Original Article

Molecular Characterization of Osteogenesis Imperfecta in Pakistani Families: Identification of COL1A1 and COL1A2 Variants

Iqra Munir, Saman Khan, Areeba Manzoor, Urooj Fatima, Muhammad Zeeshan, Moaz ur Rahman, Asma Ali Khan

Center of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan

Abstract

Objective: Osteogenesis imperfecta (OI) is a genetically heterogeneous disorder characterized by bone fragility, recurrent fractures, and skeletal and extra-skeletal deformities. Although *COL1A1* and *COL1A2* variants are well-established causes of OI, molecular data from the Pakistani population remain limited.

To identify pathogenic variants in selected regions of *COL1A1* and *COL1A2*, assess genotype–phenotype associations, and expand the molecular understanding of OI in Pakistan.

Methodology: Sixteen clinically diagnosed Pakistani patients from multiple families were recruited. Selected exons of *COL1A1* (exons 17 and 33) and *COL1A2* (exons 19 and 21) were analyzed using targeted sequencing. Identified variants were evaluated for their potential association with clinical phenotypes, supported by structural analysis.

Results: Multiple nucleotide substitutions were identified. A distinct missense variant, c.965G>T (p.Gly322Val), was detected in a member of the consanguineous family PK0I09. Structural analysis suggested increased local steric hindrance due to the replacement of glycine with the bulkier valine residue within the collagen triple-helical region.

Conclusion: The findings highlight the contribution of *COL1A1* and *COL1A2* variants to diverse OI phenotypes in Pakistani patients. The identified missense variant provides further insight into the molecular mechanisms underlying OI and contributes valuable genetic data to the international literature.

Keywords: *COL1A1* and *COL1A2* genes, Osteogenesis imperfecta, Sanger sequencing mutation analysis, Pakistani population genetic profiling

How to cite this:

Munir I, Khan S, Manzoor A, Fatima U, Zeeshan M, Rahman Mu, Khan AA. Molecular Characterization of Osteogenesis Imperfecta in Pakistani Families: Identification of COL1A1 and COL1A2 Variants. J Health Advance Res Genom. 2026;1(1):41-46

Corresponding Author: Asma Ali Khan **Email:** asma.ali@cemb.edu.pk

Received: 14-07-2026 **Revised:** 24-09-2026 **Accepted:** 04-11-2026

Introduction

Osteogenesis imperfecta (OI), commonly referred to as brittle bone disease, is one of the most extensively studied heritable skeletal disorders, yet it remains a major clinical and genetic challenge due to its heterogeneity. It is primarily characterized by bone fragility, recurrent fractures, skeletal deformities, and growth deficiency, but extra-skeletal manifestations such as dentinogenesis imperfecta, blue or gray sclera, hearing loss, joint hypermobility and cardiovascular complications are also frequently observed.[1,2] Osteogenesis imperfecta (OI) is traditionally classified according to the Sillence classification (1979) into four major clinical types (I–IV).

With advances in molecular genetics, OI is now recognized as a genetically heterogeneous group of disorders with pathogenic variants found in more than 20 genes including TENT5A, KDELR2, CRTAP, LEPRE1, FKBP10, SERPINF1, WNT1.[2,4] The current Nosology and Classification of Genetic Skeletal Disorders recognizes OI as a genetically heterogeneous group, but in clinical practice, the Sillence classification remains widely used for its simplicity and correlation with disease severity.[4]

Collagen is the most abundant structural protein in the extracellular matrix, characterized by its unique right-handed triple-helical structure formed by three left-handed polyproline II–type chains. Each chain consists of repeating Gly–X–Y motifs, where glycine, the smallest amino acid, is essential for tight packing of the helix, while

proline and hydroxyproline stabilize the structure through steric and hydrogen-bonding interactions. Type I collagen, the principal fibrillar collagen in bone, tendon, and skin, is a heterotrimer composed of two α1(I) chains encoded by COL1A1 and one α2(I) chain encoded by COL1A2. Mutations in these genes, particularly glycine substitutions within the Gly–X–Y repeat, disrupt helix stability, delay chain folding, and impair fibril assembly. Such structural defects reduce bone matrix integrity, leading to the phenotypic spectrum of osteogenesis imperfecta (OI), which ranges from mild bone fragility to perinatal lethality. The genotype–phenotype correlation underscores the central role of COL1A1 and COL1A2 in maintaining collagen architecture and skeletal strength.[5]

