Accurate genotype diagnosis of Hong Kongαα thalassemia based on third-generation sequencing

Jun Li[1#], Guimei Ye[1#], Dan Zeng[1], Baodong Tian[1], Wenjuan Wang[1,2], Qiao Feng[1], Chunjiang Zhu[1]

[1]Genetics and Precision Medicine Laboratory, Affiliated Hospital of Guilin Medical University, Guilin, China; [2]Northwest Women's and Children's Hospital, Xi'an, China

Contributions: (I) Conception and design: C Zhu, J Li; (II) Administrative support: C Zhu; (III) Provision of study materials or patients: C Zhu; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: C Zhu, J Li, G Ye; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Chunjiang Zhu. Genetics and Precision Medicine Laboratory, Affiliated Hospital of Guilin Medical University, Guilin 541001, China. Email: zcjiang2003@qq.com.

Background: The Hong Kongαα (HKαα) allele is a complex structural rearrangement of the α-globin gene containing $-\alpha^{3.7}$ and $\alpha\alpha\alpha^{anti\ 4.2}$ crossover junctions. Clinically, individuals carrying the HKαα allele are often misdiagnosed or missed using conventional thalassemia gene detection technology. This study aims to identify and validate different HKαα thalassemia subtypes using third-generation sequencing (TGS) technology.

Methods: Between January 2015 and June 2021, 32 patients suspected of having HKαα thalassemia were included in this study. Genomic DNA was extracted, and gap-polymerase chain reaction (PCR), two-round nested PCR, multiplex ligation-dependent probe amplification (MLPA), and TGS were used for thalassemia gene detection.

Results: The results of HKαα/αα and HKαα/$-\alpha^{3.7}$ were similar to $-\alpha^{3.7}$/αα using the gap-PCR method. Two-round nested PCR could be used to verify the HKαα gene, but could not distinguish the subtypes of HKαα thalassemia. The MLPA assay was used to detect the change in the copy number of the α-globin gene, but it could not determine whether $-\alpha^{3.7}$ and $\alpha\alpha\alpha^{anti\ 4.2}$ were in cis or in trans. Long-read TGS technology could accurately detect the HKαα allele and distinguish the genotypes of HKαα/αα, HKαα/$-\alpha^{3.7}$, HKαα/$-\alpha^{4.2}$, and HKαα/$--^{SEA}$ without pedigree analysis. The contiguous sequence of the HKαα allele was detected using the TGS approach. This study also demonstrated that individuals with HKαα/αα and β^N/β^N genotypes tended to have normal hematological phenotypes.

Conclusions: Long-read TGS is a reliable and efficient approach for accurate detection of HKαα thalassemia, which can be widely used in clinical practice. Accurate molecular diagnosis of HKαα thalassemia will benefit clinical genetic counseling and prenatal diagnosis.

Keywords: HKαα thalassemia; genotype detection; third-generation sequencing

Submitted Aug 19, 2022. Accepted for publication Sep 16, 2022.
doi: 10.21037/atm-22-4309
View this article at: https://dx.doi.org/10.21037/atm-22-4309

Introduction

Thalassemia is one of the most common genetic diseases worldwide, characterized by a variable degree of chronic progressive hemolytic anemia (1). According to hemoglobin chain damage, thalassemia can be categorized into α-, β-, δ-, γ-, δβ-, and εγδβ-thalassemias (2). Thalassemia is highly prevalent in the Mediterranean, central Asia, Middle East, India, and southern China (3). Of all thalassemia types, α-thalassemia is the most widely distributed and is common in Southeast Asia and China (2,4,5). This thalassemia type is caused by disturbances in α-globin chain synthesis. The gene for synthesizing α-like globin peptides is located on 16p13.3. Normally, with 2 α-globin genes on each chromosome, an individual would have 2 pairs of α-globin genes, defined as αα/αα. In the α-globin gene cluster, *HBA1* and *HBA2* have a high degree of homology (homologous X, Y, Z regions), which can lead to unequal crossover.

The most common type of α-thalassemia is caused by deletions of different lengths in the α-globin genes, with -$\alpha^{3.7}$ and -$\alpha^{4.2}$ variants being the more common forms (5,6). The reciprocal recombination between Z segments during meiosis leads to the -$\alpha^{3.7}$ and $\alpha\alpha\alpha^{\text{anti 3.7}}$ chromosomes. The crossover between impaired X-boxes leads to -$\alpha^{4.2}$ and $\alpha\alpha\alpha^{\text{anti 4.2}}$ chromosomes (4). On this basis, further misalignment and unequal crossover of homologous regions in α-globin gene clusters may generate other special crossover events. Wang *et al.* (7) first reported a complex structural rearrangement containing both -$\alpha^{3.7}$ and $\alpha\alpha\alpha^{\text{anti 4.2}}$ crossover junctions on the same chromosome, designated as Hong Kongαα (HKαα). The HKαα allele contains neither a single gene deletion nor triplication. Several studies have reported that individuals with the HKαα allele present with normal hematological phenotype (8-10). Individuals with compound heterozygote of HKαα and --$^{\text{SEA}}$ showed a typical α-thalassemia trait (11). Although the HKαα allele does not have any deleterious effect on the clinical phenotype of carriers, genetic testing of the HKαα allele has important implications for genetic counseling and prenatal diagnosis. Because routine molecular testing reagents for deletional α-thalassemia cannot directly detect $\alpha\alpha\alpha^{\text{anti 4.2}}$ and $\alpha\alpha\alpha^{\text{anti 3.7}}$, it is easy to misjudge HKαα/αα or anti-HKαα/αα as -$\alpha^{3.7}$/αα or -$\alpha^{4.2}$/αα. Such misjudgments may lead to incorrect genetic counseling by clinical geneticists and may lead to incorrect prenatal diagnoses for patients.

