Open Access Review Article

Molecular Diagnosis of Genetic Haemoglobin Disorders

Suhaib Ahmed

Address for Correspondence
Maj Gen (Dr) Suhaib Ahmed (R), HI (M)
Professor of Haematology
Riphah International University, Islamabad
Genetics Resource Centre (GRC), Rawalpindi
suhaib955@hotmail.com

In 1949 Pauling used electrophoresis to study sickle haemoglobin and showed that it was different from the adult haemoglobin. He used the term "molecular disease" for the first time to describe sickle cell anaemia.[1] Haemoglobin research through physical, chemical, physiological, and genetic methods has played a leading role in the development of "Molecular Medicine". Haemoglobin genes were one of the first to be sequenced. Most of the common and uncommon genetic defects and several single nucleotide polymorphisms linked to the β-globin genes were discovered in the late seventies and early eighties.[2] An almost parallel development in fetal sampling techniques paved the way for prenatal diagnosis .[3] The availability of polymerase chain reaction (PCR) and chorionic villus sampling (CVS) in the mid-eighties revolutionized prenatal diagnosis.[4] This led to a steep decline in the births of children with thalassaemia major in the high risk developed countries.[5] Although prenatal diagnosis is available in several developing countries but most of these are still struggling to benefit from its full potential.[6,7] This review focuses on the molecular basis of genetic haemoglobin disorders in the developing countries, currently available molecular genetic techniques and their applications in various clinical settings.

Structure of globin genes

The globin chains are encoded by an α- and a β-globin gene cluster located on the chromosomes 16 and 11 respectively (Figure 1). The α-globin gene cluster includes an embryonic gene (ζ), two adult genes (α1 and α2), three pseudo-genes (ψζ1, ψα2, ψα1), and a gene of unknown significance (θ1). The β-globin gene cluster comprises of an embryonic gene (ε), the duplicated fetal genes (γ), a pseudo β-gene, a minor adult gene (δ), and an adult β-gene.[8]

Intron-exon structure of the globin genes remains constant throughout the vertebrates. These are one to two kb in size and have three coding regions (exons) that are interrupted by two intervening sequences (IVS or introns).

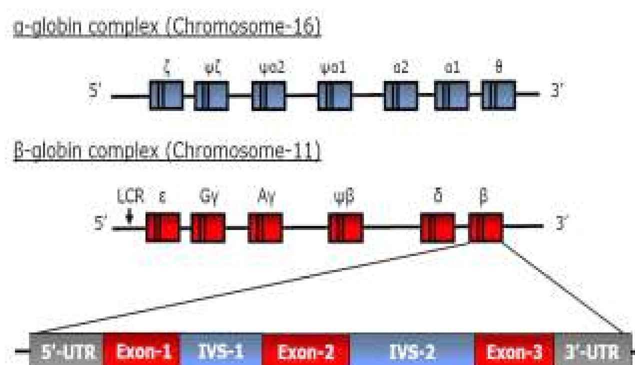

Figure 1. Organization of α- and β-globin gene complexes on the chromosomes 16 and 11 respectively. The structure of the human globin genes remains the same; each gene has three coding regions (exons) separated by two intervening sequences (IVS). The 5' un-translated region (UTR) contains the transcriptional promoter region and the 3' UTR contains the termination and the poly-A tail.

Function of globin genes

The globin genes are expressed in a developmental sequence that shifts from yolk sac to liver in the fetal life and bone marrow in the postnatal life. A locus control region (LCR) has been identified in the cis-

acting sequences for controlling the sequential expression of β-like globin genes (Figure 1).[9]

The region located upstream to the coding sequences of globin genes is called promoter region. It plays an important role in locating the site at which the transcription is initiated. The transcription initiation site for the globin genes correspond to the Cap site located 50 bp upstream from the initiation codon (AUG). The transcription includes exons, introns and sequences beyond the highly conserved 3' "AATAAA" polyadenylation site. An important post transcriptional event is the splicing in which the introns are removed from the transcript and the three exons unite to form mRNA transcript. The latter is transported out of the nucleus and acts as template on which the amino acid residues are sequentially added to form the globin chain. The process is initiated at the initiation codon (AUG) and is terminated when a stop codon (UAA) is encountered.

Types of genetic haemoglobin disorders

Genetic haemoglobin disorders are broadly classified into quantitative defects called thalassaemia syndromes and qualitative defects known as abnormal haemoglobins. Thalassaemias are best classified on the basis of the type of globin gene(s) affected, for example α, β, or δ-thalassaemia etc. There is a long list of thalassaemias involving one or more globin genes. The types are further complicated by the frequent finding of coinheritance of thalassaemias and abnormal haemoglobins. Table I gives a simplified genetic classification of thalassaemia syndromes. The hereditary persistence of fetal haemoglobin is a unique category of genetic haemoglobin disorders that is characterized by persistence of Hb-F in to adult life. But it usually lacks the imbalance in the globin chain production.[10]

A very large number of abnormal haemoglobins have been reported but most of these are of no clinical significance. Relatively few abnormal haemoglobins like Hb-S, Hb-E, and Hb-C are of clinical significance. The abnormal haemoglobins either alone or in combination with thalassaemia may lead to a variety of clinical syndromes.

Table I: Genetic classification of thalassaemias and HPFH.