COL1A1 is located on chromosome 17q21.33, spans approximately 18 kb, and contains 51 exons. COL1A2 is located on chromosome 7q21.3, spans about 38 kb, and comprises 52 exons. Both genes are highly conserved and extensively expressed in osteoblasts and fibroblasts. Mutations in these loci—especially glycine substitutions within the triple-helical domain or nonsense mutations leading to haploinsufficiency—are the predominant cause of osteogenesis imperfecta, accounting for nearly 90% of cases.[5,6] (NCBI Gene, 2026; Saravanan et al., 2024). Pathogenic variants destabilize collagen fibril formation and secretion, ultimately impairing bone strength. The structural and functional importance of COL1A1 and COL1A2 highlights them as central targets for both diagnostic genetic testing and future therapeutic strategies in OI.[2,5]

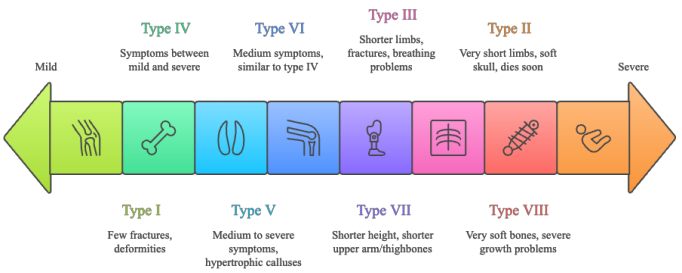


Figure 1: *Types of osteogenesis imperfecta with their severity range*

Epidemiologically, OI is considered a rare disorder, with an estimated incidence of 6–7 per 100,000 live births globally. However, the distribution of OI types and mutations varies across populations, influenced by genetic background, founder effects, and rates of consanguinity. In countries such as Pakistan, where consanguineous marriages are common, recessive forms of genetic disorders are expected to occur more frequently than in outbred populations. Yet, despite the high burden of inherited conditions, there is a striking scarcity of published molecular data on OI from Pakistan. Most clinical diagnoses are made based on radiological and phenotypic features, while genetic confirmation remains limited due to resource constraints. As a result, the true mutational spectrum of OI in this region remains under-characterized, creating a significant gap inboth local healthcare and global understanding of the disease.[2,7] In this context, genetic testing offers immediate benefits by enabling precise diagnosis, facilitating informed family planning, and supporting genetic counseling to reduce recurrence risk in high-risk families.[8]

