



OPEN ACCESS

EDITED BY

Cunyou Zhao,
Southern Medical University, China

REVIEWED BY

Diyu Chen,
Guangzhou Institutes of Biomedicine
and Health (CAS), China
Xiuqin Bao,
Guangdong Province Women and
Children Hospital, China

*CORRESPONDENCE

Xiaoyuan Wang,
wangxiaoyuan1025@163.com
Tizhen Yan,
luobu_0971@163.com

[†]These authors have contributed equally
to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Genetics of
Common and Rare Diseases,
a section of the journal
Frontiers in Genetics

RECEIVED 21 June 2022

ACCEPTED 01 August 2022

PUBLISHED 02 September 2022

CITATION

Luo S, Chen X, Zeng D, Tang N, Yuan D,
Liu B, Chen L, Zhong Q, Li J, Liu Y,
Chen J, Wang X and Yan T (2022),
Detection of four rare thalassemia
variants using Single-molecule
realtime sequencing.
Front. Genet. 13:974999.
doi: 10.3389/fgene.2022.974999

COPYRIGHT

© 2022 Luo, Chen, Zeng, Tang, Yuan,
Liu, Chen, Zhong, Li, Liu, Chen, Wang
and Yan. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Detection of four rare thalassemia variants using Single-molecule realtime sequencing

Shiqiang Luo^{1,2†}, Xingyuan Chen^{3,4†}, Dingyuan Zeng⁵,
Ning Tang⁵, Dejian Yuan^{1,2}, Bailing Liu², Lizhu Chen^{1,2},
Qingyan Zhong^{1,2}, Jiaqi Li⁶, Yinyin Liu⁶, Jianping Chen⁷,
Xiaoyuan Wang^{7*} and Tizhen Yan^{1,2*}

¹Liuzhou Key Laboratory of Birth Defects Prevention and Control, Department of Medical Genetics, Liuzhou Maternity and Child Healthcare Hospital, Liuzhou, China, ²Liuzhou Key Laboratory of Reproductive Medicine, Liuzhou, China, ³Department of Laboratory Medicine, The People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, China, ⁴Guangxi Academy of Medical Sciences, Nanning, China, ⁵Guangxi Health Commission Key Laboratory of Birth Cohort Study in Pregnant Women of Advanced Age, Liuzhou, China, ⁶Berry Genomics Corporation, Beijing, China, ⁷Liuzhou Maternity and Child Healthcare Hospital, Liuzhou, China

Conventional methods for the diagnosis of thalassemia include gap polymerase chain reaction (Gap-PCR), reverse membrane hybridization (RDB), multiplex ligation-dependent probe amplification (MLPA) and Sanger sequencing. In this study, we used single molecule real-time technology (SMRT) sequencing and discovered four rare variants that have not been identified by conventional diagnostic methods for thalassemia. We also performed genotype and phenotype analyses on family members of thalassemia patients. The SMRT technology detected five cases in which the proband had abnormal results by conventional diagnostic methods or inconsistencies between the genotype and phenotype. The variants included two cases of an α -globin gene cluster 27,311 bp deletion, $--^{27.3}/\alpha\alpha$ (hg38 chr16:158664-185974), one case of an HS-40 region 16,079 bp deletion (hg38 chr16:100600-116678), one case of a rearrangement of $-\alpha^{3.7}\alpha1\alpha2$ on one allele and one case of a β -globin gene cluster *HBG1-HBG2* 4,924 bp deletion (hg38 chr11:5249345-5254268). This study clarified the hematological phenotypes of four rare variants and indicated the application value of SMRT in the diagnosis of rare α -globin and β -globin gene cluster deletions, gene recombination and deletion breakpoints. The SMRT method is a comprehensive one-step technology for the genetic diagnosis of thalassemia and is particularly suitable for the diagnosis of thalassemia with rare deletions or genetic recombination.

KEYWORDS

single molecule real-time technology, carrier screening, athalassemia, β -thalassemia, rarevariant

Introduction

Thalassemia is a hereditary blood disease that seriously threatens human health, causing death and disability. Many countries around the world are concerned about this disease that is identified as a birth defect (Bajwa and Basit, 2021; Li et al., 2014). The WHO has estimated that thalassemia gene carriers account for approximately 1.67% of the total world population. Thalassemia is one of the most common genetic diseases in southern China and the high incidence provinces of Guangdong, Guangxi, Yunnan and Hainan have a thalassemia gene carrier rate of between 11% and 25% (Xiong et al., 2010; Zhang et al., 2012; Yao et al., 2014). Thalassemia is a hemoglobinopathy that disrupts the balance of α - and β -globin chains due to mutations, deletions, duplications and gene rearrangements in the α - and β -globin gene clusters. The global carrier rate of thalassemia is approximately 1.67%. The phenotype of individuals with thalassemia varies widely, ranging from asymptomatic carriers, mild anemia, moderate anemia, severe anemia, which requires prolonged blood transfusions to maintain life, more severe cases and even pre or postnatal death (Modell and Darlison, 2008). The severity of the disease is often closely related to the number of globin genes that are nonfunctional. In addition, duplication of the α -globin gene cluster, which exacerbates the α/β imbalance, can aggravate the clinical manifestations of heterozygous β -thalassemia. At present, the important thalassemia control measures are carrier screening and prenatal diagnosis of high-risk couples (Goonasekera et al., 2018). The range of α -globin gene deletions is heterogeneous, ranging from deletions of several thousand base pairs, which only involve the vicinity of two α -genes, to those that involve more than two α -genes. In addition to cases caused by deletion in the α -globin gene, a small number of α -thalassemia cases are caused by point mutations. To date, more than 120 α -globin gene mutations have been identified in the Chinese population (data from the Human Variome Project China, <http://www.genomed.zju.edu.cn/LOVD3/genes>). Six mutations ($-\alpha^{SEA}$, $-\alpha^{3.7}$, $-\alpha^{4.2}$, $HBA2:c.369C > G$, $HBA2:c.427T > C$ and $HBA2:c.377T > C$) account for more than 98% of all α -gene mutations (Harewood and Azevedo, 2021). Therefore, the existing clinical diagnosis routine is mainly aimed at identifying these mutations. β -thalassemia is mainly caused by mutation of the β -globin gene. Most of these mutations are point mutations or small fragment deletions, with a few large fragment deletions (Mettananda and Higgs, 2018). To date, 129 β -globin point mutations and 16 β -thalassemia deletions have been found in the Chinese population (data from HVP China; <http://www.genomed.zju.edu.cn/LOVD3/genes>) and named in accordance with the Human Genome Variation Society rules (den Dunnen et al., 2016). Eight of these mutations ($HBB:c.124_127delTTCT$, $HBB:c.52A > T$, $HBB:c.316-197C > T$, $HBB:c.-78A > G$, $HBB:c.216_217insA$,

$HBB:c.79G > A$, $HBB:c.92+1G > T$ and $HBB:c.-79A > G$) account for more than 95% of the β -thalassemia mutations in the Chinese population (Needs et al., 2021).

In recent years, detection methods for α - and β -globin deletions have included multiplex amplifiable probe hybridization (MAPH) (Kousoulidou et al., 2008), MLPA (Shen and Wu, 2009), quantitative multiplex PCR of short fluorescent fragments (QMPSF) (Feuk et al., 2006), real-time quantitative PCR (RT-PCR) (Weaver et al., 2010) and PCR melting curve analysis (PMCA) (Xiong et al., 2011). Each of these methods has advantages and disadvantages for screening and diagnostic purposes. The molecular diagnostic technologies used for the clinical detection of α - and β -globin point mutations include Gap-PCR and RDB. These methods are mainly used to detect four common deletions of α -thalassemia ($-\alpha^{SEA}$, $-\alpha^{3.7}$, $-\alpha^{4.2}$, $-\alpha^{THA1}$), three non-deletion types of α -thalassemia ($\alpha^{WS}\alpha$, $\alpha^{CS}\alpha$ and $\alpha^{QS}\alpha$) and 17 common β -thalassemia mutations in the Chinese population. The MLPA technology is a highly efficient and specific method for screening α - and β -globin copy number variations and can simultaneously detect changes in the copy number of more than 40 target sequences in one reaction. It has been popularized and applied as a routine technology in many large hospitals in economically developed areas of China (Liu et al., 2008). Research teams have also used array comparative genomic hybridization (aCGH) to detect α -thalassemia copy number variation (Phylipsen et al., 2012). High-throughput sequencing technology (NGS) has been used for genetic screening of thalassemia and can effectively detect the genotype, in addition to finding new genetic mutations. This method has the advantages of simple collection of specimens and high accuracy of results (Achour et al., 2021). However, due to the expensive equipment and consumables required for NGS technology, its clinical application has been limited.