In recent years, researchers have developed several molecular detection methods to detect HKαα thalassemia. Wang *et al.* (7) first developed a two-round nested PCR strategy to detect the presence of the novel HKαα allele. Long *et al.* (12) designed a qPCR system to screen for $\alpha\alpha\alpha^{\text{anti 3.7}}$, $\alpha\alpha\alpha^{\text{anti 4.2}}$, and HKαα genes. Huang *et al.* (13) also proposed that PCR-based multicolor melting curve analysis could identify HKαα-derived genotypes. However, these methods have certain limitations.

In clinical practice, the use of conventional thalassemia gene analysis methods to detect the HKαα allele presents certain challenges. In the α-globin gene cluster, *HBA1* and *HBA2* are highly homologous, and the fragment of the homologous region is relatively long, thus the sequence detection of highly homologous regions of α-globin genes remains a major problem for short-read technology. Long reads can resolve tandem repeats and complex structural rearrangements (14). Long-molecule sequencing (LMS) is the dominant feature of third-generation sequencing (TGS). Studies have proven that the LMS-based PacBio TGS platform is an accurate test with potential clinical utility as an alternative for screening thalassemia carriers (15,16). In this study, TGS based on single-molecule real-time (SMRT) technology was performed to detect the HKαα allele, and this was compared with the other molecular techniques. We present the following article in accordance with the MDAR reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-22-4309/rc).

Methods

Patients

Between January 2015 and June 2021, 32 individuals suspected of having HKαα thalassemia at the Affiliated Hospital of Guilin Medical University (Guilin, People's Republic of China) were included in the study. Fresh peripheral blood samples were collected and then stored at 4 ℃. The basic information, routine hematology examination, and hemoglobin electrophoresis of all patients were collected from medical records. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Affiliated Hospital of Guilin Medical University (No. 2020GZRLL-46) and informed consent was obtained from all individual participants.

DNA extraction

Genomic DNA was extracted using the Nucleic Acid Extraction System (Zeesan Biotech, Xiamen, China). DNA concentration and quality were determined by the NanoDrop 2000 spectrophotometer. DNA concentration ranged from 40 to 60 ng/μL, and the A260/A280 ranged from 1.80 to 1.88. DNA was stored at −20 ℃ in a freezer until use.

Gap-PCR

Gap-PCR and agarose gel electrophoresis for -$\alpha^{3.7}$, -$\alpha^{4.2}$, --$^{\text{SEA}}$, and --$^{\text{THAI}}$ were performed using a commercial deletional α-thalassemia gene detection kit (Yishengtang

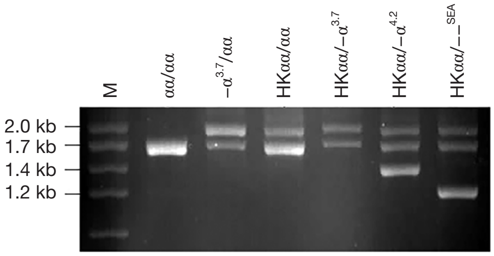

Figure 1 Molecular detection of HKαα alleles by gap-PCR. HKαα, Hong Kongαα; PCR, polymerase chain reaction.

Biological Products Co. Ltd., Shenzhen, China). The two-round nested PCR was performed to determine the presence of the HKαα allele as our previously described methods (10).

PCR-reverse dot blot (RDB) assay

PCR-RDB analysis was used to detect 3 common point mutations in α-thalassemia ($\alpha^{CS}\alpha$, $\alpha^{QS}\alpha$, and $\alpha^{WS}\alpha$) and 17-point mutations in β-thalassemia using a commercial kit (Yaneng BioSciences Co., Shenzhen, China).

Multiplex ligation-dependent probe amplification (MLPA) assay

MLPA was used to detect copy number variations of α-globin genes. The MLPA-PCR reaction was performed in a 50 µL reaction system with the SALSA MLPA Probemix P140-C1 kit (MRC-Holland, Amsterdam, Netherlands). In short, the thermocycler program for the MLPA reaction was as follows: (I) DNA denaturation; (II) hybridization reaction; (III) ligation reaction; and (IV) PCR reaction. The PCR products were analyzed using the 3500Dx Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

TGS and data analysis

The long-read sequencing-based comprehensive analysis of thalassemia alleles (CATSA) assay was carried out similarly as previously described (16). In brief, genomic DNA was extracted and subjected to multiplex long-range PCR with primers covering the full length of the *HBA* and *HBB* genes and the majority of known structural variations (SVs). End-repair and ligation reactions were added to PCR amplicons with unique PacBio barcoded adaptors, and the exonucleases were added to digest failed ligation products to obtain an individual prelibrary for each sample. After purification and quantification, the individual prelibrary was pooled by equal mass and converted to an SMRT dumbbell library with the Sequel Binding and Internal Ctrl Kit 3.0 (Pacific Biosciences, Menlo Park, CA) and sequenced on the Sequel II platform (Pacific Biosciences) for 30 h. High fidelity circular consensus sequencing (CCS) reads were generated from raw subreads, demultiplexed by barcodes, and aligned to genome build hg38 in the SMRT Link system (Pacific Biosciences). Single-nucleotide variations (SNVs) and indels were called by FreeBayes1.3.4 (Biomatters, Inc., San Diego, CA), and SVs were called according to read length. SNVs, indels, and SVs were annotated according to HbVar, Ithanet, and LOVD databases. CCS reads generated by CATSA were displayed in the Integrative Genomics Viewer (IGV) to show different variants.

