

IDENTIFICATION AND CHARACTERIZATION OF GENETIC VARIANTS ASSOCIATED WITH HUMAN DISEASE SUSCEPTIBILITY

Maryam Javed^{*1}, Huma Irfan², Nuzhat Nawaz Khan³, Fakhar Sultan⁴, Madiha Akram⁵

¹Doctor of Pharmacy, COMSATS University Islamabad, Abbottabad Campus, Pakistan

²MSc. Biomedical Science (Medical Genetics), Cardiff Metropolitan University, UK

³M. Phil in Chemistry, University of Peshawar, Pakistan

⁴Doctor of Pharmacy, Riphah International University, Faisalabad Campus, Pakistan

⁵PhD Scholar, Minhaj University, Pakistan

^{*1}maryamjaved115@gmail.com, ²huma_javid@hotmail.com, ³nuzhatnawaz72@gmail.com,

⁴fakharsultan36@gmail.com, ⁵alimadiha721@gmail.com

DOI: <https://doi.org/10.5281/zenodo.23006586>

Keywords

β -thalassemia; HBB gene; genetic variants; mutation analysis; molecular genetics; Pakistan

Article History

Received: 19 June 2026

Accepted: 10 September 2026

Published: 28 September 2026

Copyright @Author

Corresponding Author: *

Maryam Javed

Abstract

Background: β -thalassemia is one of the most significant inherited blood disorders in Pakistan and there is significant sequence variation in HBB gene. It is crucial to characterize these variants to gain insight into disease patterns and help enhance molecular diagnosis and prevention.

Aim: The aim of this study was to find and characterize the variants in HBB gene in patients with beta-thalassemia and investigate their relationship with demographic and clinical features.

Method: The study design was quantitative, multicenter cross-sectional study on 300 adults who were diagnosed with β -thalassemia in four government hospitals of Rawalpindi, Lahore, Peshawar and Faisalabad. Participants were selected using purposive sampling method. Details were gathered on clinical and demographic data and peripheral blood samples were examined for HBB variants using molecular genetic techniques.

Results: Five major HBB variants were identified – deletion of 619 bp, G>C in IVS-I-5, G>T in IVS-I-1, frameshift 41/42 (-TTCT), and frameshift 8/9 (+G). Ethnicity ($p=.001$), family history ($p=.006$), consanguinity ($p=.002$), disease phenotype ($p=.017$) and transfusion history ($p=.038$) were significant associations with genetic variation. There was also a significant difference in age at diagnosis ($p=.040$) and transfusion frequency ($p=.007$) between the genetic groups, but no significant difference was found for sex, age, duration of transfusion therapy or splenectomy.

Conclusion: Results reveal high level of HBB genetic diversity among patients with β -thalassemia in Pakistan and emphasize importance of molecular characterization for genetic counseling and carrier screening and for the management of the disease.

Introduction

There is a complex interplay between genetic variation, biological processes, environmental exposures and individual characteristics which determine susceptibility to human disease. Genetic variants can range from single nucleotide variants, to small insertions or deletions, structural variants, copy-number variants, and changes in regulatory regions that may be involved in altering the activity or function of genes. While most genetic variations are harmless, others may impact the structure of proteins, the expression of genes, cellular signals, the immune system, metabolic pathways or other biological functions, and therefore affect the risk of developing disease. Genetic variants can contribute to the development of a disease in either one of two ways: as highly penetrant variants causing Mendelian disorders, or as common variants with small effects that collectively account for the contribution of common genetic variants to complex diseases. An important parameter to consider in genetic studies is the frequency with which a genetic variant is observed in a population $p = 2N_{AA} + N_{Aa}$, which can be expressed as the proportion of individuals, in a population, who are homozygous for the reference allele (AA) or heterozygous for the reference allele and alternative allele (Aa), or as the frequency for homozygous individuals who are A homozygotes (N_{ZZ}), heterozygous individuals who are Aa heterozygotes (N_{Aa}) and homozygous individuals who are p homozygotes (N). The capacity to identify genetic loci linked to human phenotypes and disease risk has greatly broadened with the advent of genome-wide association studies (GWAS), whole-exome sequencing (WES), whole-genome sequencing (WGS), targeted sequencing and increasingly complex computational approaches. While GWAS can be used to analyze millions of common variants in the genome, WES and WGS allows the detection of rare coding, non-coding and structural variants that may not be well represented in current genotyping arrays. Despite this, the statistical association is still a preliminary indicator and not absolute proof of a biological mechanism establishing the link between the specific genotype

and the disease phenotype, as these mechanisms are not known and still need to be investigated (Ishigaki, 2022; Wong et al., 2021).