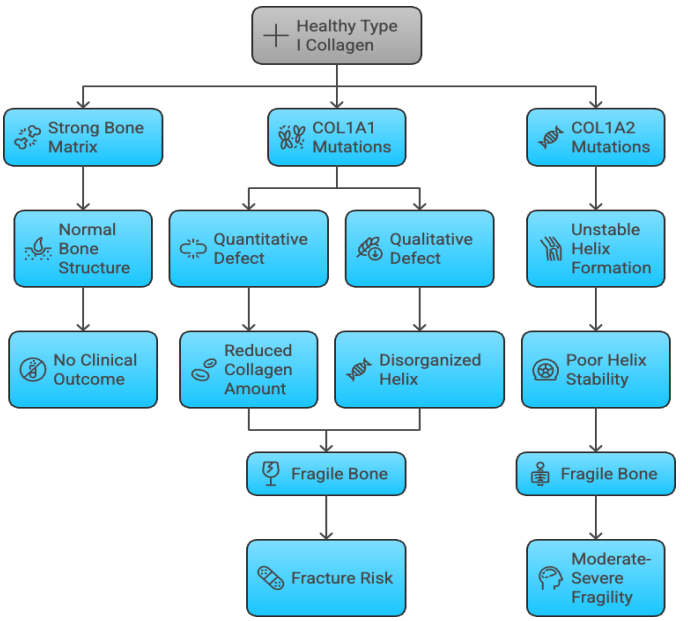


Figure 2: *Role of COL1A1 and COL1A2 genes in osteogenesis imperfecta.*

This study was therefore designed to address the paucity of molecular data on OI in Pakistan. By recruiting clinically diagnosed patients from multiple families and sequencing key exons of COL1A1 and COL1A2, we aimed to identify causative mutations, assess genotype–phenotype relationships, and contribute new data to the international knowledge base. This work represents an important step toward integrating molecular genetics into the diagnostic and counseling framework for OI in Pakistan, with potential implications for neighboring regions sharing similar genetic and sociocultural contexts.

Methodology

Institutional Review Board (IRB) of Centre of Excellence in Molecular Biology (CEMB), University of the Punjab, Lahore approved this study. Sixteen clinically diagnosed patients with Osteogenesis Imperfecta (OI) were recruited

from 12 unrelated families across different regions of Punjab, Pakistan. Whole Blood Sample was withdrawn from the affected and available unaffected family members in EDTA coated sterile vials and stored at -4°C. Informed written consent were taken from the participating individuals or their guardians.

Genomic DNA was isolated from the whole blood samples by using organic extraction method standardized in Genetic Disease lab, CEMB. DNA purity was checked by gel electrophoresis and NanoDrop quantification.

Primers targeting the *COL1A1* (exons 17 and 33) and *COL1A2* (exons 19 and 21) regions were designed using Primer3 and UCSC in-silico PCR, with SNPs excluded via SNPCheck to ensure specificity. PCR reactions were set up in 25 µL volumes containing genomic DNA, primers, dNTPs, $MgCl_2$, Taq polymerase, and buffer. Amplification was performed using a thermal cycler (Bio-Rad), with optimized cycling conditions; initial denaturation at 95 °C for 5 min; 35 cycles of 95 °C for 30 sec, annealing at optimized temperatures (58–62 °C depending on exon) for 30 sec, and extension at 72 °C for 45 sec, final extension at 72 °C for 10 min. PCR products werevisualized on 2% agarose gels with ethidium bromide. Amplicons were purified using a commercial PCR purification kit (Qiagen) and sequenced bidirectionally via Sanger sequencing.

Chromas 2.6.2 was used to visualize and edit raw sequence. Consensus sequences were aligned against reference genomes in BLAST (Basic Local Alignment Search Tool), UCSC genome browser, and Ensembl Genome Browser. Mutations were annotated following HGVS nomenclature. Structural significance of glysine residues was studied using AlphaFold2. Pathogenicity predictions were obtained using MutationTaster, PolyPhen-2, and SIFT. Structural consequences were evaluated through ExPASy Swiss-Model, comparing wild-type and mutant protein conformations.

Results

Sixteen individuals with clinically diagnosed osteogenesis imperfecta (OI) from twelve unrelated Pakistani families were studied at clinical and molecular levels. The studied cohort included individuals from both sexes with substantial age and phenotypic heterogeneity ranging from mild features as blue sclera, recurrent fractures, preserved stature to severe functional impairments and skeletal deformities. Extra-skeletal indications included craniofacial abnormalities, ocular impairments and dentinogenesis imperfecta (DI).

Around 30% of the studied individuals had a positive family history while the rest 70% were isolated cases of OI. Targeted sequencing of *COL1A1* (exons 17 and 33) and *COL1A2* (exons 19 and 21) yielded robust amplicons across all samples, and bidirectional Sanger sequencing produced high-quality chromatograms. Multiple nucleotide substitutions were observed after alignment with corresponding reference sequences. Certain substitutions corresponded to known pathogenic variants already catalogued in dbSNP or ClinVar whereas other variants need further evaluation due to absent or limited disease related evidence.

Family PKOI09, recruited from Lahore, presented with multiple affected individuals within a consanguineous pedigree. The index case (V:6), a 29-year-old male, had a history of recurrent fractures since childhood, blue sclerae, dentinogenesis imperfecta, short stature, and mild skeletal and chest deformities. Dual-energy X-ray absorptiometry revealed reduced bone mineral density (BMD), and molecular analysis by Sanger sequencing of exons 19 and 21 of ***COL1A2*** and exon 33 of ***COL1A1*** revealed distinct pathogenic variants in exon 19 of ***COL1A2***.

