BIOSCIENCES BIOTECHNOLOGY RESEARCH ASIA, September 2024. Vol. **21**(3), p. 1043-1053

Genotype–Phenotype Correlation of β-Thalassemia in the Indian College Student Population: Insights and Challenges in Screening and Molecular Diagnosis

Abhishek Samanta[1] and Nandan Bhattacharyya[2]

[1]Department of Zoology, Panskura Banamali College (Autonomous), Vidyasagar University, Panskura R.S., West Bengal, India.
[2]Department of Biotechnology, Panskura Banamali College (Autonomous), Vidyasagar University, Panskura R.S., West Bengal, India.

http://dx.doi.org/10.13005/bbra/3284

(Received: 24 April 2024; accepted: 16 June 2024)

Thalassemia, a widespread global health issue stemming from abnormal haemoglobin levels, affects approximately 4.5% individuals worldwide. Despite advances in treatment, this study investigates ß-thalassemia among 3,000 college students from West Bengal, India, examining genotype-phenotype correlations and silent carrier prevalence. Methodologically, blood analyses and DNA extraction were used to screen participants. PCR amplification of four primers covering the HBB gene, followed by amplicon purification and nucleotide sequencing, was employed. Bioinformatics tools, evolutionary conservation analysis, and machine learning-assisted variant categorization were utilized to provide insights for clinicians. Results revealed 385 thalassemia carriers among 2,984 individuals, exhibiting various traits including ß-thalassemia, haemoglobin E-heterozygotes, haemoglobin D Punjab heterozygotes, hereditary persistent foetal haemoglobin (HPFH), and borderline ßthalassemia traits. Significant phenotypic variations were observed. Genotype analysis identified six mutations, each associated with distinct ethnic prevalences and clinical presentations. The discussion addressed diagnostic challenges, emphasizing the need for accurate diagnosis through haemoglobin analysis and DNA testing. Pathogenicity assessments provided insights into mutation impact. Overall, this study contributes to ß-thalassemia management by offering guidance for research and clinical practice.

Keywords: Genetic Mutations; Genotype-phenotype correlation; Molecular Diagnosis; Silent Carrier Prevalence; ß-thalassemia.

Thalassemia is an autosomal recessive genetic disorder stemming from abnormal haemoglobin levels, a significant global health concern. Individuals carrying one altered allele, known as thalassemia carriers, typically lead normal lives without symptoms, constituting around 4.5% of the world's population. However, when a child inherits defective globin genes from both parents, thalassemia major manifests, necessitating lifelong reliance on blood transfusions and posing challenges in maintaining an adequate national blood supply. Thalassemia complications include iron overload, cardiovascular issues, and bone deformities. [2-8]

*Corresponding author E-mail: bhattacharyya_nandan@rediffmail.com

Modern treatments such as hematopoietic stem cell transplantation (HSCT) are complex and costly due to limited medical facilities and a scarcity of HLA-matched donors. [8]Consequently, we propose that preventing thalassemia through proactive measures is a more effective strategy.

In India, where approximately 3-4% percent of the population are carriers, 10,000–15,000 children with Thalassemia Major are born annually due to insufficient awareness. This knowledge gap leads to a critical survival rate, as highlighted by the Mumbai carrier screening program's 20year report. Successful premarital screening programs in various countries, such as Italy, Greece, Canada, Great Britain, and Cyprus have been effective in preventing thalassemia. While Cyprus has seen a reduction in thalassemia prevalence through compulsory prenatal screening, cultural and ethnic factors worldwide contribute to the increasing population of affected individuals.[9]

The Ithagenes (https://www.ithanet.eu/db/ithagenes) database has reported nearly 3447 entries involving the â-globin gene. However, small deletions that remove all or part of the â-globin gene are rare. Some âthalassemia are caused by deletions or mutations in the regulatory regions upstream of the â-globin gene complex. The â-globin gene mutation was detected using PCR to amplify fragments and sequence analysis. Automated DNA sequencing employing the dye terminator method is preferably used for the complete sequencing of the whole â-globin gene.

Notable mutations in India are IVS1-5 (G'!C), which are often found in a homozygous state or in combination with other â-thalassemia mutations.[10]Another prevalent mutation is codon 15 (G'!A), which is observed in both homozygous and compound heterozygous states with other â-thalassemia mutations.[11] The codon 41/42 (-TCTT) mutation is also common and presents in homozygous or compound heterozygous forms. In addition, the IVS1-1 (G'!T) mutation is noteworthy and is often found in a homozygous state or in combination with other â-thalassemia mutations.[12]Frameshift mutations, such as Codon 8/9 and Codon 8/9 (-AA), are prevalent and occur in homozygous states or in combination with other âthalassemia mutations.[13]

The objective of this study was to evaluate the phenotype-genotype correlation of âthalassemia and to assess the prevalence of silent â-thalassemia carriers in the Indian college student population.

MATERIALS AND METHODS

Thalassemia screening

A total of 2984 blood samples were collected from students at PanskuraBanamali College in West Bengal, India, between 2019 and 2022 with written consent and ethical clearance (Memo no. PBC/003/IE/2020) from institution, analysed the red blood cell indices, and phenotypic verification carried out using high-performance liquid chromatography (HPLC) [14]. The phenotype parameters, primarily erythrocytic parameters such as haemoglobin (Hb) concentration, mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC), and red cell distribution width (RDW), were described [14-15] with the collaboration of the School of Tropical Medicine, Department of Health and Family Welfare, Government of West Bengal, India.

DNA extraction

Based on thalassemia screening results and prior consent, twelve mutant participants are taking part in further genotype analysis. The phenol-chloroform method was used to extract pure DNA from whole blood without a commercial kit, using lab-prepared reagents. First, Reagent A was prepared with sucrose, magnesium chloride (MgCl,), Tris-HCl (a buffering agent), and Triton X (a detergent), all mixed in double-distilled autoclaved water. To start the extraction, 5 ml of blood was added to Reagent A in a polypropylene tube and mixed gently until the solution was clear. The mixture was then centrifuged at 2500 rpm for 10 minutes to form a pellet and separate out the red blood cells (RBCs). If any RBCs remain, the process is repeated to ensure complete removal.

