Journal of Assisted Reproduction and Genetics (2025) 42:2565–2574
https://doi.org/10.1007/s10815-025-03552-z

GENETICS

A novel targeted long-read sequencing-based preimplantation genetic testing method for α-thalassemia (tlrPGT-α-thal)

Qiuwen Shi[1] · Tiantian Xie[2] · Jianmin Li[1] · Ning Li[1] · Changlong Xu[1] · Zhong Lin[3] · Ying Huang[1] · Ling Zhou[1] · Aiping Mao[2] · Jiaqi Li[2] · Rong Li[1]

Received: 24 November 2024 / Accepted: 2 June 2025 / Published online: 8 July 2025

Abstract

Purpose The purpose is to develop a preimplantation genetic testing (PGT) method for α-thalassemia in one reaction, without requiring a proband, using targeted long-range polymerase chain reaction (LR-PCR) followed by long-read sequencing (LRS).

Methods We developed a targeted LRS-based PGT method for α-thalassemia (tlrPGT-α-thal). This novel approach detects both *HBA*-associated variants and informative linked fragments in peripheral regions by targeted LR-PCR in one reaction.

Results A total of 115 samples, including 10 couples and 95 embryos, were subjected to this method. Available heterozygous SNPs were used to distinguish four parental haplotypes. Haplotype linkage analysis was successfully performed in all families, and the results of embryos were all consistent with control results obtained through conventional methods (next-generation sequencing, Sanger sequencing and gap-PCR). This confirmed that tlrPGT-α-thal achieved 100% sensitivity and accuracy.

Conclusion This study has demonstrated the accuracy and clinical utility of tlrPGT-α-thal. This method holds significant promise for clinical application for patients requiring PGT for α-thalassemia. Furthermore, the methodology could potentially be extended to PGT for monogenic disorders (PGT-M) in other diseases.

Keywords Targeted long-range PCR · Targeted long-read sequencing · Haplotype linkage analysis · Informative linked fragment · Preimplantation genetic testing for α-thalassemia

Introduction

Alpha-thalassemia (α-thalassemia), caused by the defect of α-globin chain synthesis, is one of the most common blood disorders worldwide [1, 2]. There is a series of α-globin genes located on chromosome 16. The most important genes among them are *HBA1* and *HBA2*, due to their ability to encode globin in both embryo and adult. Phenotypes of α-thalassemia range from asymptomatic to severe (even lethal) anemia, and the severity is determined by the number of mutated genes inherited from the parents. According to the literature and public databases, the variants of *HBA* are predominantly structural variants (SVs), mainly large deletions, and single-nucleotide variants (SNVs). The most common SVs include the 19.3-kb $-^{SEA}$ deletion, as well as the smaller 3.7-kb ($-\alpha^{3.7}$) and 4.2-kb ($-\alpha^{4.2}$) deletions [3, 4]. The accumulation of the mutant alleles from parents to children can lead to hemoglobin H (Hb H) disease and hemoglobin Bart hydrops fetalis (Hb Bart) syndrome (a lethal hemolytic condition with no expression of *HBA* genes) [4].

Thalassemia is a congenital hereditary condition with long-lasting health impact. Currently, there is no radical treatment for α-thalassemia; thus, patients have to suffer from the long treatment courses and face the risk of lethal hemolytic anemia, which leads to serious economic and social burdens for families [5]. Therefore, couples who are α-thalassemia carriers and at risk of conceiving pregnancies affected by Hb Bart's and Hb H disease may benefit from preimplantation genetic testing for monogenic diseases

Qiuwen Shi and Tiantian Xie are co-first authors.

✉ Rong Li
LiRong@sr.gxmu.edu.cn

[1] Reproductive Medicine Center, The Third Affiliated Hospital of Guangxi Medical University, Nanning 530031, Guangxi, China

[2] Berry Genomics Corporation, Beijing 102200, China

[3] The Reproductive Hospital of Guangxi Zhuang Autonomous Region, Nanning 530218, Guangxi, China

(PGT-M) [6]. With PGT-M, embryos with lower genetic risks can be selectively transferred, thereby completely preventing or significantly reducing the transmission or accumulation of the mutant alleles from parents to offspring.

Due to the limited genetic material available from embryos, which typically consisting of 1–2 polar body cells or 5–10 trophectoderm (TE) cells, whole genome amplification (WGA) is commonly employed to amplify the genetic material for genetic analysis [7]. At present, there are several WGA methods available [8, 9], including multiple displacement amplification (MDA), multiple annealing, looping-based amplification cycles, and degenerate oligonucleotide-primed polymerase chain reaction. However, allele drop-out (ADO), false SNVs and indels, copy number variant, and SVs are prone to occur during the DNA amplification process by all the amplification kits available [9], which causes some allelic genes to be undetectable. Normal ADO would lead to a false-positive homozygous allelic result, while mutant ADO would lead to a false-negative result.

In order to improve the accuracy of diagnosis, simultaneous haplotype linkage analysis has become a standard method for PGT-M [10, 11]. Due to the abundance of single-nucleotide polymorphisms (SNPs) in human genome, SNP-based haplotype analysis is more widely applied than the short tandem repeat (STR)-based haplotyping for genetic testing [7, 12]. The SNP array can only design probes based on known SNPs [13], potentially missing some unknown SNPs that could be used for haplotype analysis. Next-generation sequencing (NGS) is able to detect more SNPs for haplotype linkage analysis and has made great progress in the field of PGT [14], including PGT-M for α-thalassemia [15]. Due to the limitation of its sequencing principle, NGS can only produce short reads (approximately 300 bp in length) [16]. However, there is about one SNP per 1000 bp in human genome [17]. Therefore, NGS typically yields fewer than one SNP per DNA fragment on average, making it difficult to identify continuous SNPs. Moreover, only a limited number of SNPs are suitable for haplotype analysis. Consequently, NGS and SNP array normally need a large number of informative SNPs for linkage analysis, which is a challenge for some specific genomic regions, such as *HBA* located in the proximal telomere of chr16. Thus, combination of several methods might be needed for PGT-M for α-thalassemia in clinical applications [18]. However, all these methods require the existence of a proband or sibling for SNP linkage analysis. The proband's genetic material serves as the template for constructing the disease-associated haplotype. However, the proband or the genetic material of the proband might be unavailable in some families. Recently, there have been several attempts to perform haplotype analysis using the affected embryos as probands [15], which relies on the SNP diversity of embryos and may yield unreliable genetic results for some families. Therefore, the application of NGS and SNP array to PGT-M has technical limitations for certain clinical cases.

