Erythroid Krüppel-like factor mutations are relatively more common in a thalassemia endemic region and ameliorate the clinical and hematological severity of \( \beta \)-thalassemia

Dun Liu,1 Xinhua Zhang,2 Lihua Yu,1 Ren Cai,3 Xiaoxia Ma,1 Chengguang Zheng,4 Yuqiu Zhou,5 Qiji Liu,6 Xiaofeng Wei,1 Li Lin,1 Tizhen Yan,3 Jiwei Huang,1 Narla Mohandas,7 Xiuli An,7 and Xiangmin Xu1

1Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China; 2Department of Hematology, 303rd Hospital of the People's Liberation Army, Nanning, Guangxi, China; 3Department of Birth Health and Heredity, Liuzhou Women and Children Care Hospital, Liuzhou, Guangxi, China; 4Prenatal Diagnostic Center, Guangxi Zhuang Autonomous Region Women and Children Care Hospital, Nanning, Guangxi, China; 5Department of Birth Health and Heredity, Zhuhai Women and Children Care Hospital, Zhuhai, Guangdong, China; 6Key Laboratory for Experimental Teratology of the Ministry of Education and Department of Medical Genetics, School of Medicine, Shandong University, Jinan, Shandong, China; 7Red Cell Physiology Laboratory, New York Blood Center, New York City, NY, United States;

Correspondence author: Xiangmin Xu, Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, Guangdong, P.R. China. Tel: +86-020-61648293. Fax: +86-020-87278766. e-mail: gzxuxm@pub.guangzhou.gd.cn.

Running title: KLF1 mutations and \( \beta \)-thalassemia

Scientific Heading: Red Cells, Iron, and Erythropoiesis
Key Points

- The prevalence of KLF1 mutations is significantly higher in a thalassemia endemic region of China than in a non-endemic region.
- KLF1 mutations ameliorate the clinical and hematologic features of β-thalassemia.

ABSTRACT

Mutations in human Krüppel-like factor 1 (KLF1) have recently been reported to be responsible for increased fetal hemoglobin (HbF) and hemoglobin A2 (HbA2). As increased HbF and HbA2 levels are important features of β-thalassemia, we examined whether there is any relationship between KLF1 mutation and β-thalassemia in China. For this, we first studied the incidence of KLF1 mutations in two Chinese populations: 3839 individuals from a thalassemia endemic region in south China and 1190 individuals from a non-thalassemia endemic region in north China. Interestingly we found that the prevalence of KLF1 mutations is significantly higher in thalassemia endemic region than that in non-thalassemia endemic region (1.25% versus 0.08%). Furthermore, we identified seven functional variants including four previously reported (p.Gly176AlafsX179, p.Ala298Pro, p.Thr334Arg and c.913+1G>A) and three novel ones (p.His299Asp, p.Cys341Tyr and p.Glu5Lys) in southern China. The two most common mutations, p.Gly176AlafsX179 and p.His299Asp, accounted for 90.6% of the total. We found that zinc-finger mutations in KLF1 were selectively represented in 12 β-thalassemia intermedia patients, and resulted in significantly different transfusion-free survival curves. Our findings suggest that KLF1 mutations occur selectively in the presence of β-thalassemia to increase the production of HbF which in turn ameliorates the clinical severity of β-thalassemia.
INTRODUCTION

The clinical severity of β-thalassemia varies widely from mild to severe forms and genotype-phenotype associations in β-thalassemia have been extensively studied by systematic analysis of defective β-globin genes. However, the same genotype of β-thalassemia can manifest with highly variable patterns of clinical severity ranging from moderate to severe disease due to various genetic modifiers of disease severity. Several genetic modifiers unlinked to the β-globin gene locus have been identified, including genetic variants modulating fetal hemoglobin (HbF) levels, α-globin gene copy numbers/α-thalassemia, alpha hemoglobin stabilizing protein (AHSP), and heme-regulated eIF2alpha kinase (HRI) in either human patients or murine models of the disease. Increased HbF levels and concomitant α-thalassemia are two main modifiers that can ameliorate the clinical and hematological severity of β-thalassemia.

The erythroid transcriptional factor KLF1 has recently emerged as one of the key regulators of the γ- to β-globin gene switching. It is now known that the dominant mutation (p.E325K) in KLF1 can cause congenital dyserythropoietic anemia (CDA, type IV) and a compound heterozygous mutation in KLF1 lead to chronic hemolytic anemia but most monoallelic KLF1 mutations give rise to benign phenotypes, including significant increases in levels of HbF and adult hemoglobin A2 (HbA2). However, little is known about the modification of β-thalassemia phenotypes in the context of co-inheritance of KLF1 mutations with β-globin gene mutations. In addition, currently there is little information regarding the population genetics of KLF1 mutations. Here, we investigated the incidence and spectrum of KLF1 mutations in the Chinese population, and show that KLF1 mutations are common in a thalassemia endemic region in southern China and that the hematologic features of β-thalassemia are modulated by different KLF1 mutations.
DESIGN AND METHODS