The resulting pellet was then mixed with Reagent B, which included Tris-HCl, sodium chloride (NaCl), EDTA (a chelating agent), and SDS (a detergent). Proteinase K was added to a concentration of 100 µg/l and SDS to make a final concentration of 1%. This mixture was gently inverted for 3-4 minutes until it became viscous, then incubated at 56°C for 3-4 hours to ensure complete digestion.

After digestion, reagent C (sodium perchlorate in double-distilled water) was added and mixed gently for 3-4 minutes. An equal volume of a phenol, chloroform, and isoamyl alcohol mixture (25:24:1) was then added, and the solution was centrifuged at 4000 rpm, resulting in three layers: an aqueous layer containing DNA, a protein layer, and a solvent layer. The aqueous layer was carefully transferred to another tube, avoiding the protein layer.

To further purify the DNA, an equal volume of chloroform and isoamyl alcohol (24:1) was added to the supernatant, mixed gently for 3-4 minutes, and centrifuged again at 4000 rpm for 15 minutes. The aqueous phase was transferred to a fresh tube, and DNA was precipitated by adding two volumes of chilled absolute alcohol and mixing gently. The DNA was spooled out, placed in a fresh 1.5 ml tube, and the alcohol was decanted.

The DNA was washed twice with 70% alcohol, dried thoroughly to remove all alcohol, and then re-suspended in 200 µl of TE buffer. The tubes were kept at 56°C for 30-45 minutes to enhance DNA dissolution. The DNA sample was then stored at -80°C.

For quantification and quality checks, the extracted DNA was measured using a UV-Visible spectrophotometer at 260 nm. An absorbance value of 1 at 260 nm corresponds to 50 ng/µl of double-stranded DNA. The DNA concentration was calculated using the formula: Concentration DNA = 260nm absorbance * 50 ng/µl * Dilution Factor. The purity of the DNA sample was assessed by the absorbance ratio at 260 nm/280 nm, with a ratio of 1.8 indicating pure DNA, free from proteins, organic solvents, RNA, or significant degradation.

Thermal cycle amplification of DNA

A set of four primers, including Front-P (ACGGCTGTCATCACTTAGAC) and Int-1-P (CACTGTTATTCTTTAGAATGGTGCAAAG) as forward primers, and Int-2-P (GGAATATATG TGTGCTTATTTGCATATTC) and Rev-P (CAGATTCCGGGTCACTGTG) as reverse primers were carefully chosen to cover the entirety of the HBB gene. The PCR amplification of the â-globin gene used PCR Master Mix (Bioline) in a 25 µl reaction volume. Thermocycling conditions involved a denaturing cycle at 95°C for 5 min, followed by 35 cycles of denaturing at 94°C for 30 seconds, annealing at 62°C for 40 seconds, and extension at 72°C for 20 s. The final extension step was performed at 72°C for 10 min using a thermal cycler. After amplification, the resulting amplicons underwent a quality assessment through electrophoresis on a 1% agarose gel.

Nucleotide Sequencing

The PCR products amplified with set-1 primers (Front-P & Int-1-P and Rev-P & Int-2-P) were purified before nucleotide sequencing. Sequencing was performed using a Big-Dye Terminator cycle at a commercial facility (Barcode BioSciencesPvt.Ltd,Bangalore, Karnataka).

Subsequently, the raw data obtained will be analysed using bioinformatics software, including Chromas 2.6.6 and Unipro UGENE.

Mutations were identified for each genotype study.

Pathogenicity of mutation through in-silico study

Pathogenicity scores, such as GERP++ and phyloP, were adopted from Ithagenes database and were used with a focus on evolutionary conservation. Additional tools, including SpliceAI and CADD, were considered, taking functional annotations into account. Population frequencies play a crucial role in differentiating common polymorphisms from potentially harmful rare variations.

Machine learning tools, specifically Meta SVM and REVEL, were used to capture complex relationships between genomic features and disease association. The in-silico predictions were used to categorize variants, providing valuable guidance for clinicians and researchers in decision-making and prioritizing variants for experimental validation. It is important to emphasize that while these predictions offer valuable insights, confirmation through experimental validation and correlation with clinical cases is essential for a comprehensive understanding of the impact of genetic variants on human health.

RESULT AND DISCUSSION

Phenotype study

Of the 2,984 people screened, 385 were found to be carriers of thalassemia mutations, making up 12.89% of the group. Among these carriers, 289 had â-thalassemia traits, 63 had haemoglobin E-heterozygotes, 15 had haemoglobin

D. Punjab heterozygotes, four had hereditary persistent foetal haemoglobin (HPFH), two Hb E/â-thalassemia and 14 had borderline â-thalassemia traits with HbA2 levels ranging from 3.3% to 3.8%. Between 385 carriers 12 participants are taking part in genotype analysis (Table 1).

Phenotypes of each group, such as variations in Red Blood Cell (RBC) counts, Haemoglobin levels, Haematocrit (HCT), Mean Corpuscular Volume (MCV), Mean Corpuscular Haemoglobin (MCH), Red Cell Distribution Width (RDW), and Mean Corpuscular Haemoglobin Concentration (MCHC) and additionally, the percentages of HbA0, HbF, and HbA2, critical indicators in thalassemia diagnosis, exhibit distinct patterns among the groups.

Deviation of each parameter from the normal is measured with regression equation (Table 2).

Table 2 depict the mean ± SD (min-max). RBC, red blood cells; haemoglobin, haemoglobin; HCT, haematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; RDW, red cell distribution width; MCHC, mean corpuscular hemoglobin concentration; HbA0, hemoglobin A0; HbF, hemoglobin F; HbA2, hemoglobin A2; SE, standard error and regression equation with normal and different mutations parameter.

Various haematological parameters for different study groups, including individuals with normal haemoglobin, â thalassemia carrier, HbE thalassemia carrier, HbD thalassemia carrier, HPFH thalassemia carrier, borderline cases, and Hb E/â-thalassemia, have been analyzed. The phenotypic analysis includes mean values, standard deviations, ranges, standard error, p-values, and regression equations, showing the relationship between these haematological parameters.