Long-read sequencing (LRS) can generate continuous contigs or even scaffolds, enabling genome assemblies of high quality [19]. Some current dilemmas of PGT-M based on NGS and SNP array are expected to be solved by performing LRS. Preliminary results have been achieved on PGT by LRS without requiring a proband [20–22], which implies the possibility of LRS technology for PGT applications. However, the haplotype construction using LRS-based whole-genome sequencing (WGS) with parental genomic DNA [20] remains high-cost. And the informative linked parental SNPs were detected in embryos by Sanger sequencing or NGS [20, 22], which suggests that a combination of multiple methods is required. Moreover, obtaining results by more than one reaction [21] is still inefficient. Here, we constructed a thalassemia genetic testing approach termed comprehensive analysis of thalassemia alleles (CATSA) based on LRS [23–25].

In this study, we established a targeted LRS-based PGT method for α-thalassemia (tlrPGT-α-thal) (Fig. 1). This method analyzed both *HBA*-associated variants directly and informative linked fragments in peripheral regions by targeted long-range PCR in one reaction, integrating haplotype construction, linkage analysis, and embryo genotype determination without requiring a proband.

Method

Study subjects

A total of 115 samples from 10 couples who underwent routine PGT-M for α-thalassemia at Nanning Second People's Hospital from 2021 to 2023, along with their corresponding embryos were enrolled in this retrospective study. These 10 couples carried diverse SVs and SNVs including $--^{SEA}$ deletion, $-\alpha^{3.7}$ deletions, $-\alpha^{4.2}$ deletions, and $\alpha^{cs}\alpha$ (Table S1). Genomic DNA was extracted from their peripheral blood with the QIAamp DNA blood mini kit (Qiagen, Hilden, Germany). For embryos, MDA was performed with several trophectoderm cells with the REPLI-g® Single Cell Kit (Qiagen, Hilden, Germany).

Control method

The genotypes of all these couples and their corresponding embryos have been identified by control methods (Fig. 2). Briefly, genomic DNA from the parents and their relatives was used to confirm variant mutation by Sanger sequencing and Gap-PCR, and haplotypes were constructed by NGS. Then, MDA products of embryos were screened for aneuploidy through NGS using the MGISEQ-2000 platform (MGI Tech

