RESEARCH

Development of preimplantation genetic testing for monogenic reference materials using next-generation sequencing

Weihua Zhao[1†], Yanyan Song[2†], Chuanfeng Huang[3†], Shan Xu[2], Qi Luo[1], Runsi Yao[1], Nan Sun[3], Bo Liang[4,5], Jia Fei[6], Fangfang Gao[7], Jie Huang[3*] and Shoufang Qu[3*]

Abstract

Objective Preimplantation genetic testing for monogenic disorders (PGT-M) has been used for over 20 years to detect many serious genetic conditions. However, there is still a lack of reference materials (RMs) to validate the test performance during the development and quality control of PGT-M.

Method Sixteen thalassemia cell lines from four thalassemia families were selected to establish the RMs. Each family consisted of parents with heterozygous mutations for α- and/or β-thalassemia and two children, at least one of whom carried a homozygous thalassemia mutation (proband). The RM panel consisted of 12 DNA samples (parents and probands in 4 families) and 4 simulated embryos (cell lines constructed from blood samples from the four nonproband children). Four accredited genetics laboratories that offer verification of thalassemia samples were invited to evaluate the performance of the RM panel. Furthermore, the stability of the RMs was determined by testing after freeze–thaw cycles and long-term storage.

Results PGT-M reference materials containing 12 genome DNA (gDNA) reference materials and 4 simulated embryo reference materials for thalassemia testing were successfully established. Next-generation sequencing was performed on the samples. The genotypes and haplotypes of all 16 PGT-M reference materials were concordant across the four labs, which used various testing workflows. These well-characterized PGT-M reference materials retained their stability even after 3 years of storage.

Conclusion The establishment of PGT-M reference materials for thalassemia will help with the standardization and accuracy of PGT-M in clinical use.

Keywords Preimplantation genetic testing, Reference materials, Next-generation sequencing, Thalassemia, Monogenic disorders

---

†Weihua Zhao, Yanyan Song and Chuanfeng Huang contributed equally to this work.

*Correspondence:
Jie Huang
jhuang5522@126.com
Shoufang Qu
qsfnifdc@126.com
Full list of author information is available at the end of the article

Introduction

Thalassemia, one of the most prevalent monogenic diseases worldwide, is a hereditary hemolytic anemia caused by the reduction or absence of α- or β-globin chain synthesis [1]. Every year, at least 330,000 babies with thalassemia are born worldwide, and the carrier rate of the thalassemia gene in pregnant women is as high as 7% [2]. In southern China, the prevalence of α-thalassemia and β-thalassemia are 8.53% and 2.54%, respectively [3, 4]. Thalassemia is a complex autosomal recessive genetic disorder; α-thalassemia is mainly caused by deletions, while β-thalassemia frequently results from single-nucleotide variations (SNVs) or oligonucleotide insertion/deletions (Indels). Most fetuses with severe α-thalassemia die in utero or soon after birth, and although patients with severe forms of β-thalassemia can survive, they require regular blood transfusions and may gradually accumulate too much iron, leading to significant economic and mental burdens on their family [3].

Preimplantation genetic testing for monogenic disorders (PGT-M) has become an increasingly popular method to reduce the risk of couples having children with monogenic conditions. In vitro fertilization (IVF) combined with PGT-M is widely used to identify whether embryos carry a genetic disorder before implantation and pregnancy [5]. Since 1998, an increasing number of couples have delivered thalassemia-free babies through PGT-M [6, 7]. Several approaches have been successfully used to test for thalassemia in PGT-M, including polymerase chain reaction (PCR)-based detection [8–10] and short tandem repeat (STR) haplotyping [11, 12]. However, these approaches are labor intensive and are restricted by the adverse influence of allele dropouts (ADOs) and limited STR loci. Recently, next-generation sequencing (NGS) combined with single-nucleotide polymorphism (SNP)-based haplotyping has shown favorable results when used in PGT-M [13, 14]. The use of NGS reduces ADOs, improving the accuracy of monogenic disease detection. Although PGT-M has been widely used over the past 20 years, there is a lack of pedigree-based reference materials that can be used to evaluate the accuracy and efficiency of the technology.

Genomic reference materials (RMs) are urgently needed for the validation of PGT-M development and quality control of its performance. With the continuous development of genetic tests used in clinical and public health practice, there has been an overall trend of increased detection accuracy compared with earlier PGT-M techniques. With the standardization and rapid evolution of PGT-M, it has been recommended that DNA RMs be used in clinical genetic testing laboratories, especially for complex recessive diseases [15]. To further improve PGT-M, genomic DNA RMs used in clinical practice should be thoroughly characterized using multiple commonly used PGT-M detection methods and should contain a complete pedigree for haplotyping.

In this study, samples from four thalassemia families were used to establish PGT-M RMs, and these RMs were validated using an NGS platform combined with SNP-based haplotyping. The validation was carried out in four research institutions, and all of the results support the feasibility of these reference materials for use in PGT-M testing.