Regression analysis in this context helps to understand how different types of thalassemia

Table 1. Distribution of β Thalassemia Mutations Among Affected Participants

β Thalassemia mutation	Affected participants
β-thalassemia traits	289
Haemoglobin E-heterozygotes	61
Haemoglobin D Punjab Heterozygotes	15
Hereditary persistent foetal haemoglobin (HPFH)	4
Borderline β-thalassemia traits	14
Hb E/β-thalassemia	2

Table 2. Haematological parameters of the study groups

Groups	Normal	β thalassemia carrier	HbE Thalassemia carrier	HbD thalassemia carrier	HPFH thalassemia carrier	Boderline case	Hb E/ β-thalassemia
RBC	5.18 ± 0.61 (4.15 - 7.11)	5.14 ± 0.8 (3.14 - 7.59)	5.12 ± 0.8 (3.95 - 7.27)	4.82 ± 0.78 (4.23 - 5.87)	4.84 ± 0.26 (4.65 - 5.02)	4.87 ± 0.92 (3.21 - 6.32)	5.57 ± 1.3 (5.5–6.5)
Hb	13.97 ± 1.72 (10.9 - 19.1)	11.27 ± 1.6 (7.9 - 15.2)	12.08 ± 1.42 (9.2 - 14.6)	13.65 ± 2.24 (11.5 - 16.8)	13.45 ± 0.07 (13.4 - 13.5)	11.17 ± 1.47 (8.2 - 12.3)	11.80 ± 2.34 (10.30-12.5)
HCT	41.86 ± 4.44 (34.7 - 55.8)	40.57 ± 6.22 (26.8 - 52.3)	39.92 ± 4.02 (31.6 - 46.2)	43.55 ± 4.19 (38.2 - 48)	43.15 ± 4.31 (40.1 - 46.2)	42.73 ± 6.41 (32 - 50.2)	32.9 ± 3.0 (31.0–34.7.0)
MCV	80.9 ± 3.87 (72.9 - 89.6)	78.26 ± 14.76 (39.2 - 104.1)	78.73 ± 9.83 (64.4 - 99)	91.43 ± 6.98 (81.8 - 98.2)	87.2 ± 10.32 (79.9 - 94.5)	88.81 ± 13.01 (62.2 - 100.4)	59.1 ± 7.3 (58.5–61.08)
MCH	26.94 ± 1.61 (21.5 - 30.6)	22.02 ± 3.36 (16.3 - 29.2)	23.76 ± 2.09 (18.2 - 30.2)	27.45 ± 1.16 (26.4 - 28.6)	25.95 ± 1.06 (25.2 - 26.7)	24.91 ± 2.61 (21.2 - 29.2)	23.8 ± 7.0 (13.8–33.0)
RDW	12.88 ± 0.82 (11.5 - 14.7)	15.23 ± 3.05 (10.8 - 31.9)	14.07 ± 1.58 (11.2 - 19.7)	13.55 ± 1 (12.5 - 14.5)	14.75 ± 0.64 (14.3 - 15.2)	15.71 ± 2.41 (12.6 - 19.2)	17.5 ± 23.0 (2.6–68.0)
MCHC	27.1 ± 9.53 (19.9 - 74.7)	26.5 ± 8.21 (19.9 - 74.7)	26.83 ± 2.25 (24.2 - 32)	28.97 ± 1.2 (27.8 - 30.2)	30.3 ± 4.38 (27.2 - 33.4)	24.32 ± 1.98 (21.2 - 28.2)	23.83 ± 2.05 (24.2 - 32)
HbA0	84.21 ± 1.43 (81.1 - 88.2)	83.04 ± 2.53 (76.4 - 87)	62.93 ± 3.74 (55.6 - 79.6)	49.73 ± 3.5 (44.8 - 52.9)	60.75 ± 3.75 (58.1 - 63.4)	86.97 ± 0.35 (86.5 - 87.6)	128.2 ±12.8 (129.9-132.3)
HbF	0.59 ± 0.45 (0.2 - 2.8)	1 ± 0.97 (0 - 6.3)	0.58 ± 0.4 (0.2 - 2.5)	0.45 ± 0.13 (0.3 - 0.6)	29.25 ± 4.6 (26 - 32.5)	0.38 ± 0.22 (0.2 - 0.9)	0.5 ±0.79 (0.3-0.8)
HbA2	3.51 ± 0.33 (2.6 - 3.9)	5.1 ± 0.96 (0.6 - 6.8)	27.65 ± 2.38 (23.2 - 30.6)	2.13 ± 0.61 (1.6 - 3)	2.25 ± 0.49 (1.9 - 2.6)	3.69 ± 0.11 (3.6 - 3.9)	8.4 ± 1.54 (8.32-9.54)
SE		2.222	9.211	12.323	11.837	2.764	18.573
p value of <0.05		0	0.000173	0.001	0.001	0	0.001074
Regression equation		Y=”0.0527+ 0.9709Å.X	Y=7.3198+ 0.7558Å.X	Y=2.8275+ 0.8407Å.X	Y=9.2605+ 0.7662Å.X	Y=”1.5886+ 1.0720Å.X	Y=”2.5481+ 1.1267Å.X

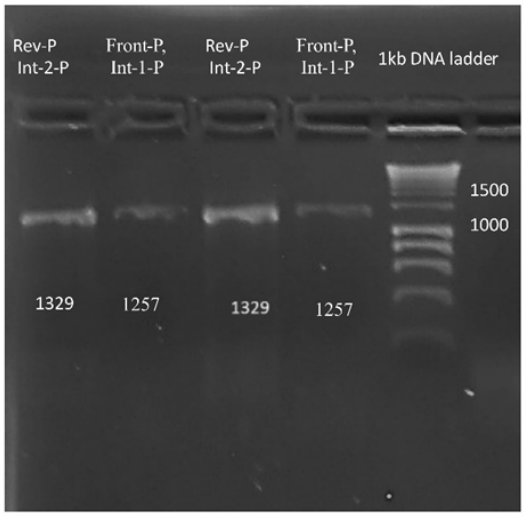

Fig. 1. PCR-amplified products using the primers Front-P, Int-1-P, Rev-P, and Int-2-P, with corresponding product sizes of 1257 and 1329 base pairs demonstrated on an agarose gel with a 1KB DNA ladder.

carriers affect haematological parameters such as RBC count, haemoglobin levels, HCT, MCV, and others. A regression equation like Y=”0.0527+0.9709Å”XY = -0.0527 + 0.9709 \ cdotXY=”0.0527+0.9709Å”X for normal versus â thalassemia carriers indicates how a specific parameter (Y) changes with respect to the independent variable (X), which represents the type of thalassemia. This allows for a detailed examination of the impact of each type of thalassemia on various blood parameters.