Patient V:6 carried a missense substitution, **c.965G>T (p.Gly322Val)**. The variant interchanges a highly conserved glycine residue within the collagen triple-helical domain of the α2(I) chain. The c.965G>T (p.Gly322Val) variant was not found in dbSNP or ClinVar during the original variant analysis and was considered a potentially novel disease-associated variant. In silico evaluation using VarSome supported a deleterious effect of the mutation.

These findings demonstrate intrafamilial heterogeneity within PK0109, with multiple affected siblings harboring distinct glycine substitutions in **COL1A2 exon 19**. Both variants disrupt the highly conserved Gly-X-Y motif of the collagen α2(I) chain, destabilizing triple helix formation and thereby accounting for the severe skeletal phenotype observed. The identification of **c.965G>T (p.Gly322Val)** as an unreported mutation expands the mutational spectrum of **COL1A2** and provides further evidence of exon 19 as a mutational hotspot in OI across populations.

A shows the full length AlphaFold model of the protein. Most of the model is made of long, extended loops with very little regular secondary structure. A compact folded domain rich in beta strands (red and orange) sits near the center. A short alpha helix (blue) lies close to it. The circledregion is enlarged on the right. It shows the mutation site (yellow) on a flexible cyan loop. The mutated residue is therefore located in a loop segment and not

inside a defined helix or sheet. The next figures compare its local interactions in the wild type and mutant.