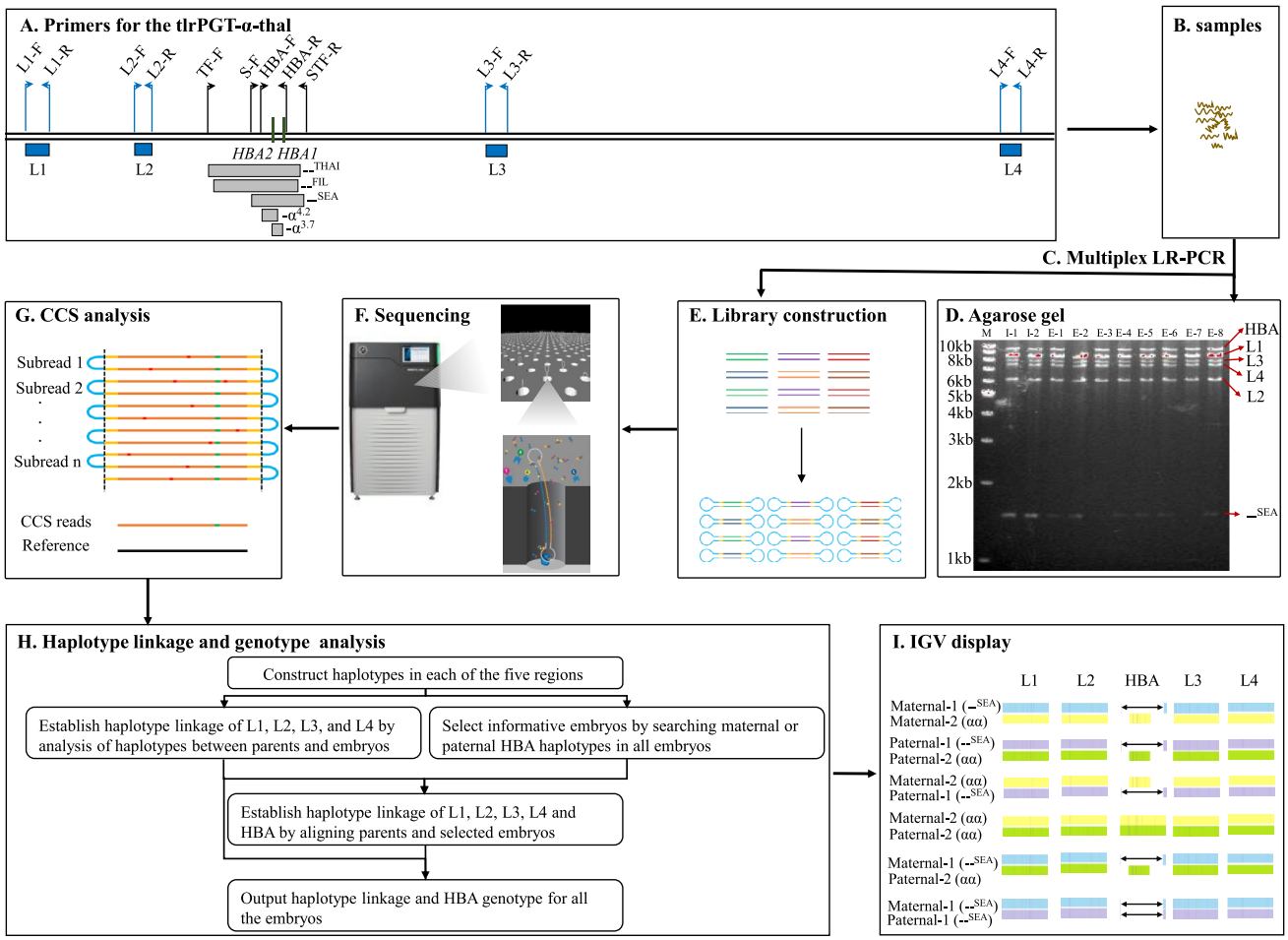

Fig. 1 Targeted long-read sequencing preimplantation genetic testing for α-thalassemia (tlrPGT-α-thal). **A** Design of multiplex long-range primers for this study. The first set of primers marked black arrows was used to detect SVs and SNVs involved in the *HBA* gene directly, which were modified from previous multiplex long PCR analysis [23]. The second set of primers marked blue arrows was designed to amplify the peripheral region of the *HBA* gene. The amplicons (blue boxes) named L1, L2, L3, and L4 are approximately 8.5 kb, 6.2 kb, 8.0 kb, and 7.5 kb, which were designed to construct haplotypes. **B** Samples for tlrPGT-α-thal, including genomic DNA from parental peripheral blood and MDA (multiple displacement amplification) products of embryos. **C** Multiplex LR-PCR, multiplex long-range PCR. **D** Agarose gel electrophoresis showing the amplicons of one family. M, marker. I-1 and I-2 represent the couples in this family. E-1, E-2, E-3, E-4, E-5, E-6, E-7, E-8, and E-9 represent their embryos. **E** The single-molecule real-time (SMRT) dumbbell (SMRTbell) library construction: the amplicons were ligated to barcoded adaptors. **F** Use PacBio Sequel® II systems to sequence. SMRT® Cells contain millions of zero-mode waveguides (ZMWs). A single molecule of DNA is immobilized in each ZMW. SMRTbell templates enable repeated sequencing of circular template with real-time detection of base incorporation. **G** The subreads sequencing data are aligned, sequencing errors are removed, and a final consensus sequence for each long read is generated. CCS, circular consensus sequencing. **H** Customized bioinformatics pipeline for haplotype linkage and *HBA* variants analysis. **I** IGV plots showing the results of haplotype linkage and genotype analysis by tlrPGT-α-thal. IGV, Integrative Genomics Viewer

Co. Ltd., Shenzhen, China) at approximately × 0.1 genome depth. Next, linkage analysis was conducted on MDA products of embryos by NGS, and mutation was validated by Sanger sequencing and Gap-PCR. Finally, the genotype of embryo was determined.

Design of the LRS-based PGT for α-thalassemia (tlrPGT-α-thal)

We aimed to establish a PGT-M method for α-thalassemia, which consists of haplotype construction, linkage analysis,

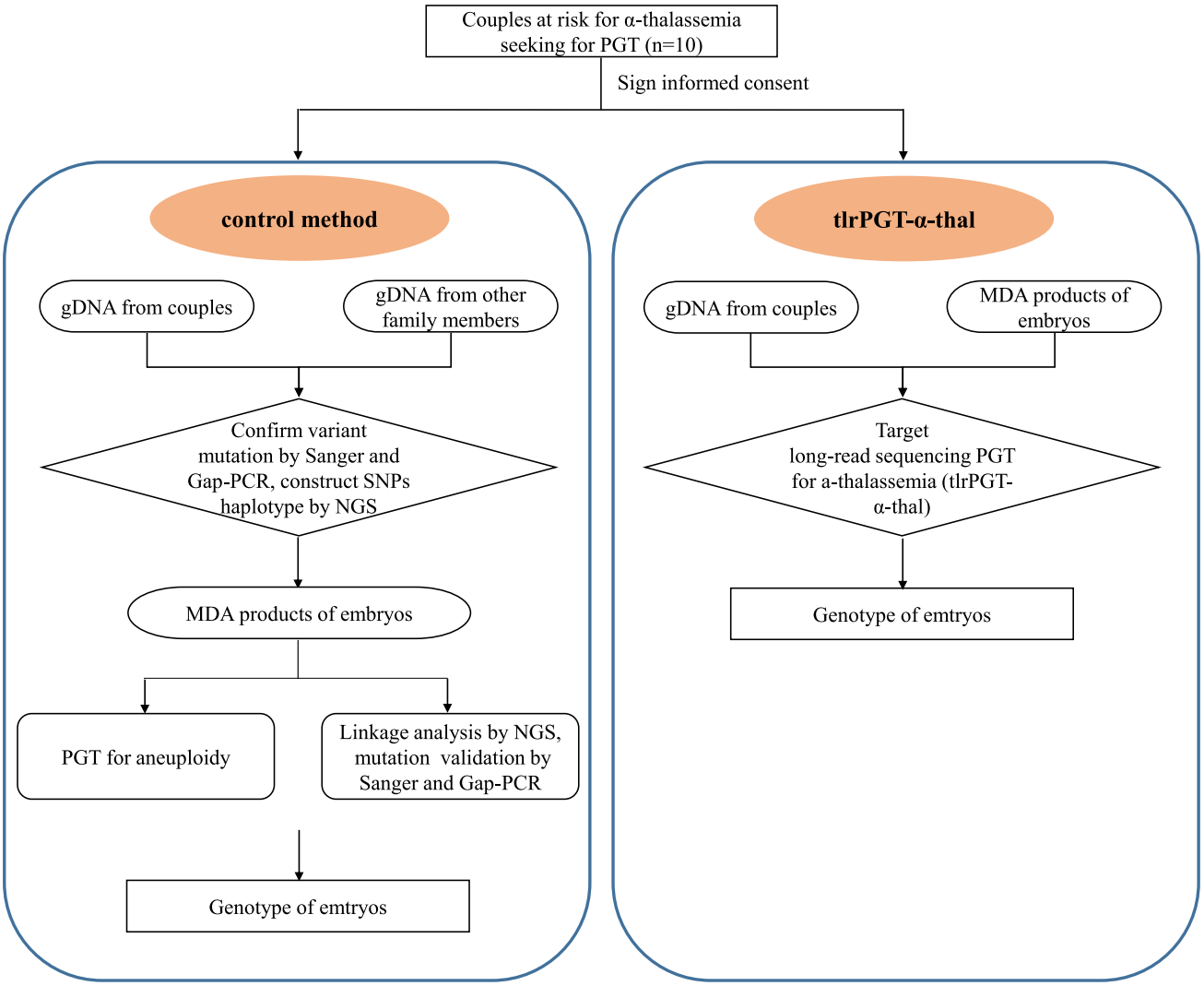

Fig. 2 Flowchart and comparison between tlrPGT-α-thal and control methods. PGT, preimplantation genetic testing. gDNA, genomic DNA. MDA, multiple displacement amplification. NGS, next-generation sequencing

and genotype determination of the embryo without requiring a proband. Two sets of primers were designed in order to avoid the influence of ADO (Fig. 1A). The first set of primers was used to detect SVs and SNVs involved in *HBA* genes, named, HBA-F, HBA-R, S-F, STF-R, and TF-F, which was modified from our previous approach (CATSA) [23] to adapt for PGT-M applications. The second set consisted of four pairs of primers (L1-F and L1-R; L2-F and L2-R; L3-F and L3-R; L4-F and L4-R) targeting regions within 1 Mb upstream and downstream of the *HBA* core region to analyze the haplotypes of the peripheral region.