The regression analysis aimed to uncover the relationships between different groups (Normal, Â Thalassemia carrier, HbE Thalassemia carrier, HbD thalassemia carrier, HPFH thalassemia carrier, Borderline case, Hb E/â-thalassemia) and various blood parameters. The outcomes revealed distinct regression equations for each parameter, providing insights into the influence of different groups on specific haematological measures. For instance, the equation for the red blood cell count (RBC) was Y = -0.0527 + 0.9709X, demonstrating the impact of the group (X) on RBC levels. Similarly, the equation for haemoglobin was Y = 7.3198 +

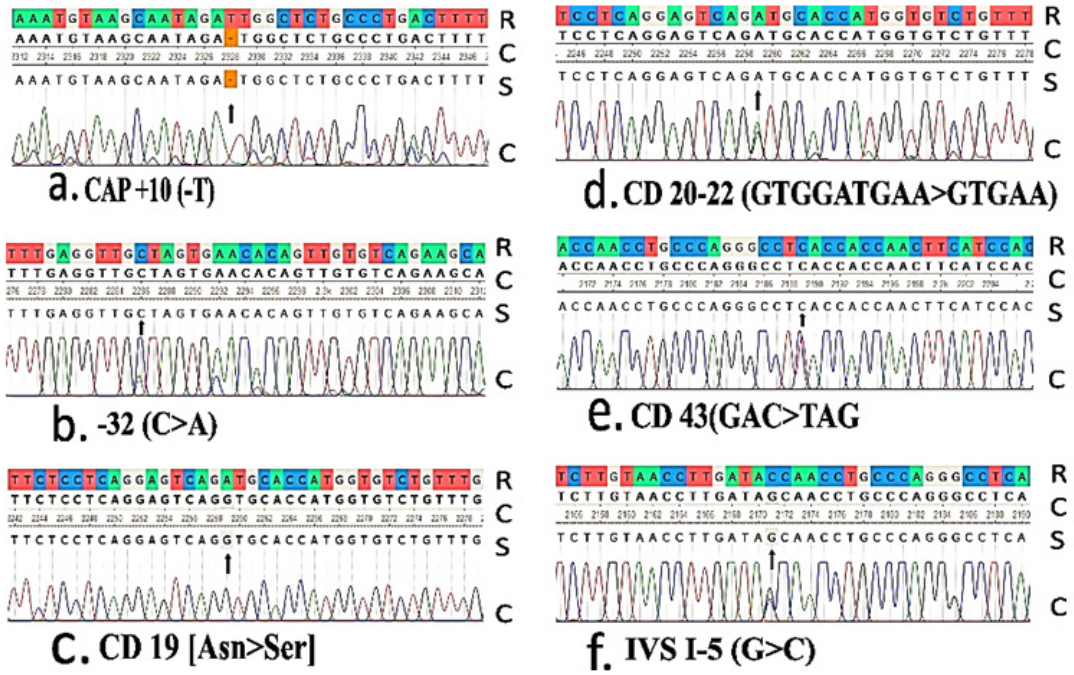

Fig. 2. The mutations associated with â-thalassemia include a promoter mutation (-32 C>A), 5’UTR mutation (CAP +10 -T), an exon 1 mutation (CD 19 AAC>AGC), another exon 1 mutation (CD 20-22 GTGGATGAA>GTGAA), an exon 2 mutation (CD 43 GAG>TAG), and an intron 1 mutation (IVS 1-5 C>G). In the figure R stands for reference sequence, C stands for Chromatograms

Table 3. Mutations with their location and allele phenotype on the β-globin gene of the studied groups

Mutation	Location	Chromosome	Locus	Allele Phenotype	Types of Mutation
-32 (C>A)	Promoter	11	70513	β^{+}	Point-Mutation (Substitution)
CAP +10 (-T)	5'UTR	11	70554	β^{++} (silent)	Point-Mutation (Deletion)
CD 19 (AAC>AGC) [Asn>Ser]	Exon 1	11	70653	β++	Cryptic splice site (mRNA Processing)
CD 20-22 (GTGGATGAA>GTGAA)	Exon 1	11	70658	β_____	Point-Mutation (Deletion)
CD 43 (GAG>TAG)	Exon 2	11	70854	β0_____	Point-Mutation (Substitution)
IVS I-5 (G>C)	Intron 1	11	70691	β	Point-Mutation (Substitution)

0.7558X, indicating a correlation between the group and haemoglobin levels. Haematocrit (HCT) exhibited an equation of Y = 2.8275 + 0.8407X, demonstrating the influence of the group on HCT levels. The mean corpuscular volume (MCV) had the equation Y = 9.2605 + 0.7662X, illustrating the relationship between the group and MCV. Mean corpuscular haemoglobin (MCH) showed an equation of Y = -1.5886 + 1.0720X, signifying the influence of the group on MCH levels. The red cell distribution width (RDW) displayed an equation of Y = -2.5481 + 1.1267X, indicating the correlation between the group and RDW. The regression analysis identified significant associations ($p < 0.05$) with the HbE thalassemia carrier, HbD thalassemia carrier, HPFH thalassemia carrier, and Hb E/â-thalassemia groups, underscoring their impact on specific haematological parameters. The coefficients in the regression equations provided valuable insights into the strength and direction of these associations, contributing to a deeper understanding of how different thalassemia carrier groups may influence distinct blood parameters.

