BLOOD RESEARCH
VOLUME 56 · NUMBER S1
April 2021
REVIEW ARTICLE

Molecular basis and diagnosis of thalassemia

Jee-Soo Lee, Sung Im Cho, Sung Sup Park, Moon-Woo Seong

Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

p-ISSN 2287-979X / e-ISSN 2288-0011
https://doi.org/10.5045/br.2021.2020332
Blood Res 2021;56:S39-S43.

Received on December 25, 2020
Revised on April 14, 2021
Accepted on April 16, 2021

Correspondence to
Moon-Woo Seong, M.D., Ph.D.
Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea
E-mail: mwseong@snu.ac.kr

Abstract

Thalassemia is characterized by the impaired synthesis of globin chains due to disease-causing variants in α- or β-globin genes. In this review, we provide an overview of the molecular basis underlying α- and β-thalassemia, and of the current technologies used to characterize these disease-causing variants for the diagnosis of thalassemia. Understanding these molecular basis and technologies will prove to be beneficial for the accurate diagnosis of thalassemia.

Key Words Thalassemia, α-globin gene, β-globin gene

INTRODUCTION

Inherited hemoglobin (Hb) disorders are classified into two disease groups: hemoglobinopathy caused by the structural defects in Hb and thalassemia caused by the impaired synthesis of globin, usually of normal structure; both are typically inherited in an autosomal recessive manner [1]. It has been estimated that 70,000 children are born with various forms of thalassemia annually [2]. The term "thalassemia" is derived from the Greek word "thalassa (sea)," as this inherited disease was first described in the Mediterranean region. Thalassemia is also highly prevalent in other regions of the world (e.g., Southeast Asia, the Indian subcontinent, and Africa), where malaria is endemic. Moreover, the prevalence of thalassemia has increased in recent years, due to an increase in migrant populations in other regions, including North America, Northern Europe, Australia, and South Korea. It is estimated that approximately 7% of the global population are thalassemia carriers [3].

Thalassemia is classified into α-, β-, δβ-, γδβ-, δ-, and γ-thalassemia, which are identified by the specific type of globin chain(s) where impaired synthesis occurs. For example, α- and β-thalassemia are deficient in the synthesis of α- and β-globin chains, respectively [4]. Among these various forms, the two most common forms are α- and β-thalassemia [5]. This review describes the genetic basis and molecular diagnosis of thalassemia, focusing on α- and β-thalassemia.

GENETIC BASIS OF HEMOGLOBIN

Human Hb consists of proteins with symmetric pairing of α-like and β-like globin dimers, which form a tetrameric structure, as well as functional units [6]. Individual α-like and β-like globin chains are encoded by two distinct gene clusters: the α-globin gene cluster on the short arm of chromosome 16 and the β-globin gene cluster on the short arm of chromosome 11. The α-globin gene cluster comprises three functional globin genes, the embryonic ζ gene (*HBZ*), and two fetal/adult α (α1 and α2) genes (*HBA1* and *HBA2*) (Fig. 1A). The β-globin gene cluster contains five functional genes, the embryonic ε gene (*HBE*), two fetal Gγ and Aγ genes (*HBG2* and *HBG1*), and adult δ and β (*HBD* and *HBB*) genes (Fig. 1B). These genes are arranged along each chromosome, and are differentially expressed at each stage of development to produce different Hb tetramers [7]. The embryonic Hb includes Hb Portland ($\zeta_2\gamma_2$), Hb Gower-1 ($\zeta_2\varepsilon_2$), and Hb Gower-2 ($\alpha_2\varepsilon_2$), and the fetal Hb consists of $\alpha_2\gamma_2$. In adults, Hb A ($\alpha_2\beta_2$) accounts for 95% of the total Hb, while Hb A2 ($\alpha_2\delta_2$) constitutes the remaining 5% (Fig. 1C) [5]. The upstream region of each α-globin and β-globin gene

