



OPEN

Analysis of rare thalassemia genetic variants based on third-generation sequencing

Cuiting Peng^{1,2}, Haixia Zhang^{1,2}, Jun Ren^{1,2}, Han Chen^{1,2}, Ze Du^{1,2}, Tong Zhao^{1,2}, Aiping Mao³, Ruofan Xu³, Yulin Lu³, He Wang^{1,2}, Xinlian Chen^{1,2,4}✉ & Shanling Liu^{1,2,4}✉

Thalassemia is a group of common hereditary anemias that cause significant morbidity and mortality worldwide. However, precisely diagnosing thalassemia, especially rare thalassemia variants, is still challenging. Long-range PCR and long-molecule sequencing on the PacBio Sequel II platform utilized in this study could cover the entire *HBA1*, *HBA2* and *HBB* genes, enabling the diagnosis of most of the common and rare types of thalassemia variants. In this study, 100 cases of suspected thalassemia were subjected to traditional thalassemia testing and third-generation sequencing for thalassemia genetic diagnosis. Compared with traditional diagnostic methods, an additional 10 cases of rare clinically significant variants, including 3 cases of structure variants and 7 cases of single nucleotide variations (SNVs) were identified, of which a case with $-\alpha^{3.7}$ subtype III ($-\alpha^{3.7III}$) was first identified and validated in the Chinese population. Other rare variants of 11.1 kb deletions ($-11.1/\alpha$), triplicate α -globin genes ($aaa^{3.7}/\alpha$) and rare SNVs have also been thoroughly detected. The results showed that rare thalassemia variants are not rare but have been misdiagnosed by conventional methods. The results further validated third-generation sequencing as a promising method for rare thalassemia genetic testing.

Thalassemia is a group of common hereditary anemias that cause significant morbidity and mortality worldwide, especially in Southeast Asian, Middle Eastern and Mediterranean populations^{1,2}. α -Thalassemia (α -thal) and β -thalassemia (β -thal) are two main types of thalassemia caused by mutations in the *HBA1/2* and *HBB* genes respectively, which result in abnormal α - and β -globin synthesis and defective hemoglobin structure^{3,4}.

The clinical manifestations of thalassemia vary greatly and are usually directly related to the degree of globin chain reduction. There are three types of thalassemia carrier states (silent, minor and intermedia) and one disease state (major) according to the amount of globin chain and the disease severity^{5,6}. α -Thalassemia major is also known as Hb Bart's hydrops fetalis, and affected fetuses usually die in utero or shortly after birth due to severe anemia and lack of oxygen⁷. Whereas, affected individuals with the most severe form of β -thalassemia usually can survive, but only with regular blood transfusions and iron chelation therapy^{8,9}. It has been estimated that approximately 5% and 1.5% of the population worldwide are carriers of α - and β -thalassemia genetic mutations. Thus, a large number of children are still born annually with hemoglobin disorders, which cause severe birth defects and place heavy burdens on society and families^{3,10}.

Due to the complexity of thalassemia genetics and genotype–phenotype correlation, precisely diagnosing thalassemia patients and carrier status are still challenging¹¹. Effective and accurate molecular diagnosis methods are urgently needed to identify rare clinically significant variants of thalassemia genes. For traditional thalassemia genetic diagnosis, the reverse dot blot hybridization, Sanger sequencing, GAP-PCR and Multiple ligation-dependent probe amplification (MLPA) can be utilized to detect the prevalent variants, including SNVs, indels and copy number variations in *HBB* and *HBA1/2*¹². More recently, next-generation sequencing (NGS) based on PCR for targeted exons and selected intronic regions has shown advantages in thalassemia gene screening¹³. However, for rare variants not located in regular regions or variants in homologous regions of *HBA1* and *HBA2*, conventional methods and short-read based NGS methods may lead to missed diagnosis or even misdiagnosis. Currently, a methodology based on third-generation sequencing (TGS) named Comprehensive Analysis of Thalassemia Alleles (CATSA) was developed and validated for comprehensive thalassemia screening^{14,15}. Based on

¹Center of Prenatal Diagnosis, Department of Medical Genetics, West China Second University Hospital, Sichuan University, Chengdu, China. ²Key Laboratory of Birth Defects and Related Diseases of Women and Children, Sichuan University, Ministry of Education, Chengdu, China. ³Berry Genomics Corporation, Beijing 102200, China. ⁴Department of Obstetrics and Gynecology, West China Second University Hospital, Sichuan University, 17 South Renmin Road, Chengdu, China. ✉email: chenxinlian11@qq.com; sunny630@126.com

long-range PCR and long-molecule sequencing on the PacBio Sequel II platform, the CATSA method is adequate to cover the entire gene region and enable the diagnosis of common and rare types of thalassemia variants. To further investigate the potential of long-molecule sequencing in rare thalassemia carrier testing, we enrolled 100 cases that either showed an abnormal hematology phenotype or hemoglobin electrophoresis, but had negative conventional genetic diagnosis results. In this study, the CATSA method detected an extra 10 cases of clinically significant variants from *HBA1/2* and *HBB*, including 3 cases of SVs and 7 cases of SNVs, of which a very rare subtype of $-\alpha^{3.7}$ named $-\alpha^{3.7}$ subtype III ($-\alpha^{3.7\text{III}}$) was first identified in the Chinese population and validated by Sanger sequencing. Other rare genotypes with 11.1 kb deletions ($-11.1/\alpha\alpha$) or triplicate α -globin genes ($\text{aaa}^{3.7}/\alpha\alpha$) and rare SNVs can be thoroughly detected in one test. The study further validated long-molecule sequencing based CATSA as an efficient and valuable method in the diagnosis of rare thalassemia.

Patients and methods

Patients. A total of 100 patients highly suspected of having thalassemia gene mutations were included in this study. Informed consents were obtained from all participants or from a parent and/or legal guardian if participants are under 18. The inclusion criteria for the enrolled patients were: either (1) routine hematology examinations showed abnormal hemoglobin of $\text{HbA}_2 < 2.3\%$ or $\text{HbA}_2 \geq 3.2\%$ or elevated HbF; or (2) the blood test showed abnormalities with mean corpuscular volume (MCV) ≤ 80 fL for adults and/or mean corpuscular hemoglobin (MCH) ≤ 27 pg; or (3) the genotype identified by conventional genetic testing could not adequately explain their clinical phenotype. The age range for the included 18 children was 25 days to 2.5 years and the mean age for the other 82 adults was 30.5.