Genotype study

The Front-P and Int-1-P primer pairs amplified 1257 base pairs of DNA fragments, confirmed by comparison with a 1kb DNA ladder on the agarose gel. Likewise, the Rev-P and Int-2-P primer pair amplified another band measuring approximately 1329 base pairs, with its size validated against the same 1kb DNA ladder on the gel. The successful amplification of specific DNA regions and the observed sizes of the amplified fragments were visually confirmed through gel documentation (Figure 1).

Through chromatogram analysis and using the Ithagenes[16] database, six distinct mutations were identified, each revealing diverse clinical presentations. Several causative mutations associated with â-thalassaemia were outlined, each contributing to the development of haemolytic anaemia and ineffective erythropoiesis. The IVS I-5 (G>C)[17,18] mutation, also known as HBB: c.92+5G>C, is linked to â0 thalassaemia, disrupting normal splicing and activating cryptic donor sites in Intron 1. The -32 (C>A)[19] mutation, denoted as HBB: c.82C>A, falls within the â+ allele phenotype, occurs in the promoter region and impacting transcription. The CAP +10 (-T)[20] mutation, or HBB: c.-41delT, contributes to â++ thalassaemia, involving a deletion in the 5' untranslated region (5'UTR). The CD 19 (AAC>AGC) [21-22]mutation, HBB: c.59A>G, leads to a cryptic splice site in exon 1 and is associated with the â-thalassaemia and structural haemoglobinopathy group. Finally, the CD 20-22 (GTGGATGAA>GTGAA) [23] mutation, designated as HBB: c.64_67del, induces a frameshift in exon 1. The CD 43 (GAG>TAG)[24] mutation in the HBB gene is a causative mutation associated with â-thalassaemia, characterized by a point mutation (substitution) at position c.130G>T. This pathogenic mutation leads to a frameshift in the translation of the globin gene, resulting in â0

Table 4. Mutation Analysis using CADD Scores

Mutation	Tool	Score	P/LP threshold	B/LB threshold	Predicted class	Score Range
CD 43 (GAG>TAG)	CADD	38	>11.5 (Supporting)	<=11.5 (Supporting)	P/LP	[0, 47]
CAP +10 (-T)	CADD	17.85	>11.5 (Supporting)	<=11.5 (Supporting)	P/LP	[0, 47]
CD 20-22 (GTGGATGAA>GTGAA)	CADD	23.5	>11.5 (Supporting)	<=11.5 (Supporting)	P/LP	[0, 47]
IVS I-5 (G>C)	CADD	23.5	>11.5 (Supporting)	<=11.5 (Supporting)	P/LP	[0, 47]
CD 19 (AAC>AGC) [Asn>Ser]	CADD	20.3	>11.5 (Supporting)	<=11.5 (Supporting)	P/LP	[0, 47]
-32 (C>A)	CADD	17.85	>11.5 (Supporting)	<=11.5 (Supporting)	P/LP	[0, 47]

thalassaemia, a recessively inherited condition prevalent in Chinese and Thai populations. Individuals carrying this mutation may exhibit associated phenotypes, including haemolytic anaemia and ineffective erythropoiesis. These mutations exhibit specific ethnic prevalences, such as Asian Indian, Southeast Asian, Melanesian, Pakistani, Taiwanese, Greek, Southeast Asian, Chinese, Malaysian, Singaporean, Thai, and Spanish populations. Detailed genetic information, including specific nucleotide sequences, protein effects, inheritance patterns, and associated phenotypes, has been elucidated through molecular studies, providing valuable insights into the molecular basis of â-thalassaemia in diverse populations. (Table 3) (Figure 2)

Table 3 depicts the mutations include -32 (C>A) in the promoter, leading to a â+ phenotype due to a point mutation. CAP +10 (-T) in the 52 UTR results in a silent â++ phenotype through a deletion. CD 19 (AAC>AGC) in Exon 1 causes a â++ phenotype by affecting a cryptic splice site during mRNA processing. CD 20-22 (GTGGATGAA>GTGAA) in Exon 1 leads to a â phenotype through a deletion. CD 43 (GAG>TAG) in Exon 2 results in a â0 phenotype due to a substitution. IVS I-5 (G>C) in Intron 1 leads to a â phenotype through substitution. These mutations, which occur in different gene regions, result in varying degrees of impact on â-globin synthesis, offering insights into the diverse molecular manifestations of âthalassemia.

Pathogenicity study

For the CD 43 (GAG>TAG) mutation, SpliceAI indicated a benign/likely benign (B/LB) prediction with a score of 0, BayesDel suggested a possibly pathogenic (P/LP) outcome with a score of 0.486378, Eigen-PC-phred predicted a P/LP outcome with a score of 3.86254, GERP++ labelled it as a variant of uncertain significance (VUS) with a score of 4.19, MetaSVM and REVEL did not provide predictions, and CADD suggested a P/LP outcome with a score of 38. PhyloP100 and phastCons30 supported the B/LB prediction. For the CAP +10 (-T) mutation, CADD indicated a P/LP outcome with a score of 17.85, whereas other tools did not provide predictions. The CD 20-22 (GTGGATGAA>GTGAA) mutation showed a B/LB prediction by SpliceAI with a score of 0, and CADD suggested a P/LP outcome with a score of 23.5. For the IVS I-5 (G>C) mutation, CADD indicated a P/LP outcome with a score of 23.5. Finally, the CD 19 (AAC>AGC) [Asn>Ser] mutation showed varying predictions, with SpliceAI suggesting a B/LB outcome, BayesDel, Eigen-PC-phred, and CADD indicating a P/LP outcome, GERP++ suggesting a VUS, and MetaSVM and REVEL predictions. These scores and predictions contribute to understanding the potential pathogenicity of each mutation.[25] (Table 4)

Table 4 summarizes the predicted pathogenicity of various genetic mutations as assessed by the CADD tool. The score thresholds

for categorizing mutations as Pathogenic/Likely Pathogenic (P/LP) and Benign/Likely Benign (B/LB) are indicated, along with the predicted class for each mutation based on its score. Scores ranged from 0 to 47, with higher scores suggesting higher potential pathogenicity.