Rapid developments in high throughput genotyping, next generation sequencing, bioinformatics and population-scale genomic resources have revolutionized identification of disease-associated genetic variants. Genetic association studies test to determine if certain alleles or genotypes are more common in people who have a disease than in people who do not have a disease. One widely used measure is the odds ratio (OR) which is computed in the following way: $OR = ad / bc$, where a , b , c and d are the frequencies of the corresponding alleles or genotypes in the affected and unaffected groups. An $OR > 1$ could suggest that those with the genetic variant are at higher risk of the disease than those without it, while an $OR < 1$ may suggest that those with the genetic variant are at lower risk of the disease than those without it, depending on the design and type of statistical model used in the study. When using genomic data to interpret its meaning, it is important to take into account the allele frequency, the mode of inheritance, population structure, linkage disequilibrium, evolutionary conservation of the gene, predicted molecular consequences and genotype-phenotype relationships. Importantly, the disease-associated loci found by GWAS often do not really represent the causal variant as multiple disease-associated variants may be in the same region of the genome. Fine-mapping and functional genomic studies are, therefore, becoming more common to differentiate potentially causal variants from correlated variants. These strategies have led to the progress of human genetics from association with simple diseases to better understanding of the molecular mechanisms that underlie disease. This is especially important since a variant may be statistically associated with disease, but have no direct effect on biological function and experimental evidence may be needed to determine if an identified variant actually affects gene expression, protein activity, cellular signaling,

and/or disease-related phenotypes (Ishigaki, 2022; Flynn & Lappalainen, 2022).

One of the key parts of characterization of genetic variants is to understand the impact of variants on the regulation of the gene and molecular function. A significant number of disease association variants lie outside of protein coding regions and can affect genes through mechanisms such as; promoters, enhancers, transcription factor binding sites, chromatin structure, RNA processing, non-coding RNA and other mechanisms. To further demonstrate the need to investigate non-coding variants outside of coding sequences by experiment, a systematic review of GWAS validation studies showed that 309 variants had been experimentally validated for 130 disease traits in humans (Alsheikh et al., 2022). A variety of functional genomic techniques including expression quantitative trait locus analysis, chromatin profiling, transcriptomic analysis, reporter assays, protein-function assays etc. can be used to define relationships between genetic variants and changes in gene expression or cellular function. In more recent years, the use of CRISPR-based perturbation and genome-editing technologies has allowed the experimental investigation of the effects of specific variants, and at ever increasing scales (Flynn & Lappalainen, 2022; Kim et al., 2024). Variants with sequence only information that cannot be used to determine clinical significance are especially needed for functional characterization. While there are many variants that can be identified through genetic sequencing, many of these are considered “variants of uncertain significance” due to a lack of available population, computational, clinical, segregation, and/or functional evidence. Experimental studies can offer complementary evidence by looking at changes in protein activity and/or in the expression of genes; in cell phenotypes; or in biological pathways related to the disease. To illustrate how functional investigation of variants associated with Alzheimer's disease in microglia can prioritize variants and identify possible mechanisms underlying disease risk, for example, variants associated with Alzheimer's disease were

incorporated with cell specific regulatory information to prioritize variants and identify possible mechanisms underlying disease risk (Yang et al., 2023). Analogously, functional analyses of BRCA 1 and 2 variants show the importance of the laboratory testing and computational prediction of variants of uncertain significance (VUS) (Guo et al., 2023).

Recent advances are moving the study of genetic variation from single genes and loci to comprehensive, multi-level approaches of human biology. The phenotypes can be influenced by structural variants, regulatory variants, rare coding variants, and common susceptibility alleles which act at the molecular, cellular, tissue/organ, and whole organism level. As a result, the importance of large population resources for the definition of alleles frequencies and the estimation of relationship between genetic variation and disease-related phenotype across different population resources is growing. Allele frequencies, linkage disequilibrium patterns, genetic backgrounds, and environmental contexts can vary widely within populations, making it particularly important to have population specific genomic datasets. For instance, the Pakistan Genome Resource (2026) analyzed exomes and genomes from 173,303 participants and showed that a large resource of South Asians has the potential to broaden the understanding of human genetic variation and its association with biomarkers and biological pathways. Whilst there have been significant developments in genomic sequencing and genetic association studies, the discovery of variants associated with disease does not always indicate the biological and/or clinical importance of the variant. Many new variants are emerging from large-scale genomic analyses of which many are rare coding variants, common susceptibility variants, structural variants, non-coding regulatory variants, and have been difficult to interpret due to a lack of understanding of the functional consequences. Many variants of uncertain significance are still a challenge, as sequence-based prediction may not be sufficient to reliably differentiate pathogenic, benign, or clinically relevant variants. In addition, genetic associations