The selection of the second primer set was based on population allele frequencies of SNPs and their amplification efficiency. First, the population allele frequencies of all SNPs located within the 1 Mb region of the upstream and downstream of *HBA* gene were listed according to gnomAD (https://gnomad.broadinstitute.org/). Secondly, SNVs with appropriate population allele frequency (40–60%) were considered to be optimal. Thirdly, 6–10 kb genomic areas with maximum number of SNPs were selected as amplicon regions. Fourthly, primers were designed to avoid regions containing heterozygous SNPs to avoid allele-specific amplification failure caused by primer-template mismatches. Fifthly, the selected genomic regions were prioritized, and the primers compatible with the entire system were subsequently chosen through the amplification validation. Finally, the four optimal regions were selected based on both assay performance and cost-effectiveness.

By analyzing the haplotypes of parents and embryos in the five regions above, the haplotype linkage between the *HBA* and the peripheral regions was established to determine the genotypes of all embryos.

Long-range PCR and PacBio sequencing

A minimum of 10 ng of genomic DNA and 100 ng of MDA products (Fig. 1B) were co-amplified in a single multiplex long-range PCR (LR-PCR) reaction using both primer sets (mentioned above), following a two-step amplification method (Fig. 1C). Multiplex LR-PCRs were performed in 50-μL reactions containing 10 to 100 ng of genomic DNA or MDA products, 1 × PCR KOD FX Neo buffer, 0.4 mM of each dNTP, 1 μM of primer mixture, and 1 μL of KOD FX Neo (Toyobo, Osaka, Japan). PCR program for optimal fragment amplification was 94 °C for 2 min (1 cycle), 98 °C for 10 s, and 68 °C for 14 min (30 cycles) and 68 °C for 10 min (1 cycle) (Fig. 1D).

The single-molecule real-time (SMRT) dumbbell (SMRTbell) libraries were prepared and sequenced as previously described [23] (Fig. 1E,F). Briefly, the targeted amplicons (about 200 ng per sample) were subjected to a one-step end-repair, phosphorylation and ligation to ligate barcoded adaptors using T4 DNA polymerase, T4 polynucleotide kinase, and T4 DNA ligase. The fragments failed to ligation were digested by exonucleases to reserve only the dumbbell-shaped fragments. The individual pre-libraries of each sample were purified by PB beads (Pacific Biosciences, Menlo Park, CA, USA) and pooled together by equal mass. The mixed library was combined with sequencing primer and Sequel® II Polymerase 2.0 (Pacific Biosciences, Menlo Park, CA, USA) to obtain a sequencing library. The polymerase-bound SMRTbell complexes were sequenced on the Sequel® II LRS system under the circular consensus-sequencing (CCS) mode for 30 h.

PacBio date analysis and the haplotype linkage analysis

CCS reads were split by barcoded adaptors for different samples and aligned to reference genome hg38with pbmm2 (https://github.com/PacificBiosciences/pbmm2) (Fig. 1G). The quality control criterion required a minimum sequencing depth of > 100 × for aligned reads. FreeBayes1.3.4 (https://www.geneious.com/plugins/freebayes; Biomatters, Inc., San Diego, CA, USA) was used to call SNVs and indels. SVs of *HBA* were called based on HbVar databases (https://globin.bx.psu.edu/hbvar/menu.html), Ithanet databases (https://www.ithanet.eu/), and LOVD databases (https://www.lovd.nl/).

In addition to the haplotype calling of *HBA* gene, the linkage relationship analysis was also used to determine the genotypes of embryos (Fig. 1H). First, haplotypes for all samples in the five regions were determined using whatshap (https://github.com/whatshap/whatshap), based on the informative heterozygous SNPs within each fragment. Then, by comparing haplotypes between parents and embryos in the four peripheral regions, a linked haplotype profile spanning the four regions was established. In order to avoid the impact of ADO, a retrieval-based method was developed to search the haplotypes with specific SNP markers from the father or mother in all embryos of one family in the *HBA* region, and the genotypes of the selected embryos can link the haplotypes of *HBA* to the four peripheral regions. Finally, the genotypes of the remaining embryos were determined based on the haplotype linkage relationship.

Sequences were visualized by Integrative Genomics Viewer (IGV) (Fig. 1I).

Results

Validation of tlrPGT-α-thal

To evaluate the performance of this PGT-M method for α-thalassemia among clinical samples, a total of 115 samples from 10 unrelated pedigrees were enrolled, including 10 couples who had undergone routine PGT-M, as well as their corresponding embryos. They were renamed and subjected to blinded analysis using the tlrPGT-α-thal assay developed in this study. All the samples passed quality control, and the average sequencing depth was > 2000 ×. The average sequencing depth for *HBA* region, L1, L2, L3, and L4 was 2127 × (range 436 × to 5064 ×), 2116 × (range 558 × to 4664 ×), 2558 × (range 1319 × to 5003 ×), 3009 × (1060 × to 5910 ×), and 2277 × (663 × to 4112 ×), respectively. Haplotype linkage relationships were established successfully in all the families, with embryo results fully consistent with conventional control method combining NGS, Sanger sequencing, and gap-PCR (Table S1 to S11), demonstrating 100% of sensitivity and accuracy for this approach. Additionally, this method also enabled identification of the genotypes of two triploids initially detected by PGT for aneuploidy (PGT-A), such as E-1 in family 2 and E-8 in family 8 (Table S1).