In recent years, third-generation sequencing technology has been applied to the detection of the thalassemia gene. This technology is also called SMRT. The SMRT method performs separate sequencing of each DNA molecule and PCR amplification is not required. The technology is able to produce ultra-long read lengths (between 30 and 100 kb) and high accuracy (QV30 > 99.8%), characterized by GC-free preference and single-molecule resolution (Wenger et al., 2019). Using SMRT, one-time full coverage of all mutation types of α -thalassemia and β -thalassemia genes has been achieved in the Chinese population, with coverage of other abnormal hemoglobin variant sites on $HBA1$, $HBA2$ and HBB genes (Wu et al., 2022). This method can help with the accurate diagnosis and treatment of the disease and avoids the risk of missed detection. In this study, SMRT technology was used to detect four rare variants (large fragment deletions and gene recombination). Pedigree and phenotype analyses were also performed to explore the value of SMRT technology in the study of large fragment deletion and gene recombination of the globin gene cluster.

Materials and methods

Research subjects

Family A

The proband, registered in Liuzhou, Guangxi, China, was female, 40 years old. The genetic testing results were normal for common type α -thalassemia and 17 types of common Chinese β -thalassemia. Her hemoglobin level was 99 g/L, which indicated microcytic hypochromic anemia, of unknown cause. The copy number variation of the α -globin gene cluster was assessed by MLPA technology and revealed an HBZP1 region-HBQ1-3 heterozygous deletion but the breakpoint was not clear. The family members of the proband were also analyzed. In June of the same year, one sporadic case was also found, registered in Liuzhou City, Guangxi. The patient was female, 30 years old and test results were normal for the common type of α -thalassemia and 17 types of common Chinese β -thalassemia. Her hemoglobin level was 98 g/L, with small cell hypochromic anemia of unknown cause. The MLPA results showed an HBZP1 region-HBQ1-3 heterozygous deletion. We collected peripheral venous blood samples from all subjects in the study, using EDTA-K2 anticoagulation tubes. Signed informed consent was obtained from all subjects.

Family B

The proband, registered in Liuzhou, Guangxi, China, was female, 25 years old. The genetic testing results were normal for common type α -thalassemia and 17 common Chinese β -thalassemia types. The hemoglobin level was 111 g/L, with microcytic hypochromic anemia of unknown cause. The HS-40 deletion was found by copy number analysis of the α -globin gene cluster but the deletion breakpoint was not clear. The family members of the proband were also analyzed.

Family C

The proband, registered in Laibin City, Guangxi, China, was male and 27 years old. He came to our hospital for a pregnancy test with his wife in 2020. The genetic testing results were normal for common type α -thalassemia and 17 types of common Chinese β -thalassemia. The Hb F level was increased but the reason for this was not known. The *HBG1-HBG2* deletion was found by copy number analysis of the β -globin gene cluster. The deletion breakpoint was not clear. His wife was a β -thalassemia minor gene carrier. As there was a potential risk of having a hemoglobin intermediate β -thalassemia fetus, the wife of the proband underwent prenatal diagnosis by amniocentesis at 22 weeks of gestation. The amniotic fluid of the fetus was tested by the conventional reverse dot blot technique for β -globin gene mutation. The fetus was tested by MLPA for copy number

variation of the β -globin gene cluster. The family members of the proband were also analyzed. The wife chose to continue with the pregnancy and the child was followed up six months after birth.

Family D

The proband's household registration was in Liuzhou City, Guangxi Province, China. She was female, 40 years old and came to our hospital for a pregnancy test in 2021. The routine genetic test found $-\alpha^{3.7}/\alpha\alpha$ in the proband and her husband was $--^{SEA}/\alpha\alpha$. As the parents were carriers of homotype α -thalassemia and at risk of giving birth to an intermediate α -thalassemia baby, we collected amniotic fluid and umbilical cord blood for prenatal diagnosis at 20 weeks and three days into pregnancy. Due to the abnormal results from the routine techniques, gene recombination was assessed using the α -globin gene cluster copy number variation test. The family members of the proband were also analyzed.

Hematology and hemoglobin electrophoresis analysis

An automatic blood cell analyzer was used for routine blood analyses and high-performance liquid chromatography was used for hemoglobin analysis to detect hemoglobin F (Hb F), hemoglobin A2 (Hb A2), hemoglobin H (HbH) and other hemoglobin variants. Umbilical cord blood samples were analyzed for hemoglobin composition using a Capillarys 2 capillary electrophoresis analyzer (Sebia, France). The instrument was operated in accordance with the instruction manual and internal quality control was performed for each experiment.

Genomic DNA extraction

The magnetic bead method was used to extract nucleic acids (LabAid820, Xiamen Zhishan Biotechnology, Xiamen, China). We then used a nucleic acid analyzer (ASP-2680; ACTGene, Piscataway, NJ, United States) to detect DNA concentration and purity. The A_{260}/A_{280} was between 1.6 and 1.9 and the concentration was 20-30 ng/ μ L.

α -thalassemia and β -thalassemia genotyping analysis

Genomic DNA was extracted from peripheral blood using the thalassemia detection kit (Yishengtang, Shenzhen, China). The routine analysis for the four common α -thalassemia deletions [$--^{SEA}$ (Southeast Asia), $-\alpha^{3.7}$ (rightward), $-\alpha^{4.2}$

(leftward) and --^{THAI} (Thailand)] was performed using Gap-PCR and three common non-deletion α -thalassemia mutations, Hb Constant spring (*HBA2*: c.427T > C), Hb Quong Sze (Hb QS, *HBA2*: c.377T > C) and Hb Westmead (*HBA2*: c.369G > C), were assessed. We also analyzed 17 known β -thalassemia mutations; -28 (A > G) (*HBB*: c.-76A > G), -29 (A > G) (*HBB*: c.-79A > G), -30 (T > C) (*HBB*: c.-80T > C), -32 (C > A) (*HBB*: c.-82C > A), 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 (-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), IVS-II-654 (C > T) (*HBB*: c.316-197C > T), CAP+1(A > C) (*HBB*: c.-50A > C) and initiation codon (T > G) (*HBB*: c.2T > G). These mutations were detected by PCR and a reverse dot-blot assay (Shenzhen Yishengtang Biological Products Co., Ltd., Shenzhen, China). The MLPA detection method was performed, using the P102 and P140 probe kit (MRC-Honand, Netherlands), to analyze the copy number variation in deletional thalassemia genes. Capillary electrophoresis was performed on amplified products using a genetic 3500Dx analyzer (Applied Biosystems, Foster City, CA, United States).

Third generation sequencing and data analysis

Experiments were conducted as described in a previous study. In brief, genomic DNA was amplified by PCR with primers that covered the majority of known structural variations, single nucleotide variations (SNVs), insertions and deletions (indels) in the *HBA1*, *HBA2*, *HBB*, *HBG* and *HS40* genes. The PCR products were ligated to barcoded adaptors by one-step end-repair and ligation reactions to construct pre-libraries. These were pooled together by equal mass and converted to a SMRT bell library by the Sequel Binding and Internal Ctrl Kit 3.0 (Pacific Biosciences). The SMRT bell library was then sequenced under the CCS mode on the Sequel II platform (Pacific Biosciences). After sequencing, subreads were converted to CCS reads, by CCS software (Pacific Biosciences), and debarcoded by lima in the Pbbioconda package (Pacific Biosciences). After alignment of the processed reads to genome build hg38, by pbmn2 (Pacific Biosciences), structural variations were identified in accordance with the HbVar, Ithant and LOVD databases. The SNVs and indels were identified by FreeBayes1.3.4 (<https://www.geneious.com/plugins/freebayes>; Biomatters, Inc., San Diego, CA). Haplotype analysis was performed with WhatsHap (version 0.18) and pbampliconclustering software.

The α -thalassemia and β -thalassemia gene sequencing

Four primer pairs were designed and used to amplify and sequence the *HBA1* and *HBA2* genes: Sequences were as follows:

$\alpha 1$ -F: 5'-TGG AGG GTG GAG ACG TCC TG-3'; $\alpha 1$ -R: 5'-TCC ATC CTC TCC CGC CCC TGC CTT TTC-3'; $\alpha 2$ -F: 5'-TGG AGG GTG GAG ACG TCT TG-3'; and $\alpha 2$ -R: 5'-CCG TTG GCA CAT TCC GG-3'. β -FP: 5'-AAC TCC TAA GCC AGT GCC AG-3'; AvaII- β -FP: 5'-TTG GGG ATC TGT CCA CTC CT-3'; AvaII- β -RP: 5'-CCA GCC TTA TCC CAA CCA TAA AAT AA-3'; β -RP: 5'-ATG CAC TGA CCT CCC ACA TTC CCT-3'. DNA sequencing was performed using the Sanger dideoxy termination method and the reference sequence was NM_000517. Amplification was performed using 50 ng of genomic DNA and 20 pmol of forward (F) and reverse (R) primers, on the C1000 thermal cycler (C1000 Thermal Cycler; Bio-Rad Laboratories, Hercules, CA, United States). The PCR products were sequenced on the ABI PRISM[®] 3130 automated sequencer (Applied Biosystems).