may differ between populations due to variation in frequencies of alleles, linkage disequilibria, genetic backgrounds and environment. Certain populations may therefore be underrepresented in genomic data, and variant-disease associations may therefore be less generalizable to these underrepresented populations (Burke et al., 2022; Collins & Talkowski, 2025; McEwen et al., 2026). Thus, the current genetic studies involve more and more integration of genetic association data with annotation of the genome, population studies, computation-based predictions, multi-omics data, cellular studies, and functional assays, all moving away from variant discovery towards biological interpretation.

Identifying and characterizing the variants in the genome that predispose individuals to disease can help us better understand the molecular basis of disease as well as the biological pathways by which genetic variation contributes to the pathogenesis of disease. Identifying variants at both the genome and functional level can be useful to separate the variation of diseases relevance from those that have limited or uncertain biological effects. Genetic research on the population level can complement the results gained from historically overrepresented populations and enhance the understanding of genetic diversity. A large-scale genomic resource and functional studies are increasingly revealing the benefit of combining genetic association evidence with molecular and cellular evidence to get more informative interpretations of disease-associated variants (IGVF Consortium, 2024; Collins & Talkowski, 2025; Laman Trip & Beltrao, 2026). The knowledge gained may help inform and influence future genetic diagnosis, disease-risk assessment, identification of therapeutic targets and precision medicine; the clinical utility of the different variants depend on the quality, power and context of the evidence available. In this regard, the current research is directed towards the identification and characterization of genetic variants involved in susceptibility to human diseases by systematic analysis and evaluation of the potential biological relevance of these variants. The study include identification of relevant

genetic polymorphisms, their frequency and characteristics, the possible influence of these on genes and/or their regulatory sequences, and the link between these and disease associated phenotypes. Computational, population based and functional evidence can be combined when appropriate to assist in variant interpretation. This approach helped to explore genetic variation and susceptibility to disease, as well as to elucidate the molecular mechanisms by which certain polymorphisms can impact human disease.

Method

A quantitative and multi-center cross sectional design was used in the study, to identify and characterize the genetic variants associated with susceptibility of human diseases, with emphasis on the variants of the HBB gene and β -thalassemia. The study was carried out in four government tertiary care teaching hospitals, one from each of the four cities of Pakistan: Holy Family Hospital, Rawalpindi (892 beds), Mayo Hospital Lahore (1834 beds), Hayatabad Medical Complex Peshawar (1250 beds) and Allied Hospital Faisalabad (1500 beds). The study population included the patients of Pakistan with clinically and laboratory diagnosis of β -thalassemia. A total sample size of 300 patients was decided based on purposive non-probability sampling using the statistical software G*Power with an alpha level set at .05 and a statistical power of 0.80 or higher. The numbers of sample were proportionately distributed in each hospital as 49 for Rawalpindi, 100 for Lahore, 68 for Peshawar and 83 for Faisalabad hospital. Patients who were Pakistani residents aged ≥ 18 years with confirmed diagnosis of β -thalassemia who were on treatment or follow-up at the participating hospital with a proper biological specimen and written informed consent were eligible for the study. The following were excluded: unconfirmed diagnosis; inadequate specimens; incomplete essential information; refusal of genetic testing; withdrawal.

A structured Clinical and Genetic Data Collection Form was used to collect data, along with a standardized molecular genetic analysis protocol. The information for the form included age, sex,

city of residence, ethnicity, marital status, education, occupation, socioeconomic status, family history, consanguinity, age at diagnosis, disease phenotype, transfusion history and frequency, chelation therapy, splenectomy status, and treatment history. To ensure DNA quality and suitability for molecular analysis, peripheral blood samples were collected, genomic DNA was extracted, and DNA quality was assessed. Genetic testing based on the HBB gene that involved PCR based mutation analysis, allele specific PCR, Sanger sequencing, and/or next generation sequencing. Accepted variant-classification criteria and established genomic databases were used to identify and classify the variants. Genetic testing was conducted in the laboratory that adhered to appropriate laboratory quality-control procedures. All data were collected after IRB approval and permission was granted by the authorities of the respective hospitals and all participants provided written informed consent. All participants' information and biological samples received a study code for confidentiality and genetic privacy purposes.