Results

HKαα allele detection by gap-PCR and two-round nested PCR

By using the deletional α-thalassemia gene detection kit, the 32 samples were detected by the gap-PCR method. The test results showed that individuals with $-\alpha^{3.7}/\alpha\alpha$, HKαα/αα, and $HK\alpha\alpha/-\alpha^{3.7}$ were positive for the $-\alpha^{3.7}$ allele and the normal allele (*Figure 1*). The electrophoresis bands of HKαα/αα and $HK\alpha\alpha/-\alpha^{3.7}$ were similar to $-\alpha^{3.7}/\alpha\alpha$. However, for HKαα/αα, the electrophoresis band of the $-\alpha^{3.7}$ allele was significantly weaker than that of the normal allele. Individuals with $HK\alpha\alpha/-\alpha^{4.2}$ were positive for the $-\alpha^{3.7}$ allele, $-\alpha^{4.2}$ allele, and the normal allele, while individuals with $HK\alpha\alpha/--^{SEA}$ were positive for the $-\alpha^{3.7}$ allele, $--^{SEA}$ allele, and the normal allele (*Figure 1*). Two-round nested PCR was performed to confirm the HKαα allele, which was published by our research group in previous papers (10).

HKαα allele detection by the MLPA assay

The MLPA assay can be used to detect the deletion/duplication of large fragments in the *HBA* gene region. The copy number of α-globin genes in individuals with the HKαα allele was detected using MLPA (*Figure 2*). For the HKαα/αα genotype, the probe signals in the $-\alpha^{3.7}$ region were decreased (ratio 0.5), while the probe signals in the $-\alpha^{4.2}$ region were increased (ratio 1.5) (*Figure 2A*). For the $HK\alpha\alpha/-\alpha^{3.7}$ genotype, the probe signals in the $-\alpha^{3.7}$ region were completely absent (ratio 0), while the probe signals