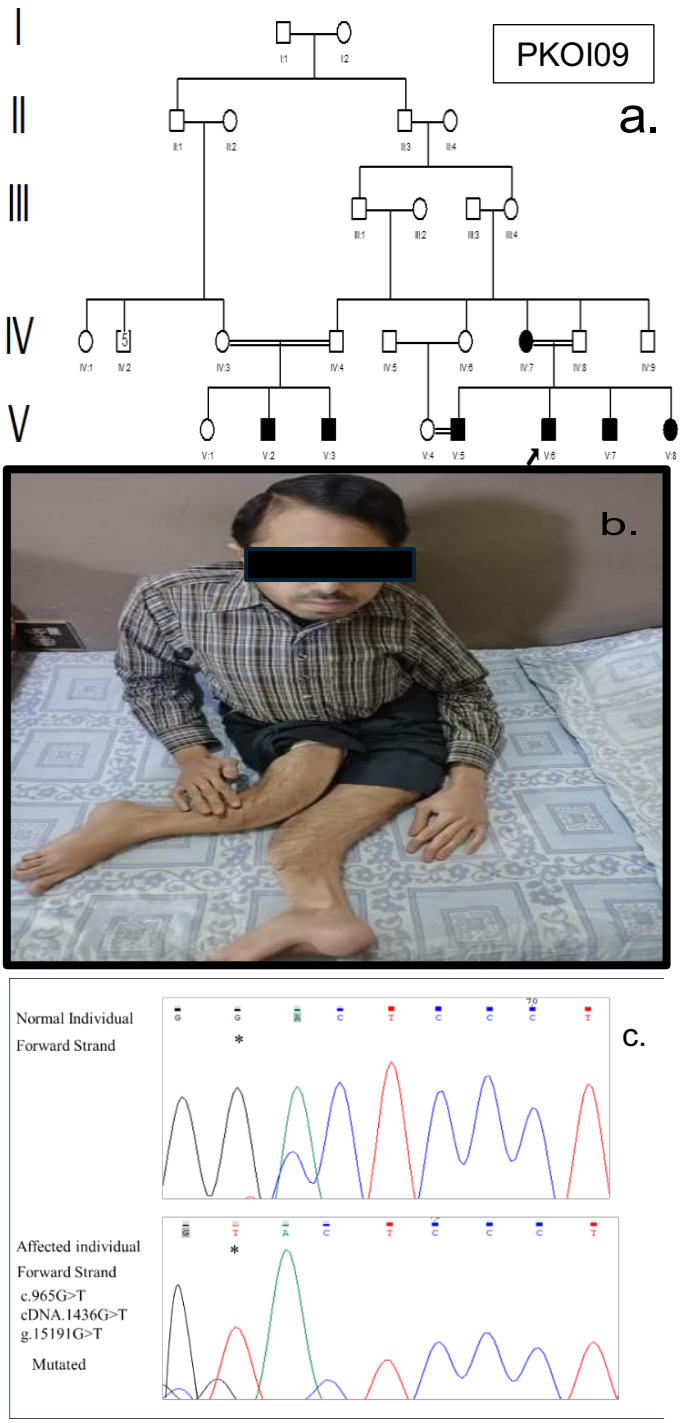


Figure 3: *Pedigree, clinical phenotype and genetic findings of Family affteced with OI(PKO109) a) pedigree of Family PKOI09; b) photograph of V:3 member of family PKOI09; c) Chromatogram showing mutation in exon 19 of COL1a2 c.965G>T*

Figure B shows the wild type region around the mutation site (residues Ala 320 to Pro 324, chain A) as a backbone ribbon. Gly 322 is colored purple. The residues around it lie on a loop, and no secondary structure is seen in this stretch. Figure C shows the same region with the atoms displayed. Gly 322 has no side chain, so only its backbone atoms are present. No hydrogen bonds (blue lines) were found around Gly 322 in the wild type. Two clash contacts (yellow lines) were found. Both involve backbone atoms of Pro 324, which are marked in red. One connects a Pro 324 atom to the carbonyl oxygen of Gly 322. The other connects the Pro 324 nitrogen to the carbonyl oxygen of Gly 322. These contacts are close to van der Waals distance and were picked up by the -0.4 Å cutoff. They are therefore near-contacts and not serious clashes.

Figure D shows the mutant region as a backbone ribbon. Residue 322 is now valine and is colored red. The backbone follows the same loop as in the wild type, so the mutation does not visibly change the overall path of the chain here. Figure E shows the same region with atoms displayed. Valine has a branched side chain, while glycine had none. This new side chain sits close to Leu 323 and the backbone of Pro 324. No hydrogen bonds (blue lines) were seen around residue 322 in the mutant, which is the same as the wild type. The number of clash contacts (yellow lines) went up. The wild type had 2, and the mutant shows several more. Most of them link atoms of the Val 322 side chain and backbone to atoms of Leu 323 and Pro 324, and they cluster around the mutated residue. This suggests that the added valine side chain crowds the neighboring residues in this tight loop.

Discussion

This study represents one of the few molecular investigations of osteogenesis imperfecta (OI) in Pakistan, and provides important insights into both the clinical spectrum and underlying genetic architecture of the disease in this high-consanguinity population. OI is now studied as a genetically heterogeneous group of skeletal abnormalities with defects in structure and quantity of type I collagen, folding and post-translational modifications of collagen protein, bone mineralization,osteoblast differentiation, abnormalities in intracellular signaling pathways and extracellular matrix.[7,9]

The cohort showed a clinical variability consistent with other genotype-phenotype studies. In a large study of 671 patients, Lin et al. reported pathogenic variants in fifteen OI-associated genes and mentioned that ***COL1A1*** and ***COL1A2*** majorly explained genetic basis of OI. Notably, substitutions of glycine in ***COL1A1*** and ***COL1A2*** were linked with more prominent skeletal deformities than collagen haploinsufficiency, demonstrating the biological significance of qualitative and quantitative defects in collagen.[10] In 2024, a study of 294 pediatric patients found

that *COL1A1* or *COL1A2* variants were present in approximately 91% of genetically characterized patients, while the rest 9% carried variants in other OI-related genes. Dentinogenesis imperfecta (DI) and blue sclerae were mainly associated with collagen-gene-related disease.[11]

