Analysis of Rare Thalassemia Genetic Variants Based on Third Generation Sequencing

Cuiting Peng

West China Second University Hospital of Sichuan University

Haixia Zhang

West China Second University Hospital of Sichuan University

Han Chen

West China Second University Hospital of Sichuan University

Jun Ren

West China Second University Hospital of Sichuan University

Ze Du

West China Second University Hospital of Sichuan University

Tong Zhao

West China Second University Hospital of Sichuan University

Aiping Mao

Berry Genomics Corporation

Ruofan Xu

Berry Genomics Corporation

Yulin Lu

Berry Genomics Corporation

He Wang

West China Second University Hospital of Sichuan University

Xinlian Chen (✉ chenxinlian11@qq.com)

West China Second University Hospital of Sichuan University

Shanling Liu

West China Second University Hospital of Sichuan University

Research Article

Keywords: Thalassemia genetic screening, Third generation sequencing, - α 3.7 subtype III, Rare thalassemia variants

Posted Date: March 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1365771/v1>

Abstract

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

1. Introduction

Thalassemia is a group of common hereditary anemia that cause significant morbidity and mortality around the world, especially in South East Asian, Middle East and Mediterranean populations[1, 2]. α -thalassemia (α -thal) and β -thalassemia (β -thal) are two main types of thalassemia caused by mutations in *HBA1/2* and *HBB* gene respectively, thus resulted in abnormal α - and β -globin synthesis and defective hemoglobin structure[3, 4].

The clinical manifestations of thalassemia vary greatly and are usually directly related to the degree of globin chain reduction. There are three types of thalassemia carrier states (silent, minor and intermedia) and one disease states (major) according to the amount of globin chain and severity of the disease[5, 6]. The α -thalassemia major is also known as Hb bart's hydrops fetalis, of which the affected fetuses usually die in utero or shortly after birth due to severe anemia and lack of oxygen[7]. While, the affected individual with the most severe form of β -thalassemia usually can survive but only with regular blood transfusions and iron chelation therapy[8, 9]. It has been estimated that about 5% and 1.5% of the population worldwide are carriers of α - and β -thalassemia genetic mutation, thus a large number of children are still born annually with hemoglobin disorders, which cause severe birth defect and places a heavy burden on society and families[3, 10].

Due to the complexity of the thalassemia genetics and genotype-phenotype correlation, precisely diagnosing of thalassemia patients and carrier status are still challenging[11]. Effective and accurate molecular diagnosis methods are urgently needed to identify rare clinically significant variants from thalassemia genes. Conventional genetic diagnostic technology such as reverse dot blot hybridization, Sanger sequencing, GAP-PCR and Multiple Ligation-Dependent Probe Amplification (MLPA) can be utilized to detect those most prevalent variants including SNVs, indels in *HBB* and deletions in

HBA1/2[12]. More recently, the next generation sequencing (NGS) based on PCR for targeted exons and selected intronic regions, has showed advantages in thalassemia gene screening[13]. Yet, for those rare variants not located in regular regions or variants in homologous regions of *HBA1* and *HBA2*, conventional methods and short-read based NGS method may lead to missed diagnosis or even misdiagnose. Nowadays, a methodology based on third generation sequencing (TGS) named Comprehensive Analysis of ThalaSsaemia Alleles (CATSA) was developed and validated for comprehensive thalassemia screening[14, 15]. Based on long range PCR and long-molecule sequencing on the PacBio Sequel II platform, CATSA method is adequate to cover the entire gene region and enable to diagnose common and rare types of thalassemia variants. To further investigate the potential of long-molecule sequencing in rare thalassemia carrier testing, we enrolled 100 cases either showed abnormal hematology phenotype or hemoglobin electrophoresis, but with negative conventional genetic diagnosis results. In this study, CATSA method detected extra 10 cases of clinically significant variants from *HBA1/2* and *HBB*, including 3 cases of SVs and 7 cases of SNVs, of which a very rare subtype of $-\alpha^{3.7}$ named $-\alpha^{3.7}$ subtype III ($-\alpha^{3.7III}$) was first identified in Chinese population and validated by Sanger sequencing. Other rare genotype with 11.1 kb deletions ($-11.1/\alpha\alpha$) or triplicate α -globin genes ($aaa^{3.7}/\alpha\alpha$) and rare SNVs can be thoroughly detected in one test. The study further validated long-molecule sequencing based CATSA as an efficient and valuable method in the diagnosis of rare thalassemia.