Thalassemia screening primarily focuses on anticipating and checking during pregnancy.[26] However, the existing method for identifying carriers of â-thalassemia has limitations, potentially missing mutations when blood measurements appear normal. To confirm a â-thalassemia diagnosis, a comprehensive approach involving haemoglobin analysis using high-performance liquid chromatography (HPLC) or capillary electrophoresis (CE) and DNA analysis of the globin genes is necessary. Recent clinical guidelines have provided specific protocols for â-thalassemia diagnosis.[27] The condition manifests in various clinical phenotypes, including â-thalassemia major, â-thalassemia intermedia, and â-thalassemia trait. Notably, there are silent â-thalassemia carriers—healthy individuals with the â-globin gene mutation but without identified hepatological abnormalities.[28] Silent carriers, often discovered during family genetic studies, can interact with the â-thalassemia trait, resulting in a milder form known as silent â-thalassemia. Detecting silent â-thalassemia is critical for effective carrier identification and preventing â-thalassemia, presenting a diagnostic challenge for clinicians. Silent â-thalassemia carriers display no symptoms, have nearly normal â/á chain ratios, and exhibit no haematological abnormalities, including normal levels of Hb A2.[29] The phenotypic consequences of this study based on regression analysis to understand the impact of different thalassemia groups (Normal, â-Thalassemia carrier, HbE Thalassemia carrier, HbD Thalassemia carrier, HPFH Thalassemia carrier, Borderline case, Hb E/â-thalassemia) on various blood parameters. The analysis produced distinct regression equations for parameters like red blood cell count (RBC), haemoglobin (HCT), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), and red cell distribution width (RDW). Significant associations were found particularly with the HbE Thalassemia carrier, HbD Thalassemia carrier, HPFH Thalassemia carrier, and Hb E/â-thalassemia groups.

Additionally, the study highlighted the importance of pathogenicity descriptions in â-thalassemia research. It provided detailed information on genetic mutations, their pathogenicity scores, and classifications. Specific mutations such as CD 43 (GAG>TAG), CAP +10 (-T), CD 20-22 (GTGGATGAA>GTGAA), IVS I-5 (G>C), CD 19 (AAC>AGC) [Asn>Ser], and -32 (C>A) were analysed using similar tools like SpliceAI and BayesDel. These scores help determine the likelihood of a mutation causing disease, aiding in their classification as pathogenic, likely pathogenic, benign, likely benign, or of uncertain significance. This information is crucial for diagnosing and managing â-thalassemia, guiding decisions on carrier detection, prenatal diagnosis, and patient care.

Across India, the carrier rate for â-thalassemia varies from 3-17%, with a few common mutations accounting for over 95% of severe cases. The most prevalent mutation is the IVS-I-5 (G '! C), with regional variations, such as the high prevalence of the 619 bp deletion among Sindhis and Lohanas.[30] Understanding regional and ethnic mutation profiles is crucial for tailoring molecular analysis and management strategies effectively. In the Garhwal region of India, the most prevalent â-thalassemia mutation was IVS 1-5 (G-C) at 18.75%, followed by Codon 8/9 at 12.5%, and IVS 1-1 (G-T) at 6.25%. Codon 41/42 (-TCTT) and the 619 bp deletion were not detected in this population.[31-32]

In Fujian province, China, Zhuang *et al.* in 2021, identified 4,884 cases of á-thalassemia and 2,056 cases of â-thalassemia, with –SEA/áá and âIVS"II"654/âN being the most common mutations and notable regional differences in mutation patterns.[33] Khan *et al.*in 2021, identified 105 HBB gene mutations in 20 Arab countries, with six unique to Arabs and the most common being splice site and frameshift mutations.[34]Karnpean*et al.*in 2022, analysed 163 â-thalassemia subjects, noting severe anaemia in compound heterozygous â-thalassemia/Hb E individuals and minimal anaemia in homozygous Hb E and â-thalassemia carriers, revealing significant allele frequency differences in intragenic â-globin gene polymorphisms.[35]Tamaddoni*et al.*in 2021, examined 50 patients, finding a high prevalence of blood transfusions (60%) and splenomegaly

(80%), with the HBB. 315+1G>A [IVS II-1 (G>A)] mutation being most common and associated with thalassemia intermedia phenotypes.[36]Pooladi*et al.* in 2022 screened 340 Iranian-Kurd couples, identifying 20 beta-thalassemia and nine alpha-thalassemia mutations, with significant portions having consanguineous marriages and detailed associations between blood indices and mutation types.[37] Farra *et al.*in 2021, analysed 126 Lebanese, 26 Iraqi, and 18 Syrian individuals for beta-thalassemia, identifying 28 different mutations, with ten mutations newly identified in Lebanon.[38] Zhu *et al.* in 2021, found that 33.8% of 1,706 subjects in Hubei province, China, had thalassemia mutations, with —SEA/áá, "á3.7/áá, and -á4.2/áá being predominant in á-thalassemia, and âIVS-II-654/âN in â-thalassemia, noting the efficacy of MCV and MCH as markers for detecting these mutations.[39] Lama *et al.*in 2021, reported on 61 thalassemia patients from 17 ethnicities across Nepal, identifying nine â-thalassemia mutations, with the highest mutation rates in Brahmin and Chhetri, and Huang *et al.*in 2020, found a carrying rate of 31.92% among 1,130 thalassemia gene carriers, with the most common mutations being —SEA/áá in á-thalassemia and CD17 (A>T) and CD41–42 (–TTCT) in â-thalassemia.[40-41]

Similarly, in studies based in Malaysia, a significant proportion of healthy blood donors were identified as silent â-thalassemia carriers, highlighting the need for clinician awareness and timely genomic DNA extraction for precise â-globin sequencing.[41] Patients with â-thalassemia major were diagnosed at a younger age and required more frequent transfusions compared to those with Hb E/â-thalassemia. Other studies on Haemoglobin E â-thalassemia revealed that patients with milder forms of the disease were diagnosed later and had higher haemoglobin levels at diagnosis compared to those with severe forms. Common â mutations varied between Malay and Chinese patients, with most patients exhibiting â° gene mutations, particularly IVS1-5 and CD41/42, which were associated with moderate to severe phenotypes.[42-43]