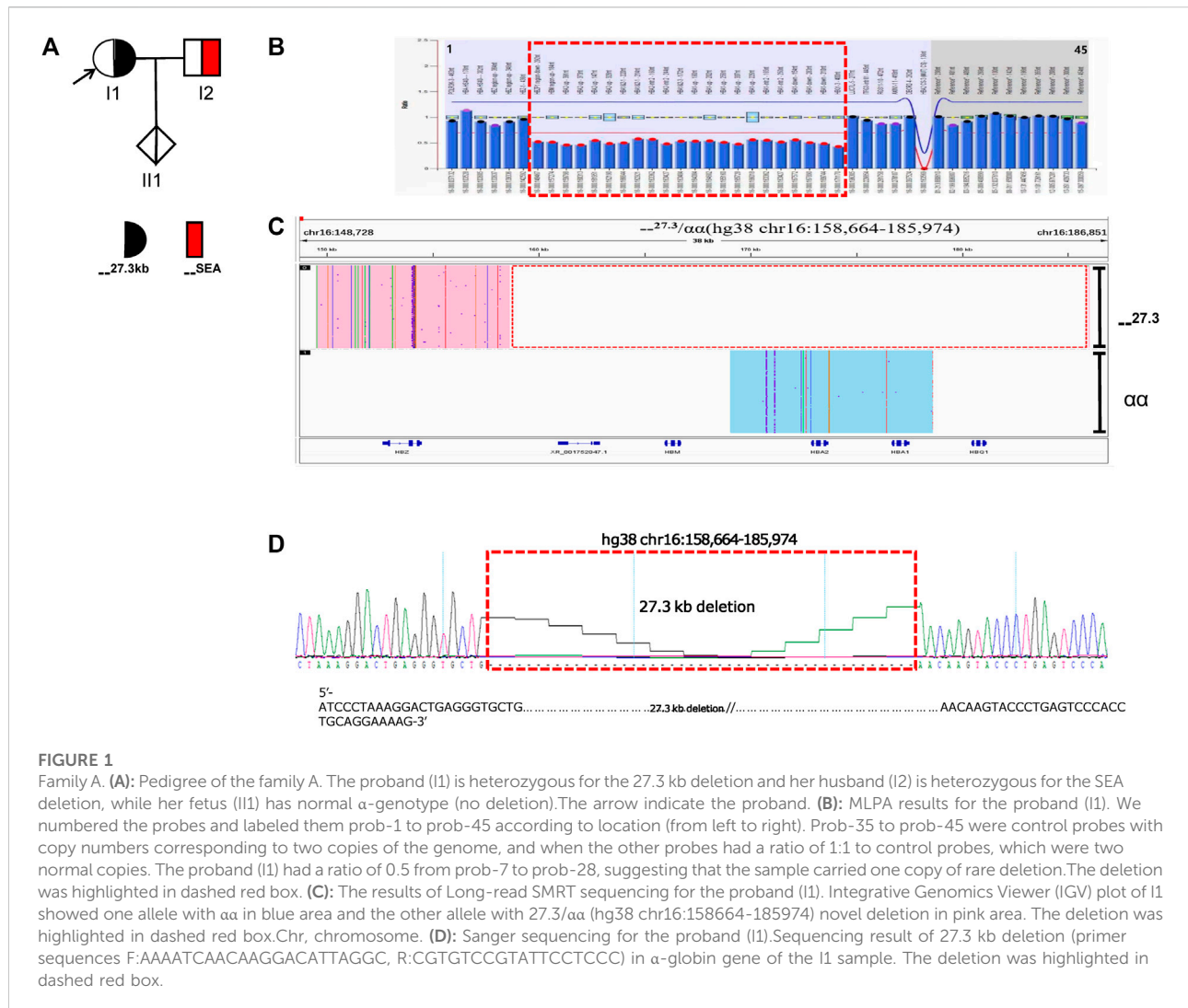
Data analysis

The results of the two methods were assessed and individuals with negative test results were classed as negative for thalassemia. Those with positive test results were diagnosed as thalassemia carriers. We calculated the coincidence rate for the results of the two methods. For the samples with inconsistent results between the two methods, verification by Sanger sequencing was performed (deletion and duplication mutations were verified by Q-PCR and MLPA technology) to determine which result was correct.

Results

Identification of a novel 27,311 bp deletion (--^{27.3}/) that caused α^0 -thalassemia in family A and the sporadic case

The proband of family A had microcytic hypochromic anemia but no abnormality was found in routine genetic testing and their ferritin result was normal. The pedigree chart showed in Figure 1A. Due to the discordance between the thalassemia genetic testing and the clinical hematologic phenotype, other types of mutations were considered. No mutations were found in the *HBA* and *HBB* genes by Sanger sequencing. An MLPA analysis was performed to identify the presence of deletions in the α -globin gene cluster. This showed that the proband carried a large deletion that contained the *HBZP1*, *HBM*, *HBA1*, *HBA2* and *HBQ1* genes (Figure 1B). Due to the limited number of MLPA probes, the deletion breakpoint could not be identified and, subsequently, SMRT was performed to pinpoint the deletion extent and



breakpoint. SMRT confirmed the MLPA results and indicated the presence of a heterozygous deletion in the mother, which had been inherited by the proband (Figure 1C). Bioinformatics analysis successfully positioned the breakpoints at chromosome 16: 158664-185974 (GRCh38) and predicted the deletion to be approximately 27311 bp in size and Sanger sequencing through designed primers were consistent with the SMRT sequencing (Figure 1D). The proband (I1) had a heterozygous 27.3 kb deletion in the α -globin gene cluster ($-^{27.3}/\alpha\alpha$ (hg38 chr16: 158664-185974), which resulted in the expression of only two functional α -globin genes, reduced mean corpuscular volume (MCV) and reduced mean corpuscular hemoglobin (MCH). The proband had α^0 -thalassemia characteristics and a hematological phenotype that was similar to $-^{SEA}$ (NC_000016.10: g.165401_184701del) carriers. Similarly, we found a sporadic case of the α -globin gene cluster 27,311 bp deletion, $-^{27.3}/\alpha\alpha$, that also had α^0 -thalassemia features, reduced MCV and reduced MCH (Table 1).

Identification of a novel 16.1 kb deletion of HS-40 that caused thalassemia in family B

The family B proband (II) and her son (III) had microcytic hypochromic anemia. Routine genetic testing revealed no abnormalities, which suggested the possibility of an unknown mutation. The pedigree chart showed in Figure 2A. No mutations were found in the *HBA* and *HBB* genes by Sanger sequencing. The MLPA analysis revealed that the proband was heterozygous for a deletion in the upstream regulatory element, HS-40, of the *HBA* gene. The son (III) also had a heterozygous deletion that involved HS-40, with a missing breakpoint (Figure 2B). Subsequently, SMRT was used to confirm MLPA results and indicated the presence of a heterozygous deletion in the mother, which had been inherited by the proband. Bioinformatics analysis successfully positioned the breakpoints at chromosome 16:100600-116678 (GRCh38) and predicted the deletion to be approximately 16.1 kb in size (Figure 2C). The

TABLE 1 Hematological data and α/β -globin genotype of case.