Population, patients and phenotypic tests

We designed a large population-based cohort study that consisted of a total of 3918 randomly selected unrelated subjects living in southern China. This cohort included 3839 consecutive subjects who were investigated to determine the incidence of \( KLF1 \) mutations and 79 healthy subjects with borderline HbA\(_2\) (3.3%–4.1%) and/or elevated HbF levels (>1.5%) to determine \( KLF1 \) mutation frequency in this selected population. All subjects in whom \( KLF1 \) mutations were identified were available for study of both the spectrum of mutations and of the hematological features of \( KLF1 \) variants (Figure 1A). The cohort of 3839 consecutive subjects (aged 1–50 years; 1987 males and 1852 females) was grouped into three subgroups based on the \( \beta \)-globin genotypes: 1971 non-thalassemics (cohort A), 946 \( \beta \)-thalassemia heterozygotes (cohort B), and 922 \( \beta \)-thalassemia homozygotes or compound heterozygotes (cohort C). Cohort A subjects were randomly drawn from 5789 samples used in a previous study. Consecutive samples in cohort B and cohort C were recruited from November 2008 to June 2013 from southern China. The subset of 79 subjects (cohort D) was also selected from a large random Guangxi population screened for the above two positive indices of suspected \( KLF1 \) mutations. All participants are ethnic southern Chinese, whose registered parental origin is from either Guangxi or Guangdong Provinces in southern China where thalassemia is highly prevalent. In addition, we recruited 1190 healthy Chinese subjects from Shandong Province in northern China where thalassemia is much less prevalent. Clinical data was collected retrospectively on all 922 thalassemia patients in cohort C. Diagnosis of thalassemia was made through a comprehensive assessment of clinical presentation, hematological and molecular analysis based on criteria previously described. A summary of clinical and genetic characterizations of 922 patients, 568 thalassemia major (TM) and 354 thalassemia intermedia (TI) in cohort C, is presented in Table 1.

Hematological parameters were measured using an automated hematology analyzer (Sysmex F-820, Sysmex Co. Ltd, Kobe, Japan) and hemoglobin analysis was carried out using the high-performance liquid chromatography (Variant II, Bio-Rad Laboratories, Hercules, CA, USA). Flow-cytometric analysis of CD44 expression on mature erythrocytes was performed with a FACSCanto II flow cytometer (BD Biosciences, San Jose, CA, USA) by immunolabeling of red blood cells (RBCs) with reagents from BD Biosciences-Phar. (BD Biosciences) according to standard protocols. The study protocol was reviewed and approved by the local medical ethics committee at Nanfang Hospital, China, and written consent was obtained from parents and/or patients prior to commencing the study. The study was conducted in accordance with the Declaration of Helsinki.
**Molecular genetic tests**

Genomic DNA was extracted from peripheral blood using the standard phenol/chloroform method. We used the high-resolution melting (HRM) assay to scan for *KLF1* mutations and followed by Sanger sequencing for confirmation of *KLF1* variants (Supplemental Figure 1). A total of nine pairs of primers were designed to amplify the coding region (exons 1-3) and the conserved non-coding regions (promoter region, flanking splice junctions, and 5'- and 3'-untranslated regions) of the *KLF1* gene (Figure 1B and Supplemental Table 1). XmnI PCR-RFLP and HRM assays were used to genotype four primary HbF-associated SNPs (rs7482144, XmnI, in *HBG*, rs11886868 and rs766432 in *BCL11A*, and rs9399137 in *HBS1L-MYB*) (Supplemental Figure 1 and Supplemental Table 1). Point mutations and deletions in the α-, β-, and γ-globin genes were detected using previously described methods.

Functionally relevant *KLF1* mutations were identified on the basis of phenotypic changes (decreased erythrocytes CD44 expression and/or HbF levels >1%) and functional predictions for non-synonymous substitutions in KLF1 analyzed by SIFT (http://sift.jcvi.org) and PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2). To determine whether the mutations are hereditary or somatic, peripheral blood and DNA samples were available from six parents whose offspring were identified as carriers for *KLF1* mutations in cohort C (Supplemental Figure 2).

**Statistical analysis**

Statistical analyses were carried out using SPSS version 19.0 (IBM, Armonk, NY, USA) and *P* < 0.05 was considered statistically significant. All quantitative variables were compared by either the Mann-Whitney *U* test or the χ² test as appropriate. Multivariate analysis was performed using a backward stepwise Cox proportional hazards model in 922 β-thalassemia homozygotes or compound heterozygotes in cohort C, and hazard ratios (HR) with 95% confidence intervals (CI) were obtained to evaluate the associations between putative ameliorating alleles and the patients’ age at first transfusion. All covariates were classified as follows: α-globin gene mutations as 0, 1, 2 or 3 based on the number of mutated copies of the *HBA*; For other loci, each variable was defined with values 0, 1 or 2 according to the number of copies of putative modifying allele: *HBB* mutations (β⁺), rs7482144 (XmnI: +) in *HBG2*, rs11886868 (C) and rs766432 (C) in *BCL11A*, rs9399137 (C) in *HBS1L-MYB*; *KLF1* mutations was only scored as 0 or 1 since no homozygous mutation was noted; Gender for male was 1 and for female was 0. The performance of the model was measured by Harrell’s concordance index (C-index) by using the package of *rms* in R version 3.0.1 (http://www.r-project.org/). Univariate analysis was conducted to compare clinical features by either the Mann-Whitney *U* test or the χ² test as appropriate in β-thalassemia patients with or without *KLF1* mutations, and odds ratios (OR) with 95% CI were calculated to assess the risk of receiving systematic transfusion and TI. These two groups had same globin genotype category.
(β0/β0 and αα/αα) and similar HbF-modulating genetic variants [XmnI (−/−), rs766432 (AA or AC), rs9399137 (TT or CT)]: 7 out of 12 KLF1 mutation-positive subjects and 362 out of 910 KLF1 mutation-negative subjects in cohort C were included in both case-control and survival analysis. Transfusion-free survival curves were depicted according to the Kaplan-Meier method and compared using the log-rank test.
RESULTS

The prevalence of KLF1 mutations is significantly higher in the thalassemia endemic region in southern China.

