Golestan University of Medical Sciences
Journal of Clinical and Basic Research
Online ISSN: 2538-3736

Bioinformatic analysis of *G6PD* variants with a focus on key Iranian mutations

Romina Jafari Kalokan [1], Sepide Sharifi Tabar [1], Morteza Oladnabi [2,3]*

1. Student Research Committee, Golestan University of Medical Sciences, Gorgan, Iran
2. Gorgan Congenital Malformations Research Center, Jorjani Clinical sciences Research Institute, Golestan University of Medical Sciences, Gorgan, Iran
3. Department of Medical Genetics, School of Advanced Technologies in Medicine, Golestan University of Medical Sciences, Gorgan, Iran
Romina Jafari Kalokan and Sepide Sharifi Tabar contributed equally to the manuscript
* Correspondence: Morteza Oladnabi. Department of Medical Genetics, School of Advanced Technologies in Medicine, Golestan University of Medical Sciences, Gorgan, Iran. Tel: +9891044958390; Email: oladnabidozin@yahoo.com

Abstract

Background: Glucose-6-phosphate dehydrogenase deficiency, an enzyme disorder, can lead to red blood cell dysfunction and affect millions of individuals worldwide. In this study, we aimed to investigate the mutated variants of the G6PD gene at the global and national scales, with a focus on the Iranian population.

Methods: Bioinformatic tools, such as SIFT, PolyPhen-2, PANTHER, FATHMM, I-Mutant, and MUpro, were used to predict the effects of gene variants on the structure, function, and stability of G6PD protein. Also, STRING was utilized to assess protein-protein interactions, Consurf for conservation analysis, I-Tasser and PSIPRED for the analysis of secondary and tertiary structures, and finally, ProtScale for Hydrophobicity analysis.

Results: A total of 215 missense variants were obtained from HGMD. According to bioinformatic analyses, 69 variants were predicted to be pathogenic. Additionally, three G6PD variants were found to be prevalent in the Iranian population.

Conclusion: This study identified 69 missense variants with a possible pathogenic effect on G6PD protein. Moreover, three prevalent variants in the Iranian population were identified as Mediterranean, Chatham, and Cosenza. These findings offer valuable supportive information on the molecular basis of G6PD deficiency, needing further experimental studies to validate if these pathogenic variants have particular clinical relevance.

Article Type: Research Article

Article History

Received: 19 July 2025
Received in revised form: 20 September 2025
Accepted: 29 September 2025
Available online:
DOI: 10.29252/JCBR.9.X.X

Keywords

Glucosephosphate Dehydrogenase Deficiency
Mutation, Missense
In silico analysis
Pathogenic variants
Iranian population

Highlights

What is current knowledge?

- *G6PD* is a key gene whose mutations, such as missense ones, are associated with enzyme deficiency.
- In silico tools are commonly used to predict the pathogenicity and structural effects of genetic variants.

What is new here?

- Out of 215 *G6PD* missense variants, 69 were predicted to be pathogenic using a comprehensive in silico pipeline.
- Three common variants in the Iranian population were also analyzed and interpreted.
- A rigorous filtering strategy combining multiple bioinformatic tools was applied for variant prioritization.

Introduction

Favism or G6PD (Glucose-6-phosphate dehydrogenase) deficiency is a disease characterized by the lack of inadequacy of a key metabolic enzyme in erythrocytes, leading to hemolytic anemia and red blood cell destruction. This condition is inherited with an X-linked recessive pattern. Individuals with G6PD deficiency are susceptible to hemolytic crises when exposed to certain triggers (1,2). The G6PD gene is located at Xq28 and contains 1545 base pairs, organized in 13 exons. The protein encoded by this gene has 514 amino acids (3,4). About 338 mutations have been identified to cause G6PD deficiency, with the majority being missense mutations or small in-frame deletions (5,6) (Figure 1).