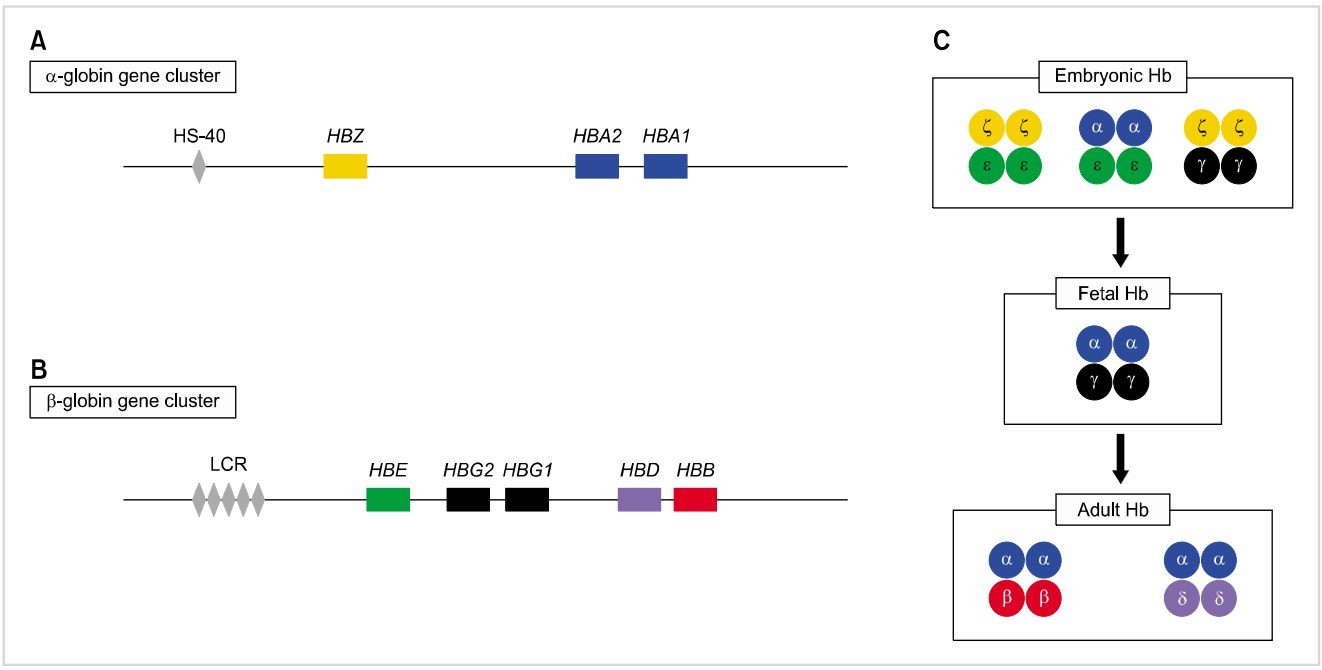

Fig. 1. The human α-globin gene cluster on chromosome 16 and the β-globin gene cluster on chromosome 11. The α-globin gene cluster contains three functional globin genes, the embryonic ζ gene (*HBZ*) and two fetal/adult α, α1 and α2, genes (*HBA1* and *HBA2*) **(A)**. The β-globin gene cluster contains five functional genes, the embryonic ε gene (*HBE*), two fetal Gγ and Aγ genes (*HBG2* and *HBG1*), and adult δ and β (*HBD* and *HBB*) genes **(B)**. HS-40 and the locus control region (LCR) regulate α- and β-globin gene expression, respectively. Hemoglobin differentially expressed at embryonic, fetal, and adult stages are represented **(C)**.

cluster contains cis-acting regulatory elements that play a role in the regulation of globin gene expression. Within 30–70 kb upstream of the α-globin gene cluster, multispecies conserved sequence (MCS) regions (MCS-R1, 2, 3, and 4) were found. MCR-R2, also known as HS-40, is a single DNase hypersensitive site that is crucial for α-globin gene expression [5]. β-globin gene expression is regulated by the locus control region (LCR), which consists of five DNase I hypersensitive sites (HS-1, 2, 3, 4, and 5). β-globin LCR (β-LCR) spans 34 kb upstream of the ε-globin gene.

Table 1. Common variants of each variant type of α-thalassemia.