Materials and methods

Selection and collection of samples

This study was approved by MGI's institutional review board on bioethics and the Ethics Committee of The Second People's Hospital of Shenzhen. Sixteen participants from four thalassemia families were recruited (Table 1). The parents in each family carry a heterozygous (Het) α- and/or β-thalassemia mutation, and each couple has two children. In each family, the child with a homozygous (Homo) mutation of a homotypic thalassemia gene was defined as the proband, and the sample from their sibling was considered a simulated embryo. All family members with α- and/or β-thalassemia mutant genotypes signed an informed consent form and agreed to donate whole blood for B lymphoblast cell line development.

Cell line generation and sample processing

Sixteen thalassemia cell lines, one from each of the four families, were established as previously described [16, 17]. The twelve cell lines constructed from the parents and probands in each family were then subjected to genomic DNA (gDNA) extraction. The extraction was conducted with an MGIEasy Magnetic Beads Genomic DNA Extraction Kit (MGI Tech Co., Ltd. Item No. 1000010524) according to the manufacturer's instructions. Extracted gDNA was diluted to 20 ng/μL and packaged in 15 μL/tube. To mimic the embryos tested in PGT-M, the cell lines constructed with blood from the four nonproband children were used as simulated embryos. These four cell lines were sorted by fluorescence activated cell sorting (FACS) (BD FACSAria™ II cell sorter, BD Biosciences) and cryopreserved for genetic testing without DNA extraction. Each sorted cellular sample contained 3–5 cells with 4 μL 1×PBS in the tube. The thalassemia PGT-M RMs, including gDNA samples and simulated embryo samples, were stored at -80 °C and shipped on dry ice when necessary.

Verification protocol

Four accredited genetics laboratories that offer verification of thalassemia samples were invited to participate in the study. Each laboratory received twelve gDNA samples extracted from thalassemia cell lines and four cellular samples sorted from the four simulated embryo

Table 1 Information on thalassemia families and established cell lines

Families	Relationship	Genotypes	HGVS nomenclature	Clinical severity	Cell line no.	Package	Sample type
Family 1	Husband 1	$-\alpha^{3.7}/\alpha\alpha$, Codons 41/42(-TTCT)/$\beta^N$	NG_000006.1: g.34164_37967del, *HBB*: c.126_129delCTTT	Thalassemia trait	CNGB030011	20 ng/µL, 15 µL/tube	gDNA
	Wife 1	Gγ(Aγδβ)^0/β^N	NG_000007.3: g.48795_127698del	Thalassemia trait	CNGB030012	20 ng/µL, 15 µL/tube	gDNA
	Offspring 1–1 (Proband)	Codons 41/42(-TTCT)/Gγ(Aγδβ)^0	*HBB*: c.126_129delCTTT, NG_000007.3: g.48795_127698del	Thalassemia major	CNGB030014	20 ng/µL, 15 µL/tube	gDNA
	Offspring 1–2 (Simulated embryo)	$-\alpha^{3.7}/\alpha\alpha$, Gγ(Aγδβ)^0/$\beta^N$	NG_000006.1: g.34164_37967del, NG_000007.3: g.48795_127698del	Thalassemia trait	CNGB030013	3–5 cells/tube	Cells
Family 2	Husband 2	$-\!-^{SEA}/\alpha\alpha$	NG_000006.1: g.26264_45564del	Thalassemia trait	CNGB030015	20 ng/µL, 15 µL/tube	gDNA
	Wife 2	$\alpha^{CS}\alpha/\alpha\alpha$	*HBA2*: c.427T > C	Thalassemia trait	CNGB030016	20 ng/µL, 15 µL/tube	gDNA
	Offspring 2–1 (Proband)	$-\!-^{SEA}/\alpha^{CS}\alpha$	NG_000006.1: g.26264_45564del, *HBA2*: c.427T > C	Thalassemia intermedia	CNGB030018	20 ng/µL, 15 µL/tube	gDNA
	Offspring 2–2 (Simulated embryo)	$\alpha^{CS}\alpha/\alpha\alpha$	*HBA2*: c.427T > C	Thalassemia trait	CNGB030017	3–5 cells/tube	Cells
Family 3	Husband 3	Gγ(Aγδβ)^0/β^N	NG_000007.3: g.48795_127698del	Thalassemia trait	CNGB030019	20 ng/µL, 15 µL/tube	gDNA
	Wife 3	Codons 41/42(-TTCT)/β^N	*HBB*: c.126_129delCTTT	Thalassemia trait	CNGB030020	20 ng/µL, 15 µL/tube	gDNA
	Offspring 3–1 (Proband)	Codons 41/42(-TTCT)/Gγ(Aγδβ)^0	*HBB*: c.126_129delCTTT, NG_000007.3: g.48795_127698del	Thalassemia major	CNGB030328	20 ng/µL, 15 µL/tube	gDNA
	Offspring 3–2 (Simulated embryo)	Gγ(Aγδβ)^0/β^N	NG_000007.3: g.48795_127698del	Thalassemia trait	CNGB030021	3–5 cells/tube	Cells
Family 4	Husband 4	IVS-II-654 (C > T)/β^N	*HBB*: c.316-197 C > T	Thalassemia trait	CNGB030008	20 ng/µL, 15 µL/tube	gDNA
	Wife 4	$-\!-^{SEA}/\alpha\alpha$, IVS-II-654 (C > T)/β^N	NG_000006.1: g.26264_45564del, *HBB*: c.316-197 C > T	Thalassemia trait	CNGB030009	20 ng/µL, 15 µL/tube	gDNA
	Offspring 4–1 (Proband)	$-\!-^{SEA}/\alpha\alpha$, IVS-II-654 (C > T)/IVS-II-654 (C > T)	NG_000006.1: g.26264_45564del, *HBB*: c.316-197 C > T	Thalassemia intermedia	CNGB030004	20 ng/µL, 15 µL/tube	gDNA
	Offspring 4–2 (Simulated embryo)	IVS-II-654 (C > T)/β^N	*HBB*: c.316-197 C > T	Thalassemia trait	CNGB030010	3–5 cells/tube	Cells