2. Patients And Methods

2.1 Patients

A total of 100 patients highly suspected with thalassemia gene mutations were included in this study and signed an informed consent form. The inclusion criteria for the enrolled patients are: either 1) the routine hematology examinations showed abnormal hemoglobin of $HbA2 < 2.3\%$ or $HbA2 \geq 3.2\%$ or elevated HbF. Or 2) the blood test showed abnormalities with the mean corpuscular volume (MCV) ≤ 80 fL and/or mean corpuscular hemoglobin (MCH) ≤ 27 pg. Or 3) the genotype identified by conventional genetic testing could not well explain their clinical phenotype.

2.2 GAP-PCR testing for large deletion of α -thalassemia

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

2.3 Reverse dot blot hybridization assays

The PCR reverse dot-blot (PCR-RDB) assay was performed for three common non-deletional α -thalassemia mutations including Hb Constant Spring (Hb CS, HBA2: c.427T > C), Hb Quong Sze (Hb QS, HBA2: c.377T > C), and Hb Westmead (Hb WS, HBA2: c.369C > G), and 17 β -thalassemia mutations including -32 (C > A) (HBB: c.-82C > A), -30 (T > C) (HBB: c.-80T > C), -29 (A > G) (HBB: c.-79A > G), -28 (A > G) (HBB: c.-76A > G), CAP + 40–43 (-AAAC) (HBB: c.-11_8delAAAC), initiation codon (T > G) (HBB:c.2T >

G), codons 14–15 (+ G) (HBB: c.45_46insG), codon 17 (A > T) (HBB: c.52A > T), codon 26 (or Hb E) (G > A) (HBB: c.79G > A), codons 27/28 (+ C) (HBB: c.84_85insC), codon 31 (-C) (HBB: c.94delC), codons 41–42 (-TTCT) (HBB: c.126_129delCTTT), codon 43 (G > T) (HBB: c.130G > T), codons 71–72 (+ A) (HBB: c.216_217insA), IVS-I-1 (G > T) (HBB: c.92 + 1G > T), IVS-I-5 (G > C) (HBB: c.92 + 5G > C), and IVS-II-654 (C > T) (HBB: c.316-197C > T). (Yaneng Bioscience, Shenzhen, China).

2.4 long PCR based third generation sequencing and data analysis

Experiments were performed as previously described[14]. Briefly, genomic DNA was amplified by PCR with primers targeting the majority of known structural variations, SNVs and indels in the *HBA1*, *HBA2* and *HBB* genes. Barcoded adaptors were ligated to the PCR products to construct individual sequencing library. Then each library was quantified and pooled together by equal mass. After purification and quantification, the pooled library was converted to SMRT bell library with Sequel Binding and Internal Ctrl Kit 3.0 (Pacific Biosciences) and sequenced on the Sequel II platform (Pacific Biosciences) under CCS mode. Then raw subreads were analyzed by CCS software (Pacific Biosciences) to generate CCS reads, then debarcoded by lima in the Pbbioconda package (Pacific Biosciences) and aligned to genome build hg38 by pbmn2 (Pacific Biosciences). Finally, structural variations were identified according to HbVar, Ithantnet and LOVD database. SNVs and indels were identified by FreeBayes1.3.4 (<https://www.geneious.com/plugins/freebayes>; Biomatters, Inc., San Diego, CA).

2.5 Variants confirmation

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

3. Results

3.1 The detection and identification of rare clinically significant SNVs

In this study, we identified 7 cases of rare SNVs that can be ranked as clinically significant variants including *HBA2* SNVs (c.168dup, c.-59C > T, c.51G > T, c.91_93delGAG, and c.300 + 34G > A) and *HBB* c.316-45G > C (Table 1). All the long molecular sequencing data has been verified by specific PCR and Sanger sequencing (Fig. 1). Among them, the heterozygous of *HBA2*:c.168dup, also known as codons 55/56 (+ T) could cause α^+ -thalassemia due to a frameshift mutation of $\alpha 2$ -globin gene, which could better explain the phenotype of participant #2 with mild microcytosis and hypochromia (Table 2)[16]. The variant *HBA2*:c.-59C > T (participant #3) is one type of point mutations found in the promoter of $\alpha 2$ -globin gene which affects *HBA2* transcription and protein synthesis[17]. Also, heterozygous of *HBA2*:c.51G > T (CD16 (AAG > AAC or AAT)) and *HBA2*:c.91_93delGAG (codon 30(-GAG)) may influence protein translation

thus resulted in the abnormal hemoglobin and hematology phenotype for participant #4 and #5 (Table 2) [18, 19].