| Case | Sex-age | Hb (g/L) | RBC ($10^{12}/L$) | MCV (fL) | MCH (pg) | MCHC (g/L) | Hb A2 (%) | Hb A (%) | Hb F (%) | Hb Bart's (%) | α Genotype (Gap-PCR+RDB) | β Genotype (RDB) | α or β Genotype (MLPA) | α or β Genotype (SMRT) | |
|---------------|-----------------|----------|---------------------|----------|----------|------------|-----------|----------|----------|---------------|---------------------------------|------------------------------------|---|---|---|
| Family A | | | | | | | | | | | | | | | |
| I-1 | Proband(mother) | M-40 | 99 | 4.46 | 66.5 | 22.1 | 332 | 2.3 | 97.7 | - | - | $\alpha\alpha/\alpha\alpha$ | β^N/β^N | Normal | hg38 chr16:158664-185974 |
| I-2 | Father | F-39 | 152 | 7.45 | 69 | 21 | 307 | 2.3 | 97.7 | - | - | -- ^{SEA} / $\alpha\alpha$ | β^N/β^N | --SEA/ $\alpha\alpha$ | -- ^{SEA} / $\alpha\alpha$ |
| II-1 | Fetus | 25W+4D | - | - | - | - | - | - | - | - | - | $\alpha\alpha/\alpha\alpha$ | β^N/β^N | Normal | $\alpha\alpha/\alpha\alpha$ |
| Sporadic case | | | | | | | | | | | | | | | |
| I-1 | Sporadic case | M-38 | 98 | 4.28 | 72.2 | 22.8 | 316 | 2.6 | 97.4 | 0 | - | $\alpha\alpha/\alpha\alpha$ | β^N/β^N | <i>HBA1</i> 、 <i>HBA2</i> deletion | hg38 chr16:158664-185974 |
| Family B | | | | | | | | | | | | | | | |
| I-1 | Proband(mother) | M-25 | 111 | 4.69 | 71.5 | 23.6 | 330 | 2.5 | 87.1 | 0.4 | - | $\alpha\alpha/\alpha\alpha$ | β^N/β^N | <i>HS-40</i> deletion | hg38 chr16:100600-116678 |
| I-2 | Father | F-25 | 136 | 6.82 | 62.9 | 19.9 | 316 | 5.3 | 83.8 | 1.5 | - | $\alpha\alpha/\alpha\alpha$ | <i>HBB</i> : <i>c.126_129delCTTT</i> | $\alpha\alpha/\alpha\alpha$ and β^N/β^N | <i>HBB</i> : <i>c.126_129delCTTT</i> |
| II-2 | Proband | F-3 | 119 | 5.76 | 65.1 | 20.7 | 317 | 2.1 | 93.9 | 4 | - | $\alpha\alpha/\alpha\alpha$ | <i>HBB</i> : <i>c.126_129delCTTT</i> | <i>HS-40</i> deletion | hg38 chr16:100600-116678; <i>HBB</i> : <i>c.126_129delCTTT</i> |
| Family C | | | | | | | | | | | | | | | |
| I-1 | Mother | M-51 | 125 | 4.06 | 93.2 | 30.8 | 331 | 2.7 | 97.3 | 0 | - | $\alpha\alpha/\alpha\alpha$ | β^N/β^N | <i>HBG1-HBG2</i> deletion | hg38 chr11:5249345-5254268 |
| I-2 | Father | F-55 | 129 | 5.73 | 70.3 | 22.5 | 320 | 4.8 | 89.5 | 5.7 | - | $\alpha\alpha/\alpha\alpha$ | <i>HBB</i> : <i>c.-78A>G</i> | Normal | <i>HBB</i> : <i>c.-78A>G</i> |
| II-1 | Sister | M-35 | 118 | 5.16 | 74.9 | 22.9 | 306 | 4.1 | 91.6 | 4.3 | - | $\alpha\alpha/\alpha\alpha$ | <i>HBB</i> : <i>c.-78A>G</i> | Normal | <i>HBB</i> : <i>c.-78A>G</i> |
| II-2 | Proband | F-27 | 139 | 4.45 | 93 | 31.3 | 329 | 2.1 | 93.9 | 4 | - | $\alpha\alpha/\alpha\alpha$ | β^N/β^N | <i>HBG1-HBG2</i> deletion | hg38 chr11:5249345-5254268 |
| II-3 | Wife | M-21 | 110 | 5.77 | 58.8 | 19 | 325 | 5.6 | 93 | 1.4 | - | $\alpha\alpha/\alpha\alpha$ | <i>HBB</i> : <i>c.126_129delCTTT</i> | Normal | <i>HBB</i> : <i>c.126_129delCTTT</i> |
| II-4 | Younger brother | M-26 | 129 | 5.55 | 73.5 | 23.2 | 316 | 5.4 | 91.8 | 2.8 | - | $\alpha\alpha/\alpha\alpha$ | <i>HBB</i> : <i>c.-78A>G</i> | Normal | <i>HBB</i> : <i>c.126_129delCTTT</i> |
| III-1 | Fetus | 22W | - | - | - | - | - | - | - | - | - | $\alpha\alpha/\alpha\alpha$ | <i>HBB</i> : <i>c.126_129delCTTT</i> | <i>HBG1-HBG2</i> deletion | hg38 chr11:5249345-5254268; <i>HBB</i> : <i>c.126_129delCTTT</i> |
| | | F-3D | 165 | 4.75 | 103 | 34.8 | 338 | 0 | 15 | 85 | | | | | |
| | | F-6M | 106 | 5.53 | 60.6 | 19.2 | 317 | - | - | - | | | | | |
| Family D | | | | | | | | | | | | | | | |
| II | Mother | M-70 | 121 | 5.75 | 65.4 | 21.1 | 322 | 5.2 | 94.4 | 0.4 | - | $\alpha\alpha/\alpha\alpha$ | <i>HBB</i> : <i>c.-78A>G</i> | Normal | <i>HBB</i> : <i>c.-78A>G</i> |

(Continued on following page)

TABLE 1 (Continued) Hematological data and α/β -globin genotype of case.

| Case | Sex-age | Hb (g/L) | RBC ($10^{12}/L$) | MCV (fL) | MCH (pg) | MCHC (g/L) | Hb A2 (%) | Hb F (%) | Hb Bart's (%) | α Genotype (Gap-PCR+RDB) | β Genotype (RDB) | α or β Genotype (MLPA) | α or β Genotype (SMRT) |
|------|---------|----------|---------------------|----------|----------|------------|-----------|----------|---------------|--------------------------------------|------------------------|--------------------------------------|--|
| II1 | Sister | 132 | 4.59 | 86.8 | 28.7 | 331 | 2.5 | 96.5 | 1 | $-\alpha^{3.7}/\alpha\alpha$ | β^N/β^N | - | $-\alpha^{3.7}/\alpha1\alpha2/\alpha\alpha$ |
| II4 | Brother | 129 | 5.75 | 69.9 | 22.4 | 320 | 5.5 | 94.5 | 0 | $-\alpha^{3.7}/\alpha\alpha$ | $HBB:c.-78A>G$ | - | $-\alpha^{3.7}/\alpha1\alpha2/\alpha\alpha$; $HBB:c.-78A>G$ |
| II5 | Proband | 117 | 3.89 | 87.9 | 30 | 341 | 2.4 | 97.6 | 0 | $-\alpha^{3.7}/\alpha\alpha$ | β^N/β^N | - | $-\alpha^{3.7}/\alpha1\alpha2/\alpha\alpha$ |
| II6 | Husband | 141 | 3.92 | 75 | 23.2 | 336 | 2.4 | 96.9 | 0.7 | $--SEA/\alpha\alpha$ | β^N/β^N | $--SEA/\alpha\alpha$ | $--SEA/\alpha\alpha$ |
| III1 | Son | 113 | 5.33 | 64.5 | 21.2 | 328 | 2.2 | 97 | 0.8 | $--SEA/\alpha\alpha$ | β^N/β^N | $--SEA/\alpha\alpha$ | $--SEA/\alpha\alpha$ |
| III2 | Fetus | 20W+3D | - | - | - | - | 7.3 | 79.2 | 13.5 | $-\alpha^{3.7}/\alpha1\alpha2/--SEA$ | β^N/β^N | $-\alpha^{3.7}/\alpha1\alpha2/--SEA$ | $-\alpha^{3.7}/\alpha1\alpha2/--SEA$ |

Abbreviations: RBC, red blood cell count Hb, hemoglobin. MCV, mean corpuscular volume. MCH, mean corpuscular Hb. β^N , 17 β -thalassaemia genotypes commonly found in the Chinese population was not found. $--SEA$, Southeast Asian deletion, NG_000006.1:g.26264_45564del19301..c3.7, 3.7-kb deletion of HBA; NG_000006.1:g.34164_37967del3804.

hematological features of the mother with the 16,079 bp deletion in the HS-40 region (hg38 chr16:100600-116678) were as follows: Hb 111 g/L, MCV 71.5 fL, MCH 23.6 pg, 0.4% HbF and 2.5% HbA2) (Table 1). Both cases with this deletion presented with microcytic hypochromic anemia.

Identification of a novel 4.9 kb deletion that caused thalassemia in family C

In family C (Figure 3A), routine blood parameters of the proband were found to be normal but the HbF levels were increased. A SMRT analysis confirmed MLPA results (Figure 3B) and indicated the presence of a heterozygous deletion in the proband, which was inherited by the proband. Subsequently, third generation sequencing was performed to pinpoint the deletion extent and breakpoints. Bioinformatics analysis successfully positioned the breakpoints at chromosome 11:5249345-5254268 (GRCh38) and predicted the deletion to be approximately 4.9 kb in size (Figure 3C). The proband (II-2) and his mother (I-1) were carriers of the 4.9 kb deletion (hg38 chr11:5249345-5254268) in the β -globin gene cluster, HBG1-HBG2. The father (I-2), sister (II-1) and younger brother (II-4) carried $HBB:c.-78A > G$, whilst his wife (II-3) carried $HBB:c.126_129delCTTT$. The fetus (III-1) had $HBB:c.126_129delCTTT$, with the $HBG1-HBG2$ 4.9 kb deletion. The MCV and MCH levels in the proband were normal. The HbA2 content was decreased, whilst the HbF content was increased. The mother's hematological phenotype was normal, whilst the father (I-2), sister (II-1) and younger brother showed mild hypochromic anemia, increased HbA2 content and slightly increased HbF content in the father (I-2) and sister (II-1). The wife has microcytic hypochromic anemia and her HbA2 level was increased (Table 1). The son had microcytic hypochromic anemia after birth and his hemoglobin level reached mild anemia.