It has been recently reported that KLF1 mutations result in elevated HbF levels. As elevated HbF is beneficial in β-thalassemia, we hypothesized that some associations may exist between KLF1 mutations and β-thalassemia in regions with high prevalence of thalassemia. To test this hypothesis, we scanned for mutations of the KLF1 gene in two distinct Chinese populations: a total of 3839 unrelated individuals from three different cohorts in southern China where thalassemia is endemic and a total of 1190 individuals from northern China where thalassemia rarely occurs. We identified 48 subjects with KLF1 heterozygous mutations in the cohort from southern China, giving a heterozygote frequency of 1.25% in this population (Figure 1A). In contrast, we identified only one subject with heterozygous missense mutations in the population from north, giving a 0.08% frequency which is significantly lower than found in the population from south (p<0.001). These findings demonstrate that KLF1 mutations have a much higher prevalence in the presence of thalassemia. Comparison of the frequency of KLF1 mutations among the three cohorts from south revealed that the frequency is 1.27% (25/1971) in cohort A, 1.16% (11/946) in cohort B and 1.30% (12/922) in cohort C respectively. Statistical analysis showed that there were no statistically significant differences among these three sample groups (p=0.959), implying that higher KLF1 mutation prevalence is a common phenomenon of a southern Chinese population where thalassemia is endemic.

The spectrum of KLF1 mutations in southern China.

As described above, we have identified 48 heterozygotes with KLF1 mutations in the southern Chinese population comprising 3839 subjects. To enlarge the size of sample with KLF1 mutations for further analysis, we identified additional 16 subjects with KLF1 heterozygous mutations in 79 subjects (non-thalassemics, cohort D) based on borderline increases in HbA2 (3.3%–4.1%) and/or elevated HbF levels (>1.5%) (Figure 1A). Thus a total of 64 mutant alleles were documented in the present study, of which 41 are KLF1 heterozygotes alone, 11 are KLF1 mutations co-inherited with β-thalassemia heterozygotes, and 12 are KLF1 mutations co-inherited with β-thalassemia homozygotes or compound heterozygotes. We identified four previously reported KLF1 gene mutations (p.Gly176AlafsX179, p.Ala298Pro, p.Thr334Arg and c.913+1G>A) and three novel mutations. Among these are five missense mutations (p.Glu5Lys, p.Ala298Pro, p.His299Asp, p.Thr334Arg and p.Cys341Tyr), one frameshift mutation (p.Gly176AlafsX179), one splicing mutation (c.913+1G>A) and the number and percentage of mutant alleles that carried...
these seven different types of KLF1 mutations are listed in Figure 1B and Supplemental Table 2. We determined that the six mutations (p.Gly176AlafsX179 frameshift, p.Ala298Pro, p.His299Asp, p.Thr334Arg, p.Cys341Tyr and c.913+1G>A) disrupting the zinc-finger (ZF) domain of KLF1 accounted for the vast majority of the molecular changes with two mutations (p.Gly176AlafsX179 and p.His299Asp) being most frequent, accounting for 90.6% of the KLF1 mutant alleles in this population (Figure 1B and Supplemental Table 2). In addition, we found 13 novel neutral KLF1 variants including 4 missense (tolerated) and 4 synonymous mutations in the coding region and 5 variants in the 3’-UTR in our population (Supplemental Table 2). Our findings, together with previously published studies12,18,19,29-32 demonstrate the broad spectrum of KLF1 mutations.

The effects of KLF1 mutations on hematologic parameters of non-thalassemic subjects.

Having demonstrated that KLF1 mutations occur at about the similar frequency in all three cohorts (non-thalassemics, β-thalassemia heterozygotes and β-thalassemia homozygotes/compound heterozygotes), we then examined the effect of these mutations on the hematological parameters of these three cohorts. Figure 2A shows the MCV and MCH values of normal individuals with or without KLF1 mutations. Both MCV and MCH were mildly lower in 41 normal subjects with KLF1 mutations (25 from cohort A and 16 from cohort D), compared to the normal subjects without KLF1 mutations (n=1946, P<0.001), demonstrating that mutations of KLF1 result in MCV and MCH values at the lower end of the normal range. In normal subjects without KLF1 mutations, MCV and MCH values were 89.29±4.77 fL (n=1946, range, 79.00–110.10 fL) and 29.46±1.44 pg (n=1946, range, 25.00–36.30 pg), respectively. In contrast, MCV and MCH values were 82.45±5.39 fL (n=41, range, 67.50–91.00 fL) and 26.95±2.11 pg (n=41, range, 21.20–30.00 pg), respectively, in normal subjects with the KLF1 mutations. Furthermore, as shown in Figure 2B, we confirmed borderline increases in HbA2 levels (n=41, 3.28±0.45%, range, 2.50–4.87%) and of slightly elevated HbF levels (n=41, 3.44±1.75%, range, 1.00–7.40%) in KLF1 heterozygous mutants in our normal population, in the two groups (25 from cohort A and 16 from cohort D) that expected to show the same effects of elevated HbA2 and/or HbF.

The effects of KLF1 mutations on hematological parameters of β-thalassemia heterozygote subjects.

We then examined the effects of KLF1 mutations in modulating hematological parameters of β-thalassemia heterozygotes. For this, we analyzed β0-thalassaemia heterozygotes with or without KLF1 mutations. MCV and MCH values were 63.53±3.94 fL (n=14, range, 57.00–72.00 fL) and 20.26±1.42 pg (n=14, range, 18.20–22.90 pg), for 14 β0-thalassaemia heterozygotes (β0/βN and αα/αα genotypes) with the KLF1 mutations (seven out of 11 from cohort B except for one βE/βN, one β+/βN and two β0/βN + −/αα and seven from familial cases, shown in Figure S2), while they
were 65.11±4.34 fL (n=217, range, 51.00-77.80 fL) and 20.34±1.37 pg (n=217, range, 14.60-27.60 pg), for 217 β₀-thalassemia carriers with no co-inherited KLF1 mutations. Surprisingly, in contrast to normal subjects no statistically significant differences between these two groups with respect to these two hematological parameters (P=0.112 or P=0.567, Figure 2A) were noted. This suggests that mutations of KLF1 gene in the context of β-thalassemia trait does not have significant effect on MCV and MCH. However, as shown in Figure 2B, we found that co-existence of β-thalassemia trait and KLF1 mutations results in higher levels of HbA₂, of over 6.5% (n=17, 6.47±0.81%, range, 5.10–8.00%), implying additive effects of these two factors on δ-globin gene expression. In addition, a highly variable increase in HbF levels (n=17, 7.67±4.11%, range, 3.60–18.20%) was noted in individuals with co-inheritance of KLF1 mutations in the context of β-thalassemia trait. The basis of reason of this variability needs to be further studied.