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

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

Table 2

Hematology examination and hemoglobin electrophoresis results in patients with detected clinically significant variants.

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

3.2 Multiple mutations in the polyadenylation signal site and compound –SEA/αα

The participant #1 showed a Hb H disease-like abnormal hematology phenotype of significantly decreased MCV and MCH, together with abnormal hemoglobin (0.8% of HbA2). Through screening of the common thalassemia deletions or mutations, only heterozygous Southeast Asian deletion (–SEA/αα) was identified, which may be not able to explain the abnormal Hb electrophoresis. In this study, the third generation sequencing further identified multiple mutations at the polyadenylation signal site in the α2-globin gene, including HBA2:c.*64(T > C), c.*68(A > C), c.*71(G > A), c.*74(C > A), c.*82(G > A), c.* 92(A > G) and c.*98(T > C) (Table 1 and Fig. 2). The results were further verified by specific PCR and agarose gel electrophoresis (Yaneng Bioscience) (Fig. 2C). These multiple mutations in polyadenylation signal site in combination with –SEA/αα genotype may led to the phenotype of Hb H disease[20, 21].

3.3 The identification of -α^{3.7} subtype III

Through third generation sequencing, we identified a rare deletion of 3.8 kb involving *HBA2* and *HBA1* (chr16:173707–177518 based on sequencing analysis) in participant #8. The exact deletion breakpoints were also inferred by Sanger sequencing (Table 1 and Fig. 3). Since the short segment of ‘TGGTCTTTGAATAAAGTCTGAGTGGGC’ could locate at chr16:173680–173706 and chr16:177492–177518 based on sequence alignment, the breakpoint should be at chr16:173707 and chr16:177518. The corresponding HGVS nomenclature for this type of deletion was NG_000006.1: g.34570_38382del 3812bp, which is a very rare type of $-\alpha^{3.7}$, that is $-\alpha^{3.7}$ subtype III[22]. To our knowledge, this is the first reported case of $-\alpha^{3.7}$ subtype III in Chinese population. The results also verified third generation sequencing as a valuable method for rare thalassemia diagnosis.

3.4 The identification of rare deletion of 11.1 kb in *HBA1/2*

Another rare deletion of 11.1 kb involved *HBA2* and *HBA1* gene was identified for participant #9 which could not be easily identified by conventional methods (Table 1 and Fig. 4). This deletion with the HGVS nomenclature of NC_000016.9:g.(220831_220860)_(231920_232003)del was one type of α^0 -thalassemia that has been previously reported in Chinese population[19]. The carriers with this deletion usually showed mild α -thalassemia with hypochromic microcytic phenotypes (Table 2).

3.5 The detection of α -Globin gene triplication

One case of α -Globin gene triplication ($\alpha\alpha\alpha^{3.7}/\alpha\alpha$) was found in this study (participant #10). The results was further validated by specific PCR and agarose electrophoresis (Yaneng Bioscience)(Table 1 and Fig. 5). The carriers with this α -Globin gene triplication usually showed normal results for routine hematology examination but abnormal HbA2 content[23]. Besides, if compound with β -thalassemia, patients usually show intermediate β -thalassemia phenotypes.