Gene	Variant type	Common variant	Effect
α-globin gene	--/	$--^{SEA}$, $--^{MED}$, $--^{FIL}$	α^0
	-α/	$-\alpha^{3.7}$ and $-\alpha^{4.2}$	α^+
	$\alpha^T\alpha$/	$\alpha^{IVS1(-5nt)}\alpha$, $\alpha^{PA(AATAAG)}\alpha$, $\alpha^{CS}\alpha$	α^+
	$\alpha\alpha^T$/	*HBA1*: c.223G>C ($Hb^{Q\text{-}Thailand}$)	α^+

MOLECULAR BASIS OF THALASSEMIA

The major pathophysiology of thalassemia is an imbalance in the ratio of globin chains, which is normally well controlled. Unbound globin chains (i.e., α-globin in β-thalassemia and β-globin in α-thalassemia) precipitate, leading to the destruction of erythroid precursors. RBC precursor damage leads to ineffective erythropoiesis in the bone marrow, and RBC hemolysis in circulation [8, 9].

α-thalassemia

The major cause of α-thalassemia is deletions involving one or more α-globin genes with variable lengths of the α-globin locus, which account for approximately 95% of α-thalassemia cases [10]. α-globin genes are duplicated and localized into two highly homologous units, and unequal crossover between these units during meiosis is likely to be the underlying mechanism of gene deletion [1]. Normal individuals have two α-globin genes on each chromosome. In diagnostic practice, α-thalassemia is classified into 1) α^0-thalassemia, in which both α-globin genes are deleted (--/), and 2) α^+-thalassemia, in which one of two α-globin genes is deleted (-α/) [11]. Less frequently, α^+-thalassemia results from non-deletional variants [i.e., single nucleotide variants (SNV) or short insertion/deletions] [$\alpha^T\alpha$/ (α2-globin gene is affected) or $\alpha\alpha^T$/ (α1-globin gene is affected)] (Table 1).

In α^0-thalassemia, the two most common deletion forms are (/$--^{SEA}$, South Asia) and (/$--^{MED}$, Mediterranean). Another type of rare deletion leading to α^0-thalassemia is the deletion of the MCS, in which the α-globin genes remain intact, but are completely inactivated [5, 11]. In α-thalassemia, $-\alpha^{3.7}$ and $-\alpha^{4.2}$ are the most common forms. Unequal recombination between two homologous segments (Z boxes) that are 3.7 kb apart results in the formation of a chromosome with one α-globin gene (/$-\alpha^{3.7}$), and similarly, the other mispaired homologous segment (X boxes), 4.2 kb apart, produces the /$-\alpha^{4.2}$ allele [12]. The non-deletional type includes SNVs

or short insertion/deletions in the α-globin gene, or in regions that affect α-globin expression. More than two-thirds of these variants are observed in the α2-globin gene, while less than one-third of the variants are observed in the α1-globin gene. The products of α2-globin genes account for the majority (~two-thirds) of the total α-globin, while the α1-globin gene accounts for the remainder. Therefore, non-deletional variants in the α2-globin gene would elicit more severe effects than non-deletional variants in the α1-globin gene [13]. In addition, non-deletional variants lead to a greater reduction in α-globin chain expression than the single α-globin gene deletion form of thalassemia [12, 13]. The most common forms that occur in the α2-globin gene are $\alpha^{IVS1(-5nt)}\alpha$ (Mediterranean, 5 nucleotide deletion in IVS1), $\alpha^{PA(AATAAG)}\alpha$ (Middle East Asia, 3′ untranslated region [UTR] polyadenylation site variant), and $\alpha^{CS}\alpha$ (South Asia, stop codon variant resulting in protein extension by an additional 32 amino acids) [14]. Table 2 presents the distribution of the α-thalassemia genotypes identified in the Seoul National University Hospital, South Korea.

β-thalassemia

Unlike α-thalassemia, which is primarily caused by deletions, the majority of β-thalassemia-causing variants are non-deletions, including single nucleotide substitutions and short insertion/deletions leading to frameshift [15, 16]. This disorder is heterogeneous at the molecular level, and more than 300 variants of the β-globin gene have been identified thus far [16]. According to the degree of quantitative reduction in the production of β-globin, β-thalassemia alleles are classified into three categories: 1) the absence of β-globin (β^0); 2) β-globin is produced but reduced (β^+); and 3) β-globin production is minimally reduced (β^{++}, also known as silent).