Common thalassaemias
α-Thalassaemias
α^o–Thalassaemias
Deletion form
Non-deletion form
α^+-Thalassaemias
Deletion form
Non-deletion form
With abnormal haemoglobins like Hb-S, Hb-E etc.
With other Thalassaemias
β-thalassaemias
β^o–Thalassaemias
β^+-Thalassaemias
With abnormal haemoglobins like Hb-S, Hb-E etc.
With other Thalassaemias
Uncommon thalassaemias
$(\delta\beta)^o$–Thalassaemia
$(A\gamma\delta\beta)^o$–Thalassaemia
$(\varepsilon\gamma\delta\beta)^o$–Thalassaemia
δ-Thalassaemia
Hereditary Persistence of Fetal Haemoglobin (HPFH)
Deletion form
Non-deletion form

Molecular pathology of genetic haemoglobin disorders

The thalassaemia syndromes are quantitative defects in which one or more of the globin chains are either absent (α^o and β^o) or are produced at a reduced rate (α^+ and β^+). The abnormal haemoglobins, for example Hb-S and Hb-C, are produced in normal amount but are structurally abnormal and therefore lack the usual qualities of haemoglobin. Some abnormal haemoglobins like Hb-E, in addition to being structurally abnormal are also produced in reduced amount.[11]

α-thalassaemia

Alpha thalassaemias are mostly caused by large deletions.[12] One gene deletion (-α/αα) is the most common α-thalassaemia variant and in some populations over 50% of the people may carry the abnormality. Two gene deletions on the same chromosome (--/αα) is mostly seen in the Fareast and the Mediterranean region and it is responsible for the high frequency of Hb-H disease and hydrops fetalis.[13]. In Pakistan and the its neighboring countries $-\alpha^{3.7}$ kb deletion is the commonest α-thalassaemia variant whereas the $-\alpha^{4.2}$ kb deletion is less common. The numbers of people with non-deletional α-thalassaemia in

Pakistan remains unknown and the triplicated α-gene (ααα/αα) is reported in about 1% of the population.[14]

β-thalassaemia

In contrast to α-thalassaemia majority of the β-thalassaemias are caused by point variants.[10] There are two types of β-thalassaemias; one with some residual β-globin synthesis (β^+-thalassaemia) and the second with no β-globin synthesis (β^o-thalassaemia). Over 200 different variants have been identified to cause β-thalassaemia. Each ethnic population usually has its own set of common and uncommon gene variants (Table II).

Table II: Common β-thalassaemia variants found in the major world populations.[10]

Region	Variants	Old names
Mediterranean	HBB:c.316-106C>G) HBB:c.92+1G>A HBB:c.118C>T HBB:c.92+6T>C HBB:c.316-106C>G Others	IVSI-110 (G-A) IVSI-1 (G-A) Cd 39 (C-T) IVSI-6 (T-C) IVSI-745 (C-G) Others
Middle East	HBB:c.92+5G>C HBB:c.27dupG HBB:c.315+1G>A HBB:c.92+6T>C Others	IVSI-5 (G-C) Fr 8-9 (+G) IVSII-1 (G-A) IVSI-6 (T-C) Others
Indian Subcontinent	HBB:c.92+5G>C HBB:c.27dupG HBB:c.126_129delCTTT HBB:g.71609_72227del 619 HBB:c.92+1G>T Others	IVSI-5 (G-C) Fr 8-9 (+G) Fr 41-42 (-TTCT) Del 619 bp IVSI-1 (G-T) Others
Far East	HBB:c.126_129delCTTT HBB:c.79G>A HBB:c.52A>T HBB:c.316-197C>T Others	Fr 41-42 (-TTCT) Hb-E Cd 17 (A-T) IVSII-654 (C-T) Others

These variants affect the gene expression by a variety of mechanisms. Gene variants in the promoter or the termination region of the β-globin gene are usually mild in nature (β^+). HBB:c.-50A>C (Cap+1 A-C) and HBB:c.-138C>T (-88 C-T) are typical β^+ variants in Pakistan.[15] Point variants in the coding region of the gene (exons) may shift the reading frame (genetic code) forward or backward resulting in creation of a new stop signal (TAA) with premature termination of transcription. HBB:c.27_28insG (Fr 8-9 (+G) and HBB:c.124_127delTTCT (Fr 41-42 (-TTCT) are the common frame shift variants that cause β^o-thalassaemia in the Indian subcontinent.[16] Variants close to the exon-intron junctions create alternate splice sites. This causes abnormal and ineffective splicing of RNA. HBB:c.92+5G>C (IVSI-5 (G-C) is the most common splice junction variant with a very wide distribution extending from the Middle East to the Indian subcontinent.[10] HBB:c.92+1G>T (IVSI-1 (G-T) is another common splice junction variant widely distributed in the Indian subcontinent. The HBB:c.79G>A (Hb-E), in addition to producing a structurally abnormal haemoglobin, also causes abnormal splicing and thalassaemia effect. Hb-E has a very wide distribution in the Fareast.[11] Large deletions are unusual to cause β-thalassaemia. The best known example is of the HBB:g.71609_72227 (619 bp deletion) involving the terminal portion of β-gene. The variant is fairly common in the Sindhis and the Gujrati population .[16]

Rarely variants in the 3rd exon of β-globin gene cause a dominant form of β-thalassaemia, in which the heterozygotes also become symptomatic. HBB:c.332T>C (Cd 110 (CTG-CCG), a dominant β-thalassaemia variant, was recently reported in a Pakistani patient.[17] Occasional patients of typical β-thalassaemia may be seen without any detectable variant in the β-globin gene or its immediate flanking regions. The thalassaemia phenotype in such patients may be caused by variants in the genes located elsewhere that are important in the expression of β-globin gene.[10]

β-thalassaemia variants in Pakistan

The spectrum of β-thalassaemia variants varies widely in different world populations and each ethnic population has its own set of common and uncommon variants. The spectrum of β-thalassaemia variants in Pakistan is given in Tables III.

Molecular genetic techniques

The molecular genetic techniques are mostly based on modifications of polymerase chain reaction (PCR). The most popular methods for the detection of known point variants include Amplification Refractory Mutation System (ARMS), Dot Blot and real time PCR [18]. Genomic sequencing may also be required when the variant is not identified by these methods.