cell lines. gDNA samples were directly subjected to targeted amplification and high-throughput sequencing for genotype identification, whereas the cellular samples had to undergo whole-genome amplification (WGA) to obtain enough DNA for genotype verification. The four laboratories genotyped each sample using their standard protocols; the methods, reagent kits and sequencing platforms are summarized in Table 2. In general, copy number variations (CNVs) and deletions (del) of hemoglobin genes were detected by both Gap-PCR and NGS, while point mutations (single-nucleotide variants, SNVs) were detected better by NGS. To further ensure the detection accuracy of simulated embryo RMs, validations by family-based linkage analysis were also used. SNP-based haplotyping was used to identify alleles linked to the mutations and decrease the ADO misdetection rate.

Laboratory 1 (Lab 1)

Target region capture library preparation and high-throughput sequencing were used to genotype the samples for thalassemia in Lab 1 (Clinical Laboratory of BGI, BGI-Shenzhen, China). Multiple displacement amplification (MDA)-based WGA was performed on the four cellular samples using the MGIEasy Single Cell Whole Genome Amplification Kit (MGI Tech Co., Ltd. Item No.: 1000007744). NGS libraries of the 12 gDNA samples and 4 WGA products were constructed using the MGICare Monogenic Disease Haplotype Phasing Kit (MGI Tech Co., Ltd. Item No.: 1000005287). The libraries were sequenced on the DNBSEQ-G400RS platform (MGI Tech Co., Ltd. China) using a DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE100, MGI Tech Co., Ltd. Item No.: 1000012554). Genotype verification

Table 2 Characterization methods in different laboratories

Laboratories	Sample type	Methods and reagents	Sequencing platform
Lab 1	Cells	WGA: MGIEasy Single Cell Whole Genome Amplification Kit (MGI)	DNBSEQ-G400RS (MGI Tech Co., Ltd. China)
	WGA products & gDNA	NGS library: MGICare Monogenic Disease Haplotype Phasing Kit (MGI); NGS	
Lab 2	Cells	WGA: REPLI-g Single Cell Kit (QIAGEN)	MiSeq (Illumina, USA)
	WGA products & gDNA	Gap-PCR; Sanger sequencing; NGS library: NEXTflex™ DNA-Seq Kit (Bioo Scientific); NGS	
Lab 3	Cells	WGA: PicoPLEX® WGA Kit (TAKARA)	Ion Torrent DA8600 (Thermo Fisher Scientific, USA)
	WGA products & gDNA	Gap-PCR, NGS library: Preimplantation Genetic Diagnosis Kit for Thalassemia (BASECARE); NGS	
Lab 4	Cells	WGA: REPLI-g Advanced DNA Single Cell Kit (QIAGEN);	NextSeq 550 (Illumina, USA)
	WGA products & gDNA	Gap-PCR; NGS library: *HBA*&*HBB* Pedigree Verification Kit (YIKON); NGS	

and subsequent genetic linkage analysis were performed. Genetic linkage analysis was performed on the alpha-globin (*HBA)* and beta-globin (*HBB)* genes, and the haplotypes were determined according to the detected number of SNPs near the mutation locus (+/− 1 Mb).

Laboratory 2 (Lab 2)
Gap-PCR combined with NGS was used to genotype the samples in Lab 2 (Jabrehoo, Pekin, China). An MDA-based WGA protocol was carried out on the simulated embryo samples using a REPLI-g Single Cell Kit (QIAGEN. Cat. No./ID: 150343). Multiplex PCR primers (Table S1) were used to amplify pathogenic genes from all gDNA samples and WGA products. The libraries were then prepared using a NEXTflex™ DNA-Seq Kit (Bioo Scientific, USA) and sequenced on the MiSeq platform (Illumina, USA) using a MiSeqDx Universal Kit v3 SBS (Illumina, Part Number: 15072805). Furthermore, multiplex PCR (PCR primers are shown in Table S2) and Sanger sequencing were used to ensure accurate detection of single-nucleotide variants (SNVs). To detect deletion mutations, Gap-PCR and gel electrophoresis were performed to further confirm the mutant types. NGS data were analyzed using a script written in the lab (02_STA_TRIM_FASTq.pl). Genetic linkage analysis was performed on the *HBA* and *HBB* genes, and haplotypes were defined as having at least 3 valid SNPs closely linked upstream or downstream of the target gene.