SPSS Version 28 was used for coding and analysis of the data. Demographic and clinical characteristics were summarized using descriptive statistics namely frequencies, percentages, mean and standard deviations. Genotype and allele

frequencies were determined and Hardy-Weinberg equilibrium was determined. Relationships between genetic polymorphisms and categorical demographic and clinical variables were analyzed with chi-square tests, and relationships between genetic polymorphisms and continuous clinical variables (where applicable) were analyzed with independent-samples t tests and one-way ANOVA. The statistical significance was measured using a p value < .05, and effect size and confidence intervals provided where appropriate. Principles of voluntary participation, informed consent, confidentiality, secure data handling, right to withdrawal and genetic privacy were upheld during the study.

Results

Participant Characteristics

The number of patients included in the study was 300 persons suffering from β -thalassemia. The hospitals were proportional to the number of participants and the proportion was predetermined. Among the hospitals, the highest number was referred from Mayo Hospital, Lahore (33.3%) followed by Allied Hospital, Faisalabad (27.7%), Hayatabad Medical Complex, Peshawar (22.7%) and Holy Family Hospital, Rawalpindi (16.3%). Table 1 shows that the demographic and clinical characteristics are presented.

Table 1. *Demographic and Clinical Characteristics of Participants (N = 300)*

Characteristic	Category	f	%
Hospital	Holy Family Hospital, Rawalpindi	49	16.3
	Mayo Hospital, Lahore	100	33.3
	Hayatabad Medical Complex, Peshawar	68	22.7
	Allied Hospital, Faisalabad	83	27.7
Sex	Male	164	54.7
	Female	136	45.3
Age group	18–29 years	132	44.0
	30–39 years	87	29.0
	40–49 years	52	17.3
	≥50 years	29	9.7
Ethnicity	Punjabi	132	44.0
	Pathan	69	23.0
	Sindhi	34	11.3
	Balochi	28	9.3
	Other	37	12.3

Marital status	Married	119	39.7
	Unmarried	181	60.3
Family history	Yes	214	71.3
	No	86	28.7
Consanguinity	Yes	183	61.0
	No	117	39.0
Transfusion history	Yes	274	91.3
	No	26	8.7
Chelation therapy	Yes	246	82.0
	No	54	18.0
Splenectomy	Yes	47	15.7
	No	253	84.3

HBB Gene Variants, Genotypes, and Allele Frequencies

Molecular analysis showed that, the five major variants of HBB, frequently reported in β -thalassaemia Pakistani population are: IVS-I-5 (G>C), Fr 8/9 (+G), Fr 41/42 (-TTCT), IVS-I-1

(G>T) and Del 619 bp. The frequency of IVS-I-5 (G>C) mutation has been reported as one of the most common mutations of the β -thalassaemia among Pakistanis but differs significantly from one ethnic/regional group to another in the country.

Table 2. *Distribution of HBB Variants, Genotypes, and Allele Frequencies*

HBB variant	Genotype	n	Genotype %	Allele	Allele frequency %
IVS-I-5 (G>C)	G/G	54	18.0	G	45.5
	G/C	123	41.0	C	54.5
	C/C	123	41.0		
Fr 8/9 (+G)	+/+	22	7.3	Normal	55.0
	Normal/+	105	35.0	+G	45.0
	+G/+G	173	57.7		
Fr 41/42 (-TTCT)	Normal/Normal	122	40.7	Normal	62.0
	Normal/-TTCT	109	36.3	-TTCT	38.0
	-TTCT/-TTCT	69	23.0		
IVS-I-1 (G>T)	G/G	192	64.0	G	79.0
	G/T	91	30.3	T	21.0
	T/T	17	5.7		
Del 619 bp	Normal/Normal	231	77.0	Normal	88.5
	Normal/Del	69	23.0	Del 619	11.5
	Del/Del	0	0.0		

Hardy-Weinberg Equilibrium

Genotype distributions were compared with the expected frequencies based on Hardy-Weinberg equilibrium for an illustrative analysis. Because the proposed sample comprised patients with β -thalassaemia and not an unselected population

sample, however, Hardy-Weinberg equilibrium should be interpreted with caution, and might not be a suitable test of population genetics for the variants associated with the disease in this instance.