Table III: The spectrum of β-thalassaemia variants in the Pakistani population.[15]

Common variants:		Uncommon variants:		Rare variants:	
IVSI-5 (G-C)	457 (36.9%)	Cd 15 (G-A)	51 (4.1%)	Cd 30 (G-A)	11 (0.9%)
Fr 8-9 (+G)	317 (25.6%)	Cd 30 (G-C)	43 (3.5%)	IVSII-1 (G-A)	10 (0.8%)
Del 619	85 (6.9%)	Cd 5 (-CT)	31 (2.5%)	Hb-S	9 (0.7%)
Fr 41-42 (-TTCT)	82 (6.6%)	Fr 16 (-C)	29 (2.3%)	-88 (C-T)	3 (0.3%)
IVSI-1 (G-T)	65 (5.2%)	Cap +1 (A-C)	20 (1.6%)	IVSI-1 (G-A)	2 (0.2%)
		Hb-E	13 (1.1%)	Fr 47-48 (+ATCT)	2 (0.2%)
				Fr126-131 (-17bp)	2 (0.2%)
				Cd 39 (C-T)	1 (0.1%)
				IVSI minus 25	1 (0.1%)
				Unknown	6 (0.5%)
Total 1240 (100%)	1006 (81.1%)		187 (15.1%)		47 (3.8%)

Amplification Refractory Mutation System (ARMS)

DNA is amplified by primers that are specific for the variant allele. ARMS PCR is also known as allele specific PCR. In PCR a mismatch at the 3’ end of a primer and its target can drastically reduce annealing of the primer and hence the amplification.[19] This is due to the absence of 3’ to 5’ exonuclease proofreading activity of Taq polymerase. High fidelity DNA polymerases, having 3’ to 5’ exonuclease activity, cannot be used in ARMS PCR. Heterozygotes and homozygotes of an ARMS positive sequence variant can be differentiated by an additional step of amplification with a primer specific for the normal allele.

ARMS PCR is done with a pair of primers including a common and an ARMS primer. The common primer is like any other PCR primer. The ARMS primer is usually 25-30 bases in length. The nucleotide at the 3’ end of the primer is complementary to the target nucleotide i.e. G for C and T for A and vice versa. Mismatch at this position can drastically reduce the amplification. For example in a variant with A>T substitution the ARMS primer for the variant allele (T) the last nucleotide should be complementary to the nucleotide T i.e. it should have A. The primer for the normal allele at the same position should be complementary to the nucleotide A i.e. it should have T. An additional mismatch at one of the last five nucleotides of the ARMS primer further increases its specificity.[20]

It is essential to include an internal PCR control in ARMS reactions. A pair of primers is included to amplify an unrelated region of the genome. Amplification of the internal control region and no amplification by the ARMS primer indicate a true negative. In a false negative result neither the ARMS primers nor the internal control primers show any amplification. There could be several reasons for the false negative result e.g. too little or too much DNA, poor quality of target DNA, failure to add primer, Taq, or other reagents and presence of PCR inhibitors.[21] Sensitivity and specificity of an ARMS PCR can be ensured by using stringent reaction conditions. Good primer design, higher annealing temperature and limited number of cycles are important in avoiding false results. The number of cycles should be just enough to give a clear positive result. Increasing the number of cycles un-necessarily can cause false positives. The usual length of the ARMS primer should be 25-30 bases and these should have high Tm and annealing temperature. ARMS PCR may be done in separate PCR tubes for each variant or as a multiplex for more than one variant in the same PCR tube.[22] The ARMS PCR products are usually separated on agarose or polyacrylamide gels (Figure 2).

Table IV: Multiplex ARMS primer combinations for β-thalassaemia variants found in the Pakistani population.[22]

Primer ID:	Variant	Product size
AD-1	Fr 8-9 (+G)	215 bp
	IVSI-5 (G-C)	285 bp
	Fr 41-42 (-TTCT)	439 bp
	IVSI-1 (G-T)	280 bp
	Del 619bp	242 bp
AD-2	Cd 5 (-CT)	205 bp
	Fr 16 (-C)	238 bp
	IVSI-1 (G-T)	280 bp
	Cd 30 (G-C)	280 bp
	Cd 30 (G-A)	280 bp
	IVSII-1 (G-A)	634 bp
AD-3	Cd 15 (G-A)	500 bp
	Cap+1 (A-C)	567 bp

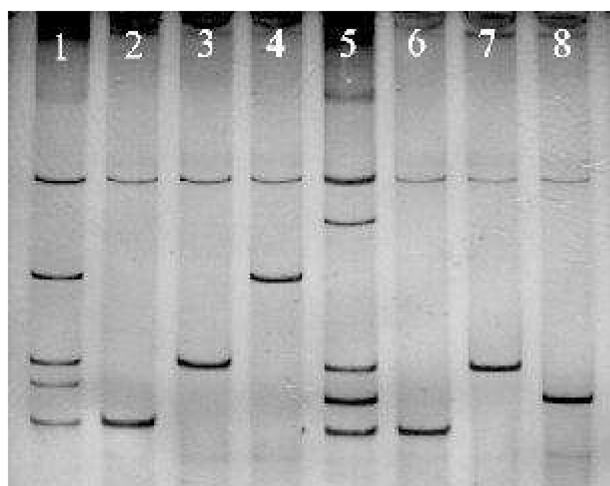

Figure 2. Silver stained mini polyacrylamide gel electrophoresis of multiplex ARMS PCR products. Lane 1 and 5 show allelic ladders for the AD-1 and AD-2 multiplexes. All lanes show the 861bp internal control fragment. The sample in lane 2, 3 and 4 are positive for Fr 8-9 (+G), IVSI-5 (G-C) and Fr 41-42 (-TTCT) variants respectively. Lanes 6, 7 and 8 show Cd5 (-CT), IVSI-1 (G-T) and Fr 16 (-C) variants respectively.

