

Evaluation of Ion Torrent next-generation sequencing for thalassemia diagnosis

Journal of International Medical Research

48(12) 1–9

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0300060520967778

journals.sagepub.com/home/imr



Peisong Chen[#] , Xuegao Yu[#], Hao Huang, Wentao Zeng, Xiaohong He, Min Liu and Bin Huang

Abstract

Introduction: To evaluate a next-generation sequencing (NGS) workflow in the screening and diagnosis of thalassemia.

Methods: In this prospective study, blood samples were obtained from people undergoing genetic screening for thalassemia at our centre in Guangzhou, China. Genomic DNA was polymerase chain reaction (PCR)-amplified and sequenced using the Ion Torrent system and results compared with traditional genetic analyses.

Results: Of the 359 subjects, 148 (41%) were confirmed to have thalassemia. Variant detection identified 35 different types including the most common. Identification of the mutational sites by NGS were consistent with those identified by Sanger sequencing and Gap-PCR. The sensitivity and specificities of the Ion Torrent NGS were 100%. In a separate test of 16 samples, results were consistent when repeated ten times.

Conclusion: Our NGS workflow based on the Ion Torrent sequencer was successful in the detection of large deletions and non-deletional defects in thalassemia with high accuracy and repeatability.

Keywords

Next-generation sequencing, Ion Torrent, Thalassemia, Sequencing

Date received: 24 November 2019; accepted: 21 September 2020

Department of Clinical Laboratory, First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[#]Contributed equally.

Corresponding author:

Bin Huang, Department of Laboratory Medicine, First Affiliated Hospital of Sun Yat-sen University, 58 Zhongshan road II, Guangzhou, Guangdong, China. Email: Hb906@163.com

Introduction

Thalassemia is one of the most common genetic disorders worldwide.^{1,2} In southern China, the prevalence of thalassemia has



been estimated to be approximately 10–15%, which is much higher than other areas.^{3,4} However, because of economic improvements and population migration, thalassemia is thought to be spreading to other parts of China and so in the future is likely to become a major public health concern.⁵

Screening and prevention programmes for thalassemia have been widely adopted,¹ but they require accurate identification of couples at high risk of this genetic disorder. Thalassemia is classified into two types, namely, alpha- and beta-, according to defects in the globin genes (HBA and HBB, respectively).^{4,6} Although more than 200 different thalassemia-causing mutations have been identified in the HBB gene,⁷ usually only 6–20 mutations account for the large majority of HBB disease-causing alleles in each at-risk population.⁶ This finding has greatly facilitated molecular genetic testing. Mutations of the HBB gene can be detected by a number of polymerase chain reaction (PCR)-based procedures including, restriction fragment length polymorphism (RFLP) analysis, denaturing gradient gel electrophoresis,⁹ reverse dot-blot hybridization (RDB),¹⁰ amplification refractory mutation system (ARMS),¹¹ and real-time PCR.¹²

Alpha-thalassemia most frequently is caused by deletions of one ($-\alpha$) or both ($-\alpha\alpha$) HBA genes from the normal chromosome ($\alpha\alpha$).¹³ In Southern China, the Southeast Asian deletion ($-\text{SEA}$) is the most common, followed by the rightward deletion ($-\alpha^{3,7}$), leftward deletion ($-\alpha^{4,2}$), non-deletional Haemoglobin (Hb)^{Constant Spring} and Hb^{Quong Sze}.¹³ PCR related techniques such as Gap-PCR and multiplex ligation-dependent probe amplification (MLPA) are typically used for detection of common deletions in alpha-thalassemia.¹³

Currently, separate tests are required for alpha- and beta-thalassemia and so testing

is complex, time consuming, and cumbersome. However, over recent years, tremendous improvements in sequencing technology and computational methods have led to the emergence of next-generation sequencing (NGS) platforms that have drastically decreased the time and cost associated with comprehensive genome analysis.^{12,14,15} This process allows for the simultaneous evaluation of many genes and the generation of millions of DNA fragments in parallel. In addition, the application of NGS in many clinical laboratories has facilitated the detection of novel genetic mutations and rare aberrations.^{16,17}

The purpose of this study was to establish and evaluate a NGS workflow using an Ion Torrent next-generation sequencer (Life Technologies) as a routine method for comprehensive genetic screening of thalassemia.