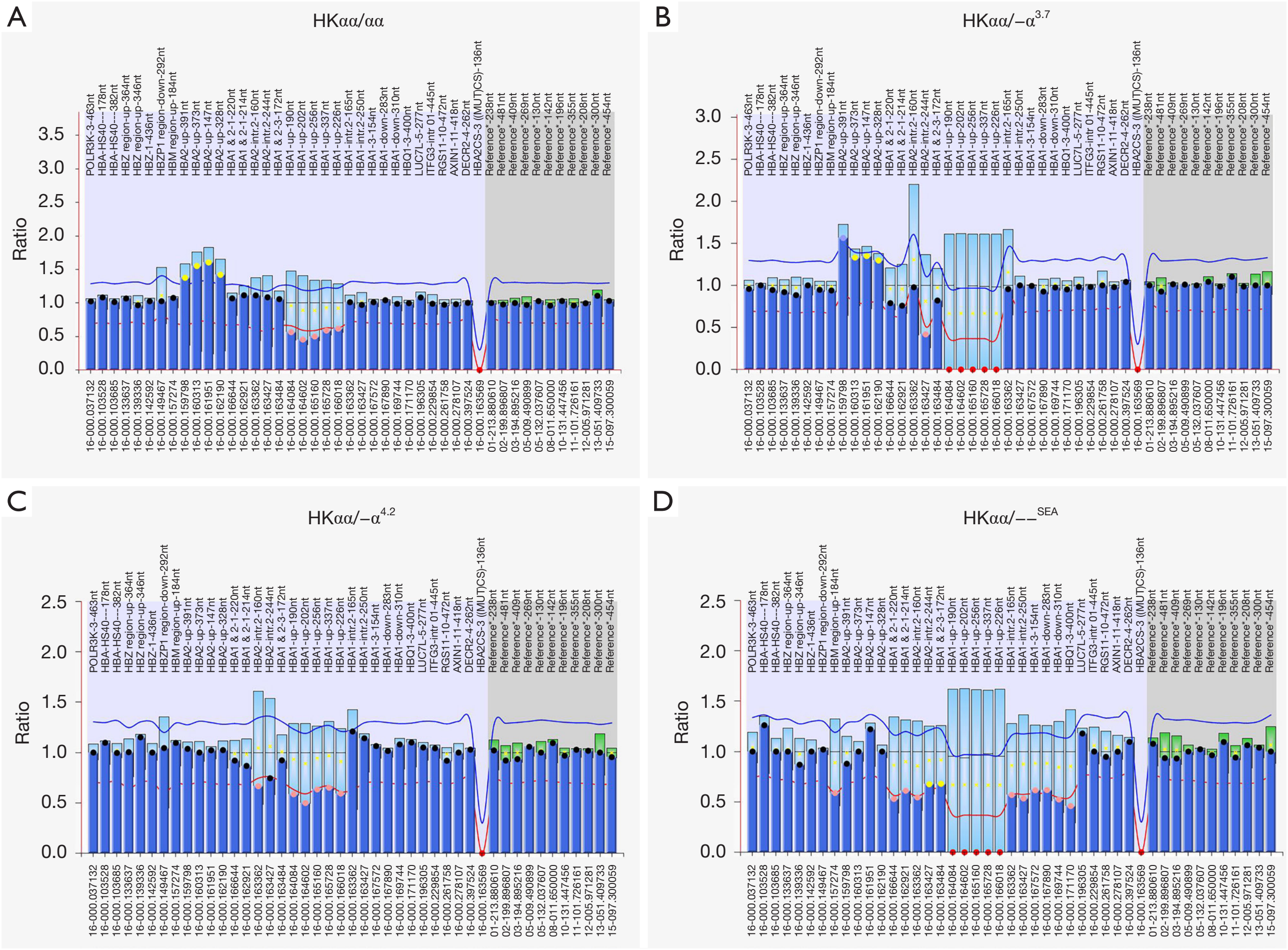

Figure 2 Molecular detection of HKαα alleles by MLPA. (A) HKαα/αα. (B) HKαα/-α$^{3.7}$. (C) HKαα/-α$^{4.2}$. (D) HKαα/--^{SEA}. The x-axis represents MLPA probe location (hg18). The y-axis represents the final ratio. HKαα, Hong Kongαα; MLPA, multiplex ligation-dependent probe amplification.

in the -α$^{4.2}$ region were increased (ratio 1.5) (*Figure 2B*). In samples with HKαα/-α$^{4.2}$, only the probe signals in the -α$^{3.7}$ region were reduced, which would cause the sample to be misdiagnosed as -α$^{3.7}$/αα (*Figure 2C*). In samples with HKαα/--^{SEA}, the signal ratio of the probe in the *HBA* gene region was the result of the sum of -α$^{3.7}$, --^{SEA}, and ααα$^{anti\ 4.2}$ (*Figure 2D*). It should be noted that a single MLPA test is not sufficient to identify the HKαα allele.

HKαα allele detection by TGS

TGS based on long-read SMRT technology was used to detect the complex structural rearrangements of α-globin genes. The results of TGS technology for individuals with the HKαα allele are shown in *Figure 3*. Recombinant SVs involving -α$^{3.7}$, -α$^{4.2}$, --^{SEA}, and HKαα variants on 2 DNA strands were well identified. The TGS technology could directly detect the HKαα allele and correctly distinguish various types of genotypes. In addition, 2 rare variants in *HBB* [-50 (G>A) and -198A>G] were discovered in 2 HKαα samples by the TGS assay.

In this study, the contiguous sequence of the HKαα allele was detected using long reads of TGS technology. Sequence data are provided in the Appendix 1.

Hematological characteristics of individuals with the HKαα allele

The hematological characteristics of individuals with the HKαα allele are shown in *Table 1*. Most individuals with

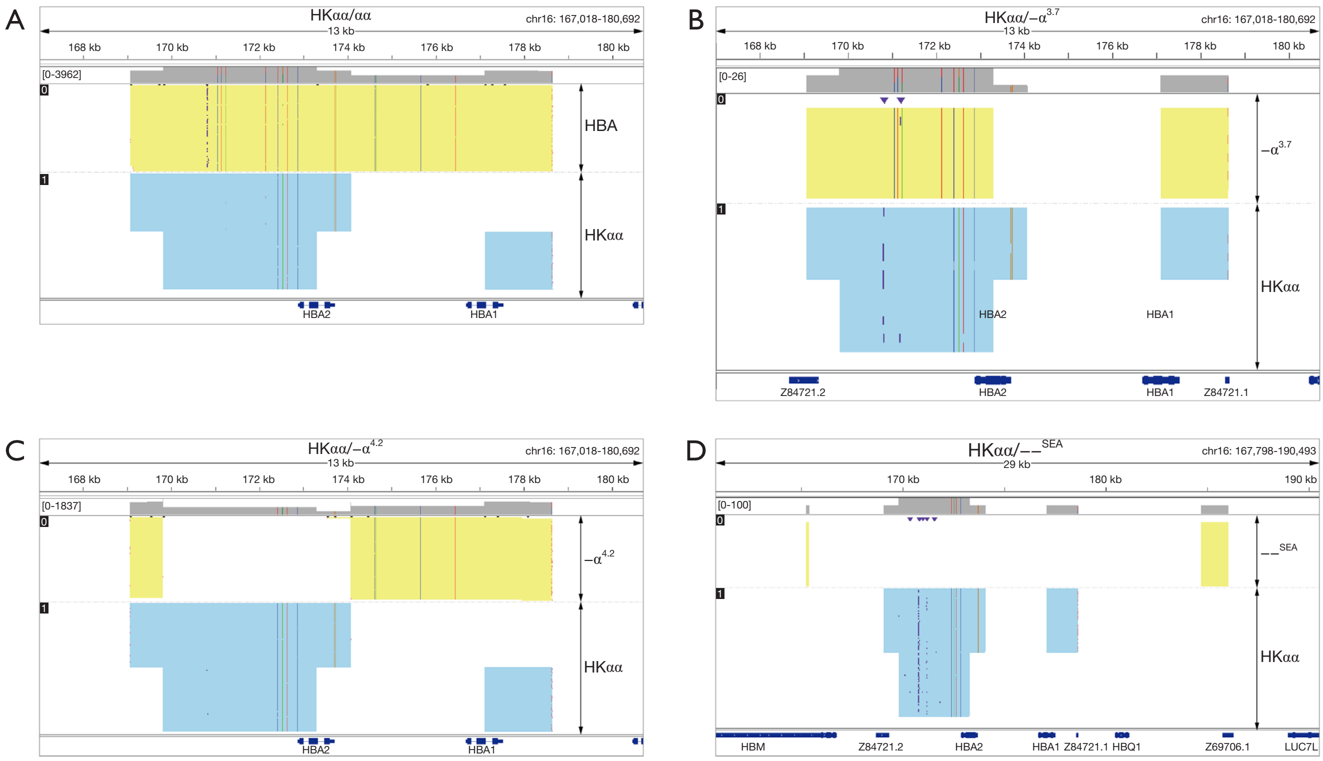

Figure 3 Molecular detection of HKαα alleles by TGS. (A) HKαα/αα. (B) HKαα/-α$^{3.7}$. (C) HKαα/-α$^{4.2}$. (D) HKαα/--^{SEA}. HKαα, Hong Kongαα; TGS, third-generation sequencing.