Real Time PCR

High Resolution Melting Curve (HRMC) analysis

DNA sequence containing the variant(s) is amplified in multiple overlapping fragments by the real time PCR method. The PCR is done in the presence of DNA intercalating dyes like SYBR green. At the end of the PCR high resolution melting curve analysis (HRMCA) step is added that measures fluorescence reading at temperature increments of 0.1°C. The Tm of an amplified DNA fragment containing a point variant differs from that of its normal sequence. The difference in the Tm of the two fragments generates different melting curves.[23]

In another type of HRMCA the region of DNA containing the variant(s) is amplified by a pair of primers.[24] Instead of the SYBR green the reaction tube contains two fluorescently labeled probes that hybridize to the adjacent sequences in the target DNA. One of the probes (sensor probe) covers the region expected to contain the variant(s) and is labelled with a fluorochrome at its 3' end. The second probe (anchor probe) lies a few bases upstream to the sensor probe. Close proximity of the annealed probes facilitates fluorescence resonance energy transfer between them. Probes are designed to have different melting temperatures (Tms), so that the sensor probe with lower Tm lies over the variant site(s). Monitoring of the emitted fluorescent signals as the temperature increases detects loss of fluorescence as the probe with the lower *T*m melts off the template. A single base mismatch under this probe produces a Tm shift of 5–10°C, allowing easy distinction between the wild-type and the variant alleles. The ability to detect base mismatches under the probe with the lower Tm (variant detection probe) and the use of two, different-colored probes allows more than one variant to be screened in a single PCR reaction. Because of its low costs, reproducibility, and ease of handling, the assay is potentially suitable for a routine clinical laboratory.

HRMC analysis software is used to generate DNA melting curves. Presence of each point variant generates its own specific melting curve that can be identified by simultaneous running of known positive and negative DNA controls. Extensive optimization of the procedure is required to define HRM curves of known variants in the target population. The patterns generated by heterozygotes, homozygotes and all possible compound heterozygote combinations in the target population are required.

Hydrolysis probe method

Segment of a gene containing the variant is amplified by a pair of primers and a hydrolysis probe (Taqman® probe). A separate pair of probe is required for identification of each variant (one for the variant and the second for the normal allele). The hydrolysis probes are typically labelled at the 5'-end by a fluorescent dye and a quencher at the 3'-end. When a probe anneals to its target the fluorochrome is detached from the probe and starts emitting fluorescence. The procedure can be multiplexed for several variants in one reaction if multi coloured fluorescent dyes are used on each probe. Real time PCR by hydrolysis probes is efficient but expensive because it requires two separate probes for each variant i.e. one for the variant and the other for the normal allele.[25]

Reverse dot blot (RDB)

In the reverse dot blot (RDB) method multiple allele specific oligonucleotide (ASO) probes for the known variants and their normal alleles are immobilized on a nylon membrane.[18] PCR is done with biotin labelled primers and the amplified product is layered on the ASO

probes bound to the nylon membrane. The results are read after incubation and washing at 45°C. ASO probes immobilized on the nylon membrane bind to their target allele if it is present otherwise the amplified product is washed away.[26]

Restriction enzyme method

Restriction endonucleases are enzymes of bacterial origin that can cut DNA at specific sequences called "restriction sites". As a result of digestion by a restriction enzyme the DNA sequence is cut into fragments of varying length. This forms the basis of restriction fragment length polymorphism (RFLP).[18] It is a very useful technique in identification of single nucleotide variants (SNV). The SNVs may be recognized by the presence or the absence of a restriction enzyme site (restriction site).

By convention the presence of a restriction site is written as "+" and its absence as "-". Any DNA sequence containing the SNV is amplified by a pair of flanking primers and the amplified DNA is incubated with the restriction enzyme. If the SNV is present the amplified DNA is cut in to two pieces at a point where the SNV is present (+). If the SNV is not present the DNA fragment remains as single piece (-). The case would be reversed if the SNV abolishes the restriction site. A heterozygote of SNV would have the cut and the uncut fragments (+/-). The homozygote would have only the cut (+/+) or the uncut (-/-) fragments.

Detection of gene deletions and inversions (Gap PCR)

Large deletions in a gene can be detected by gap PCR. The target DNA is amplified by a pair of primers flanking the deletion. In the absence of deletion PCR primers are far apart and no amplification is observed. But the presence of deletion brings the same pair of primers close enough to enable amplification. Most gene deletions causing α-thalassaemia can be detected by gap PCR.[27]

If the gene deletion is present the flanking PCR primers generate an amplification product that is smaller than the amplified product when the deletion is absent. In a heterozygote of such deletion two amplified products are generated i.e. short (deletion fragment) and long (wild-type fragment). In a homozygous deletion only the short fragment is generated. The 619 bp deletion in the distal part of β-globin gene is a typical example.[16] PCR can also be used to detect inversions in the genome. Inv/Del $G_{\gamma}(A_{\gamma\delta\beta})^{o}$ in the β-gene complex is a typical example.[28]

Detection of unknown Variants

The unknown DNA variants may be screened by non-specific methods like Denaturing Gradient Gel Electrophoresis (DGGE) or Single Strand Conformation Polymorphism (SSCP).[21] These methods only indicate whether a variant is present in the DNA or not. The unknown variant if indicated may be confirmed by genomic sequencing.