Compared to conventional methods requiring multiple detection approaches, this study accurately determined embryo genotypes through integrated haplotype construction and linkage analysis in one reaction, and without the need for a proband. Furthermore, it successfully identified the *HBA* genotype of the triploid embryos detected by PGT-A, demonstrating the efficiency and diagnostic advantages of this method.

The analysis results of *HBA* genotype and haplotype linkage relationship in the represented families.

Five out of 10 families have both parents carrying $-^{SEA}$ alleles, including family 2, 4, 6, 7, and 9, highlighting the importance of PGT-M for couples with this genotype

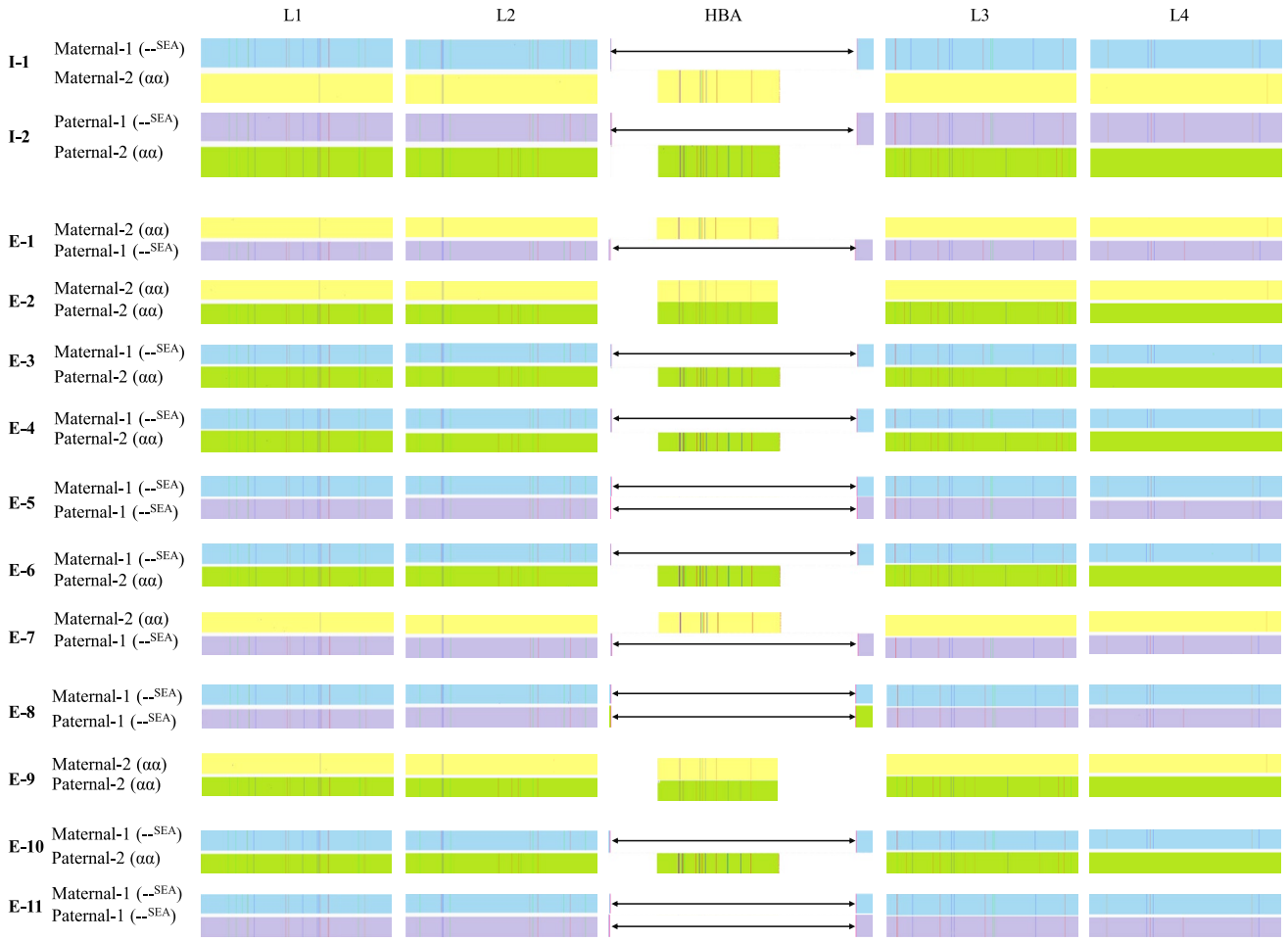

Fig. 3 IGV plots of the *HBA* gene and haplotype linkage relationships in family P06. Each colored area represents one chromosome, and each vertical line represents one SNV. The arrows show the region of the deletion. I-1 and I-2 represent the couples in this family. E-1, E-2, E-3, E-4, E-5, E-6, E-7, E-8, and E-9 represent their embryos. IGV, Integrative Genomics Viewer

combination. Family 6 including the parents and nine embryos was taken as an example to visualize the analysis results (Fig. 3; Table S7). The haplotypes and linkage relationships of four peripheral regions were constructed using informative SNPs in each fragment. A retrieval-based method was used to locate specific SNV markers from the parents in all embryos in the *HBA* region, and a haplotype linkage relationship between *HBA* and peripheral regions was established successfully. The result showed that there were six $–^{SEA}$/αα, three $–^{SEA}$/$–^{SEA}$, and two αα/αα in the embryos of this family, and the inheritance of each haplotype was clearly visible. At least five effective heterozygous SNPs were available in both L2 and L4, which can be used in distinguishing the four haplotypes from parents. This ensured that any haplotype inherited by the embryos could be unambiguously identified and tracked. In contrast, L1 and L3 only differentiated three haplotypes from the parental haplotypes, preventing the distinction between paternal and maternal $–^{SEA}$ alleles. However, by leveraging the informative SNPs from L2 and L4, the haplotype linkage analysis of the embryo on whether it inherited the mutation haplotype can still be performed accurately.