The pentose phosphate pathway (PPP) provides NADPH (Nicotinamide adenine dinucleotide phosphate) produced in the glucose metabolic pathway (7). G6PD initiates the pentose phosphate pathway by converting glucose-6-phosphate into 6-phosphogluconolactone and generating NADPH through the reduction of $NADP^+$ (8). NADPH helps preserve adequate levels of reduced glutathione, which is crucial for protecting cells from oxidative stress (9). Insufficient G6PD activity causes NADPH depletion, an increase in oxidized glutathione, and eventually, oxidative hemolysis of red blood cells (10). The prevalence of G6PD deficiency is approximately 400 million people worldwide; however, it is highly variable in different countries. The highest rate of this condition is found in regions such as Africa, the Mediterranean region, and Asia (10,11). As reported by the WHO (World Health Organization), G6PD deficiency is estimated to be highly prevalent (10-14.9%) in Iran (12,13), particularly in Sistan and Baluchestan Province (14). Iran is considered the second largest populated country in West Asia, with dispersed genetic, ethnic, and religious diversity and majority being Persians (65%) and Iranian Azeris (16%), while minority groups being Armenians, Georgians, Jews, Zoroastrians (All <1%) (15,16). Consanguineous marriage occurs commonly in Iran (38.6%) (15,17), contributing to the risk of recessive genetic disorders and multifactorial health conditions (16,17). A potential association between G6PD deficiency and malaria has been proposed in various studies. Individuals with G6PD deficiency are relatively protected against malaria (6), and this explains the relatively high prevalence of this condition in the populations facing malaria (18) (e.g., tropical regions like the south of Iran) (14).

Certain triggers, including specific medications like sulfamethoxazole, infections (Bacterial, viral, or fungal), and certain foods such as fava beans and henna, can lead to the manifestations of G6PD deficiency (18,19). Clinical presentation can range from asymptomatic hemolysis to severe hemolytic anemia, hyperbilirubinemia, jaundice, atherosclerosis, cardiovascular issues,

kernicterus, and even death (5,20). G6PD deficiency is classified into five categories based on severity, class I being the most severe one with serious complications, while class V is considered likely benign (19).

So far, no cure has been found for G6PD deficiency, yet screening programs can be implemented to ease the early diagnosis of this disease and consider appropriate interventions (21). Considering the associated risk of hyperbilirubinemia and kernicterus, screening for G6PD deficiency in neonates is essential for an early diagnosis and avoiding unnecessary hospitalization. In Iran, a fluorescent spot test is performed for screening (12,22). This study aimed to explore the causes of and mutations underlying G6PD deficiency in Iran using bioinformatic analyses.

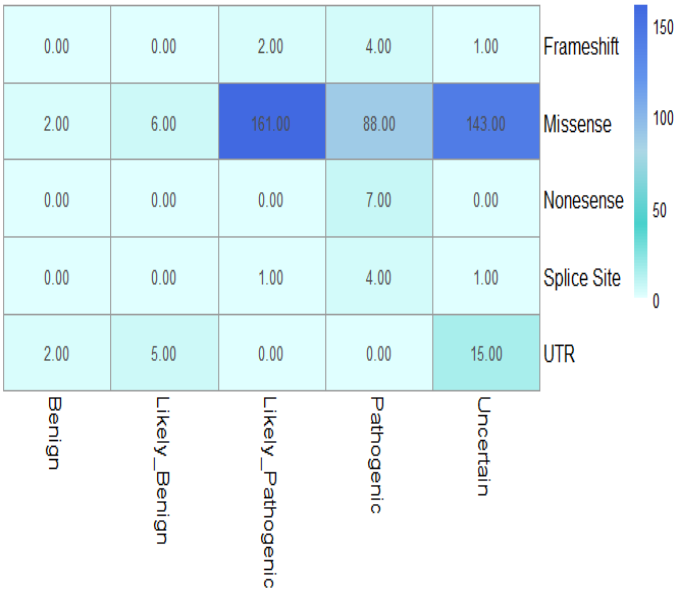

Figure 1. Mutations in the G6PD gene were obtained from the ClinVar database. The data indicates that the majority of reported variants are likely pathogenic missense mutations (161 variants), followed by uncertain missense mutations (143 variants), and pathogenic missense mutations (88 variants). In comparison, benign variants were the least common, with very few or no mutations documented in this group.

Methods

Recruiting variants: HGMD (The Human Gene Mutation Database) (https://www.hgmd.cf.ac.uk/)

This database was used to identify various types of mutations and their frequencies in the G6PD gene. Additionally, the CADD score (Combined Annotation Dependent Depletion) (https://cadd.gs.washington.edu) was utilized to identify mutations and variants. Using CADD, detrimental effect scores of these mutations and variants in the human genome can be evaluated. ClinVar (https://clinvar.ncbi.nlm.nih.gov/) was implemented to retrieve the genetic variants and their associations with diseases and other conditions. ACMG (American College of Medical Genetics and Genomics) was used to interpret the pathogenicity of genetic variants. Based on these guidelines, there are five categories of pathogenicity, including benign, likely benign, uncertain significance, likely pathogenic, and pathogenic.