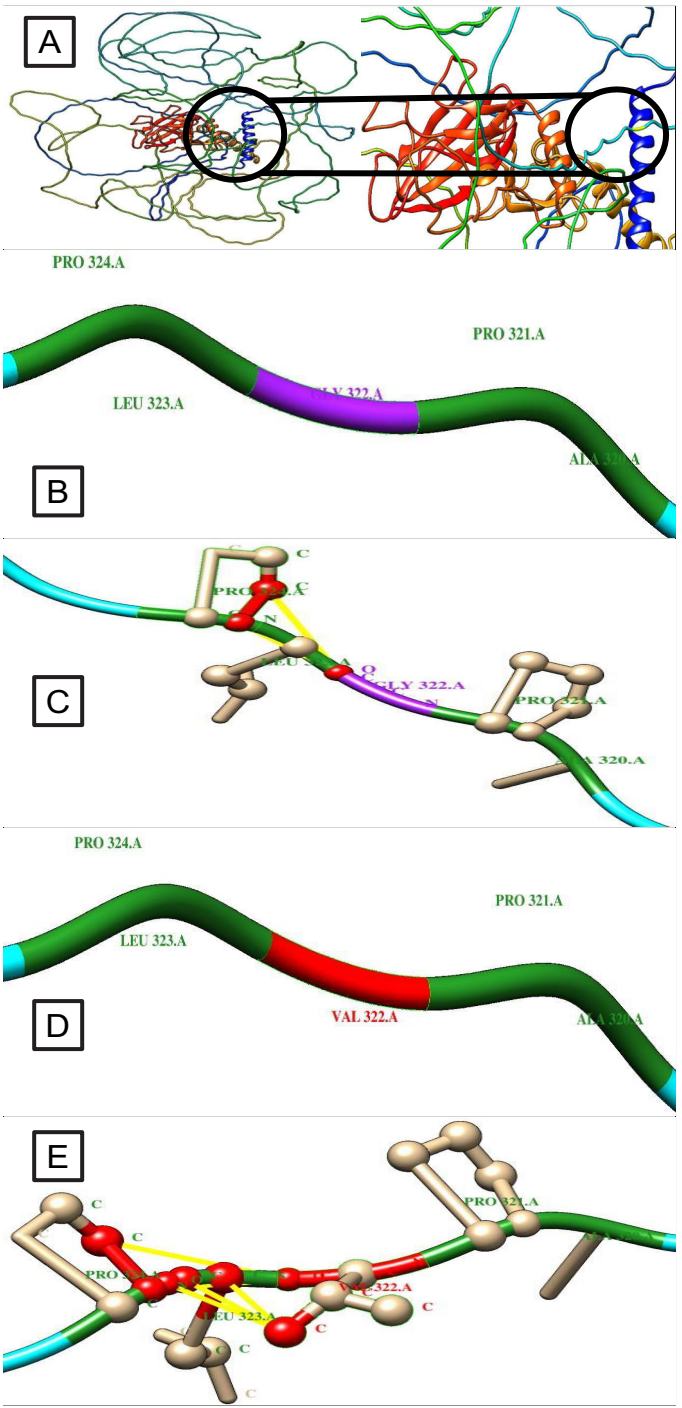


Figure 5: *structural analysis oif p.Gly322Val variant; A)AlphaFold protein showing the exact location of variant in ful length of protein, B-C) wild type Gly-322 along with its interactions, D-E) Val-322 variants and its interactions*

Collagen-related OI phenotype as blue sclerae and DI was present in members of PKO109. Type I collagen is mainly found in bone and is also a vital structural component of connective tissues as dentin, sclera, tendon, skin etc.; and hence, abnormalities in structure of collagen can result in multisystem manifestations rather than just skeletal fragility.[7,12]

The identification of mutations in *COL1A1* and *COL1A2* across multiple families confirms the central role of type I collagen defects in OI, while the discovery of potentially novel variants highlights the contribution of population-specific alleles that have not yet been incorporated into global datasets. The heterotrimer of collagen contains 2 α1(I) chains encoded by the *COL1A1* gene while *COL1A2* gene codes for 1 α2(I) chain. Mutations in these gene can potentially lead to structural disintegration of collagen of affect the amount of collagen. Variants that affect the structural integrity of collagen, particularly substitution of glycine within the triple helical domain cause comparatively severe abnormalities in folding of collagen and overall matrix organization whereas quantitative errors that produce reduce amount of structural normal collagen result in mild disease symptoms.[7,10] The studied Pakistani cohort size is not large enough to test these correlations statistically, but the extreme skeletal deformities observed in glycine substitution carrying individuals happens to be compatible with the reported data.