4. Discussion

Conventional genetic testing method including Gap-PCR, RDB, MLPA and Sanger sequencing could only detect those common variants such as $-\alpha^{3.7}$, $-\alpha^{4.2}$, $-\text{SEA}$, $-\alpha^{27.6}$, $-\alpha^{21.9}$ or other known SVs and SNVs, which is very limited. Since more than 2000 thalassemia or abnormal hemoglobin-related variant sites has been reported, it is necessary to develop new methods to implement thalassemia diagnosing especially for these rare variants. Recently, third generation sequencing technology also known as long-molecule sequencing has emerging as an incomparable method in genetic diagnosis with many advantages including long reads, high accuracy, single molecule resolution and no GC preference[24]. The long reads could cover many rare gene loci and its PCR-free characteristic made it possible to reflect the real arrangement in the genome. The method designed for thalassemia carrier screening based on third generation sequencing on Pacific biosciences Sequel II platform has been optimized these years to fulfill the detection of thalassemia gene with high efficiency and accurate, also with acceptable time and cost.

Through third generation sequencing for full length of *HBA1*, *HBA2* and *HBB* gene, multiple variants may be recognized for one sample. Thus, it is crucial to evaluate the correlation between genotype and

phenotype to further identify the clinically significant variants[25]. The SNVs reported in this study including *HBA2* SNVs (c.168dup, c.-59C > T, c.51G > T, c.91_93delGAG, and c.300 + 34G > A) and *HBB* c.316-45G > C, they all showed some evidence that relate to abnormal hematology phenotype and hemoglobin results. The carriers usually had normal phenotype, but tend to develop into intermediate thalassemia when compound with other variant types. Most of these participants are at the propiate age and have willing to procreate, thus our results could pave the way for further prenatal diagnosis or even preimplantation genetic testing (PGT) to prevent the incident of thalassemia[26].

It is worth to mention, the participant #6 showed an increased HbF content of 16.3%, yet only one clinically significant variant of *HBA2* c.300 + 34G > A was identified. We suspected that some other HPFH-causing variants in the gamma globin genes *HBG1* or *HBG2* may be responsible for the increased HbF level[27]. Thus, we conducted a full-length sanger sequencing for *HBG1* and *HBG2* gene. Two variants of -196 (C->T) and + 25 (G->A) in the promoter region of *HBG1* were found (Fig. 6). Among which, -196 (C->T) also named *HBG1*:c.-249C > T may be responsible for the high expression of the gamma-globin gene thus cause the increased HbF level according to the previous studies[28, 29]. The results suggested that it may be necessary to design more primer to detect other thalassemia related genes like *HBG1* and *HBG2* in the future, in order to implement comprehensive screening.

Moreover, we found one case with multiple mutations in polyadenylation signal site compound with $-\text{SEA}/\alpha\alpha$ which showed a rare Hb H disease. These multiple mutations led to a fusion between the $\alpha 2$ and $\psi\alpha 1$ gene, which may affect mRNA transcription and termination. Thus, $\alpha 2$ -globin gene polyA mutation in combination with $-\text{SEA}/\alpha\alpha$ may lead to Hb H disease. This patient is a 2-year-old boy with significantly decreased MCV and MCH, together with abnormal hemoglobin. Through conventional thalassemia genetic screening, only a heterozygous $-\text{SEA}/\alpha\alpha$ was found which is not consistent with his phenotype. Thus, third generation sequencing provided a more efficient method for rare thalassemia gene testing for this case. Based on its hematological phenotype and clinical manifestations for Hb H disease, it is necessary for the patient to seek further family analysis and genetic counseling.

Most importantly, here we reported a case with 3.8 kb deletion in *HBA2* that shared almost the same deletion junctions with the rare subtype of $-\alpha^{3.7}$ subtype III, which as we know was the first reported case of the $-\alpha^{3.7}$ subtype III in Chinese population. There are at least three subtypes of $-\alpha^{3.7}$ among the population based on the different recombination sites, that is $-\alpha^{3.7}$ I, $-\alpha^{3.7}$ II and $-\alpha^{3.7}$ III. The prevalence and population distribution of each subtype is different, among which $-\alpha^{3.7}$ subtype III is extremely rare in the population compared with the other two subtypes. Although relatively rare, the $-\alpha^{3.7}$ subtype III has been noted with high frequency in Pacific Island peoples like Micronesia and Polynesia, also found with about 2% of α^+ -thalassemia alleles in southern Thai population[30]. In our study, $-\alpha^{3.7}$ III was discovered for the first time in Chinese population using third generation sequencing. The $-\alpha^{3.7}$ subtype III will be prone to be undiagnosed since most of the conventional methods are designed for identification of other two types. The PCR primers used in those PCR kits are usually located around the breakpoints of $-\alpha^{3.7}$ subtype III and homologous sequences in *HBA1* could also lead to the missed diagnosed. The long reads

of third generation sequencing could better cover the whole sequences of *HBA1* and *HBA2*, thus most of the rare deletions or duplications in those two genes can be thoroughly identified.