Genomic sequencing

The process of determining the sequence of an unknown DNA is called sequencing. β-globin gene sequencing is mostly done by the automated Sanger's di-deoxy chain termination method.[21] It is done to identify an unknown variant or to confirm the results of ARMS PCR. The gene is approximately 1.5 kb in length. On a 36 cm capillary of genetic analyzer approximately 600 bases can be sequenced in one go. The entire β-globin gene can be sequenced in three overlapping segments.[17] Most of the β-thalasaemia variants are located in the first two exons and the intervening sequence.

Applications of molecular genetic analysis in genetic haemoglobin disorders

The salient applications of molecular techniques in the diagnosis of genetic haemoglobin disorders include:

Prenatal Diagnosis

Prenatal diagnosis is done to know the presence of a genetic haemoglobin disorder in a fetus before it is born. The objective is to give an informed choice to the parents. If the fetus is affected the parents have the choice to accept it or to get the pregnancy terminated. The test is offered to the couples who are both carriers of a clinically significant genetic haemoglobin disorder. It is also done in all subsequent pregnancies.

The Mediterranean thalassaemia prevention programmes based on carrier detection and genetic counselling alone were mostly ineffective in reducing the birth incidence of thalassaemia major. The introduction of prenatal diagnosis gave a new dimension to all such

programmes. The introduction of prenatal diagnosis led to a sharp decline in the birth incidence of thalassaemia major.[5] Similar results have also been observed in Iran since 2005 when the termination of pregnancy for serious defects in fetus was legally allowed.[29, 30]

Prenatal diagnosis was initially done by globin chain synthesis ratio on fetal blood obtained in mid trimester gestation by fetoscopy or placental aspiration.[31] The practice has now shifted to first trimester fetal diagnosis by chorionic villus sampling (CVS).[32] In-vitro fertilization and embryo transfer has also enabled diagnosis to be carried out on just a few cells obtained from the embryo prior to implantation (pre-implantation genetic diagnosis).[33] Fetal cell free DNA in maternal circulation is also being used for the diagnosis of fetal genetic disorders.[34] But the cell free DNA analysis for prenatal diagnosis of thalassaemia remains complex and impractical especially for a country like Pakistan.

Prenatal diagnosis of genetic haemoglobin disorders is mostly done by direct variant analysis.[32] Occasionally linkage analysis or genomic sequencing may be required when the parent's variants are not identified. Over 200 different variants are known to cause β-thalassaemia. However, each ethnic population has its own set of common and uncommon variants. Therefore, it is essential to have a knowledge about the frequency of the variants in the target population.

The first step in prenatal diagnosis of β-thalassaemia is to identify the parent's variants. In the next step the fetal DNA is tested for the parent's variants. The analysis of variants may be done by ARMS, dot blot analysis, real time PCR or genomic sequencing.[32]

In countries with high frequency of consanguineous marriage the parents usually have identical variants. Less commonly the couples have non-identical variants. The fetal diagnosis is relatively simple when the parents have non-identical variants. If both of the parental variants are present in the fetus it indicates thalassaemia major, the presence of one of the two variants indicates thalassaemia trait, and the absence of both of the variants indicates a normal fetus. When a couple has identical variants the presence of the variant in the fetus needs further testing to ascertain whether the variant is heterozygous or homozygous. It becomes mandatory to test the fetal DNA for the normal allele of the variant. A fetus having the parent's variant and the normal allele would be heterozygous whereas a fetus showing the variant but no normal allele would be homozygous (Figure 3).

ARMS is the cheapest and the most commonly used method for detection of point variants [22]. It is based on a match/mismatch between the target sequence and the 3' end of the ARMS primer. The PCR is done under stringent conditions that allow ARMS primer to anneal only to its specific target sequence.[21]

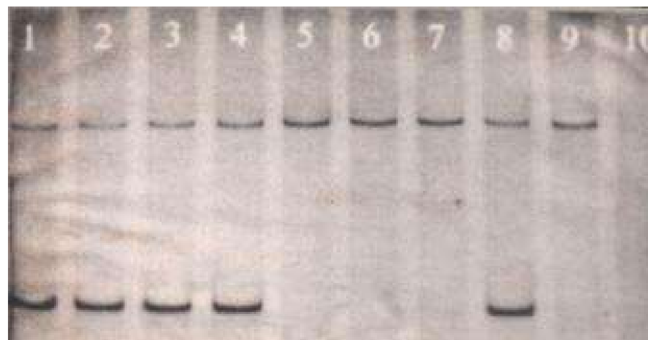

Figure 3. Polyacrylamide gel electrophoresis of prenatal diagnosis of β-thalassaemia. All lanes except lane 10 show 861bp internal control bands. Absence of any result in the reagent blank (Lane 10) excludes the false positive results. Lanes 1 & 2 shows the parent's variants (IVSI-5). Lanes 3 & 4 show the same variant (IVSI-5) tested in duplicate in the fetal DNA. Lane 5 is a negative control for IVSI-5. Lanes 6 & 7 shows testing for the normal allele of IVSI-5. Its absence in the fetal DNA confirms that the fetus has homozygous IVSI-5. Lanes 7 and 8 are negative and positive controls for the normal allele of IVSI-5 respectively. Lane 8 is a negative control for the IVSI-5 variant.

Each variant is identified by a separate allele specific primer. In the first round of PCR common variants in the target population are tested. If the variant is not identified a second round of PCR is done to identify the uncommon variants.[15] The ARMS PCR can be made cost effective by multiplexing the ARMS primers for several variants in one PCR reaction tube.[22] The multiplex ARMS PCR in addition to being cost effective also saves time. Reverse dot blot analysis is another popular method used in prenatal diagnosis.[26] Some centres also do genomic sequencing to confirm the findings of PCR. But this practice would increase the overall cost of the diagnosis and it may not be suitable for the developing countries.[32]

Fetal diagnosis by linkage analysis

In couples when the β-gene variants are not identified the prenatal diagnosis can be done by linkage analysis. There are at least seven single nucleotide variants (SNV) closely linked to the β-globin gene.[35] The SNVs are inherited en-block with the β-gene and can be used to track its inheritance. The presence of SNV on a chromosome is marked "+" whereas its absence is marked "-". The SNV genotype on a homologus pair of chromosomes can be "+/+, -/+, or -/-".