The effects of KLF1 mutations on β-thalassemia homozygotes/compound heterozygote individuals.

To define the role of KLF1 mutations in modulating the clinical and hematological severity, we studied 922 β-thalassemia patients with known genotypes of β-globin and modifier genes including KLF1 mutations (Table 1). We found it surprising that two KLF1 mutations were selectively represented in 12 TI patients who are all β-thalassemia homozygotes or compound heterozygotes in cohort C: Ten subjects had one frameshift mutation (p.Gly176AlafsX179) and the other two had the missense mutation (p.His299Asp). Both mutations disrupt the KLF1 ZF motif. The clinical features of all 12 affected β-thalassemia individuals are presented in Supplemental Table 3. We assessed the influence of different combinations of genetic factors on the entire cohort by multivariate Cox regression analysis, measured by age at first transfusion, and the results showed that the KLF1 mutations had the strongest effect on the clinical severity of β-thalassemia phenotype (HR=0.213, P<0.001), followed by β⁺ mutation (HR=0.353, P<0.001), rs7482114 (XmnI,+) in HBG2 (HR=0.512, P<0.001), HBA mutations (HR=0.671, P<0.001), rs9399137 (C) in HBS1L-MYB (HR=0.724, P<0.001) and rs766432 (T) in BCL11A (HR=0.761, P<0.001) (Table 2). In these 12 TI patients, seven subjects were β₀/β₀ and αα/αα genotypes (patients P3, P4, P5, P6, P7, P11 and P12) and three subjects (patients P1, P2 and P8) were β⁺/β⁺ and αα/αα genotypes and two subjects (patients P9 and P10) were β⁺/β⁺ co-inherited with the Southeast Asian deletion (SEA) and 3.7kb α-thalassemia deletion respectively (Supplemental Table 3). We further tested the effects of KLF1 mutations in modulating the clinical severity of β-thalassemia by univariate analysis of a matched case-control of the above seven subjects (β⁺/β⁺ and αα/αα genotypes with KLF1 mutations) and 362 thalasemic subjects without KLF1 mutation in cohort C. As shown in
Table 3, univariate analysis confirmed the links between the \textit{KLF1} mutations and decreased clinical severity with remarkably increased hemoglobin concentration ($P=0.006$) and HbF levels ($P<0.001$) for TI cases with \textit{KLF1} mutations, along with delay in mean age at onset of anemia ($P=0.001$), decreased transfusion requirements (OR=0.082, $P=0.007$), and a lower risk of exhibiting the features of TM (OR=0.009, $P<0.001$). This difference in HbF levels between 7 vs. 362 matched samples is consistent with the results of the HbF levels between the whole TI vs. TM populations (Table 1), implying a clinically homogeneous group of TI patients in our cohort C whether they had a \textit{KLF1} mutation or not. Kaplan–Meier survival curves showed significant differences between the two groups (log-rank test $P<0.001$). The median time to the first transfusion in Control Group (362 without \textit{KLF1} mutations) was 6 months, compared to a highly significant improvement in transfusion-free survival observed in Case Group (7 with \textit{KLF1} mutations) with the median time to the first transfusion at 36 months. Likewise, 342 \textit{KLF1} mutation-negative TI group received their first transfusion at 24 months, whereas it was initiated at 48 months in 12 \textit{KLF1} mutation-positive TI group, although there was no statistically significant difference between two TI subgroups (log-rank test $P=0.093$) (Figure 3).

\textbf{Prevalence of \textit{KLF1} mutations is high in individuals with increased HbA$_2$ and HbF levels.}

To further establish the link between \textit{KLF1} mutations and HbA$_2$ and HbF levels, we studied 79 non-thalassemic subjects (cohort D) with elevated levels of HbA$_2$ and HbF. In this sub-group, we found that the prevalence of \textit{KLF1} mutations to be 20.25% (16/79), suggesting elevation of HbA$_2$ and HbF levels can potentially be used for screening of \textit{KLF1} mutations in the population.
DISCUSSION

We studied the prevalence of KLF1 mutations through a population-based molecular analysis of a large number of samples comprising 3839 participants in southern China (thalassemia endemic region) and 1190 normal individuals in northern China (non-thalassemia endemic region) using a HRM assay we developed. We documented a surprisingly high frequency of KLF1 heterozygous mutations (1.25%) in the population from southern China while it was only 0.08% in northern China. These findings demonstrate that mutations in the KLF1 gene are common in the thalassemia endemic region in southern China. Interestingly, while no functionally relevant KLF1 mutations in the ZF domain were identified in the northern Chinese population such mutations occurred at a very high frequency in southern China, accounting for 98.4% of the KLF1 mutant alleles.

We further documented that KLF1 mutation is a genetic modifier ameliorating the clinical severity of β-thalassemia. Moreover, we observed the presence of hypochromic and microcytic red cells (thalassemia-like) or borderline MCV and MCH values in most non-thalassemic individuals as a result of KLF1 mutations. A recent study confirmed that compound heterozygotes for KLF1 mutations are able to cause hypochromic and microcytic anemia. Thus, we hypothesize that the frequent occurrence of the KLF1 mutations that affect protein function (particularly those disrupted the ZF domain of KLF1) might arise due to natural selection, such as in the presence of malaria in the tropical and sub-tropical areas. Our current results imply that KLF1 heterozygous mutations may be one important genetic modifier associated with the inherited hemoglobin disorders, particularly in β-thalassemia, among the southern Chinese population and potentially other ethnic populations.