Pathogenicity analysis: PANTHER, PolyPhen-2, FATHMM, and SIFT were utilized to evaluate the function and structure of G6PD, as well as the effects of missense pathogenic variants on the protein. SIFT (Sorting Intolerant from Tolerant) (https://sift.bii.a-star.edu.sg/) investigates amino acids' physical properties and also sequence homology to anticipate whether the protein function is altered by amino acid substitutions. The SIFT score can vary from 0.00 to 0.05, indicating a likely pathogenic effect, while scores above 0.05 suggest that the substitution is non-pathogenic. PolyPhen-2 (Polymorphism Phenotyping v2) (http://genetics.bwh.harvard.edu/pph2/) evaluates amino acid substitutions and their possible effects on the structure/function of the protein. A score of 0.5 or lower suggests that the mutation is tolerated, while a score above 0.5 indicates that it is likely harmful. FATHM (Functional Analysis through Hidden Markov Models) (https://fathmm.biocompute.org.uk/) is a tool used to analyze coding and non-coding variants and anticipate their possible functional impacts on the human genome. PANTHER (Protein Analysis Through Evolutionary Relationships) (http://www.pantherdb.org/) is employed to identify and classify the function of gene products.

Stability analysis: To assess the stability of the protein, we employed I-Mutant (https://folding.biofold.org/), a tool designed to analyze point mutations and their consequences for the stability of the protein. MUpro (https://mupro.proteomics.ics.uci.edu/) was also employed as a tool for assessing single-site mutations and their impacts on the stability of the protein.

Conservation analysis: We employed Consurf for conservation analyses. Consurf (https://consurf.tau.ac.il/) is a web server that displays conserved regions that are vital for protein structure and function, as well as non-conserved regions by the analysis of the evolutionary background of macromolecules.

Secondary and tertiary structure analysis: Missense pathogenic variants and their consequences for the secondary and tertiary structures of G6PD were assessed via I-Tasser and PSIPRED. I-Tasser (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) displays the predicted pathogenic variants and 3D structure of the protein. PSIPRED (http://bioinf.cs. ucl.ac.uk/psipred/) was employed to display the predicted pathogenic variants and 2D structure of the protein.

Interaction analysis: STRING (https://string-db.org) was utilized to assess protein-protein (Direct or indirect, physical or functional) interactions. For each interaction, scores varied from 0 to 1. A score of 0 reflects the weakest interaction, and a score of 1 indicates the strongest interaction.

Hydrophobicity analysis: To obtain more information about the hydrophobicity of the protein, the ProtScale tool was applied. This tool (https://web.expasy.org/protscale/) generates hydropathy plots, allowing comparison of changes in local hydrophobic or hydrophilic properties resulting from specific amino acid substitutions in different variants.

Results

Pathogenicity and stability of variants: A total of 215 missense variants were obtained from the HGMD database. The number of variants was reduced by applying various filters. First, the variants labeled as Pathogenic, according to ACMG, were selected. Following this step, variants with CADD scores over 20 were kept. In the next step, variants predicted to be harmful according to SIFT and FATHMM, and probably harmful based on PolyPhen and PANTHER, were selected. Finally, the variants associated with decreased protein stability according to I-Mutant and MUpro were chosen. After applying these filters, 69 variants were selected for further analysis (Supplementary Table 1, Figure 2).

Conservation of variants: Conservation scores for these 69 variants were calculated using ConSurf. The scores ranged from 1 to 9, indicating the level of conservation for each variant. Of these, 28 variants were located in highly conserved regions (Scores = 9), suggesting their functional importance. No variant was located in the variable region with a score of 1 or 2 (Supplementary Table 2).

Molecular interactions: The analysis of protein-protein interactions by STRING revealed that G6PD interacted with 10 other proteins with a strong connection score (≥0.8) (Figure 3).

Identification of common *G6PD* variants in Iran: Based on a study conducted by Moosazadeh et al., among all the existing *G6PD* variants, three were prevalent in Iran, including the Mediterranean, Chatam, and Cosenza (Figure 4). ConSurf analysis confirmed the scores of 4, 9, and 8 for the Mediterranean, Chatam, and Cosenza variants, respectively.