HKαα/αα and β^{N}/β^{N} genotypes had normal hematological phenotypes. Only 2 of 18 individuals with HKαα/αα and β^{N}/β^{N} genotypes manifested obvious microcytic hypochromic anemia, which might be related to the combination of iron deficiency. Individuals with HKαα/-α$^{4.2}$ or HKαα/α^{WS}α genotypes showed silent α-thalassemia, while individuals with the HKαα/--^{SEA} genotype showed mild α-thalassemia. HKαα/αα and HKαα/-α$^{3.7}$ combined with the *HBB* variant displayed obvious hematological features of β-thalassemia (*Table 1*).

Discussion

HKαα is a rare complex crossover event of the a-globin gene cluster, containing a cis-recombination of -α$^{3.7}$ and ααα$^{anti\ 4.2}$. Research has reported that the carrier rates of the HKαα allele ranged from 0.04% to 0.33% in southern China (8,12,17-20). Furthermore, the frequency of HKαα ranged from 2.27% to 8.81% in -α$^{3.7}$ carriers (9,20,21). The carrying rate of HKαα is significantly different in different regions, which might be related to different detection methods and ancestral effects of population distribution. HKαα is a rearrangement of the a-globin gene. As shown in *Figure 3*, Integrative Genomics Viewer plots of HKαα allele reveal that there is virtually no deletion or duplication of the gene. Therefore, individuals with HKαα/αα presented a normal hematological phenotype, and this structural variant does not aggravate the clinical phenotype of patients even when combined with other α/β-mutations, which is consistent with previously reported results (7,8,11,19,21). Although the carrier rate of the HKαα allele is low and the carriers do not have any clinical manifestations, its molecular diagnosis has important clinical significance.

At present, DNA testing is a regular part of the management of thalassemia. In clinical practice, the most common DNA testing methods for thalassemia are gap-PCR and PCR-RDB analysis with hot-spot mutation detection kits. The gap-PCR-based deletional α-thalassemia gene detection kit used in clinical practice cannot detect triplication, which will result in missed diagnosis or misdiagnosis of the HKαα allele. In this study, the results of gap-PCR showed that electrophoresis bands of HKαα/αα and HKαα/-α$^{3.7}$ were the same as -α$^{3.7}$/αα, resulting in both HKαα/αα and HKαα/-α$^{3.7}$ being misdiagnosed as -α$^{3.7}$/αα.

Table 1 Hematological characteristics of 32 cases with the HKαα allele

Sample	Gender	Age (years)	RBC ($\times 10^{12}$/L)	Hb (g/l)	MCV (fl)	MCH (pg)	MCHC (pg)	HbA_2 (%)	α genotype	β genotype
1	Female	26	3.78	117	91.8	31	337	3	HKαα/αα	β^N/β^N
2	Female	30	4.21	120	86.5	28.5	330	3.1	HKαα/αα	β^N/β^N
3	Female	28	4.45	135	88.3	30.4	344	2.6	HKαα/αα	β^N/β^N
4	Female	28	3.63	126	103.6	34.7	335	2.5	HKαα/αα	β^N/β^N
5	Female	31	5.06	156	88.7	30.8	347	2.8	HKαα/αα	β^N/β^N
6	Female	28	3.93	121	90.8	30.8	339	2.7	HKαα/αα	β^N/β^N
7	Female	27	4.6	123	84.3	26.7	317	3.2	HKαα/αα	β^N/β^N
8	Female	22	4.12	126	88.3	30.6	346	2.8	HKαα/αα	β^N/β^N
9	Female	30	4.13	124	87.7	30	343	2.9	HKαα/αα	β^N/β^N
10	Female	38	4.08	93	78.7	22.8	290	1.9	HKαα/αα	β^N/β^N
11	Female	32	3.44	115	97.7	33.4	342	2.8	HKαα/αα	β^N/β^N
12	Female	28	3.95	118	88.6	29.9	337	2.6	HKαα/αα	β^N/β^N
13	Female	44	4.13	116	88.6	28.1	317	2.8	HKαα/αα	β^N/β^N
14	Female	36	4.41	130	91.2	29.5	323	2.5	HKαα/αα	β^N/β^N
15	Female	29	3.83	113	88	29.5	335	2.5	HKαα/αα	β^N/β^N
16	Female	41	4.03	82	67.5	20.3	301	–	HKαα/αα	β^N/β^N
17	Female	27	5.41	156	86.9	28.8	332	–	HKαα/αα	β^N/β^N
18	Female	25	4.82	130	83.8	27	322	2.3	HKαα/αα	β^N/β^N
19	Female	25	5.09	131	80.4	25.7	320	2.3	HKαα/-$\alpha^{4.2}$	β^N/β^N
20	Male	9	5.12	126	74.2	24.6	332	2.9	HKαα/-$\alpha^{4.2}$	β^N/β^N
21	Male	33	5.94	169	90.9	28.4	312	2.5	HKαα/$\alpha^{WS}\alpha$	β^N/β^N
22	Male	35	6.98	139	64.6	19.9	308	2.6	HKαα/--^{SEA}	β^N/β^N
23	Female	30	5.85	123	70	21	296	2.5	HKαα/--^{SEA}	β^N/β^N
24	Female	25	5.46	103	60.8	18.9	310	6.2	HKαα/αα	β^{43M}/β^N
25	Female	32	4.68	97	65.8	20.7	315	6	HKαα/αα	β^{17M}/β^N
26	Female	21	5.63	85	55.4	15.1	272	–	HKαα/αα	β^{17M}/β^N
27	Female	25	5.3	103	60.9	19.4	319	5.1	HKαα/αα	$\beta^{41\text{-}42M}/\beta^N$
28	Female	31	5.58	111	61.1	19.9	326	5.4	HKαα/αα	$\beta^{41\text{-}42M}/\beta^N$
29	Male	49	6.34	126	65	19.9	306	–	HKαα/αα	β^{17M}/β^N
30	Female	31	4.11	124	87.8	30.2	343	2.7	HKαα/αα	$\beta^{-50(G>A)}/\beta^N$
31	Male	1	4.02	88	69.2	21.9	317	2.6	HKαα/αα	$\beta^{-198\ A>G}/\beta^N$
32	Male	64	5.41	104	64.9	19.2	296	6.4	HKαα/-$^{\alpha3.}$7	β^{41-42M}/β^N