Conclusions

In conclusion, in this study we identified 10 rare clinically significant variations through third generation sequencing compared to traditional thalassemia gene testing from 100 cases with clinical evidence of suspected thalassemia. The variants including SNVs, rare deletions and triplications in *HBA1* and *HBA2* genes. The results further validated third generation sequencing utilized in this study as a promising genetic testing method for thalassemia carrier diagnosis especially for rare variants identification.

Declarations

Conflicts of interest

The authors declare that they have no conflicts of interest in this article.

Funding

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

Acknowledgements

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

Ethical Approval

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

References

1. Taher AT, Weatherall DJ, Cappellini M. Thalassaemia. The Lancet, 2017, S0140673617318226
2. Weatherall, J. D. The inherited diseases of hemoglobin are an emerging global health burden. Blood, 2010, 115: 4331-4336
3. Longo D, Piel F, Weatherall D. The α -thalassemias. 2014,
4. Thein, Lay S. Molecular basis of β thalassemia and potential therapeutic targets. Blood Cells Molecules and Diseases, 2017, S1079979617302103

5. Mettananda S, Higgs DR. Molecular basis and genetic modifiers of thalassemia. *Hematology/Oncology Clinics*, 2018, 32: 177-191
6. Viprakasit V, Ekwattanakit S. Clinical classification, screening and diagnosis for thalassemia. *Hematology/Oncology Clinics*, 2018, 32: 193-211
7. Lal A, Goldrich ML, Haines DA, *et al.* Heterogeneity of hemoglobin h disease in childhood. *New Engl J Med*, 2011, 364: 710-718
8. Galanello R, Origa R. Beta-thalassemia. *Orphanet J Rare Dis*, 2010, 5: 1-15
9. Diagnosis WGFPGF, Association TOGDMGBOCM, Shang X, *et al.* Clinical practice guidelines for beta-thalassemia. *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chinese journal of medical genetics*, 2020, 37: 243-251
10. Shang X, Xu X. Update in the genetics of thalassemia: What clinicians need to know. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 2017, 39: 3-15
11. Munkongdee T, Chen P, Winichagoon P, *et al.* Update in laboratory diagnosis of thalassemia. *Frontiers in Molecular Biosciences*, 2020, 7: 74
12. Brancaleoni V, Di Pierro E, Motta I, *et al.* Laboratory diagnosis of thalassemia. *Int J Lab Hematol*, 2016, 38: 32-40
13. Shang X, Peng Z, Ye Y, *et al.* Rapid targeted next-generation sequencing platform for molecular screening and clinical genotyping in subjects with hemoglobinopathies. *EBioMedicine*, 2017, 23: 150-159
14. Liang Q, Gu W, Chen P, *et al.* A more universal approach to comprehensive analysis of thalassemia alleles (catsa). *The Journal of molecular diagnostics: JMD*, 2021, 23:
15. Xu L, Mao A, Liu H, *et al.* Long molecule sequencing: A new approach for identification of clinically significant DNA variants in alpha and beta thalassemia carriers. *Journal of Molecular Diagnostics*, 2020, 22:
16. Waye JS, Eng B, Hanna M, *et al.* A⁺-thalassemia due to a frameshift mutation of the α 2-globin gene [codons 55/56 (+t) or hba2: C.168dup]. *Hemoglobin*, 2015, 39: 1-2
17. Qadah T, Finlayson J, Dennis M, *et al.* Molecular and cellular analysis of three novel alpha2-globin gene promoter mutations [hba2. *Pathology*, 2014, 46: 46-52
18. Liang CC, Chen S, Yang K, *et al.* Hemoglobin beijing [alpha 16 (a14) lys replaced by asn]: A new fast-moving hemoglobin variant. *Hemoglobin*, 1982, 6: 629
19. Diagnosis WGFPGF, Association TOGDMGBOCM, Shang X, *et al.* Clinical practice guidelines for alpha thalassemia. *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chinese journal of medical genetics*, 2020, 37: 235-242
20. Zhang Q, Fan X, Xu M, *et al.* Hb h disease caused by multiple mutations in the polyadenylation signal site and – sea/ $\alpha\alpha$. *Hemoglobin*, 2017, 41: 189-192
21. Ju A-P, Jiang F, Li J, *et al.* Detection of an α -globin fusion gene using real-time polymerase chain reaction-based multicolor melting curve. *Hemoglobin*, 2020, 44: 427-431