Secondary and tertiary structure of variants: The PSIPRED-based secondary structure analysis of the three prevalent Iranian *G6PD* variants revealed some alterations in the mutant variants in comparison to the wild-type protein. The Mediterranean variant exhibited a coil-to-helix shift in the 8th region and a prematurely terminated helix in the 10th, suggesting localized structural stabilization and potential disruption, respectively. In the Chatham variant, a coil at the end of the 10th region was replaced by an alpha-helix, and a small change in the middle part of the sequence, causing the shortening of a helical segment compared to the wild-type protein. The Cosenza variant showed disruption of the 9th alpha-helix and coil formation, along with helix fragmentation at the C-terminal end. These changes suggested broader

structural destabilization beyond the mutation site. In summary, the mutation-induced alterations observed in the secondary structure may influence the protein's tertiary structure and compromise its biological function, highlighting the crucial role of these domains in maintaining protein stability (Figure 5). A 3D structural model of the *G6PD* Mediterranean variant was created using I-TASSER, which identified 1qki/1qkiF as the most compatible templates due to their high Z-scores and greater than 95% sequence identity. Model 1 was the top-ranked structure, with good confidence (C-score = 0.49), TM-score = 0.78, and RMSD = 6.3 Å. According to TM-aligning, 1qkiF offered the nearest structural match, showing a TM score of 0.949 and an RMSD value of 0.59 Å. Functional prediction indicated potential changes in NADP+ binding due to the mutation, with key residues identified near positions 38-43, 71-73, 110-114, and others. The enzyme was predicted to belong to class EC 1.1.1.49, with a strong confidence score (C-scoreEC = 0.689). A possible structural shift in the active site (residues 200 and 263) may affect enzymatic activity. GO analysis further supported oxidoreductase function and cytoplasmic localization. Altogether, the mutation may impair the enzymatic function and stability of *G6PD*. Regarding the Chatham variant,1qki was identified as the top threading template, with high identity (Up to 0.99) and the highest Z-score (5.94), showing strong structural reliability. Other supporting templates included 1e77A, 7d5mA, and 7snfA. Model 3 had the highest C-score (0.58), while Model 1, selected for further analysis, also showed good metrics (C-score = 0.39, TM-score = 0.77, and RMSD = 6.5 Å). TM-align validated 1qkiF as the closest analog with TM-score = 0.951 and identity = 98.2%. The mutation site was located near key ligand-binding residues, including positions 38, 40-43, and 170-171, possibly affecting interactions with NADP, the top predicted ligand. Enzyme function prediction matched EC 1.1.1.49 with a high CscoreEC of 0.725. GO scores (MF = 0.83, BP = 0.72, CC = 0.72) further supported preserved enzymatic function. Minor structural modifications near the active site could be caused by the Chatam mutation, but the essential catalytic function of *G6PD* appeared to remain intact.

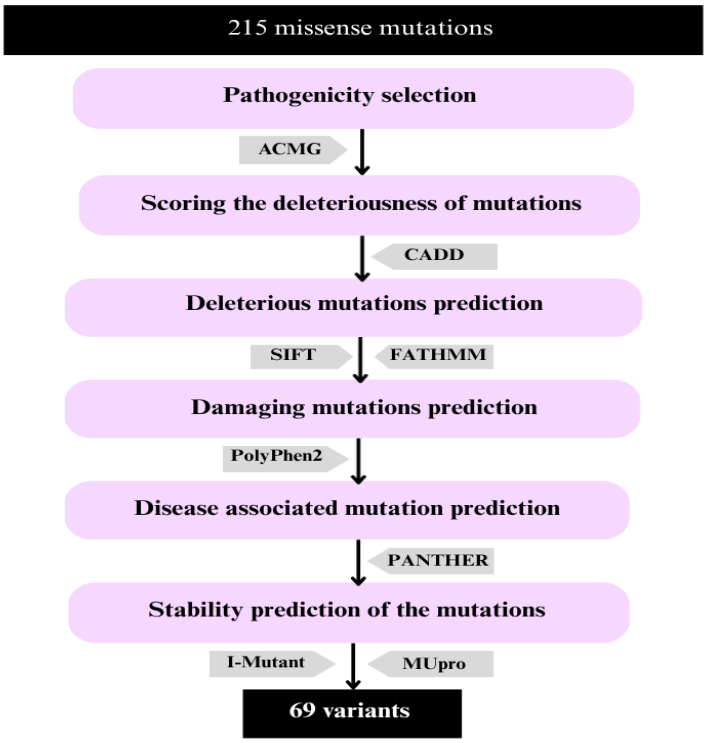

Figure 2. Filtering pipeline. Initially, all variants were evaluated using ACMG guidelines to assess their pathogenicity. Subsequently, the CADD score was used to estimate the harmfulness of the variants. In the next step, to evaluate the potential for inflicting damage, several in silico tools were employed, including SIFT, PolyPhen, PANTHER, and FATHMM. Finally, the effects of the variants on the stability of the protein were evaluated through I-Mutant and Mupro.