Laboratory 3 (Lab 3)
Gap-PCR combined with NGS was used to genotype the samples for thalassemia in Lab 3 (Basecare Medical, Suzhou, China). WGA was conducted on cell samples using a PicoPLEX® WGA Kit (TAKARA, Cat. # R30050). The NGS libraries of the gDNA samples and the WGA products were then prepared using a Preimplantation Genetic Diagnosis Kit for Thalassemia (Ion Torrent) (BASECARE, Suzhou, China). Libraries were sequenced on the Ion Torrent DA8600 platform (Thermo Fisher Scientific, USA). Bioinformatic analyses were performed to directly identify the genotype of each sample and construct the haplotypes to further reduce the errors caused by ADOs.

Laboratory 4 (Lab 4)
Lab 4 (Yikon Genomics, China) used different protocols to verify the SNV, CNV and deletion mutant genotypes. WGA was performed on the four cell samples using a REPLI-g Advanced DNA Single Cell Kit (QIAGEN. Cat. No./ID: 150365). First, Gap-PCR (primers for Gap-PCR are shown in Table S3) and gel electrophoresis were used to identify deletion mutations causing α- and β-thalassemia. Gap-PCR was carried out with the following reaction mix: 10 μL 2 × GoldStar Best Master Mix, 1 μL forward and reverse primer mix (5 μM each), 1 μL DNA template, and 8 μL H_2O. The following program was used for amplification: 95 °C for 10 min; 35 cycles of 97 °C for 45 s, 65 °C for 90 s, 72 °C for 180 s; 72 °C for 5 min; and hold at 12 °C. Amplification products were verified on a 1% agarose gel. Then, the *HBA* and *HBB* genes were amplified by multiplex PCR (Primer *HBA*&*HBB* Mix Pool Kit, YIKON) separately and sequenced on a NextSeq 550 System (Illumina, USA). The NGS results were used to directly determine genotypes and to identify haplotypes based on the SNPs near the mutation locus.

Stability tests
The gDNA samples extracted from cell lines were divided into five groups (each group contained 12 gDNA samples) for storage and freeze−thaw stability tests. Four groups of samples were kept at a constant −80 °C for 3 months, 6 months, 12 months and 3 years. The final group of samples was cycled six times between storage at room temperature (RT) and at -80 °C. The integrity of the DNA was then reevaluated by 2% agarose gel electrophoresis. The hemoglobin genotypes of these treated gDNA samples were reanalyzed by Lab 1.

For stability analysis of the four cellular samples, four tubes of each sample were randomly selected, and the resulting 4×4 tubes were divided into four groups, each containing the four simulated embryo samples. The four groups were kept at -80 °C for 3 months, 6 months, 12

months and 3 years. To quickly evaluate the storage stability of the cell samples, an identification system based on MDA rather than sequencing was used. First, the cell samples were amplified using an MGIEasy Single Cell Whole Genome Amplification Kit, and the amplification efficiency was confirmed by multiplex PCR (primers are shown in Table S4) of eight housekeeping genes. The amplicons of the housekeeping genes were then reevaluated by 2% agarose gel electrophoresis. The hemoglobin genotypes and genetic linkage analyses of these treated cellular samples were reanalyzed by Lab 1.

Results

Establishment of cell lines

Sixteen individuals from four thalassemia families containing the most common hemoglobin genotypes were recruited to construct national PGT-M RMs (Table 1). B lymphoblastoid cell lines from all sixteen participants were successfully established. The incorporated hemoglobin genotypes were as follows: $--^{SEA}/\alpha\alpha$, $\alpha^{CS}\alpha/\alpha\alpha$, $-\alpha^{3.7}/\alpha\alpha$, $--^{SEA}/\alpha^{CS}\alpha$, Codons 41/42 (-TTCT)/$\beta^N$, IVS-II-654 (C>T)/$\beta^N$, G$\gamma$(A$\gamma\delta\beta$)^0/$\beta^N$, Codons41/42(-TTCT)/G$\gamma$(A$\gamma\delta\beta$)^0, and IVS-II-654(C>T)/ IVS-II-654 (C>T).

Characterization of thalassemia cell line samples

The hemoglobin genotypes of the 12 gDNA samples and the 4 simulated embryo samples were independently characterized by four laboratories with different genetic testing methods. The test results are summarized in Table 3. The four laboratories showed 100% concordance across all samples, and the hemoglobin genotypes were perfectly consistent with expectations. The sixteen samples, including 12 gDNA samples and 4 simulated embryo samples, were well characterized as candidate thalassemia PGT-M RMs.