We also recruited four families where one parent carried $–^{SEA}$ deletion, and the other parent carried at least one $-\alpha^{3.7}$ or $-\alpha^{4.2}$ allele. These parents were at risk of pregnancies with Hb H disease. Family 8 consisting of both parents and nine embryos was taken as an example to illustrate the analysis results (Fig. 4; Table S9). Haplotype construction, linkage analysis, and genotype determination were performed by the customized pipeline (Fig. 1). One αα/αα, two $–^{SEA}$/αα, three $-\alpha^{3.7}$/αα, two $–^{SEA}$/$-\alpha^{3.7}$, and one $–^{SEA}$/$-\alpha^{3.7}$/αα were identified in the embryos of this family. Available heterozygous SNPs in L1, L3, and L4 were used to divide four parental haplotypes, and the origin of each haplotype was clearly illustrated in the embryos. In this family, L2 was unable to construct haplotypes due to the identical SNPs in two alleles of the mother, and it was impossible to recognize whether the embryo inherited the mutation haplotype. It was worth

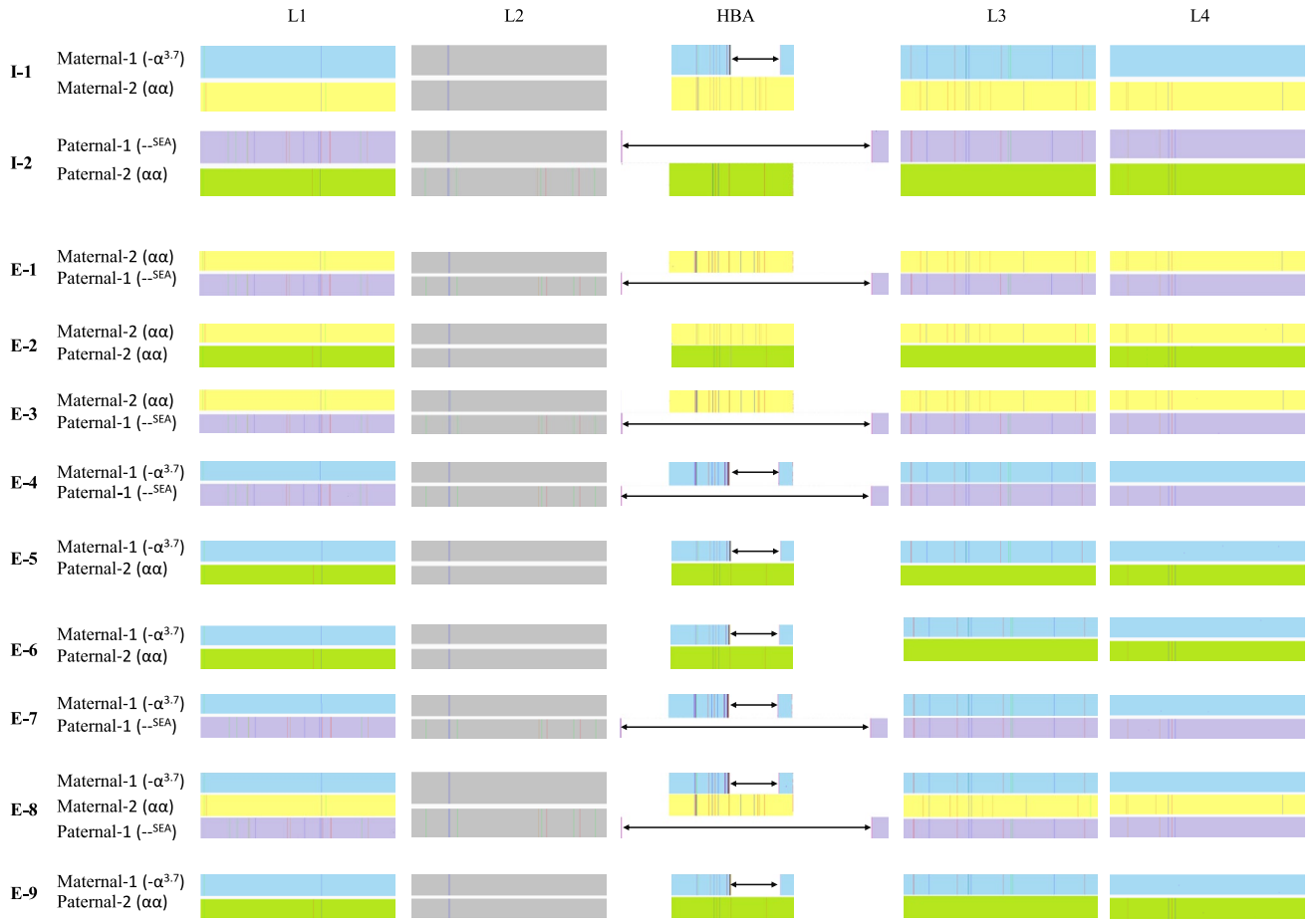

Fig. 4 IGV plots of the *HBA* gene and haplotype linkage relationships in family P08. The gray area indicates this amplicon region with no available SNPs to constructed haplotype. Each colored area (blue, yellow, purple, and green) represents one chromosome, and each vertical line represents one SNV. The arrows show the region of the deletion. I-1 and I-2 represent the couples in this family. E-1, E-2, E-3, E-4, E-5, E-6, E-7, E-8, and E-9 represent their embryos. IGV, Integrative Genomics Viewer

mentioning that there was a triploid ($–^{SEA}/\text{-}\alpha^{3.7}/\alpha\alpha$) in this family, which was consistent with PGT-A (Table S1). E-8 inherited both the $\text{-}\alpha^{3.7}$ and the αα maternally and the $–^{SEA}$ paternally.

Moreover, our study enrolled one family in which one parent was a $–^{SEA}$ deletion carrier and one parent was an $\alpha^{CS}\alpha$ carrier, which demonstrated the ability of this study to detect the combination of $–^{SEA}$ and SNVs. The results in family 5 including the parents and nine embryos illustrated the haplotypes and linkage relationships visualized by the IGV display (Figure S1).

Discussion

PGT-M is the most effective method to prevent transmission of monogenic diseases from parents to children with less invasion compared to prenatal diagnosis. Therefore, researchers have devoted to optimize PGT-M for decades.