GAP-PCR testing for large-deletion α -thalassemias. Single-tube multiplex Gap-PCR was performed for the three common α -thalassemia deletions including $-\text{SEA}$ (Southeast Asia), $-\alpha^{3.7}$ (rightward) and $-\alpha^{4.2}$ (leftward) according to the manufacturer's protocol. (Yaneng Bioscience, Shenzhen, China).

Reverse dot blot hybridization assays. A PCR reverse dot-blot (PCR-RDB) assay was performed for three common non-deletional α -thalassemia mutations including Hb Constant Spring (Hb CS, HBA2: c.427 T > C), Hb Quong Sze (Hb QS, HBA2: c.377 T > C), and Hb Westmead (Hb WS, HBA2: c.369C > G), and 17 β -thalassemia mutations including -32 (C > A) (HBB: c. $-82\text{C} > \text{A}$), -30 (T > C) (HBB: c. $-80\text{T} > \text{C}$), -29 (A > G) (HBB: c. $-79\text{A} > \text{G}$), -28 (A > G) (HBB: c. $-76\text{A} > \text{G}$), CAP+40–43 ($-\text{AAAC}$) (HBB: c. -11_8delAAAC), initiation codon (T > G) (HBB: c.2 T > G), codons 14–15 (+G) (HBB: c.45_46insG), codon 17 (A > T) (HBB: c.52A > T), codon 26 (or Hb E) (G > A) (HBB: c.79G > A), codons 27/28 (+C) (HBB: c.84_85insC), codon 31 ($-C$) (HBB: c.94delC), codons 41–42 ($-\text{TTCT}$) (HBB: c.126_129delCTTT), codon 43 (G > T) (HBB: c.130G > T), codons 71–72 (+A) (HBB: c. 216_217insA), IVS-I-1 (G > T) (HBB: c.92 + 1G > T), IVS-I-5 (G > C) (HBB: c.92 + 5G > C), and IVS-II-654 (C > T) (HBB: c.316-197C > T). (Yaneng Bioscience, Shenzhen, China).

Long PCR based third-generation sequencing and data analysis. Experiments were performed as previously described¹⁴. Briefly, genomic DNA was amplified by PCR with primers targeting the majority of known structural variations, SNVs and indels in the *HBA1*, *HBA2* and *HBB* genes. Barcoded adaptors were ligated to the PCR products to construct individual sequencing libraries. Then, each library was quantified and pooled together by equal mass. After purification and quantification, the pooled library was converted to a SMRTbell library with Sequel Binding and Internal Ctrl Kit 3.0 (Pacific Biosciences) and sequenced on the Sequel II platform (Pacific Biosciences) under CCS mode. Then raw subreads were analyzed by CCS software (Pacific Biosciences) to generate CCS reads, debarcoded by lima in the Pbbioconda package (Pacific Biosciences) and aligned to genome build hg38 by pbmm2 (Pacific Biosciences). Finally, structural variations were identified according to the HbVar, Itharet and LOVD databases. SNVs and indels were identified by FreeBayes1.3.4 (<https://www.geneious.com/plugins/freebayes>; Biomatters, Inc., San Diego, CA).

Variant confirmation. All the SNVs detected by third-generation sequencing were further confirmed by Sanger sequencing using specific primers. The deletion variants and α -globin gene triplication variants were confirmed by specific PCR assays and agarose electrophoresis according to the manufacturer's protocol (Yaneng Bioscience, Shenzhen, China).

Ethical approval. The present study was approved by the Ethics Committee of West China Second Hospital of Sichuan University. All methods were carried out in accordance with guidelines and regulations from the Declaration of Helsinki.

Results

Detection and identification of rare clinically significant SNVs. In this study, we identified 7 cases of rare SNVs that can be ranked as clinically significant variants including *HBA2* SNVs (c.168dup, c. $-59\text{C} > \text{T}$, c.51G > T, c.91_93delGAG, and c.300 + 34G > A) and *HBB* c.316–45G > C (Table 1). All the long molecular sequencing data were verified by specific PCR and Sanger sequencing (Fig. 1). Among them, heterozygous *HBA2*:c.168dup, also known as codons 55/56 (+T) could cause α^+ -thalassemia due to a frameshift mutation of the $\alpha 2$ -globin gene, which could better explain the phenotype of participant #2 with mild microcytosis and hypochromia (Table 2)¹⁶. Variant *HBA2*:c. $-59\text{C} > \text{T}$ (participant #3) is one type of point mutation found in the promoter of the $\alpha 2$ -globin gene that affects *HBA2* transcription and protein synthesis¹⁷. Additionally, heterozygosity of *HBA2*:c.51G > T (CD16 (AAG > AAC or AAT)) and *HBA2*:c.91_93delGAG (codon 30(–GAG))

Participant number	Gap-PCR and reverse dot blot hybridization assays results		CATSA results		Variant verification
	α thalassemia	β thalassemia	α thalassemia	β thalassemia	
1	-SEA/aa	N	-SEA/aa HBA2: c.*82G>A HBA2: c. +92A>G HBA2: c.*98 T>C	N	Electrophoresis
2	aa/aa	N	HBA2: c.168dup	N	Sanger sequencing
3	aa/aa	N	HBA2: c. -59C>T	N	Sanger sequencing
4	aa/aa	N	HBA2: c.51G>T	N	Sanger sequencing
5	aa/aa	N	HBA2: c.91_93delGAG	N	Sanger sequencing
6	aa/aa	N	HBA2: c.300+34G>A	N	Sanger sequencing
7	aa/aa	N	N	HBB: c.316-45G>C	Sanger sequencing
8	aa/aa	N	- α 3.7/aa	N	Sanger sequencing
9	aa/aa	N	-11.1/aa	N	-
10	aa/aa	N	aaa ^{3.7} /aa	N	Electrophoresis

Table 1. Single nucleotide variations (SNVs) and Structural variation (SV) regions in HBA2 and HBB genes detected by CASTA.