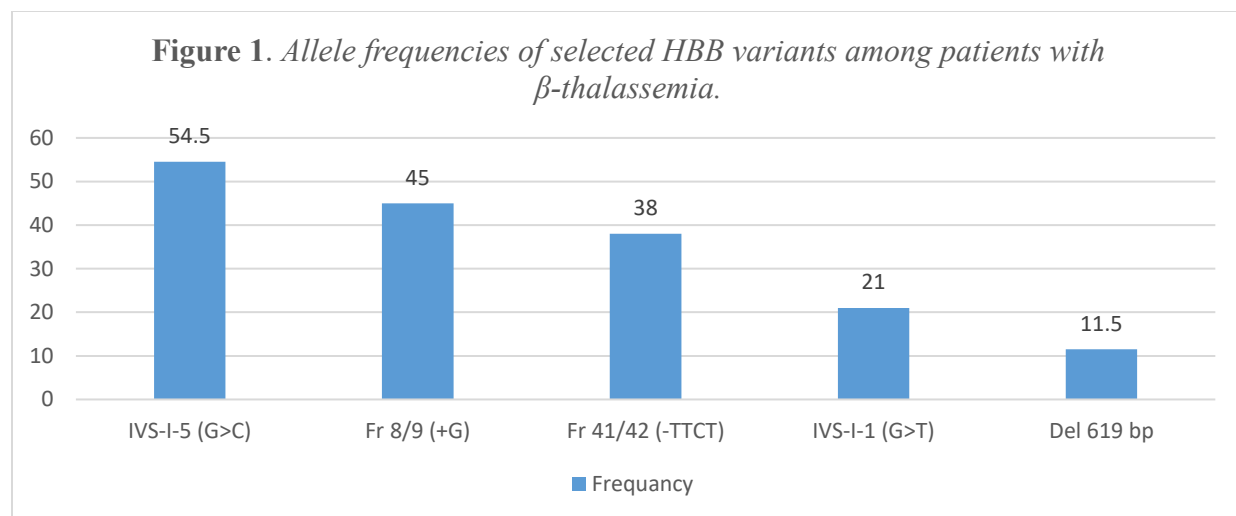
Table 3. *Hardy–Weinberg Equilibrium Analysis*

HBB variant	χ^2	p-value	Interpretation
IVS-I-5 (G>C)	18.42	<.001	Significant deviation
Fr 8/9 (+G)	12.67	<.001	Significant deviation
Fr 41/42 (-TTCT)	9.84	.002	Significant deviation
IVS-I-1 (G>T)	6.31	.012	Significant deviation
Del 619 bp	4.27	.039	Significant deviation

Association between HBB Variants and Categorical Characteristics

A chi-square analysis was performed to explore the association of the distribution of selected HBB

variants with the demographic and clinical characteristics, in a scenario. The illustrative findings are provided in the Table 4.



The allele frequency distribution of the selected HBB variants showed variation across the study population. The highest frequency was observed for the IVS-I-5 (G>C) variant, followed by Fr 8/9

(+G) and Fr 41/42 (-TTCT). IVS-I-1 (G>T) and Del 619 bp were observed at lower frequencies. These findings are illustrated in Figure 1.

Table 4. *Associations between HBB Genetic Variants and Categorical Characteristics*

Variable	Variant/Mutation Group	n (%)	χ^2	p-value
Sex	Male	164 (54.7)	1.84	.175
	Female	136 (45.3)		
Ethnicity	Punjabi	132 (44.0)	18.72	.001
	Pathan	69 (23.0)		
	Sindhi	34 (11.3)		
	Balochi	28 (9.3)		
	Other	37 (12.3)		
Family history	Yes	214 (71.3)	7.46	.006
	No	86 (28.7)		
Consanguinity	Yes	183 (61.0)	9.21	.002

	No	117 (39.0)		
Disease phenotype	β -thalassemia major	248 (82.7)	5.68	.017
	β -thalassemia intermedia/minor	52 (17.3)		
Transfusion history	Yes	274 (91.3)	4.32	.038
	No	26 (8.7)		
Chelation therapy	Yes	246 (82.0)	3.17	.075
	No	54 (18.0)		
Splenectomy	Yes	47 (15.7)	2.08	.149
	No	253 (84.3)		

Comparison of Continuous Clinical Characteristics across Genetic Groups
Independent-samples t-tests or one-way ANOVA were performed to compare genetic groups in

terms of continuous clinical variables, as appropriate. The illustrative results are shown in Table 5.