The application of the PGT-M remains challenging because ADO occurred randomly through a necessary WGA step for the very limited genetic material available from embryos. The current standard for PGT-M was simultaneous haplotype linkage analysis [10, 11]. Therefore, establishing the haplotype linkage efficiently, accurately, and quickly between causative gene regions and peripheral regions was the key to the success of any PGT-M. However, currently commercialized PGT-M based on NGS and SNP arrays requires proband samples for haplotype linkage analysis, which makes it difficult to complete the detection towards some specific clinical cases. Although, there have been a few attempts to construct haplotype inheritance through embryos or gametes when lacking related relatives or, in cases of de novo mutations, only using the affected embryos as probands [15, 26], which relies on the diversity of embryos and is still a limiting factor for some families. It may be more complex in sperms and polar bodies analysis, because it required an extra biopsy and often needed multiple single sperms to

establish the haplotypes [26]. The haplotype construction methods relying on probands have obvious limitations. In this study, we established a novel method termed tlrPGT-α-thal, in which *HBA*-associated variants were analyzed and fragments in peripheral regions were informatively linked by targeted long-range PCR in one reaction and integrated haplotype construction, linkage analysis, and genotype determination of the embryo without requiring a proband. The method was evaluated with 115 retrospective clinical samples from 10 unrelated pedigrees. Haplotype linkage relationships were successfully established in all families, and the results of all embryos achieved 100% concordance with the control results.

One advantage offered by tlrPGT-α-thal was that PGT-M was no longer dependent on genetic information from the probands and family member. This advantage expands accessibility to PGT-M for α-thalassemia, benefiting a broader population of at-risk couples. Moreover, the number of SNPs within the sequencing region was critical for haplotype linkage analysis. NGS and SNP arrays often failed to detect enough SNPs for linkage analysis in specific genomic regions, such as proximal telomere of chr16, when performing PGT-M for α-thalassemia [27]. Due to the short-read length of NGS, it was difficult to find continuous SNP sites, and only limited SNPs can be selected to haplotype analysis. As a result, the utilization efficiency of SNPs was greatly reduced, even as low as 0.6% [28]. However, linkage analysis based on LRS was no longer limited to the strict SNP usage rules; any heterozygous SNP can be used to construct haplotypes. In this study, target regions within 1 Mb upstream and downstream of the *HBA* locus were selected for amplification by a unique algorithm according to SNV population carrier frequencies from the gnomAD database, and the haplotypes were constructed using heterozygous SNPs detected in the amplicons. Our results demonstrated that SNPs suitable for linkage analysis are detectable in both upstream and downstream of *HBA* gene across all 10 families, with fragment utilization efficiency ranging from 50 to 100%. This represents a significant improvement over NGS and SNP array approaches.

In recent years, researchers used the LRS platform to perform PGT-M and have achieved some results, including PGT-M for α-thalassemia and some other monogenic diseases [20–22, 29]. In one study [20], one couple carried SNVs of the *BCKDHB* gene and the other couple carried SNVs of *SETX* turned to PGT-M. Whole-genome sequencing based on LRS was performed on PacBio sequencing platform used genomic DNA from these couples, and the detailed SNP profiles around the mutation gene were generated to establish haplotype analysis. In addition, embryos considered unsuitable for transfer were selected to detect $-^{SEA}$ deletion, and the haplotypes were constructed successfully by whole-genome sequencing with the MDA products [29]. But the cost of whole-genome sequencing based on LRS was a significant limiting factor for large-scale clinical usage [30]. Recently, long-range PCR combined with PacBio sequencing has been applied for PGT-M for β-thalassemia [21]. However, it took more than one reaction to complete the detection because the primers were overlapping, and obtaining results with more than one reaction is still inefficient and costly. Here, tlrPGT-α-thal performed LRS based on targeted multiplex long-range PCR with all regions performed in one reaction. This method not only ensured the accuracy of haplotype construction by taking full advantage of efficient SNPs, but also decreased the testing cost. The cost of tlrPGT-α-thal and control method (NGS + Sanger + Gap) is comparable, with approximately 40 US dollars per test. The turnaround time of the PGT-M for α-thalassemia would be 1–2 weeks, which saves time compared with the control method.

One of the limitations of tlrPGT-α-thal is the insufficient types of thalassemia variants. The LRS assay performance for a diversity of α-thalassemia variants including –THAI and indels in α-globin genes with genomic DNA samples have been validated in a few studies, but it is difficult to validate these variants with embryo samples because of the rarity of these genotypes. –SEA, -α3.7, and -α4.2 are the most common α-thalassemia variants in China, followed by Hb CS (*HBA2*:c.427 T > C), Hb QS (*HBA2*:c.377 T > C), and Hb WS (*HBA2*:c.369C > G). Other variants like –THAI and other SNVs/indels are very rare in China, and –FIL has been barely detected even in large-scale epidemic studies. Thus, we did not have any cohorts for these rare α-thalassemia variants in the present retrospective study. Future prospective studies with more cohorts might be able to enroll families with rare variants.

Another limitation of this study is the possibility that some peripheral regions might have no informative SNPs for haplotype analysis. Four regions are chosen out of assay performance and cost-effectiveness. In this study, all the embryos from 10 families obtained accurate genotyping results. However, if all four regions are not informative in the future clinical applications, more peripheral regions need to be added.

Moreover, there is no ADO in the *HBA* core region in this retrospective study with relatively small sample size. The appearance of ADO is related to the MDA kit, sample quality, sample size and genetic characteristics. If ADO occurred in the *HBA* core region in some embryos, the four peripheral regions are helpful for getting the right genotype with linkage analysis.

Conclusions

In this study, we developed tlrPGT-α-thal, a novel method that simultaneously analyze both *HBA*-associated variants and informative linked fragments in peripheral regions by

targeted long-range PCR in one reaction. This integrated approach enables haplotype construction, linkage analysis, and embryo genotype determination without requiring a proband. Apart from the high accuracy, tlrPGT-α-thal offered some additional advantages. First, PGT-M was performed independently of genetic information from probands and family members. Second, more information could be used by haplotype analysis through the combination of multiple SNPs. With further breakthroughs in read length performance and reduced cost of LRS, this methodology is predicted to have great clinical application to help more patients who need PGT-M for α-thalassemia. The fundamental methodology could also be potentially applied to other monogenic diseases.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10815-025-03552-z.

Acknowledgements The authors would like to thank the patients for their participation in this study.