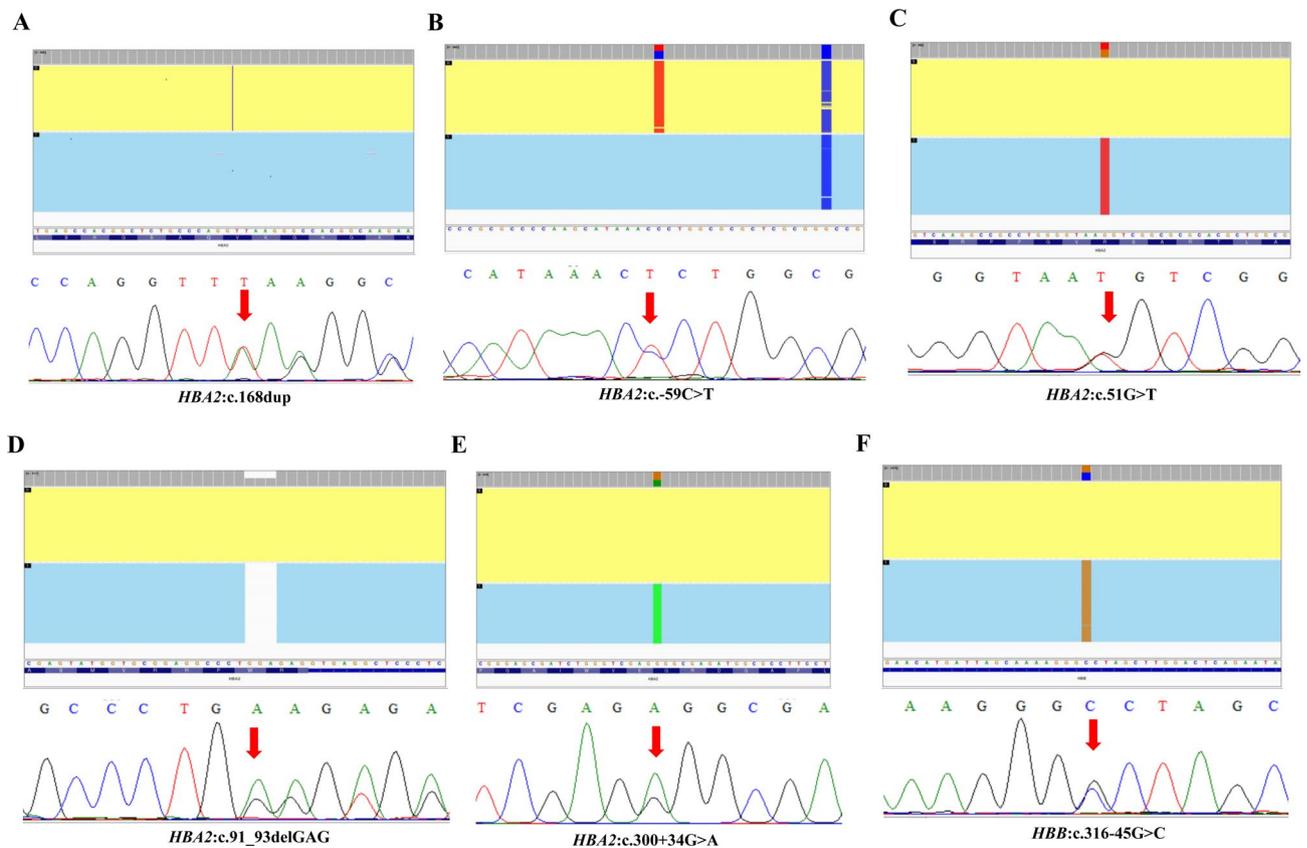


Figure 1. Identification and verification of clinically significant SNVs. (A–F) The six SNVs identified in the HBA2 and HBB genes by third-generation sequencing. The top graphs are the results exported from Integrative Genomics Viewer (IGV), and the bottom graphs are the Sanger sequencing results for each variant.

may influence protein translation, thus resulting in the abnormal hemoglobin and hematology phenotypes for participants #4 and #5 (Table 2)^{18,19}.

Multiple mutations in the polyadenylation signal site and compound -^{SEA}/aa. Participant #1 showed a Hb H disease-like abnormal hematology phenotype of significantly decreased MCV and MCH, together with abnormal hemoglobin (0.8% of HbA2). Through screening of the common thalassemia deletions or mutations, only heterozygous Southeast Asian deletion (-^{SEA}/aa) was identified, which might not be

Participant number	Genotype	Age (y)	Hb (g/L)	MCV (fL)*	MCH (pg)	HbA2 (%)	HbF (%)
1	–SEA/αα HBA2: c.*82G>A HBA2: c.*92A>G HBA2: c.*98 T>C	2	95	55.8	15.8	0.8	0.4
2	HBA2: c.168dup	28	99	79.9	25.8		
3	HBA2: c.–59C>T	28	99	74.9	23.6	2.5	0.0
4	HBA2: c.51G>T	32	135	93.4	31.9	1.8	0.7
5	HBA2: c.91_93delGAG	2	118	70.1	23.1		
6	HBA2: c.300+34G>A	30	135	87.4	30.3	1.6	16.3
7	HBB: c.316–45G>C	29	97	78.4	22.6		
8	–α3.7/αα	31	114	80.3	25.0	3.9	0.0
9	–11.1/αα	37	114	82.6	25.9	2.3	0.0
10	αα ^{3.7} /αα	30		91.5	30.0	3.3	

Table 2. Hematology examination and hemoglobin electrophoresis results in patients with detected clinically significant variants. *The reference value for MCV is age-related especially for the age of under 2. Thus, for case 1 and case 5 who are 2 years-old, the reference MCV are 70.3–87.9 according to the standard of our hospital.