We have identified 64 KLF1 mutant alleles with seven different mutations (Figure 1 and Supplemental Table 2). The detailed genetic analysis of blood samples from six available families showed germ line origin of KLF1 mutations (Supplemental Figure 2). The frameshift mutation, p.Gly176AlafsX179, was previously found in Thailand, Vietnamese and Korean populations, and its finding in Chinese populations by us and others suggests that this mutation appears to be predominant in East Asian populations. In our study, only functionally effective KLF1 mutations were scored as positive. We typed all 64 subjects with KLF1 mutations according to the following two major predictors of altered function of KLF1: (1) The presence of reduced CD44 expression on erythrocytes by flow-cytometric analysis, and this was observed in the representative samples from all six types of mutations studied (Supplemental Table 2 and Supplemental Figure 3); (2) Determination of elevated HbF levels (>1%), which was the case for all positive samples. All KLF1 variants identified in our study were submitted to the dbSNP (http://www.ncbi.nlm.nih.gov/snp). Decreased CD44 expression due to ZF domain mutations or
frameshift mutations disrupting ZF motifs of human KLF1 gene has been previously documented by several studies. \textsuperscript{1,13,16,31,32} Thus, use of a molecular assay to scan for KLF1 mutation-positive in combination with a flow-cytometric analysis for determining decreased CD44 expression will be a valuable tool for establishing functionally relevant KLF1 mutations. In addition to this, the In(Lu) phenotype had been previously used in British and France population to screen for KLF1 mutations to obtain reliable prevalence figures.\textsuperscript{31,32} By ELISA assay using labeled anti-ζ monoclonal antibody\textsuperscript{34}, elevated ζ-globin chains expression were found in four of the seven available TI patients (P4, P6, P7 and P9, Supplemental Table 3) with KLF1 heterozygous mutations, which has recently been reported in a transfusion-dependent hemolytic anemia caused by compound heterozygosity for KLF1 mutations.\textsuperscript{12}

It is well known that the clinical manifestations of thalassemia can be modified by several genetic determinants including genetic variants that modulate HbF levels, β-globin mutations and α-globin gene copy numbers.\textsuperscript{2-4} Based on previous studies, which showed that the erythroid transcriptional factor KLF1 was a key regulator of the γ to β-globin gene switching\textsuperscript{9,10} and KLF1 haploinsufficiency correlated with HPFH caused by decreased BCL11A expression,\textsuperscript{16} we decided to investigate the role of KLF1 mutations in modulating the clinical and hematological severity of thalassemia. To the best of our knowledge, this is the first study reporting the clinical features of 12 β-thalassemia patients co-inherited with KLF1 mutations, which had a mean Hb concentration of 82.57±7.81 g/L and a mean HbF level of 39.58±21.80 g/L and indicative of the amelioration of β-thalassemia severity by KLF1 mutations. Multivariate Cox regression analysis showed that the modifier of KLF1 mutations had an effect on the clinical severity of β-thalassemia phenotype (Table 2) and univariate analysis using a matched case-control (Table 3) and transfusion-free survival curve (Figure 3) supported results of the multivariate analysis. In our case, two KLF1 mutations, p.Gly176AlafsX179 and p.His299Asp, were both verified to be functional variants of KLF1, which result from generating a truncated protein with loss of the three ZF motifs downstream by frameshift mutation or abnormal structural changes in the ZF1 domain of KLF1 by missense mutation. The clinical features of our KLF1 mutation-positive patients might be ascribed to suppression of BCL11A expression and elevated γ-globin levels due to altered DNA-binding domain of KLF1 and impaired transactivation capacity of KLF1, as a similar results due to KLF1 mutations have been previously described.\textsuperscript{10,13-17} Furthermore, our functional studies in K562 cells using p.His299Asp and p.Gly176AlafsX179 mutant KLF1 constructs support such a possibility (data not shown). In addition, it needs to be noted that three subjects (patients P1, P2 and P8) with the β\textsuperscript{0}/β\textsuperscript{+} genotype and two (patients P9 and P10) who co-inherited with HBA deletions had a very mild, transfusion-independent phenotype highlighting that a combination of KLF1 mutations and known modifiers can greatly modify the clinical phenotype. In conclusion, we have defined a novel and important genetic determinant, KLF1 ZF heterozygous mutations, that contribute to TI. This knowledge should help in clinical diagnoses and genetic counseling, as well as enable
designing of appropriate and personalized transfusion program for thalassemia patients with \textit{KLF1} mutations. The use of decreased CD44 expression as a surrogate for \textit{KLF1} mutations could be a clinically important tool to evaluate individuals with unexplained TI. We suggest that it is important to test for \textit{KLF1} mutations to ensure accurate clinical assessment in those TI patients with variable HbF increases since \(\beta\)-thalassemia which benefits from increased HbF can be caused by various genetic factors.