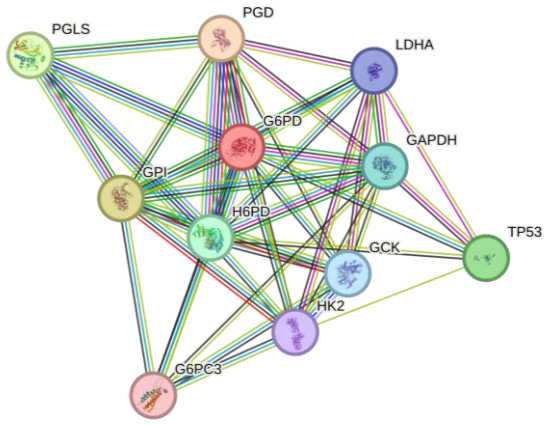

Figure 3. Protein-protein interaction of G6PD. G6PD shows high-confidence functional associations with other enzymes of the pentose-phosphate pathway, including PGD (6-phosphogluconate dehydrogenase) (0.999), GPI (glucose-6-phosphate isomerase) (0.999), and PGLS (6-phosphogluconolactonase) (0.993). The network also shows associations with TP53 (0.986) and H6PD (0.985), suggesting potential roles in redox regulation and cellular stress responses.

Considering the Cosenza variant, 1e77A was identified as the top structural template (Z-score = 5.30). Several others showed >90% identity and coverage, with secondary structure predictions confirming a classic *G6PD* fold. Among the five predicted models, Model 1 had the highest reliability (C-score= 0.49, TM-score= 0.78, and RMSD= 6.3 Å). TM-align confirmed high similarity to 1qkiF (TM-score = 0.949, 98.4% identity). Functionally, I-TASSER predicted NADP (NAP) as the primary ligand (C-score = 0.53), with key binding residues (S73, K71, R72, Y112) clustered near the NADP-binding domain, suggesting the potential disruption of cofactor binding. Enzymatic function was predicted as EC 1.1.1.49 with high confidence (C-scoreEC = 0.690), with GO analysis confirming oxidoreductase activity (GO:0050661) and participation in the pentose phosphate pathway (Figure 6).

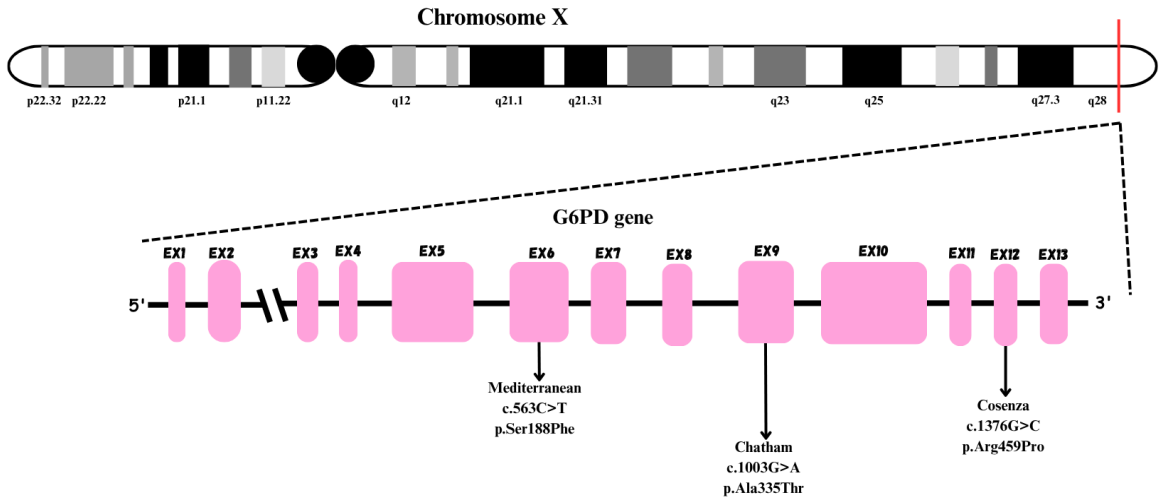

Figure 4. The structure of the X chromosome and the location of three prevalent G6PD variants in Iran. The G6PD is located on Xq28. The most common variant, known as the Mediterranean, is located at exon 6. Following this is the Chatham variant, at exon 9, and finally, the Cosenza variant, which is located at exon 12.