SNP-based haplotyping

To reduce the diagnostic errors caused by ADOs and ensure the accuracy of detecting monogenic diseases in embryos, it is vital to construct parental haplotypes for PGT-M. The four laboratories conducted genetic linkage analysis based on the NGS results to further reduce false-positive and false-negative rates caused by direct analysis of pathogenic sites. In the family-based genetic linkage analysis, the haplotypes of the husband were marked as F0 (mutant) and F1 (normal), and the haplotypes of the wife were marked as M0 (mutant) and M1 (normal). The haplotypes of the proband (Offspring N-1) were defined as M0/F0 (representing the mutant genetic haplotype from the mother and father, respectively). Haplotype phasing results are summarized in Table 4. The haplotype phasing results obtained from the four laboratories with different instruments and validation methods are completely consistent. This demonstrates the credibility of the haplotype results obtained from the constructed RMs. The haplotypes of these pedigree-based RMs are beneficial for evaluating the performance of PGT-M testing for thalassemia and help to avoid biases caused by ADOs. For more details about the haplotyping results in different labs, please see Tables S5-S11 and Figures S1-S4. All the haplotype results were in good agreement with the genotype results.

Stability tests

To evaluate the stability of our PGT-M RMs, their freeze−thaw tolerance and long-term storage capacity were tested. The 12 gDNA samples showed no obvious

Table 3 Genotype verification of thalassemia PGT-M RMs

Family	No.	Family relationship	Genotypes	Lab 1	Lab 2	Lab 3	Lab 4
Family 1	1	Husband 1	$-\alpha^{3.7}/\alpha\alpha$, Codons 41/42(-TTCT)/$\beta^N$	√	√	√	√
	2	Wife 1	Gγ(A$\gamma\delta\beta$)^0/β^N	√	√	√	√
	3	Offspring 1–1 (Proband)	Codons 41/42(-TTCT)/Gγ(A$\gamma\delta\beta$)^0	√	√	√	√
	4	Offspring 1–2 (Simulated embryo)	$-\alpha^{3.7}/\alpha\alpha$, G$\gamma$(A$\gamma\delta\beta$)^0/$\beta^N$	√	√	√	√
Family 2	5	Husband 2	$--^{SEA}/\alpha\alpha$	√	√	√	√
	6	Wife 2	$\alpha^{CS}\alpha/\alpha\alpha$	√	√	√	√
	7	Offspring 2–1 (Proband)	$--^{SEA}/\alpha^{CS}\alpha$	√	√	√	√
	8	Offspring 2–2 (Simulated embryo)	$\alpha^{CS}\alpha/\alpha\alpha$	√	√	√	√
Family 3	9	Husband 3	Gγ(A$\gamma\delta\beta$)^0/β^N	√	√	√	√
	10	Wife 3	Codons 41/42(-TTCT)/β^N	√	√	√	√
	11	Offspring 3–1 (Proband)	Codons 41/42(-TTCT)/Gγ(A$\gamma\delta\beta$)^0	√	√	√	√
	12	Offspring 3–2 (Simulated embryo)	Gγ(A$\gamma\delta\beta$)^0/β^N	√	√	√	√
Family 4	13	Husband 4	IVS-II-654 (C>T)/β^N	√	√	√	√
	14	Wife 4	$--^{SEA}/\alpha\alpha$, IVS-II-654 (C>T)/β^N	√	√	√	√
	15	Offspring 4–1 (Proband)	$--^{SEA}/\alpha\alpha$, IVS-II-654 (C>T)/IVS-II-654 (C>T)	√	√	√	√
	16	Offspring 4–2 (Simulated embryo)	IVS-II-654 (C>T)/β^N	√	√	√	√

Zhao *et al. BMC Medical Genomics*   (2024) 17:33

Page 6 of 8

Table 4 Haplotypes of four families based on genetic linkage analyses

Family	Family relationship	Haplotype of HBA	Genotypes of HBA	Haplotype of HBB	Genotypes of HBB
Family 1	Husband 1	/	/	F0/F1	Codons 41/42(-TTCT)/β^N
	Wife 1	/	/	M0/M1	$G\gamma(A\gamma\delta\beta)^0$/$\beta^N$
	Offspring 1–1 (Proband)	/	/	M0/F0	Codons 41/42(-TTCT)/$G\gamma(A\gamma\delta\beta)^0$
	Offspring 1–2 (Simulated embryo)	/	/	M0/F1	$G\gamma(A\gamma\delta\beta)^0$/$\beta^N$
Family 2	Husband 2	F0/F1	$--^{SEA}$/$\alpha\alpha$	/	/
	Wife 2	M0/M1	$\alpha\alpha^{CS}$/$\alpha\alpha$	/	/
	Offspring 2–1 (Proband)	M0/F0	$--^{SEA}$/$\alpha\alpha^{CS}$	/	/
	Offspring 2–2 (Simulated embryo)	M0/F1	$\alpha\alpha^{CS}$/$\alpha\alpha$	/	/
Family 3	Husband 3	/	/	F0/F1	$G\gamma(A\gamma\delta\beta)^0$/$\beta^N$
	Wife 3	/	/	M0/M1	Codons 41/42(-TTCT)/β^N
	Offspring 3–1 (Proband)	/	/	M0/F0	$G\gamma(A\gamma\delta\beta)^0$/ Codons 41/42(-TTCT)
	Offspring 3–2 (Simulated embryo)	/	/	M1/F0	$G\gamma(A\gamma\delta\beta)^0$/$\beta^N$
Family 4	Husband 4	/	/	F0/F1	IVS-II-654 (C > T)/β^N
	Wife 4	/	/	M0/M1	IVS-II-654 (C > T)/β^N
	Offspring 4–1 (Proband)	/	/	M0/F0	IVS-II-654 (C > T)/IVS-II-654 (C > T)
	Offspring 4–2 (Simulated embryo)	/	/	M1/F0	IVS-II-654 (C > T)/β^N