A case and pedigree analysis of the allelic $-\alpha^{3.7}\alpha1\alpha2$ gene rearrangement

In family D (Figure 4A), the routine blood parameters of the proband were normal. Routine genetic testing found that the proband was $-\alpha^{3.7}/\alpha\alpha$ and her husband was $--SEA/\alpha\alpha$. Both the proband and her husband were carriers of α -thalassaemia intermedia. When there is a risk of α -thalassaemia, prenatal diagnosis in pregnant women is performed using amniotic fluid and umbilical cord blood at 20 weeks and 3 days into pregnancy. Prenatal diagnosis by fetal Gap-PCR revealed abnormal bands (Figure 4B). The RDB results were abnormal and the MLPA analysis revealed an abnormal α -globin copy number (Figure 4C). The SMRT results showed a rearrangement of the $-\alpha^{3.7}\alpha1\alpha2$ gene on one allele but it was not possible to determine whether $-\alpha^{3.7}$ was spliced in the upstream or downstream region of $\alpha1\alpha2$ (Figure 4D). Analyses of the

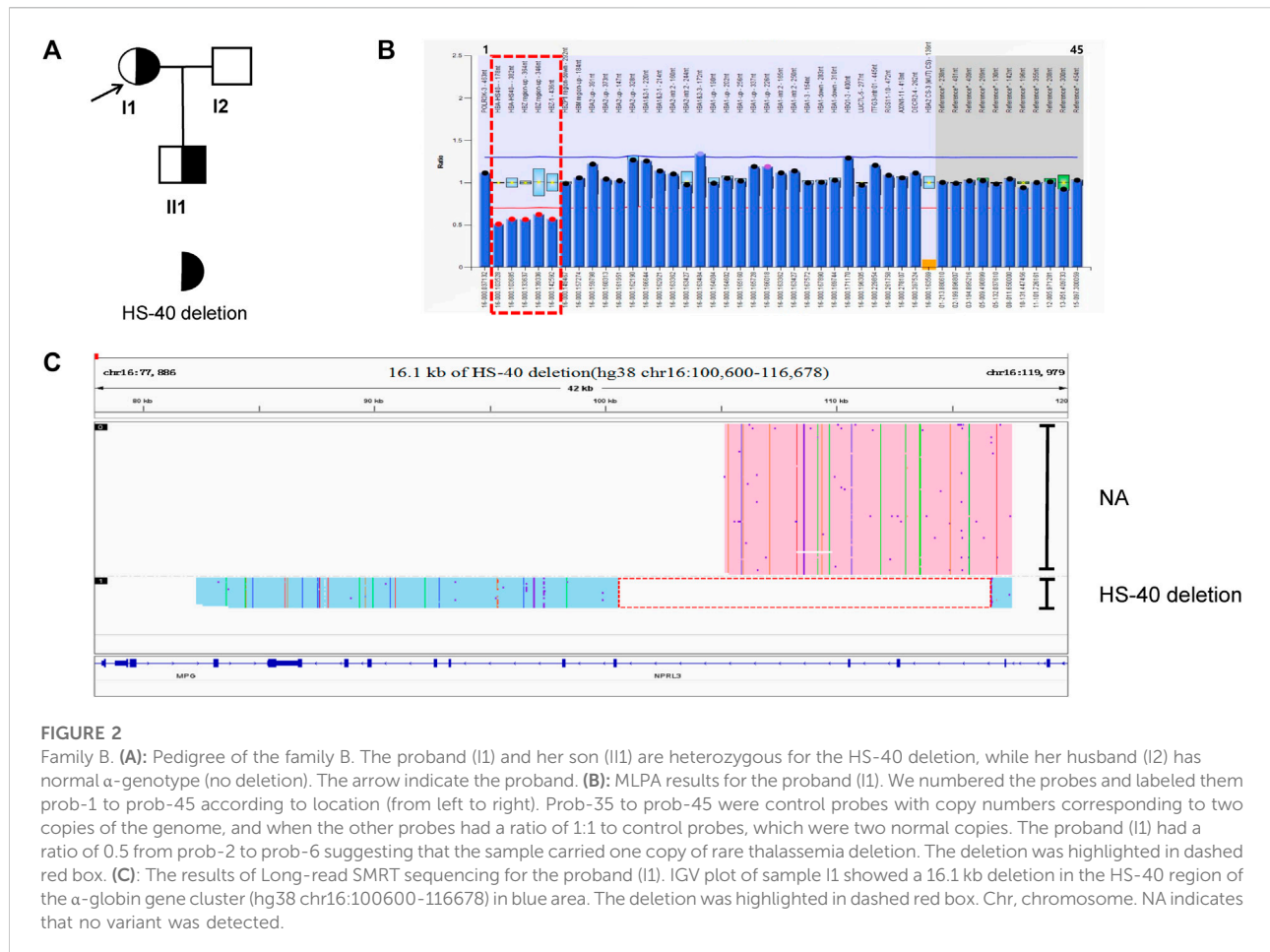


FIGURE 2

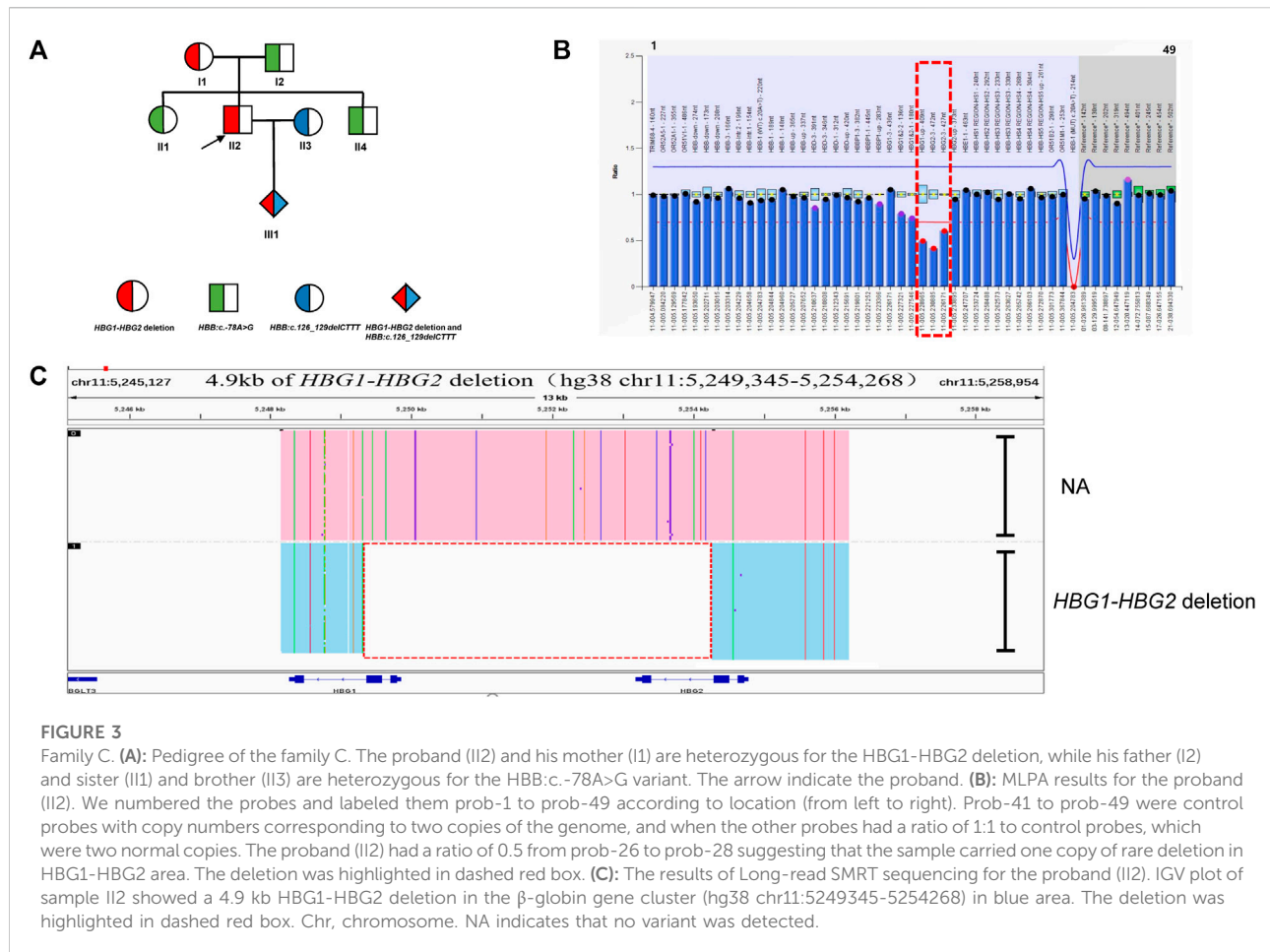
Family B. (A): Pedigree of the family B. The proband (I1) and her son (II1) are heterozygous for the HS-40 deletion, while her husband (I2) has normal α -genotype (no deletion). The arrow indicate the proband. (B): MLPA results for the proband (I1). We numbered the probes and labeled them prob-1 to prob-45 according to location (from left to right). Prob-35 to prob-45 were control probes with copy numbers corresponding to two copies of the genome, and when the other probes had a ratio of 1:1 to control probes, which were two normal copies. The proband (I1) had a ratio of 0.5 from prob-2 to prob-6 suggesting that the sample carried one copy of rare thalassaemia deletion. The deletion was highlighted in dashed red box. (C): The results of Long-read SMRT sequencing for the proband (I1). IGV plot of sample I1 showed a 16.1 kb deletion in the HS-40 region of the α -globin gene cluster (hg38 chr16:100600-116678) in blue area. The deletion was highlighted in dashed red box. Chr, chromosome. NA indicates that no variant was detected.

family and the phenotype of the rare variant were also carried out (Table 1). One $-\alpha^{3.7}\alpha 1\alpha 2$ gene rearrangement was detected in four cases (II1, II4, II5 and III2). II1 and II5 carried only gene rearrangements, and hematological parameters were normal. The II4 gene rearrangement was consistent with *HBB*:c.-78A > G and the individual had microcytic hypochromic anemia, an HbA2 content of 5.5% and an HbF content of 0%. The III2 individual was a fetus and prenatal diagnosis indicated a gene rearrangement that was consistent with $--^{SEA}/\alpha\alpha$ (Figures 4E,F). The Hb Bart's content of the umbilical cord blood was 13.5%, as shown by electrophoresis, and III2 had the Hb H disease phenotype.