Similarly, in Syria, the â0/â0 genotype was the most common, followed by â+/â+ and â0/â+ genotypes. Patients with the â0/â0 genotype received transfusions earlier and more frequently and had higher levels of Hb F, and lower levels of Hb A2 compared to other genotypes. All the patients with â-thalassemia intermedia had the â+/â+ genotype, while those with â0/â0 and â0/â+ genotypes were associated with transfusion-dependent â-thalassemia major.[44]

CONCLUSION

This study provides important information about screening for â-thalassemia, highlighting the need for specialized tests to detect silent carriers. It also offers insights into the relationships between different thalassemia groups and specific blood parameters. The regression analysis and pathogenicity table help us understand â-thalassemia better, providing valuable information for future research and clinical practice. The comprehensive approach to diagnosis and management outlined in this study can improve the prevention and treatment of â-thalassemia.

ACKNOWLEDGEMENT

The authors are thankful to Calcutta School of Tropical Medicine, 108, Chittaranjan Ave, Calcutta Medical College, College Square, Kolkata, West Bengal 700073 for providing the necessary facilities for HPLC test.

Conflict of Interest

The authors declare no conflict of interest regarding the publication of this manuscript.

Funding Source

This research received no specific funding from any external agency, commercial entity, or institution. The study was conducted as part of the authors' academic and research activities.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethical Approval

This study was approved by the Institutional Ethical Committee, PanskuraBanamali College, West Bengal, India.

Author's contribution

The study conception and design involved contributions from both authors. Abhisek Samanta played a role in study design, data collection, and manuscript preparation. Nandan Bhattacharyya contributed to the study design, supervised the

project, and prepared the final manuscript. Both authors thoroughly reviewed and approved the final manuscript.

REFERENCES

1. Sen A, Seenappa V, Chakrabarti P, Dolai TK. First Report of the 3'-Untranslated Region+ 1506 (A> C)[NM_000518.5: c.*32A> C] Mutation on the â-Globin Gene in the Indian Population. Hemoglobin. 2021;45(5):325-328.
2. Boardman FK, Clark C, Jungkurth E, Young PJ. Social and cultural influences on genetic screening program acceptability: A mixed-methods study of the views of adults, carriers, and family members living with thalassemia in the UK. J Genet Couns. 2020;29(6):1026-1040. doi:10.1002/jgc4.1231.
3. Noor FA, Sultana N, Bhuyan GS, et al. Nationwide carrier detection and molecular characterization of â-thalassemia and hemoglobin E variants in Bangladeshi population. Orphanet J Rare Dis. 2020;15:1-2.
4. Delatycki MB, Alkuraya F, Archibald A, et al. International perspectives on the implementation of reproductive carrier screening. Prenat Diagn. 2020;40(3):301-310.
5. Ali S, Mumtaz S, Shakir HA, et al. Current status of beta thalassemia and its treatment strategies. Mol Genet Genomic Med. 2021;9(12)
6. Elsadek AM, Hassan N, Shahwan M, Jairoun AA. Exploring the Gap Between Knowledge and Behavior Regarding Thalassemia Among University Students: A Cross-Sectional Study in the United Arab Emirates. J Community Health. 2022;47(3):392-399. doi:10.1007/s10900-022-01066-8.
7. Harteveld CL, Achour A, Arkesteijn SJ, et al. The hemoglobinopathies, molecular disease mechanisms and diagnostics. Int J Lab Hematol. 2022;44:28-36.
8. Hossain MS, Hasan MM, Raheem E, et al. Lack of knowledge and misperceptions about thalassaemia among college students in Bangladesh: a cross-sectional baseline study. Orphanet J Rare Dis. 2020;15(1):54. doi:10.1186/s13023-020-1323-y.
9. Alshamsi S, Hamidi S, Narci HO. Healthcare resource utilization and direct costs of transfusion-dependent thalassemia patients in Dubai, United Arab Emirates: a retrospective cost-of-illness study. BMC Health Serv Res. 2022;22(1):304.
10. Rao E, Chandraker SK, Singh MM, Kumar R. Global distribution of â-thalassemia mutations: An Update. Gene. 2023;148022.
11. Waqas M, Ahmed N, Malik HS, et al. Relationship of frequency of IVS 1-5 (GC), Fr 8-9 (+ G), Fr 41-42 and Cd-5 mutations of â-thalassemia traits with red cell indices.
12. Liao G, Zhou Y, Yin X, et al. The clinical significance of the spectrum of interactions of the rare IVS-II-5 G> C (HBB: C.315+ 5 G> C) variation with other â-thalassemia mutations in southern China. Mediterr J Hematol Infect Dis. 2022;14(1).
13. Khan AM, Al-Sulaiti AM, Younes S, Yassin M, Zayed H. The spectrum of beta-thalassemia mutations in the 22 Arab countries: a systematic review. Expert Rev Hematol. 2021;14(1):109-122.
14. Sialvi HS, Umar G, Hamid S, Khadija G. Awareness, Attitude and Knowledge about Thalassemia among Students of Different Disciplines. J Islamic Int Med Coll. 2021;16(2):117-121.
15. Rishmitha, Badagabettu S, MV A, Lakshmi RV, CU J. Awareness on thalassemia and opinion of carrier screening among young women from selected undergraduate colleges of Udupi district. Clin Epidemiol Glob Health. 2022;14:100978. doi:10.1016/j.cegh.2022.100978.
16. Xenophontos M, Minaidou A, Stephanou C, et al. IthaPhen: an interactive database of genotype-phenotype data for hemoglobinopathies. HemaSphere. 2023;7(7)
17. Wichian P, Yamsri S, Sanchaisuriya K, Fucharoen S. Direct amplification of whole blood and amniotic fluid specimens for prenatal and postnatal diagnosis of Hb E-â 0-Thalassemia diseases. Lab Med. 2021;52(5):460-468.
18. Waqas M, Ahmed N, Malik HS, et al. Relationship of frequency of IVS 1-5 (GC), Fr 8-9 (+ G), Fr 41-42 and Cd-5 mutations of â-thalassemia traits with red cell indices.
19. Jiang F, Zhou J, Zuo L, et al. Utilization of multiple genetic methods for prenatal diagnosis of rare thalassemia variants. Front Genet. 2023;14:1208102.
20. Zhao Y, Jiang F, Li DZ. Hematological Characteristics of â-Globin Gene Mutation–50 (G> A)(HBB: c.-100G> A) Carriers in Mainland China. Hemoglobin. 2020;44(4):240-243.
21. Saad HK, Taib WR, Ab Ghani AS, et al. HBB gene mutations and their pathological impacts on HbE/â-Thalassaemia in Kuala Terengganu, Malaysia. Diagnostics. 2023;13(7):1247.
22. Aziz MA, Khan NT, Khan WA, Ayub MI, Yeasmin S. Reporting two rare mutations with associated clinical features in Bangladeshi Hb E/â thalassemia patients. Bioresearch