The mutations identified in PK0109 family affect glycine residues in the triple-helical region of collagen. The right-handed triple helical structure of collagen type-I is possible because of repetitive Gly-X-Y sequence where glycine come in very third position and these small glycine residues allow tight packing of 3 collagen chains in the central axis of collagen.[7] Substitution of glycine with a larger amino acid produces steric limitations in the triple helix causing delayed or faulty folding. These substitutions can disrupt fibril stability, collagen secretion, molecular stability, post-translational modifications and its interaction with various components of the extracellular matrix.[7,10]

variant, c.965G>T (p.Gly322Val) , being identified in the exon of ***COL1A2*** demonstrate a significant molecular revelation in PK0109. This mutation disturb glycine residues within the triple-helical of collagen, explaining the functional role of this region of the α2(I) chain.

Conclusion

This study expands the molecular understanding of osteogenesis imperfecta in Pakistani population and demonstrates the role of variants within *COL1A1* and

COL1A2 genes to explain the phenotypic diversity of OI. Glycine substitutions within *COL1A2* and Insilico analysis of p.Gly322Val adds to the molecular heterogeneity of OI and highlights the importance of molecular characterization for proper diagnosis and accurate genetic counseling.

Ethical Approval: This study was approved by the Ethical Review Committee

Conflict of Interest None

Funding Source None

Authors' Contribution

IM: Conception.

SK, AM: Design of the work.

UF, MZ, MR, AAK: Data acquisition, analysis, or interpretation.

SK,UF, MZ, MR, AAK: Draft the work.

IM, AM: Review critically for important intellectual content.

All authors approve the version to be published.

All authors agree to be accountable for all aspects of the work

References

1. Marini JC, Forlino A, Bächinger HP. Osteogenesis imperfecta. Nat Rev Dis Primers. 2017;3(1):17052.
2. Jovanovic M, Marini JC. Update on the genetics of osteogenesis imperfecta. Calcif Tissue Int. 2024;115:891-914.
3. Sillence DO, Senn A, Danks DM. Genetic heterogeneity in osteogenesis imperfecta. J Med Genet. 1979;16(2):101-16.
4. Unger S, Ferreira CR, Mortier GR, et al. Nosology of genetic skeletal disorders: 2023 revision. Am J Med Genet A. 2023;191(5):1164-209.
5. Saravanan S, et al. Type 1 collagen: synthesis, structure and key functions in bone mineralization. Differentiation.2024;doi:10.1016/j.diff.2024.-100757.
6. NCBI Gene. COL1A1 collagen type I alpha 1 chain [Homo sapiens]. Current human genome annotation.
7. Jovanovic M, Guterman-Ram G, Marini JC. Osteogenesis imperfecta: mechanisms and signaling pathways connecting classical and rare OI types. Endocr Rev. 2022;43(1):61-90.
8. Umair M, Hassan A, Jan A. Homozygous sequence variants in the FKBP10 gene underlie osteogenesis imperfecta in consanguineous families. J Hum Genet. 2016;61(2):207-213.
9. Sillence DO. A dyadic nosology for osteogenesis imperfecta and bone fragility syndromes 2024. Calcif Tissue Int. 2024;115(6):873-90.
10. Lin X, Hu J, Zhou B, et al. Genotype-phenotype relationship and comparison between eastern and western patients with osteogenesis imperfecta. J Endocrinol Invest. 2024;47(1):67-77.
11. Byrd JJ, White AC, Nissen CG, et al. Genotype-phenotype correlations in 294 pediatric patients with osteogenesis imperfecta. JBMR Plus. 2024;8(11):ziae125.
12. Marom R, Rabenhorst BM, Morello R. Osteogenesis imperfecta: an update on clinical features and therapies. Eur J Endocrinol. 2020;183(4):R95-106.