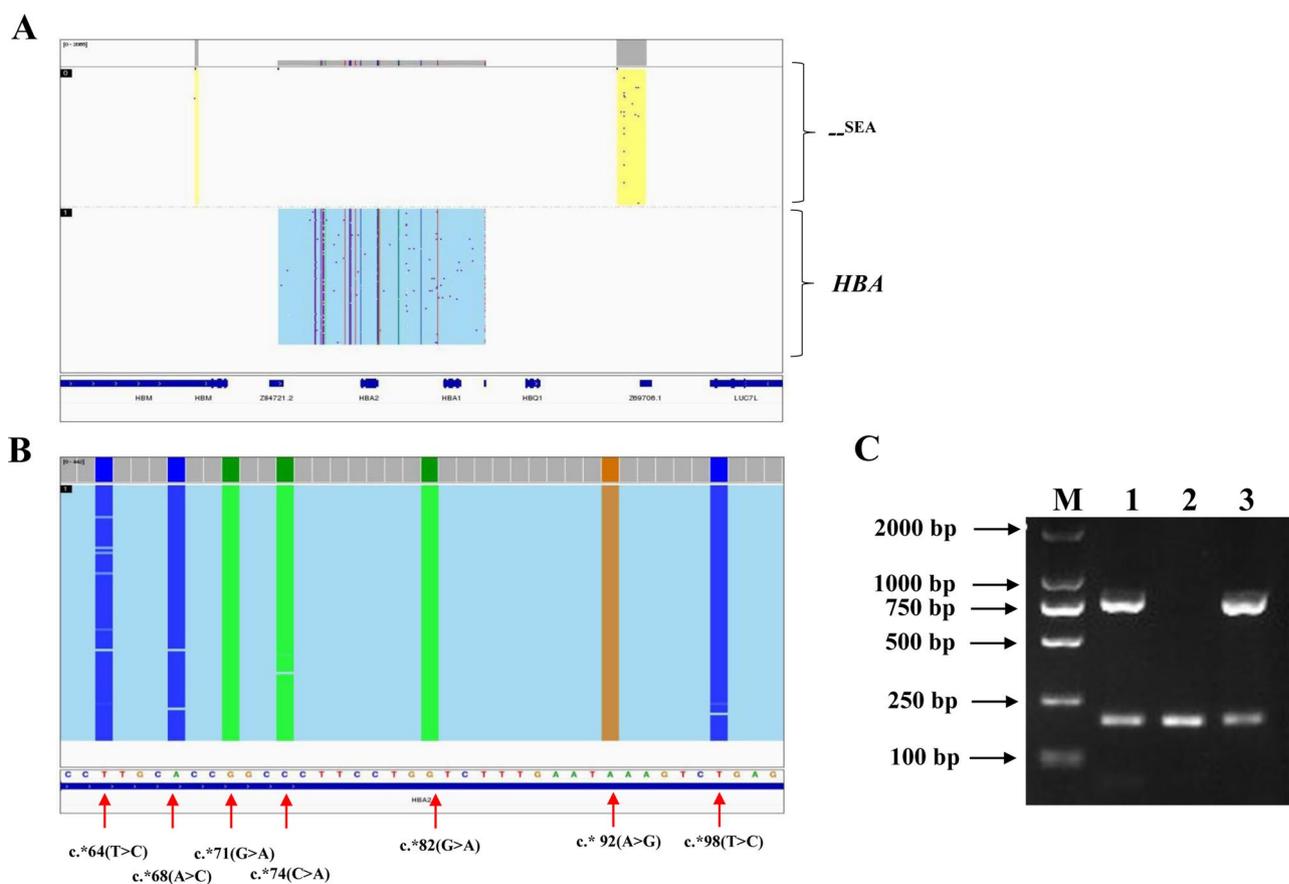


Figure 2. Identification of multiple mutations in the polyadenylation signal site and compound –^{SEA}/αα. (A) –^{SEA}/αα identified by third-generation sequencing. (B) Seven mutations in the polyadenylation signal site identified by third-generation sequencing. (C) Identification by specific PCR and agarose gel electrophoresis. M: maker; 1: positive control for the HBA2 fusion gene; 2: negative control; 3: sample for participant #1. The original gels are presented in Supplementary Fig. 1.

able to explain the abnormal Hb electrophoresis. In this study, the third-generation sequencing further identified multiple mutations at the polyadenylation signal site in the α2-globin gene, including *HBA2*:c.*64(T>C), c.*68(A>C), c.*71(G>A), c.*74(C>A), c.*82(G>A), c.*92(A>G) and c.*98(T>C) (Table 1, Fig. 2). The results were further verified by specific PCR and agarose gel electrophoresis (Yaneng Bioscience) (Fig. 2C). These multiple mutations in the polyadenylation signal site in combination with the –^{SEA}/αα genotype may lead to the phenotype of Hb H disease^{20,21}.

