the core and flank 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.

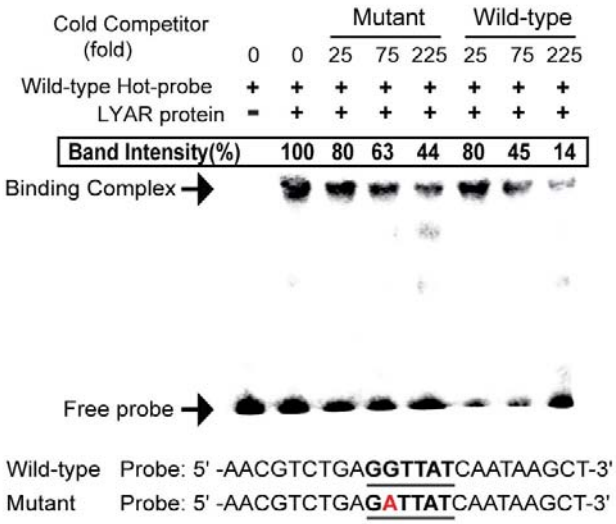


Methylation levels of the γ-globin promoter were determined by the BS-seq method for the CD235a$^+$ cells from human BM **(Figures 2A, S6A&C, S7 and S8**) or the cultured CD34$^+$ cells transduced by the LV3-LYARsiRNA lentivirus for the *LYAR* RNA interference (**Figure 4D**). Four subjects with TT-AA, 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.