Discussion

In recent years, the application of SMRT sequencing technology has been popular for the genetic detection of thalassemia (Xu et al., 2020; Liang et al., 2021; Jiang et al., 2022; Luo et al., 2022; Wu et al., 2022). The technology has a wide detection range, high accuracy and is able to detect

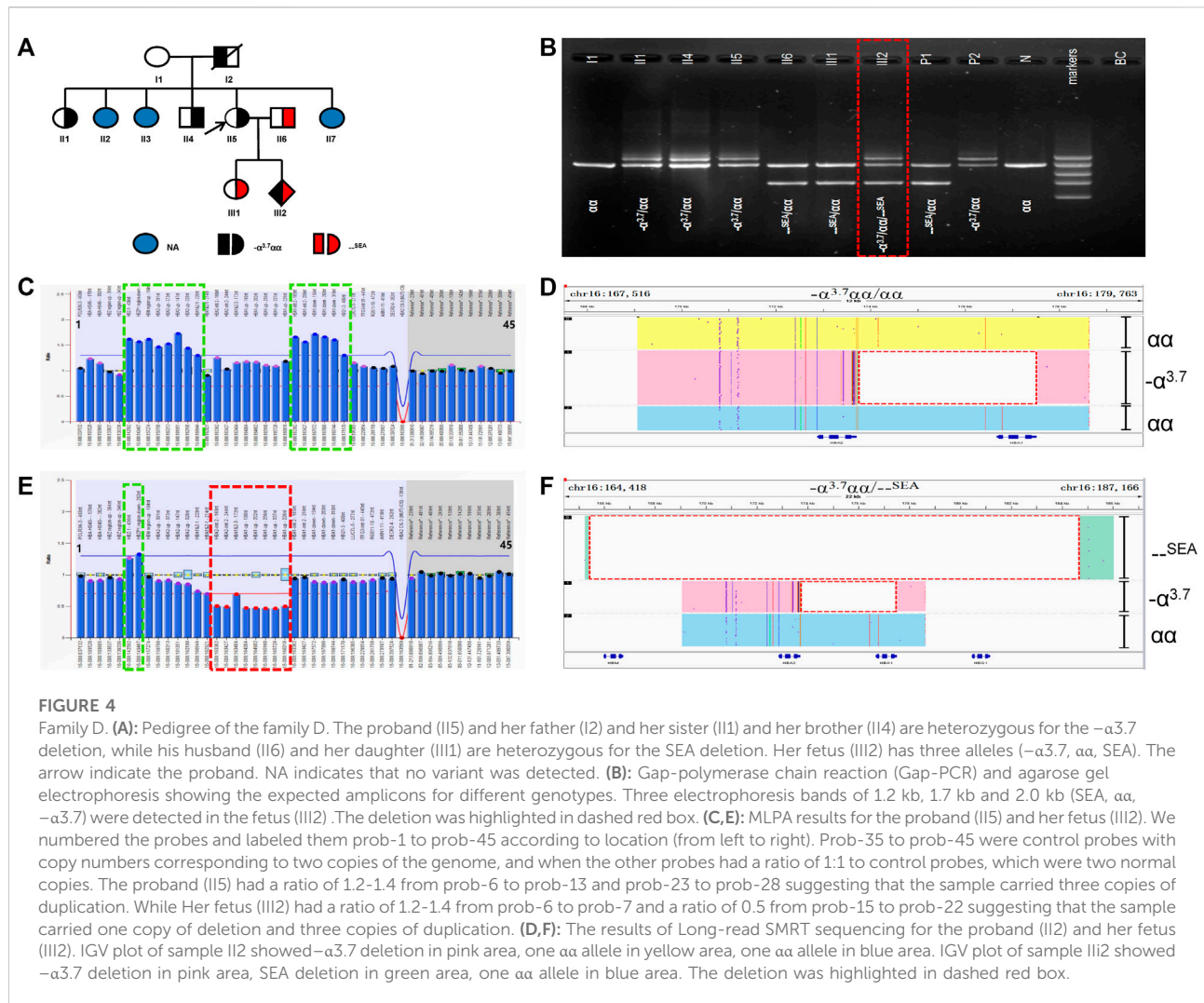
structural variations in *HBB* and *HBA1/HBA2* gene mutations. It is particularly useful for the discovery of new mutations in *HBB* and *HBA* genes and for the differential diagnosis of trans or cis variants, such as $-\alpha^{3.7}$, $-\alpha^{4.2}$, $\alpha\alpha^{anti3.7}$, $\alpha\alpha^{anti4.2}$ and HK $\alpha\alpha$. Using the powerful PacBio detection system, a high throughput level of simultaneous detection of 384 samples can be achieved, which makes the cost of a test less than \$20 (Wu et al., 2022). In this study, the 27,311 bp deletion ($--^{27.3}$) and 16,079 bp deletion in the HS-40 region of the α -globin gene cluster were reported and located for the first time by SMRT sequencing technology. One allele of the α -globin gene cluster $\alpha^{3.7}\alpha 1\alpha 2$ rearrangement and a β -globin gene cluster *HBG1-HBG2* 4,924 bp deletion were also reported. Phenotype and pedigree analyses were performed for the four rare variants (Luo et al., 2022). The limitations of conventional techniques mean that they cannot be used to accurately diagnose these four rare variants. Techniques such as Gap-PCR and PCR-RDB can only detect common types of thalassemia, which include four deletion α -thalassemias, three non-deletion α -thalassemias and 17 *HBB* gene mutations. Using MLPA, globin copy number variations, such as duplications and deletions, can be identified on the α -globin and β -globin gene



clusters but the lack of detection probes means that this technology cannot accurately locate deletion breakpoints and structures. Sanger sequencing technology is suitable for SNVs, indels and point mutations but it is very difficult to detect large fragment deletions or duplications using this method. It is possible to combine Gap-PCR and Sanger sequencing technology to design primers near the deletion breakpoint, in order to identify and locate unknown mutations. However, this is very time consuming, with a high mutation rate of primers, low detection throughput and issues with the existence of pseudogenes and highly homologous genes (Krier et al., 2016). The α -globin gene cluster 27,311 bp deletion, $-^{27.3}/\alpha\alpha$, results in the expression of only two functional α -globin genes. Individuals with heterozygous deletions (e.g., I2 and the sporadic case) have the clinical hematological features of α^0 -thalassemia, with reduced MCV and MCH and a deletion size and region similar to the $-^{SEA}/\alpha\alpha$ deletion. The carrier phenotypes are also very similar. The clinical case with the 16,079 bp deletion in the HS-40 region had microcytic hypochromic anemia. The HS-40 region contains an erythroid-specific DNase I high-sensitivity site that is 40 kb upstream of the α -globin gene cluster. It is an

important regulatory site that affects the expression of α -globin and its deletion affects downstream α -globin expression (Harteveld et al., 2002; Espéret et al., 2000). In this case, the deletion on one chromosome reduced the expression of $\alpha 1$ and $\alpha 2$ globin. The patient presented with microcytic hypochromic anemia and mild anemia, which was similar to the hematological phenotype of patients with pure $-^{SEA}/\alpha\alpha$. This indicated that deletion of this locus leads to a decrease in the expression of α -globin, as in patients with $-^{SEA}/\alpha\alpha$. Our previous studies were consistent with this result. The deletion of the HS-40 locus, combined with the Southeast Asian type deletion α -thalassemia, will manifest as papilledema fetuses (Luo et al., 2020).