degradation after six repeated freeze−thaw cycles (Figure S5A). The cell samples from the four simulated embryo cell lines were subjected to MDA-based WGA, and the coverage of MDA was characterized by amplifying housekeeping genes. The seven or eight housekeeping gene amplicon bands in the gels (Figure S5B) indicate the feasibility of the quick quality control protocol.

Moreover, the PGT-M RMs showed good stability after cryopreservation for 3 months, 6 months, 12 months and 3 years. There was no obvious degradation of gDNA samples even after 3 years of storage (Figure S6A), and the cell samples were also correctly amplified after 3 years of cryopreservation (Figure S6B). The genotypes and haplotypes of these treated gDNA samples and cell samples were consistent with the expected results (data not shown).

Discussion

In this study, we demonstrated the establishment, characterization, and stability of four thalassemia pedigree PGT-M RMs, which will help to ensure the accuracy of clinical PGT-M tests for thalassemia. Sixteen B lymphoblastoid cell lines derived from four thalassemia families were successfully established. The mutations that cause α- and/or β-thalassemia, such as deletion mutations $--^{SEA}$/ and $-\alpha^{3.7}$/ and nondeletion mutations $\alpha^{CS}\alpha$/, IVS-II-654 (C > T), 41/42(-TTCT) and $G\gamma(A\gamma\delta\beta)^0$, are widely distributed in South China [18, 19]. To faithfully simulate real sampling in PGT-M, four cell lines derived from the nonproband offspring were sorted into 3–5 cells/tube so that these cellular samples could only be analyzed using WGA. Four accredited labs correctly identified the genotypes and haplotypes of all the RMs, and there were no discordant results among the four laboratories, which each used their own workflow. Furthermore, the twelve

gDNA RMs exhibited good stability when subjected to freeze−thaw cycles and after long-term storage, and the quick quality control system based on MDA and multiplex PCR further demonstrated the stability of the cellular RMs. It should be noted that prior to our study, there were no characterized PGT-M RMs for clinical thalassemia.

Thalassemia occurs at a high frequency in tropical and subtropical regions, such as Africa and Southeast Asia. When couples give birth to a child with thalassemia or discover that they are carriers of thalassemia, they should use PGT-M to reduce the likelihood of having (more) children with thalassemia. Although PGT-M is an extremely effective detection method for thalassemia, its use is still under discussion among specialists worldwide. To meet the challenge to PGT-M posed by low levels of input DNA, it is common to amplify material from a few cells or even a single cell using WGA [20], such as Taq DNA polymerase-PCR-based WGA [21], MDA method-based WGA [22] or MDA combined with PCR-based WGA [23]. The general workflow for PGT-M includes WGA followed by multiplex PCR [24], SNP arrays [25] or NGS [26]. Subsequent haplotyping often relies on STR markers or SNP markers to increase the resolution of PGT-M. While WGA and subtesting techniques have been substantially improved over the past few years, the technique can still introduce errors due to amplification failure, contamination or errors in sequencing, which may have a severe impact on the reliability of the diagnostic result. Thus, PGT-M RMs should be used to minimize the error rate during the design and clinical implementation of tests.

In our study, PGT-M RMs for thalassemia were well characterized through NGS combined with SNP-based haplotyping in four laboratories. This NGS-based

protocol represents a truly generic method, and the residual risk may be lower than with conventional targeted amplification strategies [27, 28]. Furthermore, different NGS platforms were used by the four labs in this study, including the DNBSEQ-G400RS platform (MGI Tech Co., Ltd. China) with a state-of-the-art core technology called DNBSEQ (DNA nanoballs sequencing), the MiSeq or NextSeq 550 System (Illumina, USA) based on bridge PCR technology, and the Ion Torrent DA8600 platform (Thermo Fisher Scientific, USA) based on semiconductor sequencing theory. The well-characterized RMs constructed in our study can thus be used with many technologies. In Lab 2, the SNVs in α-thalassemia and the deletion mutations in β-thalassemia pathogenic variants were also validated using Sanger sequencing and GAP-PCR. In addition to the characterization of thalassemia PGT-M RMs, we established a rapid quality control strategy to evaluate the stability of simulated embryo RMs. This involved performing MDA followed by multiplex PCR amplification of eight housekeeping genes. Housekeeping genes have a relatively stable expression under normal and pathological conditions and are continuously expressed in almost all tissues at all stages of growth. This quality control strategy is thus suitable for the evaluation of WGA efficiency and cell sample stability.