The linkage based prenatal diagnosis is only possible when the couple has a previously affected child whose DNA is used to mark the parental chromosomes. Linkage analysis is not possible if any of the parents is homozygous (+/+) for the marker SNV. Rarely the linkage based diagnosis may turn out to be incorrect if the marker and the gene of interest dissociate from each other during the meiotic crossover.[32]

Errors in prenatal diagnosis

The technical error rate in DNA based prenatal diagnosis should be less than 1:200 (<0.5%) [36]. Common causes of errors in prenatal diagnosis include:

1. Maternal contamination in the fetal sample.
2. Errors due to PCR (false positive or negative results).
3. Sample mix-up and clerical mistakes.
4. Meiotic crossover in linkage analysis.
5. Non paternity.

Maternal contamination in CVS

Mixing of the maternal and the fetal tissue in CVS may cause error in prenatal diagnosis.[36] If the placental aspiration is done close to the decidual plate the sample may be significantly contaminated with maternal decidua. Very occasionally the CVS may comprise only of the maternal decidua. The best practice should be to meticulously dissect the CVS under a stereo microscope and only clean fetal tissue (placental villi) should be picked for DNA extraction. Fetal tissue is recognized by the branched villus structure and very fine blood vessels in the core of the villi. DNA extracted from a CVS with significant maternal contamination can result in amplification of the maternal variant and the normal alleles. This can result in fetal diagnosis of thalassaemia major to appear as thalassaemia trait.[32]

The maternal contamination in fetal DNA can be detected by Short Tandem Repeat (STR) analysis.[37] Careful dissection and cleaning of the CVS under stereomicroscope is the best method to safeguard against maternal contamination. Some labs regularly do STR analysis to look for maternal contamination. But this adds to the overall cost of testing and may not be feasible in countries with resource constraints. Experience has shown that adequate cleaning of the CVS is enough to prevent errors due to maternal contamination.[38]

Errors due to PCR

PCR is a very sensitive technique that can give false positive as well as false negative results.[40] The ARMS PCR must be done under stringent conditions. Slight increase or decrease in the Taq polymerase, problems with the ARMS primers, the PCR buffer and the number of PCR cycles are all important in causing false positive or negative PCR results.[21]

Quality assurance in prenatal diagnosis

The error rate in prenatal diagnosis should not exceed 0.5%.[36] The credibility of the service for prenatal diagnosis can be maintained only by strict quality assurance. The prenatal diagnosis report is final and the chance of repeating the test is usually not available. A missed diagnosis of thalassaemia major becomes a lifetime tragedy. Therefore every effort must be made to ensure quality. The following specific measures may help in keeping the errors to a minimum.[21, 32]

1. CVS should be aspirated from the mid portion of the placenta. A sample taken close to the decidual plate is more likely to be contaminated with maternal tissue.

2. CVS should be adequately labelled with unique identifiers especially if more than one samples are collected at the same time.

3. CVS should be meticulously cleaned/dissected to prevent maternal contamination. The samples should be transported with minimum of delay. If a delay of more than 24 hours is expected it is preferable to dissect the sample before transport.

4. Use good quality PCR consumables. The primers have a tendency to degenerate on repeat freeze thaws. It is best to aliquot the PCR reagents.

5. Use optimum DNA concentration. Too little or too

much DNA can give erroneous results.

6. Always include PCR internal controls. It is best to repeat the runs when the internal control does not work.
7. The fetal DNA should always be tested in duplicate.
8. Include known positives, negatives and reagent blanks. The reagent blank contains everything except the DNA. Amplification in the reagent blank indicates contamination with extraneous DNA and it should invalidate the result.
9. Most of the errors in prenatal diagnosis are seen in fetal diagnoses of thalassaemia trait. This is because maternal contamination in CVS can make any fetal diagnosis look like thalassaemia trait. In a lab that routinely does prenatal diagnosis the proportion of fetal diagnoses with "thalassaemia trait", "thalassaemia major" and "normal" should be checked at regular intervals. The proportion of thalassaemia traits should be very close to 50%. A proportion significantly greater than 50% should ring the alarm bell. Similarly the proportion of "thalassaemia major" and "normal" should be close to 25%. A higher proportion of "thalassaemia major" or "normal" should lead to check on the integrity of the ARMS primers for the "normal" or the "mutant" genes respectively.

Diagnosis in previously transfused patients

Repeated blood transfusions can alter the picture of thalassaemia to a great extent. In such patients PCR may be used to arrive at a diagnosis.[40] The variant analysis is not affected by a recent blood transfusion.