This study provides an overview of the important hematological features of individuals with \textit{KLF1} heterozygous mutations. Our finding of borderline MCV and MCH values in normal carriers with \textit{KLF1} mutations validates previous findings of borderline HbA\textsubscript{2} levels and elevated HbF levels resulting from \textit{KLF1} mutations.\textsuperscript{16-19} Furthermore, we noted that a heterozygote \textit{KLF1} mutation co-inherited with \(\beta\)-globin mutations can lead to a remarkable increase in both HbA\textsubscript{2} and HbF levels, compared to those with only \textit{KLF1} heterozygous mutations. These two phenotypic outcomes are most likely a result of the cumulative effect of the heterozygous \(\beta\)-thalassemia and \textit{KLF1} mutations. At the phenotypic level with respect to HbA\textsubscript{2} and HbF, there are no differences among subjects with different \textit{KLF1} and \(\beta\)-globin mutations. Of interest, in the light of re-analysis of two previously published data (Supplemental Table 4), we found that ten KLF1 target genes associated with red blood cell phenotypes might underlie the decreased MCV and MCH values. This could explain at least in part our findings in normal individuals with loss-of-function mutation in KLF1.\textsuperscript{35,36} However, the underlying mechanisms remain to be further investigated. These findings are likely to be useful for the accurate determination of \(\beta\)-thalassemia carriers in screening programs.
ACKNOWLEDGEMENTS

This study was supported by grants from the National Natural Science Foundation of China (NSFC)-Guangdong Joint Fund (NO.U1201222), NSFC (81260093) and the National Key Technology R&D Program (No.2012BAI09B01) and NIH DK26263. The authors would like to thank all of the members who participated in this study; and we thank Yaoqin Gong for assistance of sample collection in northern China, and Ping Fang, Yihong Li and Xuelian Zhang for experiment assistance.

AUTHORSHIP

Contribution: X.Z., R.C., C.Z., Y.Z., Q.L., L.L. and T.Y. collected samples and clinical data; X.Z. performed clinical classifications, diagnosis and management of thalassemia; L.Y., X.M. and D.L. performed laboratory and DNA analysis; X.W. and J.H. did molecular diagnosis of thalassemia; N.M. and X.A. designed the study, interpreted the data and edited the paper; D.L. and X.X. designed the study, analyzed and interpreted the data, and wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Xiangmin Xu, Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, 510515, Guangdong, P.R. China; e-mail: gzxuxm@pub.guangzhou.gd.cn.
REFERENCES


33. Weatherall DJ, Clegg JB. Genetic variability in response to infection: malaria and after.


Table 1. Characteristics of 922 beta-thalassemia patients in cohort C.

<table>
<thead>
<tr>
<th>Patient characteristics</th>
<th>Disease severity</th>
<th>( P^* )</th>
<th>HWP**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TM (n=568)</td>
<td>TI (n=354)</td>
<td></td>
</tr>
<tr>
<td>Gender, n</td>
<td>Males : Females</td>
<td>373:195</td>
<td>217:137</td>
</tr>
<tr>
<td>Age (years), mean ± SD</td>
<td>6.47 ± 4.24</td>
<td>11.81 ± 9.15</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hematological data, mean ± SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemoglobin (g/L)</td>
<td>64.68 ± 17.88</td>
<td>71.32 ± 16.93</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HbF (g/L)</td>
<td>9.93 ± 9.34</td>
<td>25.52 ± 19.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Requirement for systematic transfusion, n (%)</td>
<td>568 (100.0)</td>
<td>132 (37.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age at first transfusion (months), median (5 th−95 th percentile)</td>
<td>6.0 (2-24)</td>
<td>24.0 (3-144)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HBB genotype categories, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta^+ ) / ( \beta^+ )</td>
<td>0 (0.0)</td>
<td>8 (2.3)</td>
<td></td>
</tr>
<tr>
<td>( \beta^0 ) / ( \beta^+ )</td>
<td>95 (16.7)</td>
<td>197 (55.6)</td>
<td></td>
</tr>
<tr>
<td>( \beta ) / ( \beta )</td>
<td>473 (83.3)</td>
<td>149 (42.1)</td>
<td></td>
</tr>
<tr>
<td>HBA genotype categories, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha\alpha ) / ( \alpha\alpha )</td>
<td>508 (89.3)</td>
<td>271 (76.6)</td>
<td></td>
</tr>
<tr>
<td>( \alpha\alpha ) / ( \alpha\alpha )</td>
<td>28 (4.9)</td>
<td>29 (8.2)</td>
<td></td>
</tr>
<tr>
<td>( \alpha\alpha ) / ( \alpha\alpha )</td>
<td>5 (0.9)</td>
<td>9 (2.5)</td>
<td></td>
</tr>
<tr>
<td>( \alpha\alpha ) / ( \alpha\alpha )</td>
<td>27 (4.8)</td>
<td>40 (11.3)</td>
<td></td>
</tr>
<tr>
<td>( \alpha\alpha ) / ( \alpha\alpha )</td>
<td>0 (0.0)</td>
<td>1 (0.3)</td>
<td></td>
</tr>
<tr>
<td>( \alpha\alpha ) / ( \alpha\alpha )</td>
<td>0 (0.0)</td>
<td>1 (0.3)</td>
<td></td>
</tr>
<tr>
<td>( \alpha\alpha ) / ( \alpha\alpha )</td>
<td>0 (0.0)</td>
<td>3 (0.8)</td>
<td></td>
</tr>
<tr>
<td>BCL11A</td>
<td>rs11886868, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0.543</td>
</tr>
<tr>
<td>T/C</td>
<td>29 (5.1)</td>
<td>8 (2.3)</td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>539 (94.9)</td>
<td>346 (97.7)</td>
<td></td>
</tr>
<tr>
<td>rs766432, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/A</td>
<td>345 (60.7)</td>
<td>166 (49.6)</td>
<td>0.646</td>
</tr>
<tr>
<td>A/C</td>
<td>192 (33.8)</td>
<td>163 (46.0)</td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>31 (5.5)</td>
<td>25 (7.1)</td>
<td></td>
</tr>
<tr>
<td>HBS1L-MYB intergenic region</td>
<td>rs9399137, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T</td>
<td>398 (70.1)</td>
<td>220 (62.1)</td>
<td>0.627</td>
</tr>
<tr>
<td>T/C</td>
<td>155 (27.3)</td>
<td>116 (32.8)</td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>15 (2.6)</td>
<td>18 (5.1)</td>
<td></td>
</tr>
<tr>
<td>HBG2</td>
<td>rs7482144 (XmnI), n (%)</td>
<td></td>
<td>0.721</td>
</tr>
<tr>
<td>+/−</td>
<td>509 (89.6)</td>
<td>260 (73.4)</td>
<td></td>
</tr>
<tr>
<td>+/+</td>
<td>57 (10.0)</td>
<td>90 (25.4)</td>
<td></td>
</tr>
<tr>
<td>KLF1 mutations, n (%)</td>
<td>0 (0.0)</td>
<td>12 (3.4)</td>
<td></td>
</tr>
</tbody>
</table>

Definition of TM/TI in this study is based on the following four clinical indications: (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).