The categories of non-deletional variants of β-thalassemia are presented in Table 3. Non-deletional variants in β-thalassemia affect one of the following steps: transcription, RNA processing, or translation. Transcriptional variants involve promoter regions or the 5′ UTR. This category of variants generally results in mild, and occasionally silent, reduction of β-globin synthesis (β^+ and β^{++}). Variants that interfere with primary RNA transcript processing are usually found on either canonical or cryptic splice sites, and can lead to either β^0, β^+, or β^{++} thalassemia, depending on the proportion of abnormal mRNA transcripts produced. Other variants that reduce the efficiency of RNA processing include polyadenylation (poly-A) signal variants and those in the 3′ UTR, which generally cause β^+ thalassemia [15, 17]. Variants that produce a premature termination codon (e.g., nonsense or frameshift) account for the most common types of β-thalassemia, and lead to β^0 thalassemia. However, truncating variants in the last exon (exon 3) and the 3′ half of exon 2 are predicted to escape nonsense-mediated decay and produce truncated β-globin to form hemoglobin tetramers that are highly unstable and non-functional, with a dominant negative effect [16]. Variants involving the initiation codon (ATG) leading to β^0 thalassemia have also been identified. In rare cases of β-thalassemia, deletions have been reported and classified into three categories: 1) deletions restricted to the β-globin gene from 105 bp to 67 kb in size; 2) deletions of β-LCR, leaving the β-globin gene intact, yet inactivated; and 3) deletions of β-LCR, which removes most of the β-globin gene cluster, including the β-globin gene [15, 16, 18]. Table 4 presents the molecular spectrum of β-thalassemia identified at the Seoul National University Hospital, South Korea.

CURRENT MOLECULAR DIAGNOSIS

A number of molecular techniques for detecting globin gene variants have been developed. Different strategies should be applied to each variant type, which can be divided into two groups: 1) non-deletional variants, including single nucleotide substitutions and short insertion/deletions, and 2) large deletions and duplications. Disease-causing variants in thalassemia are often population specific, and each population has frequently detected thalassemia alleles [19-21]. Occasionally, the clinical manifestations of thalassemia depend on the type of variant and its location within the gene. For example, in α-thalassemia, non-deletional variants of α-globin genes are associated with more severe phenotypes compared to large deletions [7]. Thus, strategic selection according to the type of variant associated with a specific population and clinical phenotype need to be made in a diagnostic laboratory.

The following discussion focuses on the molecular techni-

Table 2. Distribution of α-thalassemia genotypes identified in Seoul National University Hospital (South Korea).

α-thalassemia genotypes	N	%
$--^{SEA}/\alpha\alpha$	8	33.3
$\alpha^{3.7}/\alpha\alpha$	5	20.8
$--^{SEA}/\alpha^{3.7}$	4	16.7
$\alpha^{3.7}/\alpha^{3.7}$	2	8.3
$\alpha^{3.7}/\alpha$212 patchwork	2	8.3
--/αα (ATRA-16 syndrome)	1	4.2
αα/αα (HS-40 deletion)	1	4.2
$--^{SEA}/-\alpha^{4.2}$	1	4.2
Total	24	

Table 3. Types of variants in β-thalassemia.

Gene	Variant type	Effect
β-globin gene	Transcriptional variants	β^+ or β^{++}
	Primary RNA transcript processing	β^0, β^+, or β^{++}
	3′ UTR or poly-A site	β^+
	Translation	
	Initiation codon	β^0
	Premature termination codon	β^0

Abbreviation: UTR, untranslated region.

Table 4. Molecular spectrum of β-thalassemia identified in Seoul National University Hospital (South Korea).