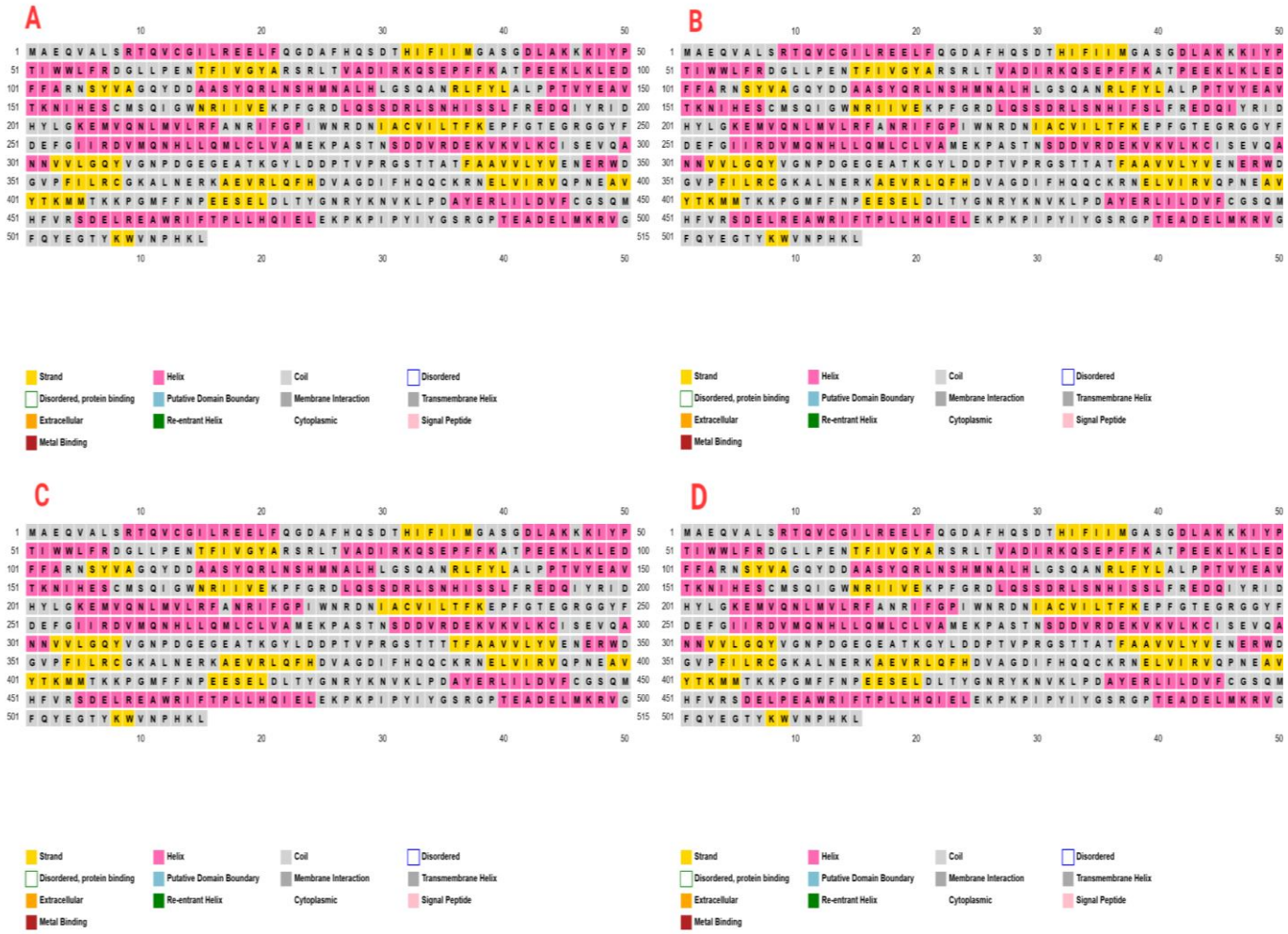

Figure 5. The 2D structure of the variants retrieved from PSIPRED. A: Wild type, B: the Mediterranean variant, in which a coil at the start of the 8th strand turns into an α-helix, and the α-helix in the 10th strand ends prematurely compared to the wild-type. C: The Chatham variant, in which a coil at the end of the 10th strand is replaced by an α-helix and a shortened helix in the central region. D: The Cosenza variant, where a continuous α-helix in the 9th strand of the wild-type protein is replaced by coils, and a rearrangement near the C-terminal causes helix shortening.