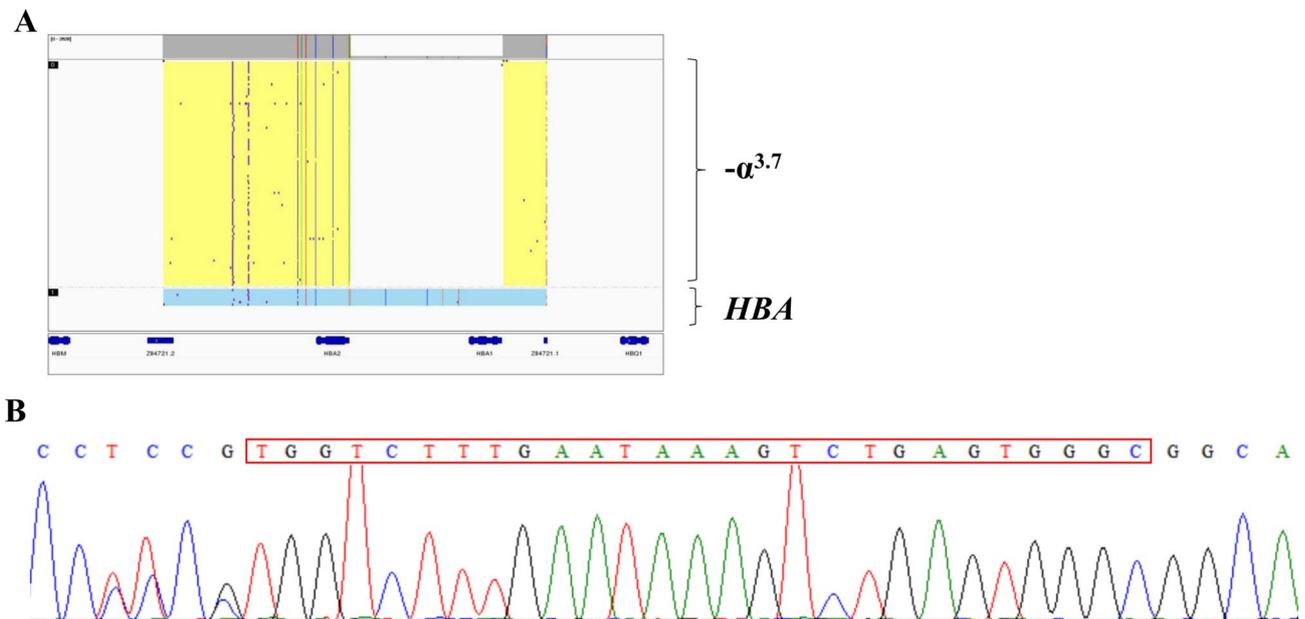


Figure 3. Identification and verification of $-\alpha^{3.7}$ subtype III. **(A)** The 3.8 kb deletion identified by third-generation sequencing. **(B)** Verification by Sanger sequencing. The sequences in the red box are these homologous sequences shared by *HBA1* (chr16:173,680–173,706) and *HBA2* (chr16:177,492–177,518).

Identification of $-\alpha^{3.7}$ subtype III. Through third-generation sequencing, we identified a rare deletion of 3.8 kb involving *HBA2* and *HBA1* (chr16:173,707–177,518 based on sequencing analysis) in participant #8. The exact deletion breakpoints were also inferred by Sanger sequencing (Table 1, Fig. 3). Since the short segment of ‘TGGTCTTGAATAAAGTCTGAGTGGGC’ could be located at chr16:173,680–173,706 and chr16:177,492–177,518 based on sequence alignment, the breakpoint should be located at chr16:173,707 and chr16:177,518. The corresponding HGVS nomenclature for this type of deletion was NG_000006.1: g.34570_38382del 3812 bp, which is a very rare type of $-\alpha^{3.7}$, that is $-\alpha^{3.7}$ subtype III²². To our knowledge, this is the first reported case of the $-\alpha^{3.7}$ subtype III in the Chinese population. The results also verified third-generation sequencing as a valuable method for rare thalassemia diagnosis.

Identification of a rare deletion of 11.1 kb in *HBA1/2*. Another rare deletion of 11.1 kb involving the *HBA2* and *HBA1* genes was identified in participant #9, which could not be easily identified by conventional methods (Table 1, Fig. 4). This deletion, which has the HGVS nomenclature of NC_000016.9:g.(220831_220860)_(231920_232003)del, resulted in a type of α^0 -thalassemia that has been previously reported in the Chinese population¹⁹. The carriers with this deletion usually show mild α -thalassemia with hypochromic microcytic phenotypes (Table 2).

Detection of an α -globin gene triplication. One case of α -Globin gene triplication ($\alpha\alpha\alpha^{3.7}/\alpha\alpha$) was found in this study (participant #10). The results were further validated by specific PCR and agarose electrophoresis (Yaneng Bioscience) (Table 1, Fig. 5). Carriers with this α -globin gene triplication usually show normal results for routine hematology examination but abnormal HbA2 content²³. In addition, if compounded with β -thalassemia, patients usually show intermediate β -thalassemia phenotypes.

Discussion

Conventional genetic testing methods including gap-PCR, RDB, MLPA and Sanger sequencing can only detect common variants such as $-\alpha^{3.7}$, $-\alpha^{4.2}$, $-\alpha^{SEA}$, $-\alpha^{27.6}$, $-\alpha^{21.9}$ or other known SVs and SNVs, which are very limited. Since more than 2000 thalassemia or abnormal hemoglobin-related variant sites have been reported, it is necessary to develop new methods to implement in thalassemia diagnosis, especially for these rare variants. Recently, third-generation sequencing technology also known as long-molecule sequencing, has emerged as an incomparable method in genetic diagnosis with many advantages, including long reads, high accuracy, single molecule resolution and no GC preference²⁴. The long reads could cover many rare gene loci, and their PCR-free characteristic made it possible to reflect the real arrangement in the genome. The method designed for thalassemia carrier screening based on third-generation sequencing on the Pacific Biosciences Sequel II platform has been optimized in recent years to detect thalassemia genes with high efficiency and accurate, also with acceptable time and cost.

Through third-generation sequencing for full-length of *HBA1*, *HBA2* and *HBB* genes, multiple variants may be recognized in one sample. Thus, it is crucial to evaluate the correlation between genotype and phenotype to further identify clinically significant variants²⁵. The SNVs reported in this study, including *HBA2* SNVs (c.168dup, c.−59C>T, c.51G>T, c.91_93delGAG, and c.300+34G>A) and *HBB* c.316–45G>C, all showed some evidence related to abnormal hematology phenotypes and hemoglobin results. Carriers usually have a normal phenotype,