RBC, red blood cell; Hb, hemoglobin; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; HbA_2, hemoglobin A_2.

Such misjudgments can lead clinicians to provide incorrect genetic counseling and prenatal diagnosis. Furthermore, α^0-thalassemia coinherited with $-\alpha^{3.7}$ ($--/-\alpha^{3.7}$) can result in hemoglobin H (HbH) disease, while α^0-thalassemia coinherited with the HKαα allele (--/HKαα) presents as mild α-thalassemia.

The MLPA technique is also a common clinical method for detecting copy number changes when rare variations in the α-globin gene are suspected. Most of the deletion subtypes of α-thalassemia can be detected by MLPA, including $-\alpha^{3.7}$, $-\alpha^{4.2}$, $--^{SEA}$, $--^{THAI}$, $--^{FIL}$, and ($-\alpha^{20.5}$), among others. The α-triplication ($\alpha\alpha\alpha^{anti\ 3.7}$ and $\alpha\alpha\alpha^{anti\ 4.2}$) can also be detected by MLPA with the P140 Probemix. However, copy number variations of the *HBA* gene detected by MLPA should be confirmed using a different technique. As shown in *Figure 2A*, both $-\alpha^{3.7}$ and $\alpha\alpha\alpha^{anti\ 4.2}$ were detected in a single sample, but it was impossible to determine whether the genotype was an HKαα/αα or $\alpha\alpha\alpha^{anti\ 4.2}/-\alpha^{3.7}$ compound heterozygous mutation. Additional experiments or pedigree analysis are required to determine the correct genotype, as genetic counseling for HKαα/αα and $\alpha\alpha\alpha^{anti\ 4.2}/-\alpha^{3.7}$ compound heterozygous mutation is completely different. In the case of $\alpha\alpha\alpha^{anti\ 4.2}/-\alpha^{3.7}$ compound heterozygosity, inheritance to the offspring could only be $\alpha\alpha\alpha^{anti\ 4.2}$ or $-\alpha^{3.7}$. In addition to compound heterozygosity of $-\alpha^{3.7}$ and α^0 mutations leading to HbH disease, α-triplication might exacerbate the proportional imbalance of α and β chains and produce more severe phenotypes of β-thalassemia when α-triplication is coinherited with *HBB* variants (12,22). In the case of HKαα/αα, it would be passed to the offspring as the αα allele or HKαα allele. Genetic modification of the HKαα allele to other thalassemia subtypes was not obvious. In this study, MLPA analysis showed no net copy number change in the 4.2 region for the HKαα/$-\alpha^{4.2}$ genotype due to a $-\alpha^{4.2}$ deletion on one chromosome and a similarly sized $\alpha\alpha\alpha^{anti\ 4.2}$ duplication on the other chromosome, which might result in a false result (*Figure 2C*). Therefore, it is difficult to identify the HKαα allele using the MLPA technique.

Two-round nested PCR is a classical method to detect the HKαα allele (7). However, HKαα/αα, HKαα/HKαα, HKαα/$-\alpha^{3.7}$, and HKαα/$\alpha\alpha\alpha^{anti\ 4.2}$ cannot be distinguished by two-round nested PCR. Chen *et al.* (21) suggested that the combination of gap-PCR, MLPA, and nested PCR could be used to detect HKαα. MLPA and the two-round nested PCR method can verify each other and help differentiate HKαα/αα, HKαα/HKαα, HKαα/$-\alpha^{3.7}$, HKαα/$-\alpha^{4.2}$, HKαα/HKαα, HKαα/$\alpha\alpha\alpha^{anti\ 3.7}$, and HKαα/$\alpha\alpha\alpha^{anti\ 4.2}$. However, this combination involves multiple sets of experiments, and the process is cumbersome and time-consuming, making it difficult to achieve large-scale population screening in clinical practice.