HGVS nomenclature	Amino acid change	N of alleles	%
c.-138C>A		1	0.6
c.-81A>G		2	1.2
c.79G>A	p.Glu27Lys	18	11.0
c.92G>A	p.Arg31Lys	1	0.6
c.92+1G>A		2	1.2
c.92+1G>T		2	1.2
c.92+2T>C		1	0.6
c.92+5G>C		3	1.8
c.93-21G>A		3	1.8
c.93-1G>C		2	1.2
c.315+1G>A		12	7.4
c.316-197C>T		1	0.6
c.316-2A>G		1	0.6
c.1A>G	Start loss	2	1.2
c.2T>G	start loss	25	15.3
c.48G>A	p.Trp16*	6	3.7
c.52A>T	p.Lys18*	18	11.0
c.114G>A	p.Trp38*	4	2.5
c.118C>T	p.Gln40*	1	0.6
c.364G>T	p.Glu122*	13	8.0
c.394C>T	p.Gln132*	1	0.6
c.25_26del	p.Lys9Valfs*14	1	0.6
c.27dup	p.Ser10Valfs*14	8	4.9
c.126_129del	p.Phe42Leufs*19	15	9.2
c.201del	p.Val68Cysfs*22	1	0.6
c.217dup	p.Ser72Lysfs*2	2	1.2
c.253_254del	p.Thr85Leufs*6	1	0.6
c.270_271del	p.Ser90Argfs*5	2	1.2
c.287dup	p.Leu97Alafs*6	1	0.6
c.378_379del	p.Val127Alafs*13	1	0.6
c.182T>A	p.Val61Glu	1	0.6
c.383A>G	p.Gln128Arg	2	1.2
HBB whole gene deletion		7	4.3
Beta-globin gene cluster deletion		1	0.6
HBG2-HBB deletion		1	0.6
Total		163	

Table 5. Molecular diagnostic methods for thalassemia.

Mutation type	Method
Deletion	Gap PCR
	MLPA[a,b]
Non-deletion	Allele-specific PCR
	Reverse dot blotting
	Denaturing gradient gel electrophoresis
	ARMS
	Sanger sequencing[a]
	Next-generation sequencing[a]

[a]Methods currently applied in clinical laboratories in Korea. [b]Commercial kits available [SALSA MLPA Probemix P140 and P102 (MRC-Holland, Amsterdam, The Netherlands)].
Abbreviations: ARMS, amplification refractory mutation system; MLPA, multiplex ligation-dependent probe amplification; PCR, polymerase chain reaction.

ques available in a clinical laboratory, depending on the type of variant and prior knowledge of the variant to be examined (Table 5).

Deletions

Gap-polymerase chain reaction can be applied to common deletions in a specific population, using primers flanking known breakpoints. Common single α-globin-gene deletions include a 3.7 kb deletion ($-\alpha^{3.7}$) and a 4.2 kb deletion ($-\alpha^{4.2}$). Moreover, two common α-globin-gene deletions include founder variants in specific populations, such as $--^{SEA}$ (Southeast Asian), $--^{FIL}$ (Filipino), and $--^{MED}$ (Mediterranean). The multiplex ligation-dependent probe amplification (MLPA) method is another technique for characterizing deletions in thalassemia, which can detect both known and unknown deletions. MLPA uses two separate oligonucleotide probes (left probe oligonucleotide and right probe oligonucleotide) that are hybridized to adjacent target sequences

and ligated. Ligated probes are amplified via PCR, and the amount of amplified probe ligation products enables the quantification of gene copy numbers. MLPA is simple to perform in clinical laboratories and is suitable for the detection of various deletions [4, 22].

Non-deletions

Several cost-effective methods, such as allele-specific PCR, reverse dot blotting, denaturing gradient gel electrophoresis, and the amplification refractory mutation system can be applied to detect common sequence variants (e.g., Hb Constant Spring in South Asia). Sanger sequencing (i.e., direct DNA sequencing) is currently the most practical method to comprehensively detect all variants without prior knowledge of variants [22]. During Sanger sequencing, the PCR product is obtained and subsequently sequenced using the Sanger dideoxy termination method [4]. However, sequencing of α-globin genes is complex, as two α-globin genes (*HBA1* and *HBA2*) are almost identical, with a length >1 kb. Moreover, sequences of α-globin genes are more guanine-cytosine-rich than those of the β-globin gene; the optimization of PCR conditions is necessary in clinical applications [22]. Recent advances in next-generation sequencing have enabled the detection of novel and structural variants by targeting specific genes or whole genomes [1].