Methods

Subjects

In this prospective study, blood samples were obtained from subjects of any age who attended the clinical laboratory department at the First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China from 01 July to 31 October 2015 for genetic testing of suspected thalassemia. Inclusion criteria were: Hb < 120g/L; mean corpuscular volume (MCV) < 28fl; HbA₂ outside the normal range of 2.5–3.5%; a family history of thalassemia. This study was approved by the Research Ethics Committee of the Sun Yat-sen University. Written informed consent for participation was obtained from subjects prior to sample collection.

DNA extraction

Venous blood samples (2ml) were collected into EDTA-anticoagulated tubes and stored at -80°C until analysis. Genomic

DNA were extracted using the QIAamp DNA Blood Kits (Qiagen, Valencia, CA). The DNA samples were quality-checked on agarose gels and quantified using a micro-volume spectrophotometer (NanoDrop 1000; Thermo Fisher Scientific Inc). For each sample, a total amount of 5ng genomic DNA was used for subsequent PCR amplification.

Amplification

To sequence the alpha-thalassemia, seven primers were designed using the freely available software Primer Premier 5.0 (Thermo Fisher Scientific Inc.). Multiplex PCR was performed to reduce the number of amplification reactions. The resulting amplicons were then purified using AMPure beads, and the concentration was measured using the Qubit dsDNA HS Assay kit (Life Technologies) and the fragment size was detected using Agilent 2100 Bioanalyzer.

Mutation analysis

A work flow diagram to illustrate our sequencing strategy is shown in Figure 1. Data analysis was performed using the Genome Analysis Toolkit (GATK) Best Practices guidelines.¹⁸ Picard (v.1.69) was used to duplicate reads for each raw Binary Alignment Map (BAM) file. For detection of variants, we applied the GATK UnifiedGenotyper, GATK HaplotypeCaller, and bcftools utility in SAMtools.¹⁹ Variant calling was performed on each sample separately as well as all samples simultaneously (combined analysis). GATK tools were used to recalibrate base mass value of reads and the local realignment near indel in the BAM file. Guidance published by the American College of Medical Genetics and Genomics (ACMG) was used in the interpretation of sequence variants.

Library preparation and Ion Torrent sequencing were performed according to

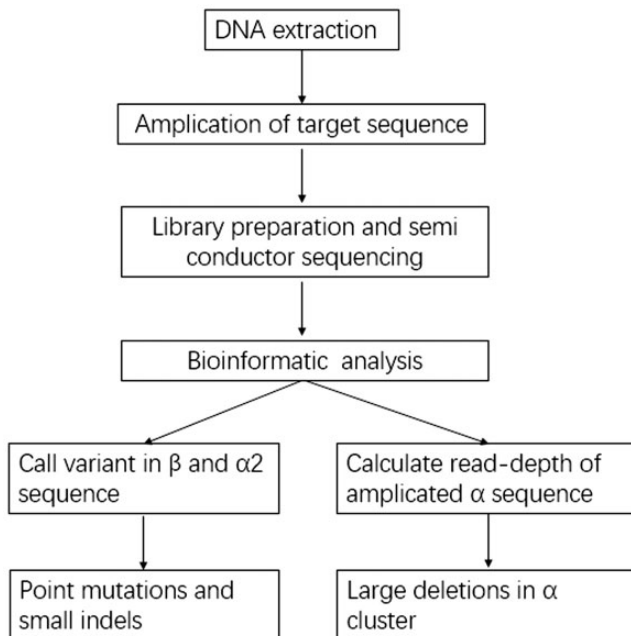


Figure 1. Workflow diagram.