Hydrophobicity of variants: Hydrophobicity analysis of the three prevalent variants using the ExPASy ProtScale tool revealed that amino acid substitutions altered the hydrophobic or hydrophilic nature of specific domains in G6PD. In the Mediterranean variant, the substitution of Serine (Hydrophobicity: -0.800) with Phenylalanine (2.800) increased hydrophobicity. Conversely, in the Chattahm variant, the replacement of Alanine (1.800) with Threonine (-0.700) led to a decrease in hydrophobicity. In the Cosenza variant, the substitution of Arginine (-4.500) with Proline (-1.600) moderately increased local hydrophobicity, indicating a shift in the protein's polarity. These hydrophobicity changes could have important implications for the structural stability and molecular interactions of the enzyme.

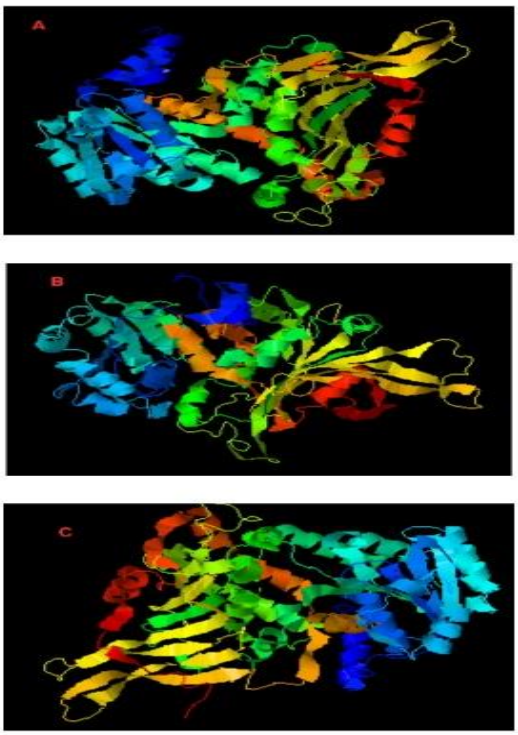

Figure 6. The 3D structure of G6PD variants predicted by I-Tasser. A: The Mediterranean variant's 3D model showed a C-score of 0.49, indicating moderate to good confidence in the predicted structure, a TM-score of 0.78 ± 0.10 that suggests the reliability of global topology, and an RMSD of 6.3 ± 3.8 Å, which is relatively high. B: The 3D model of the Chatham variant showed a C-score of 0.39 and TM-score of 0.77 ± 0.10, suggesting the model had a strong structural similarity to the template. RMSD was 6.5 ± 3.9 Å. C: The Cosenza variant, whose 3D model showed a C-score of 0.49, indicating a moderate to good confidence in the predicted structure. The estimated TM-score was 0.78 ± 0.10, and RMSD value was calculated as 6.3 ± 3.8 Å.

Discussion

G6PD deficiency is the most prevalent erythrocyte enzyme disorder. In this study, our purpose was to identify the most pathogenic and prevalent *G6PD* missense variants across global populations, with a particular focus on Iran.

Among the variants we analyzed, 69 were reported as pathogenic. Based on their research, Malik et al. reported a list of 20 high-impact variants and 16 moderate-impact *G6PD* variants. In comparison with our study, two of our variants were identical to two of the high-impact variants reported in the recent study, namely (c.404A>C / p.Asn135Thr) and (c.1057C>T / p.Pro353Ser), and two other variants identified here were identical to moderate-impact variants reported by Malik et al., including (c.536G>A / p.Ser179Asn) and (c.844G>C / p.Asp282His) (23).

Based on the study conducted by Caterina Nannelli and colleagues on *G6PD* variants, we selected the variants reported as the most frequent and globally recurrent ones in over 10 countries (24) (Table 1). Six frequent variants were identified in the G6PD gene. Among these, the Mediterranean variant showed the highest frequency. Among these, only the Seattle variant was present in our list of 69 analyzed variants.