CONCLUSION

Thalassemia is one of the most commonly inherited Hb disorders. The genetic basis of α- and β-thalassemia and molecular techniques applicable in a clinical laboratory for the diagnosis of thalassemia have been described. Understanding the genetic basis of thalassemia and these molecular techniques will have a strong impact on the accurate molecular diagnosis of thalassemia.

Authors' Disclosures of Potential Conflicts of Interest

No potential conflicts of interest relevant to this article were reported.

REFERENCES

1. Lee YK, Kim HJ, Lee K, et al. Recent progress in laboratory diagnosis of thalassemia and hemoglobinopathy: a study by the Korean Red Blood Cell Disorder Working Party of the Korean Society of Hematology. Blood Res 2019;54:17-22.
2. Musharraf SG, Iqbal A, Ansari SH, Parveen S, Khan IA, Siddiqui AJ. β-thalassemia patients revealed a significant change of untargeted metabolites in comparison to healthy individuals. Sci Rep 2017;7:42249.
3. Petrakos G, Andriopoulos P, Tsironi M. Pregnancy in women with thalassemia: challenges and solutions. Int J Womens Health 2016;8:441-51.
4. Munkongdee T, Chen P, Winichagoon P, Fucharoen S, Paiboonsukwong K. Update in laboratory diagnosis of thalassemia. Front Mol Biosci 2020;7:74.
5. Thein SL, Rees D. Haemoglobin and the inherited disorders of globin synthesis. In: Hoffbrand AV, Higgs DR, Keeling DM, Mehta AB, eds. Postgraduate haematology. 7th ed. Hoboken, NJ: Wiley-Blackwell, 2015:72-97.
6. Schechter AN. Hemoglobin research and the origins of molecular medicine. Blood 2008;112:3927-38.
7. Aliyeva G, Asadov C, Mammadova T, Gafarova S, Abdulalimov E. Thalassemia in the laboratory: pearls, pitfalls, and promises. Clin Chem Lab Med 2018;57:165-74.
8. Nienhuis AW, Nathan DG. Pathophysiology and clinical manifestations of the β-thalassemias. Cold Spring Harb Perspect Med 2012;2:a011726.
9. Mettananda S, Higgs DR. Molecular basis and genetic modifiers of thalassemia. Hematol Oncol Clin North Am 2018;32:177-91.
10. Galanello R, Cao A. Gene test review. Alpha-thalassemia. Genet Med 2011;13:83-8.
11. Vichinsky EP. Clinical manifestations of α-thalassemia. Cold Spring Harb Perspect Med 2013;3:a011742.
12. Farashi S, Harteveld CL. Molecular basis of α-thalassemia. Blood Cells Mol Dis 2018;70:43-53.
13. Shang X, Xu X. Update in the genetics of thalassemia: what clinicians need to know. Best Pract Res Clin Obstet Gynaecol 2017;39:3-15.
14. Kalle Kwaifa I, Lai MI, Md Noor S. Non-deletional alpha thalassaemia: a review. Orphanet J Rare Dis 2020;15:166.
15. Thein SL. The molecular basis of β-thalassemia. Cold Spring Harb Perspect Med 2013;3:a011700.
16. Thein SL. Molecular basis of β thalassemia and potential therapeutic targets. Blood Cells Mol Dis 2018;70:54-65.
17. Cao A, Galanello R. Beta-thalassemia. Genet Med 2010;12:61-76.
18. Varawalla NY, Old JM, Sarkar R, Venkatesan R, Weatherall DJ. The spectrum of beta-thalassaemia mutations on the Indian subcontinent: the basis for prenatal diagnosis. Br J Haematol 1991;78:242-7.
19. De Sanctis V, Kattamis C, Canatan D, et al. β-thalassemia distribution in the old world: an ancient disease seen from a historical standpoint. Mediterr J Hematol Infect Dis 2017;9:e2017018.
20. Giardine B, Borg J, Viennas E, et al. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic Acids Res 2014;42:D1063-9.
21. Yang Z, Zhou W, Cui Q, Qiu L, Han B. Gene spectrum analysis of thalassemia for people residing in northern China. BMC Med Genet 2019;20:86.
22. Sabath DE. Molecular diagnosis of thalassemias and hemoglobinopathies: an ACLPS critical review. Am J Clin Pathol 2017;148:6-15.