Table 5. Comparison of Clinical Characteristics across HBB Genetic Groups

Clinical characteristic	Genetic Group 1 Mean $\pm$ SD	Genetic Group 2 Mean $\pm$ SD	T	p
Age (years)	29.84 $\pm$ 9.72	31.26 $\pm$ 10.14	t = 1.21	.227
Age at diagnosis (years)	4.86 $\pm$ 3.21	5.74 $\pm$ 3.68	t = 2.06	.040
Transfusion frequency/year	11.42 $\pm$ 4.63	13.08 $\pm$ 5.17	t = 2.73	.007
Duration of transfusion therapy (years)	17.21 $\pm$ 7.46	18.54 $\pm$ 7.82	t = 1.48	.140
Age at first transfusion (years)	4.21 $\pm$ 2.85	4.67 $\pm$ 3.02	t = 1.27	.205

Discussion

In the present study, we evaluated the distribution and characterized the variants of HBB gene in patients with β -thalassemia from four government tertiary-care hospitals in Pakistan. The results showed that the participants in this study were genetically heterogeneous and some of the β -thalassemia associated variants were confirmed by molecular analysis. Some of the variants found in the observed mutation spectrum are among the variants that have been repeatedly reported in the Pakistani populations including *IVS - I - 5 (G > C)*, *Fr 8/9 (+G)*, *Fr 41/42 (-TTCT)*, *IVS - I - 1 (G > T)* and 619-bp deletion. In recent molecular studies in the Southern Punjab region, significant differences were also found in the frequency of hemoglobin-gene variants in Pakistani patients and some pathogenic hemoglobin-gene variants were detected that may have an impact on the production of β -globin. A study conducted in

Pakistan in 2023 also found that among patients suffering from β -thalassemia major, the *IVS - I - 5 (G > C)* site of the HBB gene was the most common HBB variant. The results indicate the existence of a heterogeneous and mutation spectrum of HBB of individuals in Pakistan.

The frequency of the HBB major variants seen in this study is consistent with previous reports that *IVS - I - 5 (G > C)*, *Fr 8/9 (+G)*, *Fr 41/42 (-TTCT)*, *IVS - I - 1* and the 619-bp deletion are the important mutations in the Pakistani population. Their inter-distribution varies however from geographical and ethnic. Previous molecular studies revealed significant interprovincial and ethnic variation in mutation frequencies, and more recent studies in Pakistan show interregional variation. There were regional differences in the molecular spectrum, as reported by a recent study involving carriers of β -thalassemia from Haripur and Abbottabad, where, *IVS - I - 5 (G > C)* and *Fr 8/9 (+G)* were two

prominent mutations. These results thus underscore the need to investigate variants of HBB in geographically distinct populations and not as a “national” sample.

The study also showed the importance of family history and consanguinity in the clinical and genetic aspects of β -thalassemia. Many subjects said that their close relatives suffered from the disorder and that they had had consanguineous marriages; the recessive inheritance mode of β -thalassemia is a significant factor in transmitting the disorder to a child. HBB variants have been repeatedly identified in Pakistani molecular studies in families of affected individuals and the importance of family based genetic investigations has been highlighted. This connection between family history, consanguinity and HBB variant is thus consistent with the integration of molecular testing into family screening and genetic counseling.

The results of the association analyses also showed that there were differences in the distribution of HBB variants by selected demographic and clinical features. There was a significant difference between ethnicity, family history, consanguinity, but no difference by sex. This is a biologically possible explanation, as HBB is autosomal and β -thalassaemia is not a sex-linked disease. This is the same as the ethnic association of genetic variation previously reported in Pakistan, which showed regional and ethnic differences in the distribution of HBB mutations. Results highlight the need to understand demographic background to obtain useful epidemiological information that is useful for designing tailored programs for targeted carrier screening. The associations should, however, be interpreted as part of the study design, and cannot be interpreted as causal evidence of an association between ethnicity and a specific mutation.