standard procedures detailed in the manufacturer's guideline and previous publication.²⁰ The Ion Xpress™ Plus Fragment Library Kit for AB Library Builder™ System, Ion PI™ Chip Kit v2 and Ion PI™ Sequencing 200 Kit v2 chemistry (200 base pairs [bp] read length, Life Technologies, Thermo Fisher Scientific, Waltham, MA) were used. For all samples, a fragment library was constructed and sequenced with 500 s flows ensuring that 200 bp read length could be achieved. The distribution of read lengths was similar on all chips with the highest peak around 150 bp. On average, 9.5 Gb sequence data, corresponding to ~73.3 million single reads with a mean read length of 130 bp was produced by each chip. For each sample, an average of 0.5 M of reads was generated per sample, and the average sequencing depth for each sample was >500x. All reads were aligned to the human reference genome sequence using TorrentSuit 3.6 software with default settings. The quality of obtained alignments was assessed using FastQC v0.7.2.

To identify the known deletions causing alpha-thalassemia, three paired primers were designed based on Gap-PCR method and included: Chr16: 215284–215308; Chr16: 235908–235888; Chr16: 221913–221936; Chr16: 227726–227702; Chr16: 219303–219327; Chr16: 225173–225148; Chr16: 223631–223612. The read depth of each base was calculated using an analysis plugin (backup depth). Samples with the read depth of each base over 500x in both regions of Chr16: 221913–221936 and Chr16: 223631–223612 and not in other regions, were considered not to have large deletions. Samples with the read depth of each base over 500x in both Chr16: 219303–219327 and Chr16: 225173–225148, and not in other regions, were considered to indicate α -thalassemia 4.2-kb ($-\alpha^{4.2}$ deletion) homozygous deletion. Samples with the read depth of each base

over 500x in both regions of Chr16: 221913–221936 and Chr16: 227726–227702 and not in other regions, were considered to indicate α -thalassemia 3.7-kb homozygous deletion ($-\alpha^{3.7}$ deletion). Samples with the read depth of each base over 500x in both regions of Chr16: 235908–235888 and Chr16: 215284–215308 and not in other regions, were considered to indicate α -thalassemia SEA homozygous deletion ($-\alpha^{SEA}$). Samples with the read depth of each base over 500x in the regions mentioned above and also in Chr16: 221913–221936 and Chr16: 223631–223612, were considered to have loss of heterozygosity (LOH).

Validation of variants by Sanger sequencing

The sequence of $\alpha 2$ cluster and β cluster were characterised by Sanger DNA sequencing analysis. All regions containing variants were amplified by PCR and sequenced by a 3730xl DNA Analyzer (Thermo Fisher Scientific, USA) according to standard procedures.

Validation of large deletion by Gap-PCR

Gap-PCR was used to detect three common α -thalassemia deletions ($-\alpha^{3.7}$, $-\alpha^{4.2}$ and $-\alpha^{SEA}$) with commercial kits from Yaneng Biosciences Ltd (Shenzhen, China) according to the manufacturer's instructions.

Results

In total, blood samples were provided by 359 subjects (156 males and 203 females; age range, 7 months to 72 years). Of the 359 samples, 82 (23%) were from the haematology department, 103 (29%) the obstetrics department, 90 (25%) the paediatrics department, and 84 (23%) the antenatal counselling department.

PCR-amplification, NGS and analysis

The results of variant detection identified 35 different types, including the most common variants which cause α and β thalassemia. Of the 359 subjects, 148 (41%) were confirmed to have thalassemia. Point mutations in the beta-cluster (48 subjects) and deletions in the alpha-cluster (49 subjects) were the main causes of thalassemia. Point mutations in the alpha-cluster were detected in 15 subjects and more than one variant was detected in 36 subjects.

Validation by Sanger sequencing and Gap-PCR

To validate point mutations and small indels detected by Ion Torrent sequencing, genomic DNA fragments were amplified and sequenced by Sanger sequencing methods. For the 359 samples, the mutational sites were consistent between the two methods (Table 1). To validate deletions in the alpha-cluster, multiplex Gap-PCR was conducted and the results were consistent with Ion Torrent sequencing (Table 2). Results indicated that the sensitivity and specificity of these methods for thalassemia diagnosis were 100%.