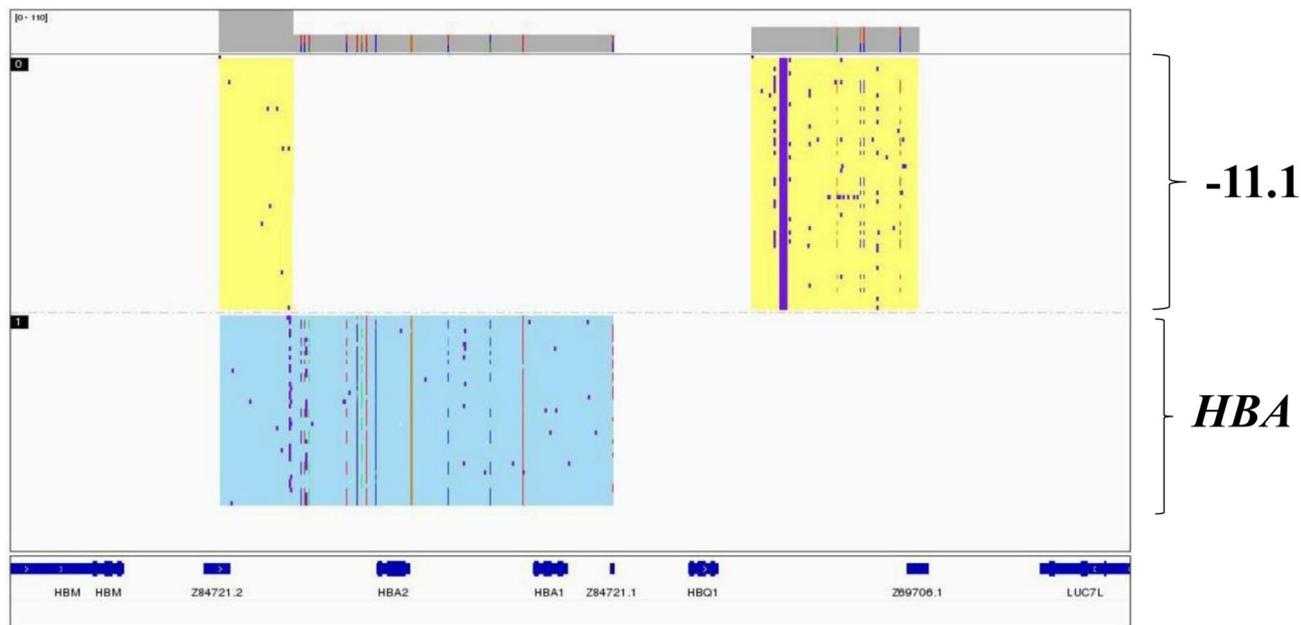


Figure 4. Identification of a rare deletion of 11.1 kb in *HBA1/2*. The 11.1 kb deletion identified by third-generation sequencing.

but tend to develop intermediate thalassemia when compounded with other variant types. Most of these participants are at the appropriate age and are willing to procreate. Thus, our results could pave the way for further prenatal diagnosis or even preimplantation genetic testing (PGT) to prevent the incidence of thalassemia²⁶.

It is worth mentioning that participant #6 showed an increased HbF content of 16.3%, yet only one clinically significant variant of *HBA2* c.300 + 34G > A was identified. We suspected that some other HPFH-causing variants in the gamma globin genes *HBG1* or *HBG2* may be responsible for the increased HbF level²⁷. Thus, we conducted full-length Sanger sequencing for *HBG1* and *HBG2* genes. Two variants of – 196 (C → T) and + 25 (G → A) in the promoter region of *HBG1* were found (Fig. 6). Among these genes, – 196 (C → T), also named *HBG1*:c. – 249C > T, may be responsible for the high expression of the gamma-globin gene, thus causing the increased HbF level according to previous studies^{28,29}. The results suggested that it may be necessary to design more primers to detect other thalassemia-related genes, such as *HBG1* and *HBG2*, in the future, to implement comprehensive screening.

Moreover, we found one case with multiple mutations in the polyadenylation signal site compound with –^{SEA}/αα, which showed a rare Hb H disease. These multiple mutations led to a fusion between the α2 and ψα1 genes, which may affect mRNA transcription and termination. Thus, α2-globin gene polyA mutation in combination with –^{SEA}/αα may lead to Hb H disease. This patient is a 2-year-old boy with significantly decreased MCV and MCH, together with abnormal hemoglobin. Through conventional thalassemia genetic screening, only a heterozygous –^{SEA}/αα was found which is not consistent with his phenotype. Thus, third-generation sequencing provided a more efficient method for rare thalassemia gene testing for this case. Based on its hematological phenotype and clinical manifestations for Hb H disease, it will be necessary for the patient to seek further family analysis and genetic counselling.

Most importantly, we reported a case with a 3.8 kb deletion in *HBA2* that shared almost the same deletion junctions with the rare –α^{3.7} subtype III, which was the first reported case of the –α^{3.7} subtype III in the Chinese population. There are at least three subtypes of –α^{3.7} in the population based on the different recombination sites, that is –α^{3.7} I, –α^{3.7} II and –α^{3.7} III. The prevalence and population distribution of each subtype are different, among which –α^{3.7} subtype III is extremely rare in the population compared with the other two subtypes. Although relatively rare, the –α^{3.7} subtype III has been noted with high frequency in people in Pacific Island nations such as Micronesia and Polynesia and was also found with approximately 2% of α⁺-thalassemia alleles in the southern Thai population³⁰. In our study, –α^{3.7} III was discovered for the first time in the Chinese population using third-generation sequencing. The –α^{3.7} subtype III is prone to be undiagnosed since most of the conventional detection methods are designed for the identification of the other two subtypes. The PCR primers used in those PCR kits are usually located around the breakpoints of –α^{3.7} subtype III and homologous sequences in *HBA1* could also lead to miss diagnoses. The long reads of third-generation sequencing could better cover the whole sequences of *HBA1* and *HBA2*, thus most of the rare deletions or duplications in those two genes can be thoroughly identified.

Conclusions

In conclusion, in this study we identified 10 rare clinically significant variations through third-generation sequencing compared to traditional thalassemia gene testing from 100 cases with clinical evidence of suspected thalassemia. The variants including SNVs, rare deletions and triplications in the *HBA1* and *HBA2* genes. The