Recently, TGS technology based on LMS has become an efficient, cost-effective, and accurate approach for comprehensive thalassemia carrier screening (15,16). Using long-read single-molecule sequencing technology, TGS can accurately distinguish the highly homologous sequences of the *HBA1* and *HBA2* genes, and correctly discern the real genetic carrier status of complex structural variants and homologous recombination. For samples with compound heterozygotes, TGS can directly determine whether the genotypes are in cis or in trans without pedigree analysis. In this study, the subtypes of α-thalassemia with the HKαα allele were correctly detected by the TGS approach, including HKαα/αα, HKαα/$-\alpha^{3.7}$, HKαα/$-\alpha^{4.2}$, and HKαα/$--^{SEA}$. In addition, a recent study also showed that more complex α-globin gene cluster variants, such as $\alpha\alpha^{anti3.7}\alpha^{anti3.7}\alpha^{17.2}/\alpha\alpha$ and $--^{SEA}/\alpha^{Westmead}\alpha^{anti3.7+Westmead}\alpha$, were identified by TGS testing (23). Therefore, long-read TGS can greatly benefit resolving complex structural rearrangements and help improve the accuracy of genetic testing.

The HKαα allele was confirmed by Sanger sequencing in our previous research (10). As described by our research group (10), the HKαα allele sequence is divided into 3 regions: $-\alpha^{3.7}$, $\alpha\alpha\alpha^{anti\ 4.2}$, and a special region about 802 base pairs long. Interestingly, in the special 802 base pairs region, the upstream area is identical to the sequence of the α-globin gene ranging from 171,116 to 171,235, the downstream area is the same as that on the α-globin gene and ranges from 171,223 to 171,882, but the part that crosses them is the undetected sequence (NCBI Reference Sequence: Ng_0000016.10), which might be caused by the shortcomings of short reads. The PacBio long-read sequencing approach can overcome the limitations (14). In this study, the contiguous sequence of the HKαα allele was detected using long reads of TGS technology. Moreover, we discovered that the sequence not detected in the previous study contained 17 T, which shared the same sequence with the α-globin gene ranging from 171,219-171,235 (hg38). The sequence was lost in the splicing, so that it could not be detected by Sanger sequencing. By using long-read contiguous sequences, TGS technology greatly makes up for the limitation of short-read sequencing and provides full-length coverage of thalassemia alleles, including *HBA* and *HBB* genes. Surprisingly, 2 rare variants in *HBB* [-50 (G>A) and -198A>G] were discovered when HKαα samples were subjected to the TGS assay. The mutation of -50 (G>A) (*HBB*: c.-100G>A) was first reported by Li *et al.* in 2009 (24). Consistent with previous studies (25,26), our result suggested that individuals with the -50 (G>A) heterozygous variant had a normal hematological phenotype. Unexpectedly, our study showed that an individual with the -198 (A>G) (*HBB*: c.-248A>G) mutation exhibited microcytic hypochromic anemia, which was inconsistent with a previously reported result that the -198 (A>G) heterozygous variant has a normal hematological phenotype with decreased HbA2 level (26). The reason was that the individual in our study had concomitant iron deficiency anemia with a serum ferritin of 9.73 ng/mL. In short, long-read TGS technology can identify both common and rare mutations in the *HBA* and *HBB* genes, thus improving the detection rate of rare thalassemia.

With sequencing of long reads, the biggest advantage of TGS is to directly identify the junction site of complex structural variants like HKαα, α-globin gene triplications and even quadruplications. The accuracy of PacBio sequencing was guaranteed by two levels. First, the PacBio platform utilized a hairpin library structure which enables it to sequence the same insert (subread) for multiple passes to generate high fidelity CCS reads. Second, a minimum 60× coverage depth of α-globin and β-globin genes was sequenced, which further proofed any sequencing errors in CCS reads. To reduce the cost per test, hundreds of barcoded samples need to be pooled together for sequencing in one flowcell, which might hinder the clinical application of TGS in small centers with limited number of samples. Pooling samples with different TGS-based tests like congenital adrenal hyperplasia, spinal muscular atrophy and fragile-X syndrome could be a solution. In addition, TGS cannot detect all the deletional thalassemia, so some difficult cases require the combination of clinical phenotype and multiple detection techniques to make a correct diagnosis.

Besides the methods mentioned above, researchers have reported that qPCR system and multicolor melting curve analysis can be used to detect the HKαα thalassemia (12,13). However, they cannot directly determine whether the genotype is in cis or in trans, and further experiments are needed to verify. Next-generation sequencing (NGS) is an effective means of detecting rare thalassemia, but it is difficult to detect HKαα due to the limitation of short-read length.

In conclusion, long-read TGS technology has significant advantages in detecting the HKαα allele, and can accurately distinguish various subtypes of HKαα thalassemia, thus achieving accurate detection. The TGS method can help to comprehensively improve the detection level of thalassemia mutant genes in the population, and will be beneficial for clinical geneticists in carrying out genetic counseling for patients and making correct prenatal diagnoses.

Acknowledgments

We thank Aiping Mao, Jiaqi Li, and Tiantian Xie from Berry Genomics Corporation for their assistance with TGS technology.
Funding: This study was supported by the National Natural Science Foundation of China (No. 82060037), the Natural Science Foundation of Guangxi Province (No. 2018JJA140062), and Guangxi Ba Gui Scholars Special Project.

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-22-4309/rc

Data Sharing Statement: Available at https://atm.amegroups.com/article/view/10.21037/atm-22-4309/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-4309/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Affiliated Hospital of Guilin Medical University (No. 2020GZRLL-46) and informed consent was obtained from all individual participants.