Repeatability and stability

To assess repeatability and stability of the method, 16 samples were chosen at random and processed ten times using Ion Torrent

sequencing. All samples provided consistent results (Table 3). The mean coefficient of variance (CV) of the mutation frequency of heterozygous mutations was under 10%. These data indicated that the method had high stability and the results were reproducible.

Discussion

Thalassemia is the most common autosomal recessive single-gene disorder in Southern China.²¹ Over the past decade, various diagnostic methods have been developed and with advances in NGS technologies, more effective methods are being tested and introduced into clinical screening of genetic diseases than were previously available. However, there is a need for evaluation of each new method that generates sequencing data comparing it against other sequencing methods as well as non-sequence-based techniques.^{22,23} The purpose of this present study was to investigate if NGS can be a suitable screening or diagnostic method for the sequencing of large numbers of samples from an area in Southern China with a high incidence of thalassemia.

We designed multiple PCR amplification of HBA and HBB genes for targeted sequencing. Our primers were designed to cover the complete HBA and HBB spectrum of mutations and deletions thought to occur in the local population.

Table 1. Analysis of (alpha- and beta-) thalassemia point mutations by Ion Torrent and Sanger sequencing.

		Sanger sequencing		
		Positive	Negative	No samples
Ion Torrent	Positive	99	0	99
	Negative	0	260	260
	No samples	99	260	359

The sensitivity of Ion Torrent was calculated as $99/(99 + 0) * 100\%$; the specificity of Ion Torrent was calculated as $260/(0 + 260) * 100\%$; the consistency of Ion Torrent was calculated as $99 + 260/(99 + 0 + 0 + 260) * 100\%$.

Table 2. Analysis of alpha-thalassemia large deletion by Ion Torrent and Gap-PCR.

Ion Torrent	Gap-PCR		
	Positive	Negative	No samples
Positive	73	0	73
Negative	0	286	286
No samples	73	286	359

The sensitivity of Ion Torrent was calculated as $73/(73+0)*100\%$; the specificity of Ion Torrent was calculated as $286/(0+286)*100\%$; the consistency of Ion Torrent was calculated as $73+286/(73+0+0+286)*100\%$.

Table 3. Repeatability and stability of the results from Ion Torrent sequencer.

Sample	Results	Repeat Count	Correct Results	Incorrect Results	Undetected	Agreement Rate
1	CD125(CTG>CCG) Heterozygous mutation	10	10	0	0	100%
2	CD122(CAC>CAG) Heterozygous mutation	10	10	0	0	100%
3	DEL (Sea Heterozygous)	10	10	0	0	100%
4	IVS-I-I(G>T) Heterozygous mutation	10	10	0	0	100%
5	CD26(GAG>AAG) Heterozygous mutation	10	10	0	0	100%
6	-28(A>G) Heterozygous mutation	10	10	0	0	100%
7	CD17(AAG>TAG) Heterozygous mutation	10	10	0	0	100%
8	Initiation codon Heterozygous mutation	10	10	0	0	100%
9	IVS-II-654(C>T) Heterozygous mutation	10	10	0	0	100%
10	CD71-72(+A) Heterozygous mutation	10	10	0	0	100%
11	CD41-42(_CTTT) Heterozygous mutation	10	10	0	0	100%
12	Wild Type	10	10	0	0	100%
13	Wild Type	10	10	0	0	100%
14	Wild Type	10	10	0	0	100%
15	DEL (3.7 Heterozygous)	10	10	0	0	100%
16	DEL (4.2 Heterozygous)	10	10	0	0	100%
Totals		$\sum a = 160$	$\sum b = 160$	$\sum c = 0$	$\sum d = 0$	100%

Agreement rate of repeatability, $\sum b/\sum a = 100\% = 160/160$.