Although PGT-M RMs of thalassemia were successfully established, there are still limitations in this study. Not all of the common variations [29] of thalassemia in China were incorporated in the final study, so it does not reflect the entire breadth of clinical patient situations. Nevertheless, the availability of these PGT-M RMs can help ensure the accuracy of PGT-M for thalassemia tests, which will help clinicians and benefit patients.

Conclusion

The PGT-M RMs for thalassemia developed in this work cover some of the common, rare and composite thalassemia genotypes in China. Validation by either genotyping or haplotyping in different labs produced consistent results. The well-characterized gDNA and cellular RMs described in this study may help reduce the burden of neonatal thalassemia to families and society by improving PGT-M testing. The establishment of these dedicated gDNA and cellular RMs is essential for case studies, quality control and technical verification in thalassemia testing. More details about the PGT-M RM panel for thalassemia are available on the NIFDC website (https://www.nifdc.org.cn/nifdc/, lot number: 360027−201901).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-024-01803-z.

Supplementary Material 1

Declarations

Ethics approval and consent to participate
This study was approved by MGI's institutional review board on bioethics (approval no. BGI-IRB 19124-T1) and the Ethics Committee of The Second People's Hospital of Shenzhen (approval no. KS20190418004-FS2019052003). All recruited members with α- and/or β- thalassemia mutant genotypes signed an informed consent form and agreed to donate whole blood for B lymphoblast cell line development.

Consent for publication
Not applicable.

Conflict of interest
Y.S. and S.X. are employed by BGI; B.L. is employed by Basecare Medical Device Co., Ltd.; J.F. is employed by Peking Jabrehoo Med Tech Co., Ltd.; and F.G. was employed by Yikon Genomics Company, Ltd. All authors declare no other competing interests.

Author details
[1]Department of Obstetrics, Shenzhen Second People's Hospital/the First Affiliated Hospital of Shenzhen University Health, Shenzhen, Guangdong, China
[2]BGI-Shenzhen, GuangdongShenzhen, China
[3]Division of Physical and Chemical Testing, Division of in Vitro Diagnostic Reagents, National Institutes for food and drug Control (NIFDC), Beijing, China
[4]Laboratory of Metabolic and Developmental Sciences, State Key Laboratory of Microbial Metabolism, Shanghai, China
[5]Basecare Medical Device Co., Ltd, Jiangsu, China
[6]Peking Jabrehoo Med Tech Co., Ltd, Beijing, China
[7]Yikon Genomics Company,Ltd, Jiangsu, China

References
1. Baird DC, Batten SH, Sparks SK. Alpha- and Beta-thalassemia: Rapid evidence review. Am Fam Physician. 2022;105(3):272–80.

2. Weatherall DJ. Thalassemia as a global health problem: recent progress toward its control in the developing countries. Ann N Y Acad Sci. 2010;1202:17–23.

3. Xu XM, Zhou YQ, Luo GX, Liao C, Zhou M, Chen PY, Lu JP, Jia SQ, Xiao GF, Shen X, et al. The prevalence and spectrum of alpha and beta thalassaemia in Guangdong Province: implications for the future health burden and population screening. J Clin Pathol. 2004;57(5):517–22.

4. Wang WD, Hu F, Zhou DH, Gale RP, Lai YR, Yao HX, Li C, Wu BY, Chen Z, Fang JP, et al. Thalassaemia in China. Blood Rev. 2023;60:101074.

5. Sullivan-Pyke C, Dokras A. Preimplantation genetic screening and preimplantation genetic diagnosis. Obstet Gynecol Clin North Am. 2018;45(1):113–25.

6. Kuliev A, Rechitsky S, Verlinsky O, Ivakhnenko V, Evsikov S, Wolf G, Angastiniotis M, Georghiou D, Kukharenko V, Strom C, et al. Preimplantation diagnosis of thalassemias. J Assist Reprod Genet. 1998;15(5):219–25.

7. Ou Z, Deng Y, Liang Y, Chen Z, Sun L. Improved non-invasive preimplantation genetic testing for Beta-Thalassemia using spent embryo culture medium containing blastocoelic fluid. Front Endocrinol (Lausanne). 2021;12:793821.

8. Piyamongkol W, Vutyavanich T, Sanguansermsri T. Preimplantation genetic diagnosis of alpha-thalassemia-SEA using novel multiplex fluorescent PCR. J Assist Reprod Genet. 2012;29(1):95–102.

9. Lee TH, Hsu YC, Chang CL. Detection of SEA-type α-thalassemia in embryo biopsies by digital PCR. Taiwan J Obstet Gynecol. 2017;56(4):487–94.