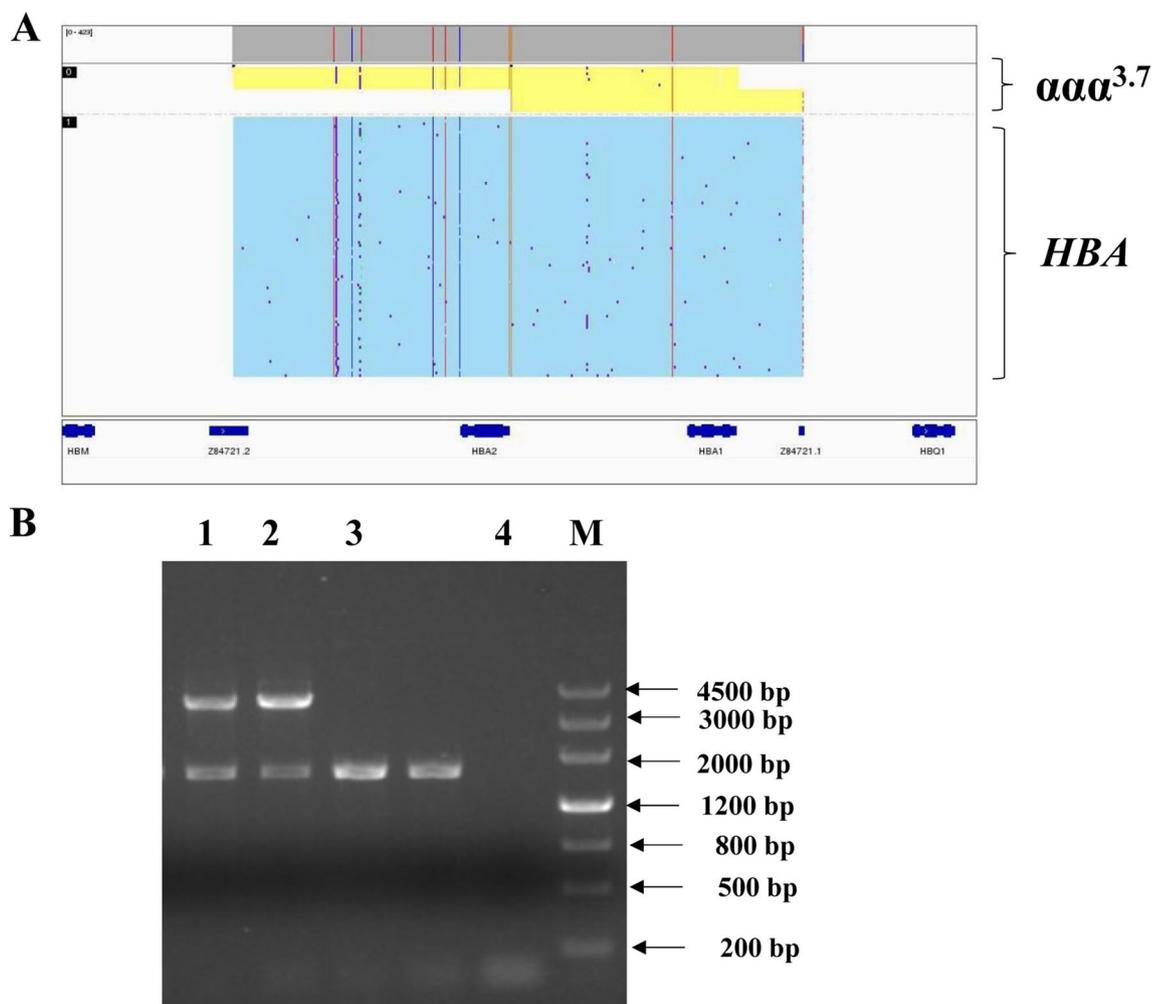


Figure 5. Identification and verification of α -globin gene triplication. (A) α -Globin gene triplication of $\alpha\alpha\alpha^{3.7}/\alpha\alpha$ identified by third-generation sequencing. (B) Identification by specific PCR and agarose gel electrophoresis. M: Maker; 1: sample for participant #10; 2: positive control for $\alpha\alpha\alpha^{3.7}/\alpha\alpha$; 3: negative control; 4: blank. The original gels are presented in Supplementary Fig. 2.

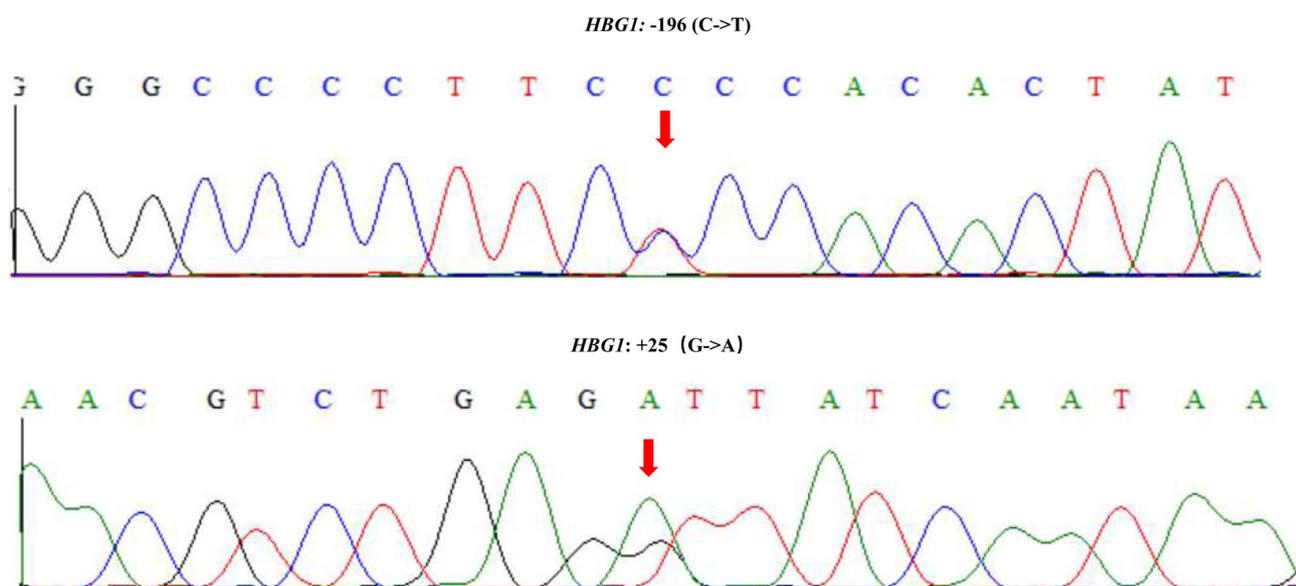


Figure 6. Identification of SNVs in *HBGI* by Sanger sequencing. The two SNVs of *HBGI*: -196 (C \rightarrow T) (top) and *HBGI*: +25 (G \rightarrow A) (bottom) identified by Sanger sequencing for the full-length of *HBGI* gene.

results further validated the third-generation sequencing utilized in this study as a promising genetic testing method for thalassemia carrier diagnosis, especially for rare variants identification.

Data availability

The datasets generated during the current study are available in the SRA database of National Center for Biotechnology Information (Accession: PRJNA836009, ID: 836009) (<https://www.ncbi.nlm.nih.gov/sra/PRJNA836009>).