We decided that targeted sequencing and multiple PCR amplification would significantly reduce costs and avoid lengthy procedures. To evaluate our methods, we

obtained blood samples from a population of 359 subjects originating from four different departments within the hospital (i.e., haematology, obstetrics, paediatrics and

psychotherapy). The selection of these departments and our chosen inclusion criteria increased our opportunity to include different types of mutations or deletions in the study.

We found that $\alpha^{3.7}$ and $\alpha^{4.2}$ heterozygosis were common in our group of subjects from Guangzhou. The sequencing data showed that 35 different types of thalassemia were retrieved that covered most of the common mutations or deletions. Results from the NGS test were consistent with Sanger sequencing and traditional Gap-PCR analysis and 100% sensitivity and specificity were achieved. In addition, the results were stable and repeatable; the mean CV for mutation frequency was less than 10%. and the results were consistent when repeated ten times. These data suggest that the NGS method is a robust method and will provide reliable results as a routine test.

An advantage of using NGS in the screening of thalassemia carriers is that it may assist in the detection of novel and rare variants.²⁴ While no unusual cases were identified in this study, this may have been related to the relatively small sample size because rare variants have been estimated to occur at less than 3%.^{4,24} An additional advantage of NGS is that compared with traditional methods such as RDB, Sanger sequencing and Gap-PCR, NGS requires less DNA, sometimes as low as 10ng/ul. This may be useful in prenatal diagnosis or other situations when only a small sample can be collected. Furthermore, another benefit of the NGS procedure in thalassemia detection is its high speed. For example, in this current study, the sequencing and data analysis procedures only required eight hours per run to analyse 100 samples; most of the time was spent on library construction. Moreover, the introduction of automated instruments for the entire workflow will decrease the test turnaround time even further.

Although other laboratories have examined the use of NGS in screening thalassemia, they did not integrate HBA deletions into their procedures.^{24,25} We suggest that compared with previous methods, our integrated sequencing strategy will be much more convenient in clinical practice. In addition, while other NGS protocols have been suggested for use in China based on protocols constructed using the Illumina or Beijing Genomics Institute (BGI) platforms,^{24,25} we suggest that the Ion Torrent platform is more flexible for small or medium size laboratories. The major limiting factor in the application of NGS for screening thalassemia is the cost. However, with the introduction of locally produced instruments and reagents, the costs associated with NGS should reduce significantly in the future. Another possible imitation of NGS, is that the detection of large deletions has been reported to be less accurate compared with MLPA or Comparative Genomic Hybridization (CGH) analysis.²⁶ However, our data showed that the detection of common deletions was accurate. We suggest that the sequencing and data analysis strategy are critical factors in the success of the method.

Our study had some limitations. For example, some deletions previously reported in the local area, such as gamma-delta-beta-thalassemia, the Thailand deletion (-THAI) and the Fil (-FIL) deletion, were not analysed. Ideally, these deletions and more loci, such as globin genes and validated modulators (KLF1, BCL11A, and MYB) should be included in the NGS diagnostic strategy, which is important for the precise diagnosis and treatment of thalassemia. Additionally, our sample size was not large enough to evaluate the performance of NGS in detecting novel or rare mutations in thalassemia. Nevertheless, our data showed that the NGS method may facilitate screening and diagnosis of

thalassemia in our region with high accuracy and repeatability.

Acknowledgements

We would like to thank Dr Liang from Darui Biotechnology for his assistance in the NGS data analysis.

Declaration of conflicting interest

The authors declare that there are no conflicts of interest.

Funding

This work was supported by a grant from the Natural Science Foundation of Guangdong province (No: 2018A0303130246).