On the β -globin gene cluster, one gene is expressed in the fetal period (ϵ), two genes are expressed in infancy ($G\gamma$ and $A\gamma$), two genes are expressed in the adult period (β and δ) and there is one pseudogene ($\psi\beta$). The order of the genes on the gene cluster is 5'- ϵ - $G\gamma$ - $A\gamma$ - $\psi\beta$ - δ - β -3'. Upstream of the ϵ gene, there is a locus control region (LCR), which is crucial for the expression of all genes on the β -globin gene cluster. The $G\gamma$ and $A\gamma$ globin genes (*HBG1* and *HBG2*) are commonly expressed in the fetal liver, spleen and bone marrow. The



predominant fetal hemoglobin (HbF, α_2/γ_2) consists of two γ chains and two α chains, which are usually replaced by adult hemoglobin (HbA, α_2/β_2) at birth. In some β -thalassemias and related diseases the production of γ chains persists into adulthood and can be phenotyped as persistent fetal hemoglobinemia. The two gamma chains of the G γ and A γ globin proteins differ at site 136, where glycine is found in the G γ protein (*HBG2*) and alanine is found in the A γ protein (*HBG1*). The G γ version is predominant at birth (Kiefer et al., 2011; Huang et al., 2017). This study reported a case of a 4,924 bp deletion (hg38 chr11:5249345–5254268) in the β -globin gene cluster, which covers the *HBG1-HBG2* region. A study has reported that this A γ -IVS2 regulatory region may affect HbF regulation in adults (Donovan-Peluso et al., 1987). The *HBG2-HBG1* fusions cause β -thalassemia and have been described in the literature (Shimizu et al., 1986; Jiang et al., 2020). However, in the cases that we reported, the blood cell

parameters were normal and the HbF content was normal or increased, which is inconsistent with the above literature. It may be that the function of the *HBG1-HBG2* gene is turned off in adults and there are differences in the impact on the hematological phenotype (Lee et al., 2010; Chaouch et al., 2020). The wife carried *HBB*: c.126_129delCTTT and showed microcytic hypochromic anemia and increased HbA2 content. Three days after the birth of the son of the proband, heel blood was collected for hemoglobin electrophoresis. No abnormality was found in HbA or HbF levels. Analysis of his genetic tests showed that he had inherited his mother's *HBB*: c.126_129delCTTT mutation and his father's *HBG1-HBG2* 4,924 bp deletion. Retrospective analysis of his blood for six months showed that he had mild microcytic hypochromic anemia. This showed that *HBB*: c.126_129delCTTT with the *HBG1-HBG2* 4,924 bp deletion did not reach intermediate anemia. The hematological

phenotype of the co-inherited cases of the two mutations was similar to the phenotype of the β -thalassemia minor genotype and this is the first report of the effect of this deletion position on the phenotype of β -thalassemia. The effect of the *HBG1-HBG2* 4,924 bp deletion on HbA2 and HbF varied among individuals, which increases the amount of data available for research.

α -globin polyploidy is ubiquitous in nature. The α -globin gene cluster has three homologous X-, Y- and Z-sequences and misplacement or unequal exchange between them, during meiosis, leads to the generation of triploids of the α -globin gene. An exchange between the homologous Z2 and Z1 boxes (cross to the right) would result in a $-\alpha^{3.7}$ monoallelic deletion and a triple $\alpha\alpha\alpha^{\text{anti-3.7}}$ allele. If there is a crossover between the X2 and X1 boxes (crossing to the left), there will be a $-\alpha^{4.2}$ monoallelic deletion and a triple $\alpha\alpha\alpha^{\text{anti-4.2}}$ allele (Wang et al., 2003). Since the triple alleles still have intact $\alpha 1$ and $\alpha 2$ genes, $\alpha\alpha\alpha^{\text{anti-4.2}}$ and $\alpha\alpha\alpha^{\text{anti-3.7}}$ will be misdiagnosed as $\alpha\alpha$, based on conventional Gap-PCR detection. Studies have shown that approximately 2% of α -globin triple alleles are present in individuals with moderate to severe β -thalassemia (Sollaino et al., 2009; Shooter et al., 2015). The coexistence of α -globin polyploidy and β -thalassemia can aggravate the β/α imbalance and results in intermediate β -thalassemia. In our case, through SMRT sequencing technology, it was clear that one chromosome of the mother was normal, whilst the other had complete $\alpha 2\alpha 1$ and $-\alpha^{3.7}$, to give a total of five copies of α -globin. Conventional Gap-PCR testing would misdiagnose this individual as $-\alpha^{3.7}/\alpha\alpha$. The MLPA technique can determine the existence of the five copies of α and the $-\alpha^{3.7}$ deletion but cannot determine the distribution. In the II1, II4, II5 and III2 cases, complete $\alpha 2\alpha 1$ and $-\alpha^{3.7}$ were detected on one chromosome and, in III1 and II5, one chromosome was normal, whilst the other had complete $\alpha 2\alpha 1$ and $-\alpha^{3.7}$, to give a total of five copies of α -globin. The hematological phenotype was normal. It is particularly noteworthy that the II4 case had a total of five copies of the α gene that were co-inherited with *HBB*: c.-78A > G, which aggravated the β/α imbalance. Previous literature has reported that this would generally cause intermediate β -thalassemia (Sollaino et al., 2009; Shooter et al., 2015). However, in this case, the individual showed only mild β -thalassemia. One chromosome of the III2 fetus was $-\alpha^{\text{SEA}}$, whilst the other was inherited from the mother's three recombined α -globin genes. Generally speaking, the presence of three normal copies of α -globin would produce a quiescent or mild α -thalassemia phenotype but, in fact, umbilical cord blood electrophoresis showed that the Hb Bart's content reached 13.5%, which can be diagnosed as Hb H disease. The SMRT sequencing technology confirms that one chromosome of the mother was normal, whilst the other had complete $\alpha 2\alpha 1$ and $-\alpha^{3.7}$. There was a total of five copies of the α -globin gene, but it is still not clear whether $-\alpha^{3.7}$ was spliced in the upstream or downstream region of $\alpha 2\alpha 1$. Through analysis of the phenotypic differences between the above two cases, we

speculated that $-\alpha^{3.7}$ recombination may have occurred at the 5' end, upstream of $\alpha 2\alpha 1$. Since the variant occurred in the upstream region, it may affect the expression of downstream $\alpha 2\alpha 1$, which requires further research.

Conclusion

In conclusion, this study demonstrated that SMRT sequencing technology is suitable for the detection of unknown deletions and complex structural variants in thalassemia cases. Compared with traditional methods, this technology is more effective and accurate. A single test can identify the breakpoints and structural variations of unknown deletions, within a short experimental period and at low cost. Due to its high technical requirements, SMRT has not yet been popularized in clinical practice. We believe that this technique shows great promise for the detection of unknown deletions and complex structural variants in thalassemia cases.

Data availability statement

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by ethics committee of the Liuzhou Maternity and Child Healthcare Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

SL designed the study, performed the experimental work and drafted the manuscript. XC compiled the data, contributed to the writing and helped in drafting the manuscript. BL, LC, and QZ contributed to revising the manuscript. DY, JL, and YL performed the experimental work and data analysis. DZ, NT, JC, XW, and TY performed the analysis and provided valuable suggestions and clinical outputs. All authors read and approved the final manuscript.

Funding

This work was supported by the Scientific Research Fund of the Health and Family Planning Commission of Guangxi Zhuang

Autonomous Region (Grant Nos. Z20170528 and Z-B20221578), the Liuzhou Science Technology Innovation Capability and Conditions Construction Project (Grant No. 2018AF10501), the Key Research and Development Program of Guangxi (Grant No. Guike AB18126056), the Key Research and Development Program of Liuzhou (Grant No. 2018BJ10301), the Scientific Research and Technology Development Project of Liuzhou (No. 2018DB20501 and No.2021YB0103A013) and the Guangxi medical high-level backbone talents '139' plan training target special (G202003028), the Liuzhou city 1/10/100 talent special project.

Acknowledgments

We are grateful to our current laboratory members for their helpful comments on the manuscript.

References

- Achour, A., Koopmann, T. T., Baas, F., and Harteveld, C. L. (2021). The evolving role of next-generation sequencing in screening and diagnosis of hemoglobinopathies. *Front. Physiol.* 12, 686689. doi:10.3389/fphys.2021.686689
- Bajwa, H., and Basit, H. (2021). "Thalassemia," in *StatPearls* (Treasure Island (FL): StatPearls Publishing).
- Chaouch, L., Sellami, H., Kalai, M., Darragi, I., Boudrigua, I., Chaouachi, D., et al. (2020). New deletion at promoter of HBG1 gene in sickle cell disease patients with high HbF level. *J. Pediatr. Hematol. Oncol.* 42 (1), 20–22. doi:10.1097/MPH.0000000000001626
- den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., et al. (2016). HGVS recommendations for the description of sequence variants:2016 update. *Hum. Mutat.* 37 (6), 564–569. doi:10.1002/humu.22981
- Donovan-Peluso, M., Acuto, S., Swanson, M., Dobkin, C., and Bank, A. (1987). Expression of human gamma-globin genes in human erythroleukemia (K562) cells. *J. Biol. Chem.* 262 (35), 17051–17057. doi:10.1016/s0021-9258(18)45490-7
- Espéret, C., Sabatier, S., Deville, M. A., Ouazana, R., Bouhassira, E. E., Godet, J., et al. (2000). Non-erythroid genes inserted on either side of human HS-40 impair the activation of its natural α -globin gene targets without being themselves preferentially activated. *J. Biol. Chem.* 275 (33), 25831–25839. doi:10.1074/jbc.M001757200
- Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7 (2), 85–97. doi:10.1038/nrg1767
- Goonasekera, H. W., Paththinige, C. S., and Dissanayake, V. H. W. (2018). Population screening for hemoglobinopathies. *Annu. Rev. Genomics Hum. Genet.* 19, 355–380. doi:10.1146/annurev-genom-091416-035451
- Harewood, J., and Azevedo, A. M. (2021). "Alpha thalassemia." in *StatPearls* (Treasure Island (FL): StatPearls Publishing).
- Harteveld, C. L., Muglia, M., Passarino, G., Kielman, M. F., and Bernini, L. F. (2002). Genetic polymorphism of the major regulatory element HS-40 upstream of the human α -globin gene cluster. *Br. J. Haematol.* 119 (3), 848–854. doi:10.1046/j.1365-2141.2002.03917.x
- Huang, P., Keller, C. A., Giardine, B., Grevet, J. D., Davies, J. O. J., Hughes, J. R., et al. (2017). Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes Dev.* 31 (16), 1704–1713. doi:10.1101/gad.303461.117
- Jiang, F., Mao, A. P., Liu, Y. Y., Liu, F. Z., Li, Y. L., Li, J., et al. (2022). TGS:A novel tool detects complex variants in the α -thalassaemia gene Detection of rare thalassaemia mutations using long-read single-molecule real-time sequencing. *Gene* 825, 146438. doi:10.1016/j.gene.2022.146438
- Jiang, F., Zuo, L., Li, D., Li, J., Tang, X., Chen, G., et al. (2020). Molecular epidemiology and hematologic characterization of $\delta\beta$ -thalassaemia and hereditary persistence of fetal