10. Shen X, Xu Y, Zhong Y, Zhou C, Zeng Y, Zhuang G, Ding C, Li T. Preimplantation genetic diagnosis for α-and β-double thalassemia. J Assist Reprod Genet. 2011;28(10):957–64.

11. He F, Zhou W, Cai R, Yan T, Xu X. Systematic assessment of the performance of whole-genome amplification for SNP/CNV detection and β-thalassemia genotyping. J Hum Genet. 2018;63(4):407–16.

12. Shen XT, Xu YW, Zhong YP, Zeng YH, Wang J, Ding CH, Xing WJ, Zhou CQ. [Combination of multiple displacement amplification with short tandem repeat polymorphismin preimplantation genetic diagnosis]. Beijing Da Xue Xue Bao Yi Xue Ban. 2013;45(6):852–8.

13. Chen L, Diao Z, Xu Z, Zhou J, Yan G, Sun H. The clinical application of NGS-based SNP haplotyping for PGD of Hb H disease. Syst Biol Reprod Med. 2017;63(3):212–7.

14. Chen D, Shen X, Xu Y, Ding C, Ye Q, Zhong Y, Xu Y, Zhou C. Successful four-factor preimplantation genetic testing: α- and β-thalassemia, human leukocyte antigen typing, and aneuploidy screening. Syst Biol Reprod Med. 2021;67(2):151–9.

15. Carvalho F, Moutou C, Dimitriadou E, Dreesen J, Giménez C, Goossens V, Kakourou G, Vermeulen N, Zuccarello D, De Rycke M. ESHRE PGT Consortium good practice recommendations for the detection of monogenic disorders. Hum Reprod Open. 2020;2020(3):hoaa018.

16. Chen B, CD OC, Boone DJ, Amos JA, Beck JC, Chan MM, Farkas DH, Lebo RV, Richards CS, Roa BB, et al. Developing a sustainable process to provide quality control materials for genetic testing. Genet Med. 2005;7(8):534–49.

17. Bernacki SH, Stankovic AK, Williams LO, Beck JC, Herndon JE, Snow-Bailey K, Prior TW, Matteson KJ, Wasserman LM, Cole EC, et al. Establishment of stably EBV-transformed cell lines from residual clinical blood samples for use in performance evaluation and quality assurance in molecular genetic testing. J Mol Diagn. 2003;5(4):227–30.

18. Yin A, Li B, Luo M, Xu L, Wu L, Zhang L, Ma Y, Chen T, Gao S, Liang J, et al. The prevalence and molecular spectrum of α- and β-globin gene mutations in 14,332 families of Guangdong Province, China. PLoS ONE. 2014;9(2):e89855.

19. Wang M, Zhang X, Zhao Y, Lu Z, Xiao M. Prevalence and genetic analysis of thalassemia in childbearing age population of Hainan, the Free Trade Island in Southern China. J Clin Lab Anal. 2022;36(3):e24260.

20. Volozonoka L, Miskova A, Gailite L. Whole genome amplification in preimplantation genetic testing in the era of massively parallel sequencing. Int J Mol Sci 2022, 23(9).

21. Coskun S, Alsmadi O. Whole genome amplification from a single cell: a new era for preimplantation genetic diagnosis. Prenat Diagn. 2007;27(4):297–302.

22. Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K. Whole-genome multiple displacement amplification from single cells. Nat Protoc. 2006;1(4):1965–70.

23. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science. 2012;338(6114):1622–6.

24. Renwick P, Trussler J, Lashwood A, Braude P, Ogilvie CM. Preimplantation genetic haplotyping: 127 diagnostic cycles demonstrating a robust, efficient alternative to direct mutation testing on single cells. Reprod Biomed Online. 2010;20(4):470–6.

25. Handyside AH, Harton GL, Mariani B, Thornhill AR, Affara N, Shaw MA, Griffin DK. Karyomapping: a universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes. J Med Genet. 2010;47(10):651–8.

26. Yan L, Huang L, Xu L, Huang J, Ma F, Zhu X, Tang Y, Liu M, Lian Y, Liu P, et al. Live births after simultaneous avoidance of monogenic diseases and chromosome abnormality by next-generation sequencing with linkage analyses. Proc Natl Acad Sci U S A. 2015;112(52):15964–9.

27. Hao Y, Chen D, Zhang Z, Zhou P, Cao Y, Wei Z, Xu X, Chen B, Zou W, Lv M, et al. Successful preimplantation genetic diagnosis by targeted next-generation sequencing on an ion torrent personal genome machine platform. Oncol Lett. 2018;15(4):4296–302.

28. Ji X, Zhang Z, Shi J, He B. Clinical application of NGS-based SNP haplotyping for the preimplantation genetic diagnosis of primary open angle glaucoma. Syst Biol Reprod Med. 2019;65(3):258–63.

29. Zheng CG, Liu M, Du J, Chen K, Yang Y, Yang Z. Molecular spectrum of α- and β-globin gene mutations detected in the population of Guangxi Zhuang Autonomous Region, people's Republic of China. Hemoglobin. 2011;35(1):28–39.

Publisher's Note