Author contribution Conceptualization, RL, AM, ZL; sequencing and data analysis, QS, TX, JL; sampling, NL, CX; data curation, YH, LZ, JL; writing and editing, QS, TX, AM, RL. All authors have read and approved the final manuscript.

Funding This work was supported by Guangxi Key Research and Development Program (Guike AB25069053), Guangxi Science and Technology Major Project (Guike AA22096032), Guangxi Science and Technology Planning Project (Guike AB22035080), Nanning Scientific Research and Technology Development Plan (20223050).

Data availability 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 author.

Declarations

Ethical approval This study was approved by the Ethics Committee of the Nanning Second People's Hospital (Y2024016). All procedures were performed in accordance with the Declaration of Helsinki and international and national guidelines for human studies. We obtained informed consent from all the couples before this study.

Conflict of interest The authors declare no competing interests.

References

1. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. Bull World Health Organ. 2008;86:480–7.
2. Williams TN, Weatherall DJ. World distribution, population genetics, and health burden of the hemoglobinopathies. Cold Spring Harb Perspect Med. 2012;2:a011692–a011692.
3. Taher AT, Weatherall DJ, Cappellini MD. Thalassaemia Lancet. 2018;391:155–67.
4. Farashi S, Harteveld CL. Molecular basis of α-thalassemia. Blood Cells Mol Dis. 2018;70:43–53.
5. Harteveld CL, Higgs DR. α-thalassaemia. Orphanet J Rare Dis. 2010;5:13.
6. Rycke MD, Berckmoes V. Preimplantation genetic testing for monogenic disorders. Genes. 2020;11:871.
7. Carvalho F, Moutou C, Dimitriadou E, Dreesen J, Giménez C, Goossens V, et al. ESHRE PGT Consortium good practice recommendations for the detection of monogenic disorders. Human Reproduct Open. 2020;2020:1–18.
8. Wen L, Tang F. Recent advances in single-cell sequencing technologies. Prec Clinic Med. 2022;5:pbac002.
9. Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-Cell Whole-genome amplification and sequencing: methodology and applications. Annu Rev Genomics Hum Genet. 2015;16:79–102.
10. Professional Committee on Reproductive Medicine CMDA and The Society of Medical Geneticists CMDA Chinese experts' consensus on preimplantation genetic testing for monogenic disorders. Chin J Reproduct Contracept 2021;41:477–485.
11. Natesan SA, Bladon AJ, Coskun S, Qubbaj W, Prates R, Munne S, et al. Genome-wide karyomapping accurately identifies the inheritance of single-gene defects in human preimplantation embryos in vitro. Genet Med. 2014;16:838–45.
12. Kruglyak L, Nickerson DA. Variation is the spice of life. Nature Genet. 2001;27:234–6.
13. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res. 2009;37:4181–93.
14. Doroftei B, Ilie O-D, Anton N, Armeanu T, Ilea C. A mini-review regarding the clinical outcomes of in vitro fertilization (IVF) following pre-implantation genetic testing (PGT)-next generation sequencing (NGS) approach. Diagnostics. 2022;12:1911.
15. Chen D, Shen X, Wu C, Xu Y, Ding C, Zhang G, et al. Eleven healthy live births: a result of simultaneous preimplantation genetic testing of α- and β-double thalassemia and aneuploidy screening. J Assist Reprod Genet. 2020;37:549–57.
16. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: an overview. Hum Immunol. 2021;82:801–11.
17. Group TISMW. A map of human genome sequence variation containing 142 million single nucleotide polymorphisms. Nature 2001;409:928–933.
18. Huang P, Lan Y, Zhou H, Lin L, Shu J, Wang C, et al. Comprehensive application of multiple molecular diagnostic techniques in pre-implantation genetic testing for monogenic. Mol Genet Genomic Med. 2023;12:e2293.
19. Van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The third revolution in sequencing technology. Trends Genet. 2018;34(9):666–81.
20. Cheng Y, Yu Q, Ma M, Wang H, Tian S, Zhang W, et al. Variant haplophasing by long-read sequencing: a new approach to preimplantation genetic testing workups. Fertil Steril. 2021;116:774–83.
21. Haitao Wu, Chen D, Zhao Q, Shen X, Liao Y, Li P, et al. Long-read sequencing on the SMRT platform enables efficient

haplotype linkage analysis in preimplantation genetic testing for β-thalassemia. J Assist Reprod Genet. 2022;39:739–46.
22. Peng C, Chen H, Ren J, Zhou F, Li Y, Keqie Y, et al. A long-read sequencing and SNP haplotype-based novel preimplantation genetic testing method for female ADPKD patient with *de novo PKD1* mutation. BMC Genomics. 2023;24:521.
23. Liang Q, Gu W, Chen P, Li Y, Liu Y, Tian M, et al. A more universal approach to Comprehensive Analysis of Thalassemia Alleles (CATSA). J Mol Diagn. 2021;23:1195–204.
24. Huang R, Liu Y, Xu J, Lin D, Mao A, Yang L, et al. Back-to-back comparison of third-generation sequencing and next-generation sequencing in carrier screening of thalassemia. Arch Pathol Lab Med. 2024;148:797–804.
25. Liu Q, Chen Q, Zhang Z, Peng S, Liu J, Pang J, et al. Identification of rare thalassemia variants using third-generation sequencing. Front Genet. 2023;13:1076035.
26. Wang Y, Zhai F, Guan S, Yan Z, Zhu X, Kuo Y, et al. A comprehensive PGT-M strategy for ADPKD patients with *de novo PKD1* mutations using affected embryo or gametes as proband. J Assist Reprod Genet. 2021;38:2425–34.
27. LI Y, Huang J, JI X, Feri J, Yu T and Huang Y. Comparative study of NGS panel and whole genome SNP microarray in pre-implantation genetic testing of thalassemia. Journal of Molecular Diagnostics 2022;14:1832–1840.
28. Deng H, LU S and Fang Y. A primer combination and method for detecting alpha-thalassemia gene mutation in human embryos. 2018;28
29. Liu S, Wang H, Leigh D, Cram DS, Wang L, Yao Y. Third-generation sequencing: any future opportunities for PGT? J Assist Reprod Genet. 2020;38:357–64.
30. Xiao T, Zhou W. The third generation sequencing: the advanced approach to genetic diseases. Translational Pediatrics. 2020;9:163–73.