22. Tca B, Ks B, Gf B, *et al.* Molecular characteristics of α + -thalassemia (3.7kb deletion) in southeast asia: Molecular subtypes, haplotypic heterogeneity, multiple founder effects and laboratory diagnostics. Clin Biochem, 2019, 71: 31-37
23. Ju L, El A. The carriage rates of $\alpha\alpha$ anti3.7 , $\alpha\alpha$ anti4.2 , and hka α in the population of guangxi, china measured using a rapid detection qpcr system to determine cnv in the α -globin gene cluster. Gene, 2020,
24. Wenger AM, Peluso P, Rowell WJ, *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol, 2019, 37:
25. 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, 2021, 1-13
26. De Rycke M, Berckmoes V. Preimplantation genetic testing for monogenic disorders. Genes, 2020, 11: 871
27. Jiang F, Zuo L, Li D, *et al.* Molecular epidemiology and hematologic characterization of $\delta\beta$ -thalassemia and hereditary persistence of fetal hemoglobin in 125,661 families of greater guangzhou area, the metropolis of southern china. BMC Med Genet, 2020, 21: 1-12
28. Giglioni B, Casini C, Mantovani R, *et al.* A molecular study of a family with greek hereditary persistence of fetal hemoglobin and beta-thalassemia. The EMBO journal, 1984, 3: 2641-2645
29. Martyn GE, Wienert B, Yang L, *et al.* Natural regulatory mutations elevate the fetal globin gene via disruption of bcl11a or zbtb7a binding. Nat Genet, 2018, 50: 498-503
30. O'Shaughnessy D, Hill A, Bowden D, *et al.* Globin genes in micronesia: Origins and affinities of pacific island peoples. Am J Hum Genet, 1990, 46: 144

Figures

Figure 1

The identification and verification of clinically significant SNVs. A-F. The identified six SNVs in *HBA2* and *HBB* gene by third generation sequencing. Top graphs are the results exported from the Integrative Genomics Viewer (IGV), the bottom graphs are the Sanger sequencing results for each variants.

Figure 2

The identification of multiple mutations in the polyadenylation signal site and compound $-\text{SEA}/\alpha\alpha$.

A. $-\text{SEA}/\alpha\alpha$ identified by third generation sequencing. **B.** Seven mutations in the polyadenylation signal site found by third generation sequencing. **C.** identification by specific PCR and agarose gel

electrophoresis. M: Maker; 1: positive control for HBA2 fusion gene; 2: negative control; 3: sample for participant #1.

Figure 3

The identification and verification of $\alpha^{3.7}$ subtype III

A. The 3.8 kb deletion identified by third generation sequencing. **B.** Verified by Sanger sequencing. The sequences in the red box are these homologous sequences shared by *HBA1* (chr16:173680-173706) and *HBA2* (chr16:177492-177518).

Figure 4

The identification of rare deletion of 11.1 kb in *HBA1/2*

The 11.1 kb deletion identified by third generation sequencing.

Figure 5

The identification and verification of α -Globin gene triplication

A. The α -Globin gene triplication of $\alpha\alpha^{3.7}/\alpha\alpha$ identified by third generation sequencing. **B.** identification by specific PCR and agarose gel electrophoresis. M: Maker; 1: sample for participant #10; 2: positive control for $\alpha\alpha^{3.7}/\alpha\alpha$; 3: negative control; 4: Blank.

Figure 6

Identification of SNVs in *HBG1* by Sanger sequencing

The two SNVs of *HBG1*: -196 (C->T) (top) and *HBG1*: +25 (G->A) (bottom) found by Sanger sequencing for the full length of *HBG1* gene.