Received: 16 February 2022; Accepted: 31 May 2022

Published online: 14 June 2022

References

1. Taher, A. T., Weatherall, D. J. & Cappellini, M. Thalassaemia. *The Lancet* **391**, 155–167 (2017).
2. Weatherall, J. D. The inherited diseases of hemoglobin are an emerging global health burden. *Blood* **115**, 4331–4336 (2010).
3. Longo, D., Piel, F. & Weatherall, D. The α -thalassemias. *N. Engl. J. Med.* **371**, 1908–1916 (2014).
4. Thein, L. S. Molecular basis of β thalassemia and potential therapeutic targets. *Blood Cells Mol. Dis.* **70**, 54–65 (2017).
5. Mettananda, S. & Higgs, D. R. Molecular basis and genetic modifiers of thalassemia. *Hematol./Oncol. Clin.* **32**, 177–191 (2018).
6. Viprakasit, V. & Ekwattanakit, S. Clinical classification, screening and diagnosis for thalassemia. *Hematol./Oncol. Clin.* **32**, 193–211 (2018).
7. Lal, A. *et al.* Heterogeneity of hemoglobin h disease in childhood. *N. Engl. J. Med.* **364**, 710–718 (2011).
8. Galanello, R. & Origa, R. Beta-thalassemia. *Orphanet. J. Rare Dis.* **5**, 1–15 (2010).
9. Diagnosis WGFPGF, Association TOGDMGBOCM, Shang, X. *et al.* Clinical practice guidelines for beta-thalassemia. *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chin. J. Med. Genet.* **37**, 243–251 (2020).
10. Shang, X. & Xu, X. Update in the genetics of thalassemia: What clinicians need to know. *Best Pract. Res. Clin. Obstet. Gynaecol.* **39**, 3–15 (2017).
11. Munkongdee, T. *et al.* Update in laboratory diagnosis of thalassemia. *Front. Mol. Biosci.* **7**, 74 (2020).
12. Brancaloni, V. *et al.* Laboratory diagnosis of thalassemia. *Int. J. Lab. Hematol.* **38**, 32–40 (2016).
13. Shang, X. *et al.* Rapid targeted next-generation sequencing platform for molecular screening and clinical genotyping in subjects with hemoglobinopathies. *EBioMedicine* **23**, 150–159 (2017).
14. Liang, Q. *et al.* A more universal approach to comprehensive analysis of thalassemia alleles (catsa). *J. Mol. Diagnost. JMD* **23**, 1195–1204 (2021).
15. Xu, L. *et al.* Long molecule sequencing: A new approach for identification of clinically significant DNA variants in alpha and beta thalassemia carriers. *J. Mol. Diagnost.* **22**, 1087–1095 (2020).
16. Wayne, J. S. *et al.* A+ thalassemia due to a frameshift mutation of the α 2-globin gene [codons 55/56 (+t) or hba2: C.168dup]. *Hemoglobin* **39**, 1–2 (2015).
17. Qadah, T. *et al.* Molecular and cellular analysis of three novel alpha2-globin gene promoter mutations [hba2. *Pathology* **46**, 46–52 (2014).
18. Liang, C. C. *et al.* Hemoglobin beijing [alpha 16 (a14) lys replaced by asn]: A new fast-moving hemoglobin variant. *Hemoglobin* **6**, 629 (1982).
19. Diagnosis WGFPGF, Association TOGDMGBOCM, Shang, X. *et al.* Clinical practice guidelines for alpha thalassemia. *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chin. J. Med. Genet.* **37**, 235–242 (2020).
20. Zhang, Q. *et al.* Hb h disease caused by multiple mutations in the polyadenylation signal site and –sea/aa. *Hemoglobin* **41**, 189–192 (2017).
21. Ju, A.-P. *et al.* Detection of an α -globin fusion gene using real-time polymerase chain reaction-based multicolor melting curve. *Hemoglobin* **44**, 427–431 (2020).
22. Charoenwittkul, T. *et al.* Molecular characteristics of α + -thalassemia (3.7kb deletion) in Southeast Asia: Molecular subtypes, haplotypic heterogeneity, multiple founder effects and laboratory diagnostics. *Clin. Biochem.* **71**, 31–37 (2019).
23. Long, J. & Liu, E. The carriage rates of $\alpha\alpha\alpha$ anti3.7, $\alpha\alpha\alpha$ anti4.2, and HK $\alpha\alpha$ in the population of Guangxi, China measured using a rapid detection qPCR system to determine CNV in the α -globin gene cluster. *Gene* **768**, 145296 (2020).
24. Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162 (2019).
25. Luo, S. *et al.* The value of single-molecule real-time technology in the diagnosis of rare thalassemia variants and analysis of phenotype-genotype correlation. *J. Hum. Genet.* **67**, 1–13 (2021).
26. De Rycke, M. & Berckmoes, V. Preimplantation genetic testing for monogenic disorders. *Genes* **11**, 871 (2020).
27. Jiang, F. *et al.* Molecular epidemiology and hematologic characterization of $\delta\beta$ -thalassemia and hereditary persistence of fetal hemoglobin in 125,661 families of greater Guangzhou area, the metropolis of southern china. *BMC Med. Genet.* **21**, 1–12 (2020).
28. Giglioli, B. *et al.* A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and beta-thalassemia. *EMBO J.* **3**, 2641–2645 (1984).
29. Martyn, G. E. *et al.* Natural regulatory mutations elevate the fetal globin gene via disruption of bcl11a or zbtb7a binding. *Nat. Genet.* **50**, 498–503 (2018).
30. O’Shaughnessy, D. *et al.* Globin genes in Micronesia: Origins and affinities of pacific island peoples. *Am. J. Hum. Genet.* **46**, 144 (1990).

Acknowledgements

We thank the participants for their cooperation and participation. We thank the team from Berry Genomics Corporation (Beijing, China) for their help with the third-generation sequencing and data analysis.

Author contributions

C.P. and H.Z. wrote the main manuscript and prepared the figures, H.C. and J.R. verified the detected variants, Z.D. and T.Z. did the traditional testing of thalassemia, A.M., R.X. and Y.L. conducted the third generation sequencing and data analysis, H.W., X.C. and S.L. reviewed the manuscript.

Funding

This study was supported by the National Key Research and Development Program of China (2021YFC1005303) and the Technology Research and Development Program of the Science and Technology Department of Sichuan Province, China (2021YFS0078).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-14038-8>.

Correspondence and requests for materials should be addressed to X.C. or S.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022