Communications-(BRC). 2024;10(01):1442-1448.
23. Waqas M, Ahmed N, Malik HS, et al. Relationship of frequency of IVS 1-5 (GC), Fr 8-9 (+ G), Fr 41-42 and Cd-5 mutations of â-thalassemia traits with red cell indices.
24. Charoenkwan P, Traisrisilp K, Sirichotiyakul S, et al. Noninvasive prenatal diagnosis of beta-thalassemia disease by using digital PCR analysis of cell-free fetal DNA in maternal plasma. Fetal Diagn Ther. 2022;49(11-12):468-478.
25. Lam TT, Nguyen DT, Le QT, et al. Combined gap-polymerase chain reaction and targeted next-generation sequencing improve á-and â-Thalassemia carrier screening in pregnant women in Vietnam. Hemoglobin. 2022;46(4):233-239.
26. Lippi G, Mattiuzzi C. Updated worldwide epidemiology of inherited erythrocyte disorders. Acta Haematol. 2020;143(3):196-203.
27. Sharma A, Uppal N, Kukreja S, et al. Screening of thalassemia in pregnant females visiting tertiary hospital in Amritsar. Int J Clin Biochem Res. 2020;7:226-231.
28. Abdullah UY, Ibrahim HM, Mahmud NB, et al. Genotype-phenotype correlation of â-Thalassemia in Malaysian population: Toward effective genetic counseling. Hemoglobin. 2020;44(3):184-189.
29. Amjad F, Fatima T, Fayyaz T, et al. Novel genetic therapeutic approaches for modulating the severity of â thalassemia. Biomed Rep. 2020;13(5):1-.
30. Colaco S, Nadkarni A. Borderline HbA2 levels: Dilemma in diagnosis of beta-thalassemia carriers. Mutat Res Rev Mutat Res. 2021;788:108387.
31. TV SB, Kabekkodu SP, Satyamoorthy K, Shantaram M. Molecular characterization and mutational analysis of beta-globin gene among Kodagu population of Karnataka, India. Biomedicine. 2023;43(01):482-487.
32. Mishra AS, Lakhera PC, Negi P, Pandey A. Molecular characterization of beta-thalassemia reveals the presence of common mutations in the population of Himalayan region: Garhwal (Uttarakhand), India. Int J Popul Stud. 2022;8(2):71-78.
33. Zhuang J, Zhang N, Wang Y, et al. Molecular characterization analysis of thalassemia and hemoglobinopathy in Quanzhou, Southeast China: A large-scale retrospective study. Front Genet. 2021;12:727233.
34. Khan AM, Al-Sulaiti AM, Younes S, et al. The spectrum of beta-thalassemia mutations in the 22 Arab countries: a systematic review. Expert Rev Hematol. 2021;14(1):109-122.
35. Karnpean R, Tepakhan W, Suankul P, et al. Genetic background studies of eight common beta thalassemia mutations in Thailand using â-globin gene haplotype and phylogenetic analysis. Genes. 2022;13(8):1384.
36. Tamaddoni A, Nesheli HM, Paniri A, et al. Genotype–phenotype association in alpha-thalassemia cohort in a population with high prevalence of alpha and beta-thalassemia: importance of genetic screening. Middle East J Med Genet. 2021;10(1):24.
37. Pooladi A, Gheshlagh RG, Kahrizi N, et al. The spectrum of alpha and beta thalassemia mutations: a 10-year population-based study of the Premarital Health Screening Program in West of Iran. Iran J Ped Hematol Oncol. 2022;12(7).
38. Farra C, Abdouni L, Souaid M, et al. The Spectrum of â-Thalassemia Mutations in the Population Migration in Lebanon: A 6-Year Retrospective Study. Hemoglobin. 2021;45(6):365-370.
39. Zhu Y, Shen N, Wang X, et al. Alpha and beta-Thalassemia mutations in Hubei area of China. BMC Med Genet. 2020;21:1-5.
40. Lama R, Yusof W, Shrestha TR, et al. Prevalence and distribution of major â-thalassemia mutations and HbE/â-thalassemia variant in Nepalese ethnic groups. Hematol Oncol Stem Cell Ther. 2021;14(2):203-210.
41. Huang TL, Zhang TY, Song CY, *et al*. Gene mutation spectrum of thalassemia among children in Yunnan province. Front Pediatr. 2020;8:159.
42. Suali L, Mohammad Salih FA, Ibrahim MY, *et al*. Genotype-Phenotype Study of â-Thalassemia Patients in Sabah. Hemoglobin. 2022;46(6):317-324.
43. Nuinoon M. Interaction of Thalassemia and Hb Variants in Southeast Asia: Genotype-Phenotype Relationship. In: Thalassemia Syndromes-New Insights and Transfusion Modalities. 2023. IntechOpen.
44. Masra F, Razak S, Murad N, *et al*. Genotype-Phenotype Correlation Among Haemoglobin E â-thalassaemia Patients.2021.