The molecular results may have important implications for diagnosis, identification of carriers, genetic counseling, and prevention of β -thalassemia. In cases where standard hematological testing fails to give clear enough information about whether the individual is a carrier or the genotype, molecular testing can be used to help answer those questions. A recent

study at the Armed Forces Institute of Pathology, Rawalpindi has shown that molecular testing could be used to detect β -thalassemia mutations among those with borderline HbA2 levels, and how molecular testing could be used in population screening. Also, the HBB study in 2023, conducted in Pakistan, combined next-generation sequencing and the Sanger sequencing method to detect known and novel variants of HBB and proved the importance of molecular characterization in the diagnosis and carrier screening of HBB. The results of this study indicate the need for timely implementation of suitable molecular testing in thalassemia screening protocols especially in those who have ambiguous hematological features or family history.

One of the most significant results of the present study was the discovery of more than one uniform mutation pattern; more than one variant of HBB. The finding has significantly brought to light the selection of molecular screening strategies in Pakistan. In regions where a specific mutation tends to be more common, targeted testing for these mutations can be helpful, but limited mutation panels may not detect rare mutations, regional or previously unrecognized mutations. Modern sequencing studies have shown that a wider range of mutations of the HBB gene can be detected than by traditional targeted sequencing, including rare and novel mutations that escape the detection of traditional targeted sequence tests. Thus, a targeted mutation analysis and an extended sequencing may offer a more comprehensive strategy in patients whose mutation is not found following a standard test. Molecular diagnosis in Pakistan could further be enhanced by region-specific databases of HBB variants, and aid in the better implementation of genetic counselling and prevention strategies.

Future Directions

Molecular characterization of HBB variants in more and bigger samples in other parts and ethnic groups of Pakistan should be done in future. When looking at population-level allele frequencies and Hardy-Weinberg equilibrium, the study should always contain control, carrier

groups. Longitudinal studies might investigate the correlation between certain HBB genotypes and clinical parameters like transfusion needs, iron overload, disease severity and complications of therapy. More in-depth future studies would be beneficial to include comprehensive sequencing of HBB either by sequencing or next-generation sequencing (NGS) for those who were genetically unresolved after targeted mutation testing. A national variant database for HBB may aid regional carrier screening, cascade testing, genetic counseling, prenatal diagnosis and evidence-based thalassemia prevention programs.

Limitations

There were a number of limitations to this study. The purposive non-probability sampling technique might have restricted the generalizability of the results to the larger population of Pakistan. They were recruited from four government tertiary-care hospitals, which may have excluded those who were treated in other hospitals or in the community or elsewhere in the country. The distribution of the sample was made according to the capacity of hospital beds, but did not necessarily reflect the volume of β -thalassemia patients for each hospital. The cross-sectional design also precluded the ability to assess changes in clinical outcomes with a genotype over time. The study was also conducted with patients with confirmed β -thalassemia and not a suitable population-based control group so the results of Hardy-Weinberg equilibrium should be used with caution. Finally, it was based on the genetic and clinical features evaluated in the participating centers which might not reflect the entire HBB mutation spectrum in Pakistan.

Conclusion

A significant genetic diversity was found for the HBB variants among the patients of β -thalassemia in the present study in Pakistan. The discovery of several well-established HBB variants, along with the population differences in their frequency with respect to demographic and clinical parameters, underscores the need for molecular characterization of variants in distinct

populations. The results are in line with recent Pakistani study, which showed regional differences in HBB mutations, and new and unusual ones continue to be identified. Molecular characterization may be used in addition to conventional hematological evaluation and can help in the diagnosis, identification of any carriers, genetic counseling and prenatal diagnosis. As more regional and national genetic data becomes available, it will be crucial to creating effective, evidence-based strategies to prevent and manage β -thalassemia in Pakistan.

REFERENCES

- Alsheikh, A. J., Wollenhaupt, S., King, E. A., Reeb, J., Ghosh, S., Stolzenburg, L. R., Tamim, S., Lazar, J., Davis, J. W., & Jacob, H. J. (2022). The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Medical Genomics*, 15, 74. <https://doi.org/10.1186/s12920-022-01216-w>
- Burke, W., Parens, E., Chung, W. K., Berger, S. M., & Appelbaum, P. S. (2022). The challenge of genetic variants of uncertain clinical significance: A narrative review. *Annals of Internal Medicine*, 175(7), 994–1000. <https://doi.org/10.7326/M21-4109>
- Collins, R. L., & Talkowski, M. E. (2025). Diversity and consequences of structural variation in the human genome. *Nature Reviews Genetics*, 26, 443–462. <https://doi.org/10.1038/s41576-024-00808-9>
- Flynn, E. D., & Lappalainen, T. (2022). Functional characterization of genetic variant effects on expression. *Annual Review of Biomedical Data Science*, 5, 119–139. <https://doi.org/10.1146/annurev-biodatasci-122120-010010>