References

1. Taher AT, Weatherall DJ, Cappellini MD. Thalassaemia. Lancet 2018;391:155-67.
2. De Simone G, Quattrocchi A, Mancini B, et al. Thalassemias: from gene to therapy. Mol Aspects Med 2022;84:101028.
3. Higgs DR, Engel JD, Stamatoyannopoulos G. Thalassaemia. Lancet 2012;379:373-83.
4. Harteveld CL, Higgs DR. Alpha-thalassaemia. Orphanet J Rare Dis 2010;5:13.
5. Piel FB, Weatherall DJ. The α-thalassemias. N Engl J Med 2014;371:1908-16.
6. Kalle Kwaifa I, Lai MI, Md Noor S. Non-deletional alpha thalassaemia: a review. Orphanet J Rare Dis 2020;15:166.
7. Wang W, Chan AY, Chan LC, et al. Unusual rearrangement of the alpha-globin gene cluster containing both the -alpha3.7 and alphaalphaalphaanti-4.2 crossover junctions: clinical diagnostic implications and possible mechanisms. Clin Chem 2005;51:2167-70.
8. Shang X, Li Q, Cai R, et al. Molecular characterization and clinical presentation of HKαα and anti-HKαα alleles in southern Chinese subjects. Clin Genet 2013;83:472-6.
9. Wu MY, Li J, Li SC, et al. Frequencies of HKαα and anti-HKαα Alleles in Chinese Carriers of Silent Deletional α-Thalassemia. Hemoglobin 2015;39:407-11.
10. Wang W, Zheng H, Zeng D, et al. The pedigree analysis and prenatal diagnosis of Hong Kongαα Thalassemia and the sequence analysis of Hong Kongαα Allele. Mol Genet Genomic Med 2020;8:e1285.
11. Wu MY, Li J, Li SC, et al. Compound Heterozygosity for HKαα and an in Cis Deletion of Double α Genes Presents as α-Thalassemia Trait. Hemoglobin. 2015;39:256-9.
12. Long J, Liu E. The carriage rates of αααanti3.7, αααanti4.2, and HKαα in the population of Guangxi, China measured using a rapid detection qPCR system to determine CNV in the α-globin gene cluster. Gene 2021;768:145296.
13. Huang Q, Wang X, Tang N, et al. Simultaneous Genotyping of α-Thalassemia Deletional and Nondeletional Mutations by Real-Time PCR-Based Multicolor Melting Curve Analysis. J Mol Diagn 2017;19:567-74.
14. Midha MK, Wu M, Chiu KP. Long-read sequencing in deciphering human genetics to a greater depth. Hum Genet 2019;138:1201-15.
15. Xu L, Mao A, Liu H, et al. Long-Molecule Sequencing: A New Approach for Identification of Clinically Significant DNA Variants in α-Thalassemia and β-Thalassemia Carriers. J Mol Diagn 2020;22:1087-95.
16. Liang Q, Gu W, Chen P, et al. A More Universal Approach to Comprehensive Analysis of Thalassemia Alleles (CATSA). J Mol Diagn 2021;23:1195-204.
17. Wang M, Zhang X, Zhao Y, et al. 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:e24260.
18. He S, Qin Q, Yi S, et al. Prevalence and genetic analysis of α- and β-thalassemia in Baise region, a multi-ethnic region in southern China. Gene 2017;619:71-5.
19. Luo X, Zhang XM, Wu LS, et al. Prevalence and clinical phenotype of the triplicated α-globin genes and its ethnic and geographical distribution in Guizhou of China. BMC Med Genomics 2021;14:97.
20. Zhang M, Huang H, Chen M, et al. Frequencies and hematological manifestations of the HKαα allele in southern Chinese population. Int J Clin Exp Pathol 2019;12:3058-62.
21. Chen DM, Ma S, Tang XL, et al. Diagnosis of the accurate genotype of HKaa carriers in patients with thalassemia using multiplex ligation-dependent probe amplification combined with nested polymerase chain reaction. Chin Med J (Engl) 2020;133:1175-81. Correction appears in: Chin Med J (Engl) 2020;133:1631.
22. Ma SK, Au WY, Chan AY, et al. Clinical phenotype of triplicated alpha-globin genes and heterozygosity for beta0-thalassemia in Chinese subjects. Int J Mol Med 2001;8:171-5.
23. Long J, Sun L, Gong F, et al. Third-generation sequencing: A novel tool detects complex variants in the α-thalassemia gene. Gene 2022;822:146332.
24. Li DZ, Liao C, Xie XM, et al. A novel mutation of -50 (G-->A) in the direct repeat element of the beta-globin gene identified in a patient with severe beta-thalassemia. Ann Hematol 2009;88:1149-50.
25. Zhao Y, Jiang F, Li DZ. Hematological Characteristics of β-Globin Gene Mutation -50 (G>A) (HBB: c.-100G>A) Carriers in Mainland China. Hemoglobin 2020;44:240-3.

26. Luo S, Chen X, Zeng D, et al. The value of single-molecule real-time technology in the diagnosis of rare thalassemia variants and analysis of phenotype-genotype correlation. J Hum Genet 2022;67:183-95.

(English Language Editor: C. Betlazar-Maseh)

Cite this article as: Li J, Ye G, Zeng D, Tian B, Wang W, Feng Q, Zhu C. Accurate genotype diagnosis of Hong Kongαα thalassemia based on third-generation sequencing. Ann Transl Med 2022;10(20):1113. doi: 10.21037/atm-22-4309