**P** for the Hardy-Weinberg equilibrium test of rs11886868, rs766432, rs9399137 and rs7482144 in 922 individuals.

*P* were determined using either a Mann-Whitney *U* test or the the $\chi^2$ test as appropriate in comparing clinical data between TM and TI.

Hemoglobin levels were untransfused or pre-transfusion data.

HbF (g/L) was calculated from total Hb level and HbF (%).

A very few patients were classified as TI according to clinical diagnostic criteria although they underwent their first transfusion before 6 months. In these cases, the reason for early transfusion was recorded to be to manage an acute hemolysis due to infections during early infancy.

*HBB* genotype categories were defined as follows: ($\beta^0$) CD41/42 –TTCT (40.0%), CD17 A>T (28.2%), IVS2-654 C>T (6.2%), CD71/72 +A (4.7%), IVS1-1 G>T (2.5%), CD43 G>T (1.1%), CD27/28 +C (0.2%), CD30 A>G (0.1%), CD14/15 +G (0.1%), CD38 –A (0.1%), IVS2-1 G>A (0.1%), IVS1-130 G>C (0.1%); ($\beta^+$) -28 A>G (9.0%), CD26 G>A (5.6%), -29 A>G (1.1%), IVS2-5 G>C (0.7%), -90 C>T (0.2%), -31 A>G (0.1%), IVS1-5 G>C (0.1%).

*HBA* genotype categories were defined as follows, -$\alpha$: -$\alpha^3.7$, -$\alpha^{4.2}$; --: --SEA; $\alpha^T$: $\alpha^CS$, $\alpha^{WS}$. 
Table 2. Results of the stepwise Cox proportional hazards analysis for 922 β-thalassemia patients in cohort C.

<table>
<thead>
<tr>
<th>Ameliorating alleles</th>
<th>P</th>
<th>Hazard ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KLF1 mutations</strong></td>
<td>&lt;0.001</td>
<td>0.213</td>
<td>0.119-0.379</td>
</tr>
<tr>
<td><strong>HBB mutations (β⁺)</strong></td>
<td>&lt;0.001</td>
<td>0.353</td>
<td>0.305-0.409</td>
</tr>
<tr>
<td><strong>HGB2</strong></td>
<td>&lt;0.001</td>
<td>0.512</td>
<td>0.432-0.608</td>
</tr>
<tr>
<td>rs7482114 (XmnI, +)</td>
<td>&lt;0.001</td>
<td>0.671</td>
<td>0.600-0.750</td>
</tr>
<tr>
<td><strong>HBA mutations</strong></td>
<td>&lt;0.001</td>
<td>0.724</td>
<td>0.644-0.813</td>
</tr>
<tr>
<td><strong>HBS1L-MYB intergenic region</strong></td>
<td>&lt;0.001</td>
<td>0.761</td>
<td>0.680-0.851</td>
</tr>
<tr>
<td>rs9399137 (C)</td>
<td>&lt;0.001</td>
<td>0.761</td>
<td>0.680-0.851</td>
</tr>
<tr>
<td><strong>BCL11A</strong></td>
<td>&lt;0.001</td>
<td>0.761</td>
<td>0.680-0.851</td>
</tr>
<tr>
<td>rs766432 (T)</td>
<td>&lt;0.001</td>
<td>0.761</td>
<td>0.680-0.851</td>
</tr>
</tbody>
</table>

Gender (HR=0.893, P=0.107) and rs11886868 in BCL11A (HR=0.871, P=0.415) were removed from the model. The discriminative ability of the model was high (C-index=0.714, R²=0.344).
Table 3. Case-control univariate analysis of KLF1 mutations.

<table>
<thead>
<tr>
<th>Patient characteristics</th>
<th>KLF1&lt;sup&gt;μ/μ&lt;/sup&gt; (case, n = 7)</th>
<th>KLF1&lt;sup&gt;WT/WT&lt;/sup&gt; (control, n = 362)</th>
<th>( P )</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males : Females</td>
<td>5 : 2</td>
<td>233 : 129</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Age of onset (years), mean ± SD</td>
<td>13.71 ± 10.03</td>
<td>5.88 ± 4.10</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Hemoglobin (g/L), mean ± SD</td>
<td>82.57 ± 7.81</td>
<td>66.33 ± 17.53</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>HbF (g/L), mean ± SD&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>39.58 ± 21.80</td>
<td>9.17 ± 11.05</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Requirement for systematic transfusion, n</td>
<td>3 : 4</td>
<td>21 : 341</td>
<td>0.007</td>
<td>0.082 (0.017–0.391)</td>
</tr>
<tr>
<td>Category of anemia</td>
<td>TI : TM</td>
<td>7 : 0</td>
<td>&lt;0.001</td>
<td>0.009 (0.000–0.153)</td>
</tr>
</tbody>
</table>

<sup>1</sup>Similar genetic variants between case group and control group: The impact of five major genetic modifiers of β-thalassemia analyzed under Cox proportional hazards model was excluded from univariate analysis of KLF1 mutations.

<sup>∗</sup> \( P \) were determined using either a Mann-Whitney \( U \) test or the \( \chi^2 \) test as appropriate.