Conflict of interest

JL and YL were employed by the company Berry Genomics Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

hemoglobin in 125, 661 families of greater Guangzhou area, the metropolis of southern China. *BMC Med. Genet.* 21 (1), 43. doi:10.1186/s12881-020-0981-x

Kiefer, C. M., Lee, J., Hou, C., Dale, R. K., Lee, Y. T., Meier, E. R., et al. (2011). Distinct Ldb1/NLI complexes orchestrate γ -globin repression and reactivation through ETO2 in human adult erythroid cells. *Blood* 118 (23), 6200–6208. doi:10.1182/blood-2011-06-363101

Kousoulidou, L., Männik, K., Sismani, C., Zilina, O., Parkel, S., Puusepp, H., et al. (2008). Array-MAPH: A methodology for the detection of locus copy-number changes in complex genomes. *Nat. Protoc.* 3 (5), 849–865. doi:10.1038/nprot.2008.49

Krier, J. B., Kalia, S. S., and Green, R. C. (2016). Genomic sequencing in clinical practice: Applications, challenges, and opportunities. *Dialogues Clin. Neurosci.* 18 (3), 299–312. doi:10.31887/dcms.2016.18.3/jkrier

Lee, S. T., Yoo, E. H., Kim, J. Y., and Ki, C. S. (2010). Multiplex ligation-dependent probe amplification screening of isolated increased HbF levels revealed three cases of novel rearrangements/deletions in the beta-globin gene cluster. *Br. J. Haematol.* 148 (1), 154–160. doi:10.1111/j.1365-2141.2009.07927.x

Li, B., Zhang, X. Z., Yin, A. H., Zhao, Q. g., Wu, L., Ma, Y. z., et al. (2014). High prevalence of thalassemia in migrant populations in Guangdong Province China. *BMC Public Health* 14 (1), 905. doi:10.1186/1471-2458-14-905

Liang, Q., Gu, W., Chen, P., Li, Y., Liu, Y., Tian, M., et al. (2021). A more universal approach to comprehensive analysis of thalassemia alleles (CATSA). *J. Mol. Diagn.* 23 (9), 1195–1204. doi:10.1016/j.jmoldx.2021.06.008

Liu, J. Z., Han, H., Schouten, J. P., Wang, L. R., Fan, X. P., Duarte, H. B., et al. (2008). Detection of alpha-thalassaemia in China by using multiplex ligation-dependent probe amplification. *Hemoglobin* 32 (6), 561–571. doi:10.1080/03630260802508111

Luo, S., Chen, X., Zeng, D., Tang, N., Yuan, D., Zhong, Q., et al. (2022). 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 (4), 183–195. doi:10.1038/s10038-021-00983-1

Luo, S., Chen, X., Zhong, Q., Wang, Q., Xu, Z., Qin, L., et al. (2020). Analysis of rare thalassemia caused by HS-40 regulatory site deletion. *Hematology* 25 (1), 286–291. doi:10.1080/16078454.2020.1799587

Mettananda, S., and Higgs, D. R. (2018). Molecular basis and genetic modifiers of thalassaemia. *Hematol. Oncol. Clin. North Am.* 32 (2), 177–191. doi:10.1016/j.hoc.2017.11.003

Modell, B., and Darlison, M. (2008). Global epidemiology of haemoglobin disorders and derived service indicators. *Bull. World Health Organ.* 86 (6), 480–487. doi:10.2471/blt.06.036673

Needs, T., Gonzalez-Mosquera, L. F., and Lynch, D. T. (2021). "Beta thalassaemia." in *StatPearls* (Treasure Island (FL): StatPearls Publishing).

- Phylipsen, M., Chaibunruang, A., Vogelaar, I. P., Balak, J. R. A., Schaap, R. A. C., Ariyurek, Y., et al. (2012). Fine-tiling array CGH to improve diagnostics for α - and β -thalassemia rearrangements. *Hum. Mutat.* 33 (1), 272–280. doi:10.1002/humu.21612
- Shen, Y., and Wu, B. L. (2009). Designing a simple multiplex ligation-dependent probe amplification (MLPA) assay for rapid detection of copy number variants in the genome. *J. Genet. Genomics* 36 (4), 257–265. doi:10.1016/S1673-8527(08)60113-7
- Shimizu, K., Harano, T., Harano, K., Miwa, S., Amenomori, Y., Ohba, Y., et al. (1986). Abnormal arrangements in the alpha- and gamma-globin gene clusters in a relatively large group of Japanese newborns. *Am. J. Hum. Genet.* 38 (1), 45–58.
- Shooter, C., Rooks, H., Thein, S. L., and Clark, B. (2015). Next generation sequencing identifies a novel rearrangement in the HBB cluster permitting to-the-base characterization. *Hum. Mutat.* 36 (1), 142–150. doi:10.1002/humu.22707
- Sollaino, M. C., Paglietti, M. E., Perseu, L., Giagu, N., Loi, D., and Galanello, R. (2009). Association of α globin gene quadruplication and heterozygous β thalassemia in patients with thalassemia intermedia. *Haematologica* 94 (10), 1445–1448. doi:10.3324/haematol.2009.005728
- Wang, W., Ma, E. S., Chan, A. Y., Prior, J., Erber, W. N., Chan, L. C., et al. (2003). Single-tube multiplex-PCR screen for anti-3.7 and anti-4.2 alpha-globin gene triplications. *Clin. Chem.* 49 (10), 1679–1682. doi:10.1373/49.10.1679
- Weaver, S., Dube, S., Mir, A., Qin, J., Sun, G., Ramakrishnan, R., et al. (2010). Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods* 50 (4), 271–276. doi:10.1016/j.ymeth.2010.01.003
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37 (10), 1155–1162. doi:10.1038/s41587-019-0217-9
- Wu, H., Chen, D., Zhao, Q., Shen, X., Liao, Y., Li, P., et al. (2022). Long-read sequencing on the SMRT platform enables efficient haplotype linkage analysis in preimplantation genetic testing for β -thalassemia. *J. Assist. Reprod. Genet.* 39 (3), 739–746. doi:10.1007/s10815-022-02415-1
- Xiong, F., Huang, Q., Chen, X., Zhou, Y., Zhang, X., Cai, R., et al. (2011). A melting curve analysis-based PCR assay for one-step genotyping of β -thalassemia mutations: a multicenter validation. *J. Mol. Diagn.* 13 (4), 427–435. doi:10.1016/j.jmoldx.2011.03.005
- Xiong, F., Sun, M., Zhang, X., Cai, R., Zhou, Y., Lou, J., et al. (2010). Molecular epidemiological survey of haemoglobinopathies in the Guangxi Zhuang Autonomous Region of southern China. *Clin. Genet.* 78 (2), 139–148. doi:10.1111/j.1399-0004.2010.01430.x
- Xu, L., Mao, A., Liu, H., Gui, B., Choy, K. W., Huang, H., et al. (2020). Long-molecule sequencing: A new approach for identification of clinically significant DNA variants in α -thalassemia and β -thalassemia carriers. *J. Mol. Diagn.* 22 (8), 1087–1095. doi:10.1016/j.jmoldx.2020.05.004
- Yao, H., Chen, X., Lin, L., Wu, C., Fu, X., Wang, H., et al. (2014). The spectrum of α - and β -thalassemia mutations of the Li people in Hainan Province of China. *Blood Cells Mol. Dis.* 53 (1-2), 16–20. doi:10.1016/j.bcmd.2014.01.003
- Zhang, J., Zhu, B. S., He, J., Zeng, X. H., Su, J., Xu, X. H., et al. (2012). The spectrum of α - and β -thalassemia mutations in yunnan province of southwestern China. *Hemoglobin* 36 (5), 464–473. doi:10.3109/03630269.2012.717327