Silent or atypical thalassaemia alleles

Individuals with β^{++} variants like Cap+1 (A-C) or those with co-existing β and α-globin variants may have normal haematological picture and even Hb-A_2 level. Their carrier status can only be ascertained by PCR.[32]

Distinction between abnormal haemoglobins

Differentiation between common or uncommon abnormal haemoglobins like Hb-E & Hb-C or Hb-S & Hb-D may require PCR.[32]

Thalassaemia Intermedia

Thalassaemia intermedia is a clinical entity and its diagnosis is most done by clinical assessment and history of blood transfusions. A prospective or even prenatal diagnosis of thalassaemia intermedia can be made by molecular genetic analysis including the type of globin chain variant and analysis of various quantitative trait loci.[10]

α-thalassaemia

Carriers of α-thalassaemia are best identified by PCR. The deletional types of α-thalassaemia are identified by Gap PCR whereas the non-deletional α-thalassaemias are identified by ARMS or similar techniques.[14, 41]

Rare thalassaemias

Unusual forms of thalassaemia like δβ-thalassaemia etc. are best identified by PCR.[28]

Quantitative Trait Loci (QTL) linked to β-globin genes

Many patients of thalassaemia major are able to produce greater amount of Hb-F than others. At least three major Quantitative Trait Loci (QTL) have been identified to be linked with high Hb-F production. These include −158 C>T *Xmn*-I Single Nucleotide Variant (SNV) located in the Gϒ-globin gene promoter.[10] a SNV in the *BCL11A* gene[42] and several SNVs in the *HBSB1L-cMYB* gene[43] All of these SNVs can be identified by restriction enzyme analysis or by ARMS PCR.[21]

Pre-implantation genetic diagnosis (PGD) of genetic haemoglobin disorders

The in-vitro fertilized embryos can be screened for genetic abnormalities including thalassaemia [33]. The embryos are sampled by polar body or tropho-ectoderm biopsy. The DNA from the biopsied cells is amplified by nested PCR (two step amplification) or by Next Generation sequencing (NGS).[44] The embryos found healthy after screening are used for implantation. The procedure is expensive and technically demanding therefore it is not suitable for the developing countries. It may be useful for couples who do not accept termination of pregnancy.

References

1. Strasser BJ (2004) Linus Pauling's "molecular diseases": between history and memory. Am J Med Genet 115(2): 83-93. doi: 10.1002/ajmg.10542.
2. Antonarakis SE, Kazazian HH, Jr, Orkin SH (1985) DNA polymorphism and molecular pathology of the human globin gene clusters. *Hum Genet* 69: 1-14.

3. Ward RHT, Modell B, Petrou M, Karagozlu F, Douratsos E (1983) Method of sampling chorionic villi in first trimester of pregnancy under guidance of real time ultrasound. *Br Med J* 286: 1542-1544.

4. Mullis KB and Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155: 335-350.

5. Cao A (1987) Results of programmes for antenatal detection of thalassaemia in reducing the incidence of the disorder. *Blood Rev* 1: 169-176.

6. Modell B and Darlison M (2008) Global epidemiology of haemoglobin disorders and derived service indicators. Bulletin World Health Org 86: 480-487.

7. Weatherall DJ (2005) Keynote address: The challenge of thalassemia for the developing countries. Ann N Y Acad Sci 1054, 11-17.

8. Collins FS Weisman SM (1984) The molecular genetics of human haemoglobin. Progress in Nucleic Acid Research and Molecular Biology (Ed Cohn WE and Moldave K) New York Acadamic Press.

9. Levings, P. P., & Bungert, J. (2002). The human β-globin locus control region. *European journal of biochemistry*, *269*(6), 1589-1599.

10. Weatherall DJ and Clegg JB (2001) The Thalassaemia Syndromes. Blackwell Science, Oxford.

11. Olivieri, N. F., Pakbaz, Z., & Vichinsky, E. (2011). Hb E/beta-thalassaemia: a common & clinically diverse disorder. *The Indian journal of medical research*, *134*(4), 522.

12. Higgs D (2013) The Molecular Basis of α-Thalassemia. Cold Spring Harb Perspect Med 3: a011718.

13. Piel FB, Weatherall DJ, Longo DL (ed) (2014) The α-thalassaemias. N Engl J Med 371, 1908-1916.

14. Khan SN, Hasan F, Sollaino C, Perseu L, Riazuddin S (2003) Molecular characterization of alpha-thalassemia in Pakistan. Hemoglobin 27: 161-166.

15. Ahmed S, Petrou M, Saleem M (1996) Molecular genetics of β-thalassaemia in Pakistan: a basis for prenatal diagnosis. Br J Haematol 94: 476-481.

16. Varawalla NY, Old JM, Sarkar R, Vankatesan R, Weatherall DJ (1991) The spectrum of β-thalassaemia mutations on the Indian subcontinent: the basis for prenatal diagnosis. *Br J Haematol* 78: 242-247.

17. Hussain A, Ahmed S, Ali N, Mailk HS, Anees M, Chuahdry AH, Ahmed P (2017) Rare β-Globin Gene Mutations in Pakistan. Hemoglobin 41: 100-103.

18. GU, X., & Zeng, Y. (2002). A review of the molecular diagnosis of thalassemia. *Hematology*, *7*(4), 203-209.

19. Newton CP, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 17: 2503-2516.

20. Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sininsky JJ (1989) Effect of primer template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Res* 18: 999-1005.

21. Ahmed S (2013) Manual of PCR in Diagnostic Pathology. GRC Publication (ISBN - 978-969-9829-00-0).

22. Ahmed S, Saleem M, Petrou M, Sultana N, Raashid Y, Waqar A, Anwar M, Karamat KA (2000) Prenatal diagnosis of β-thalassaemia in Pakistan: experience in a Muslim country. Prenatal Diagnosis 20: 378-383.

23. Chassanidis, C., Boutou, E., Voskaridou, E., & Balassopoulou, A. (2016). Development of a high-resolution melting approach for scanning beta globin gene point mutations in the Greek and other Mediterranean populations. *PloS one*, *11*(6), e0157393.

24. Moreno I, Bolufer P, Perez ML, Barragan E, Sanz MA. Rapid detection of the major Mediterranean β-thalassaemia mutations by real-time polymerase chain reaction using fluorophore-labelled hybridization probes. Br J Haematol 2002; 119: 554–557.