<sup>‡</sup>HbF (g/L) was calculated from total Hb level and HbF (%).
Figure Legends.

Figure 1. The prevalence and spectrum of KLF1 mutations in the Chinese population. (A) Sampling design and study outcomes. Southern Chinese population: Three consecutive cohorts (A, B and C) of a total of 3839 subjects from two thalassemia endemic regions (Guangxi and Guangdong Provinces) were designed to investigate the incidence of KLF1 mutations; 79 phenotype-oriented subjects (cohort D) selected from the same region were recruited for enrichment of KLF1 mutation-positive subjects. The β-globin gene in all 79 subjects were completely analyzed and those positive for β-thalassemia mutations were excluded from analysis. All mutant alleles (n=64) detected in these two approaches were used to assess the mutation spectrum of KLF1. The HBB genotype categories of 11 KLF1 mutation-positive subjects in cohort B were as follows: one β7/βN, one ββ/βN, seven β0/βN and two β0/βN co-inherited with α-thalassemia. Northern Chinese population: 1190 non-thalassemics were recruited from a non-thalassemia endemic region (Shandong Provinces). (B) Exons, introns, and domains are shown with the untranslated regions (white), proline-rich regions (blue), and zinc fingers (purple). The positions of the HRM primers are noted above the physical map, and four novel and four previously reported functional KLF1 mutations in south (red) and north (grey) China are shown below the map. Allelic distribution of KLF1 mutations in southern Chinese population were indicated in parentheses.

Figure 2. Characteristic hematological analyses of KLF1 heterozygotes in cohort A and B. (A) MCV and MCH values in normal individuals (n=1946 from cohort A), non-thalassemics (n=41; 25 from cohort A and 16 from cohort D) with seven KLF1 heterozygous mutations, β0-thalassaemia heterozygotes (n=217 from cohort B) and β0-thalassaemia heterozygotes co-inherited with KLF1 mutations (n=14; seven β0/βN out of 11 from cohort B and other seven β0/βN from six β-thalassemia families with KLF1 mutations, shown in Figure S2). The α-thalassemia deletions or point mutations were excluded in all non-thalassemic and heterozygous β0-thalassemia individuals with and without KLF1 mutations. The data shown are expressed as mean±SD. P were determined using the Mann-Whitney U test. (B) HbA2 and HbF levels in normal individuals (left) and β-thalassaemia heterozygotes (right) with or without KLF1 mutations (all have normal α globin genotypes). Left: the solid circles represent samples with KLF1 mutations, including 25 from cohort A (blue) and 16 from cohort D (green) and 50 without KLF1 mutations (black, HbA2, 2.81±0.20%, range, 2.50–3.20%; HbF, 0.23±0.22%, range, 0.00–0.80%). Right: the solid circles represent 14 samples with KLF1 mutations (red) and 50 without KLF1 mutations (black, HbA2, 4.96±0.49%, range, 3.70–6.10%; HbF, 0.93±0.70%, range, 0.20–4.10%). Genotype symbols of subjects are shown on the right bottom of the chart.
Figure 3. Kaplan-Meier survival curves for β-thalassemia patients with or without KLF1 mutations for transfusion need in cohort C (log-rank test $P$ value). Patients had never received a transfusion by the end of the study, which were recorded as censored data. The two sample matched pairs, 7 vs. 362 β-thalassemia cases with or without KLF1 mutations and 12 vs. 342 TI cases with or without KLF1 mutations, are shown on the top of the chart. The four different colored lines represent each of the four groups respectively. Selection of well-matched 7 vs. 362 subjects is described in a statistical analysis section. To compare between two groups of TI patients in the median time to the first transfusion, all 12 TI cases with KLF1 mutations versus all 342 TI cases without KLF1 mutations are used for Kaplan-Meier survival curve analysis.
Figure 1

A

Thalassemia endemic regions in China (Weatherall DJ, Nature Reviews Genetics, 2001; 2(4):245-255)

Northern Chinese population

Consecutive subjects (non-thalassemics, n=1190)

p.Leu14pro (n=1)

Prevalence of KLF1 mutations in northern China is 0.08% (1/1190)

Consecutive subjects (n=3839)

Cohort A
(non-thalassemics, n=1971)

Cohort B
(β-thalassemia heterozygotes, n=946)

Cohort C
(β-thalassemia homozygotes/compound heterozygotes, n=922)

Cohort D
[non-thalassemics with borderline HbA2 (3.3%–4.1%) and/or elevated HbF levels (>1.5%), n=79]

Prevalence of KLF1 mutations in southern China is 1.25% (48/3839)

Spectrum of KLF1 mutations in southern China

25/1971 (1.27%)

12/922 (1.30%)

16/79 (20.25%)

48 mutant alleles

64 mutant alleles

B

KLF1

Pro Ex1

Ex2-1 Ex2-2 Ex2-3 Ex2-4

Ex3 3UTR-1 3UTR-2

5' F R F R F R R F R

c.41T>C

c.13G>A (1.6%)

c.519_525dupCGGCGCC (75%)

c.892G>C (1.6%)

c.913+1G>A (1.6%)

c.895C>G (15.6%)

c.1001C>G (1.6%)

c.1022G>A (3.1%)

3'
Erythroid Krüppel-like factor mutations are relatively more common in a thalassemia endemic region and ameliorate the clinical and hematological severity of β-thalassemia

Dun Liu, Xinhua Zhang, Lihua Yu, Ren Cai, Xiaoxia Ma, Chengguang Zheng, Yuqiu Zhou, Qiji Liu, Xiaofeng Wei, Li Lin, Tizhen Yan, Jiwei Huang, Narla Mohandas, Xiuli An and Xiangmin Xu

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include digital object identifier (DOIs) and date of initial publication.