ORCID iD

Peisong Chen  <https://orcid.org/0000-0003-0640-3663>

References

1. Taher AT, Weatherall DJ and Cappellini MD. Thalassaemia. *Lancet* 2018; 391: 155–167.
2. Weatherall DJ and Clegg JB. Thalassemia—a global public health problem. *Nat Med* 1996; 2: 847–849.
3. Cai R, Li L, Liang X, et al. Prevalence survey and molecular characterization of alpha and beta thalassemia in Liuzhou city of Guangxi. *Zhonghua Liu Xing Bing Xue Za Zhi* 2002; 23: 281–285. [Article in Chinese, Abstract in English]
4. Lai K, Huang G, Su L, et al. The prevalence of thalassemia in mainland China: evidence from epidemiological surveys. *Sci Rep* 2017; 7: 920.
5. Weatherall DJ. Thalassemia in the next millennium. Keynote address. *Ann N Y Acad Sci* 1998; 850: 1–9.
6. Lau YL, Chan LC, Chan YY, et al. Prevalence and genotypes of alpha- and beta-thalassemia carriers in Hong Kong – implications for population screening. *N Engl J Med* 1997; 336: 1298–1301.
7. Cao A, Galanello R. Beta-thalassemia. *Genet Med* 2010; 12: 61–76.
8. Chan V, Chan TK, Chebab FF, et al. Distribution of beta-thalassemia mutations in south China and their association with haplotypes. *Am J Hum Genet* 1987; 41: 678–685.
9. Cai SP and Kan YW. Identification of the multiple beta-thalassemia mutations by denaturing gradient gel electrophoresis. *J Clin Invest* 1990; 85: 550–553.
10. Sutcharitchan P, Saiki R, Huisman TH, et al. Reverse dot-blot detection of the African-American beta-thalassemia mutations. *Blood* 1995; 86: 1580–1585.
11. Chan K, Wong MS, Chan TK, et al. A thalassaemia array for Southeast Asia. *Br J Haematol* 2004; 124: 232–239.
12. Vrettou C, Traeger-Synodinos J, Tzetzis M, et al. Rapid screening of multiple beta-globin gene mutations by real-time PCR on the LightCycler: application to carrier screening and prenatal diagnosis of thalassemia syndromes. *Clin Chem* 2003; 49: 769–776.
13. Hartevelde CL, Higgs DR. Alpha-thalassaemia. *Orphanet J Rare Dis* 2010 May 28; 5: 13.
14. Ross JS and Cronin M. Whole cancer genome sequencing by next-generation methods. *Am J Clin Pathol* 2011; 136: 527–539.
15. Goodwin S, McPherson JD and McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016; 17: 333–351.
16. Coonrod EM, Durtschi JD, Margraf RL, et al. Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. *Arch Pathol Lab Med* 2013; 137: 415–433.
17. Choi BO, Koo SK, Park MH, et al. Exome sequencing is an efficient tool for genetic screening of Charcot-Marie-Tooth disease. *Hum Mutat* 2012; 33: 1610–1615.
18. Van Der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; 43: 11.10.1–11.10.33.

19. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25: 2078–2079.
20. Viluma A, Sayyab S, Mikko S, et al. Evaluation of whole-genome sequencing of four Chinese crested dogs for variant detection using the ion proton system. *Canine Genet Epidemiol* 2015; 2: 16.
21. Huang H, Xu L, Chen M, et al. Molecular characterization of thalassemia and hemoglobinopathy in Southeastern China. *Sci Rep* 2019; 9: 3493.
22. Lam HY, Clark MJ, Chen R, et al. Performance comparison of whole-genome sequencing platforms. *Nat Biotechnol* 2011; 30: 78–82.
23. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 2012; 30: 434–439.
24. Zhang H, Li C, Li J, et al. Next-generation sequencing improves molecular epidemiological characterization of thalassemia in Chenzhou Region, P.R. China. *J Clin Lab Anal* 2019; 33: e22845.
25. He J, Song W, Yang J, et al. Next-generation sequencing improves thalassemia carrier screening among premarital adults in a high prevalence population: the Dai nationality, China. *Genet Med* 2017; 19: 1022–1031.
26. Zebisch A, Schulz E, Grosso M, et al. Identification of a novel variant of epsilon-gamma-delta-beta thalassemia highlights limitations of next generation sequencing. *Am J Hematol* 2015; 90: E52–54.