25. Kho SL, Chua KH, George E, Tan JAMA. High throughput molecular confirmation of β-thalassemia mutations using novel TaqMan probes. Sensors 2013; 13: 2506-2514. doi:10.3390/s130202506.

26. Li, D., Liao, C., Li, J., Huang, Y., Xie, X., Wei, J., & Wu, S. (2006). Prenatal diagnosis of β-thalassemia by reverse dot-blot hybridization in southern China. *Hemoglobin*, *30*, 365-370.

27. Pornprasert, S., Wiengkum, T., Srithep, S., Chainoi, I., Singboottra, P., & Wongwiwatthananukit, S. (2011). Detection of α-thalassemia-1 Southeast Asian and Thai type deletions and β-thalassemia 3.5-kb deletion by single-tube multiplex real-time PCR with SYBR Green1 and high-resolution melting analysis. *The Korean journal of laboratory medicine*, *31*(3), 138-142.

28. Ahmed S, Anwar M (2005) Xmn-I Gγ-polymorphism in six unrelated Pakistani families with Inv/Del $^{G}\gamma(^{A}\gamma\delta\beta)^{o}$ δβ-thalassaemia. Am J Haematol 80: 303-305.

29. Joulaei, H., Shahbazi, M., Nazemzadegan, B., Rastgar, M., Hadibarhaghtalab, M., Heydari, M., & Rahimi, N. (2014). The diminishing trend of β-thalassemia in Southern Iran from 1997 to 2011: the impact of preventive strategies. *Hemoglobin*, *38*(1), 19-23.

30. Abolghasemi, H., Amid, A., Zeinali, S., Radfar, M. H., Eshghi, P., Rahiminejad, M. S. & Akhavan-Niaki, H. (2007). Thalassemia in Iran: epidemiology, prevention, and management. *Journal of Pediatric Hematology/Oncology*, *29*(4), 233-238.

31. Fairweather DVL, Model B, Berdoukas V, Alter BP, Nathan DG, Loukopoulas D, Wood W, Cleg JB, Weatherall DJ (1978) Antenatal diagnosis of thalassaemia major. *Br Med J* 1: 350-353.

32. Ahmed S (2017) Genetic Haemoglobin disorders in Pakistan. GRC Publication. (ISBN 978-969-23272-0-6).

33. Kokkali, G., Traeger-Synodinos, J., Vrettou, C., Stavrou, D., Jones, G. M., Cram, D. S., & Pantos, K. (2007). Blastocyst biopsy versus cleavage stage biopsy and blastocyst transfer for preimplantation genetic diagnosis of β-thalassaemia: a pilot study. *Human Reproduction*, *22*(5), 1443-1449.

34. Yang, X., Zhou, Q., Zhou, W., Zhong, M., Guo, X., Wang, X., & Wang, Y. (2019). A Cell-free DNA Barcode-Enabled Single-Molecule Test for Noninvasive Prenatal Diagnosis of Monogenic Disorders: Application to β-Thalassemia. *Advanced Science*, *6*(11), 1802332.

35. Varawalla, N. Y., Fitches, A. C., & Old, J. M. (1992). Analysis of β-globin gene haplotypes in Asian Indians: origin and spread of β-thalassaemia on the Indian subcontinent. *Human genetics*, *90*(4), 443-449.

36. Ahmed S (2007) Prenatal diagnosis of β-thalassaemia: twelve years experience at a single laboratory in Pakistan. Prenat Diagn 27: 1224-1227.

37. Sharifi, Z., Rahiminejad, F., Joudaki, A., Bandehi, A. S., Farahzadi, H., Keshvar, Y., ... & Ghadami, S. (2019). Development and validation of a novel panel of 16 STR markers for simultaneous diagnosis of β-thalassemia, aneuploidy screening, maternal cell contamination

detection and fetal sample authenticity in PND and PGD/PGS cases. *Scientific reports*, *9*(1), 1-9.

38. Kamran, S., Ahmed, S., Ahmad, K. N., Abbasi, A. M., Bano, A., & Bano, F. (2019). MATERNAL CELL CONTAMINATION IN CHORIONIC VILLOUS SAMPLES FOR THE PRENATAL DIAGNOSIS OF THALASSAEMIA. *Pak J Pathol*, *30*(1), 20-23.
39. Kwok, S. A., & Higuchi, R. (1989). Avoiding false positives with PCR. *Nature*, *339*(6221), 237-238.
40. Ahmed S, Rehman ZU, Karamat KA (2003) Diagnosis of β-thalassaemia major in previously transfused patients. JCPSP 13: 19-20.
41. Chong, S. S., Boehm, C. D., Higgs, D. R., & Cutting, G. R. (2000). Single-tube multiplex-PCR screen for common deletional determinants of α-thalassemia. *Blood, The Journal of the American Society of Hematology*, *95*(1), 360-362.
42. Liu, D., Zhang, X., Yu, L., Cai, R., Ma, X., Zheng, C., & Yan, T. (2014). KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of β-thalassemia. *Blood, The Journal of the American Society of Hematology*, *124*(5), 803-811.
43. Rujito, L., Basalamah, M., Siswandari, W., Setyono, J., Wulandari, G., Mulatsih, S., & Sutaryo, S. (2016). Modifying effect of XmnI, BCL11A, and HBS1L-MYB on clinical appearances: a study on β-thalassemia and hemoglobin E/β-thalassemia patients in Indonesia. *Hematology/oncology and stem cell therapy*, *9*(2), 55-63.
44. Shen, X., Xu, Y., Zhong, Y., Zhou, C., Zeng, Y., Zhuang, G., & Li, T. (2011). Preimplantation genetic diagnosis for α-and β-double thalassemia. *Journal of assisted reproduction and genetics*, *28*(10), 957-964.