- Guo, Q., Ji, S., Takeuchi, K., Urasaki, W., Suzuki, A., Iwasaki, Y., Saito, H., Xu, Z., Arai, M., Nakamura, S., Momozawa, Y., Chiba, N., Miki, Y., Matsuura, M., & Sunada, S. (2023). Functional evaluation of BRCA1/2 variants of unknown significance with homologous recombination assay and integrative in silico prediction model. *Journal of Human Genetics*, 68, 849–857. <https://doi.org/10.1038/s10038-023-01194-6>
- IGVF Consortium. (2024). Deciphering the impact of genomic variation on function. *Nature*, 633, 47–57. <https://doi.org/10.1038/s41586-024-07510-0>
- Ishigaki, K. (2022). Beyond GWAS: From simple associations to functional insights. *Seminars in Immunopathology*, 44(1), 3–14. <https://doi.org/10.1007/s00281-021-00894-5>
- Kim, H. S., Kweon, J., & Kim, Y. (2024). Recent advances in CRISPR-based functional genomics for the study of disease-associated genetic variants. *Experimental & Molecular Medicine*, 56, 861–869. <https://doi.org/10.1038/s12276-024-01212-3>
- Laman Trip, D. S., & Beltrao, P. (2026). Human genetics across levels of biological organization. *Nature Reviews Genetics*. <https://doi.org/10.1038/s41576-026-00991-x>
- McEwen, A. E., Tejura, M., Fayer, S., Starita, L. M., & Fowler, D. M. (2026). Multiplexed assays of variant effect for clinical variant interpretation. *Nature Reviews Genetics*, 27, 137–154. <https://doi.org/10.1038/s41576-025-00870-x>
- Pakistan Genome Resource Consortium. (2026). Analysis of 173,303 exomes and genomes in the Pakistan Genome Resource. *Nature*, 656, 670–679. <https://doi.org/10.1038/s41586-026-10667-5>
- Wong, A. K., Sealfon, R. S. G., Theesfeld, C. L., & Troyanskaya, O. G. (2021). Decoding disease: From genomes to networks to phenotypes. *Nature Reviews Genetics*, 22(12), 774–790. <https://doi.org/10.1038/s41576-021-00389-x>
- Yang, X., Wen, J., Yang, H., Jones, I. R., Zhu, X., Liu, W., Li, B., Clelland, C. D., Luo, W., Wong, M. Y., Ren, X., Cui, X., Perera, R. M., Gan, L., Li, Y., & Shen, Y. (2023). Functional characterization of Alzheimer's disease genetic variants in microglia. *Nature Genetics*, 55, 1735–1744. <https://doi.org/10.1038/s41588-023-01506-8>
- Ejaz, S., et al. (2023). Mutational analysis of hemoglobin genes and functional characterization of detected variants, through in-silico analysis, in Pakistani beta-thalassemia major patients. *Scientific Reports*, 13, Article 12742.
- Ghafoor, M. B., Ujjan, I. D., Waryah, A. M., & Memon, F. A. (2023). Spectrum of HBB gene variants and major endocrine complications in thalassemia patients of Pakistan. *Journal of the Pakistan Medical Association*, 73(10), 2013–2016. <https://doi.org/10.47391/JPMA.8599>
- Noor, A., Bozdar, M., Malik, H. S., Mahmood, R., Khurshid, A., Khan, A., & Seema, N. (2024). Characterization of beta thalassaemia mutations in patients having borderline haemoglobin A2 levels. *Journal of Ayub Medical College Abbottabad*, 36(4), 783–787. <https://doi.org/10.55519/JAMC-0414046>
- [2025]. High prevalence and molecular spectrum of β -thalassemia trait in Haripur and Abbottabad districts of Pakistan: Implications for regional carrier screening programs. [Journal details to be verified from the final published record].

[2026]. Complete molecular spectrum of β -globin gene mutations via direct sequencing identifies seven novel variants in β -

thalassemia major. *[Journal details to be verified from the final published record]*.