Consistent with previous studies, this study also highlighted the impact of consanguineous marriage on the prevalence of homozygous pathogenic variants in the Iranian population (16). Consanguineous marriage, which is relatively common in many Iranian communities, plays a significant role in increasing the prevalence of autosomal and X-linked disorders such as G6PD deficiency. In our analysis, three variants, the Mediterranean: c.563C>T (p.Ser188Phe), prevalence = 78.2 %), Chatham: c.1003G>A (p.Ala335Thr, prevalence=9.1 %), and Cosenza: c.1376G>C (p.Arg459Pro, prevalence=0.5%) (25), were identified as frequent in Iranian populations. This was similar to an Indian population studied by Sukumar et al, where the Mediterranean variant was the most prevalent form, while the Chatham variant was rarely detected in this population (26). Additionally, based on Al-Jaouni et al.'s research, the Mediterranean variant was reported as the most common in Saudi Arabia (27). Even after applying stringent bioinformatic filters, these three common variants were not classified among 69 predicted pathogenic variants in the Iranian population. This observation is consistent with the well-established principle that notably harmful variants are generally rare in human populations because of their negative impact on survival and reproductive fitness. In contrast, variants with higher frequencies usually exert moderate or tolerable effects, with less influence on carriers. The relatively high prevalence of the Mediterranean, Chatham, and Cosenza variants in Iran strongly suggests that these alleles have been maintained through evolutionary mechanisms, most notably subjected to positive selection due to the protective role of G6PD deficiency against malaria. Moreover, several factors, including high rates of consanguinity, founder effects, and long-standing gene flow between Iranian and neighboring Mediterranean and Middle Eastern populations, have likely reinforced the persistence and widespread distribution of these variants (11,28,29). Based on ConSurf analysis, the Chatham and Cosenza variants, with respective score of 9 and 8, seem to undergo relatively strong evolutionary pressure to maintain these sites, underscoring their functional importance. The score of 4 for the Mediterranean variant shows a relatively low evolutionary constraint, suggesting a potentially lower functional significance or higher tolerance for this variation.

Table 1. Variants with the highest recurrence reported in at least 10 countries

Variant name	Nucleotide substitution	Amino acid substitution	Reported places
A−	202 G → A 376 A → G	376 A → G 126 Asn → Asp	All African countries, Brazil, France, Haiti, Honduras, Iraq, Italy, Jordan, Kuwait, Mexico, Portugal, Principe, Sao Tome, Saudi Arabia, Spain, UAE, USA, Venezuela
Mediterranean	563 C → T	188 Ser → Phe	Afghanistan, Algeria, Brazil, Bulgaria, Cambodia, Comoros, Croatia, Cyprus, Egypt, France, Greece, Indonesia, Iran, Iraq, Italy, Jordan, Kuwait, Macedonia, Malaysia, Mauritania, Mauritius, Myanmar, Oman, Pakistan, Portugal, Saudi Arabia, Singapore, Spain, Syria, Tunisia, UAE, Venezuela
Seattle	844 G → C	282 Asp→His	Algeria, Brazil, Bulgaria, Canary Islands, Croatia, France, Greece, Indonesia, Italy, Mauritius, Mexico, Myanmar, Portugal, Spain, Thailand, USA
A−	968 T → C 376 A → G	323 Leu → Pro 126 Asn → Asp	All African countries, Brazil, France, Haiti, Honduras, Iraq, Italy, Jordan, Kuwait, Mexico, Portugal, Principe, Sao Tome, Saudi Arabia, Spain, UAE, USA, Venezuela
Union	1360 C → T	454 Arg → Cys	Cambodia, Croatia, Italy, Laos, Mauritius, Mexico, Myanmar, Spain, Thailand, Vanuatu, Vietnam
Chatham	1003 G → A	335 Ala→Thr	Algeria, India, Indonesia, Iran, Iraq, Italy, Jordan, Kuwait, Oman, Pakistan, Spain

Structurally, the G6PD enzyme has two main domains: the $NADP^+$ binding (Residues ~27-200) and substrate-binding (Residues ~201-515) domains (30). Pathogenic mutations in these domains can impair enzymatic function, decrease protein stability, and lead to clinical manifestations such as hemolytic anemia. In the present study, three potentially pathogenic and prevalent variants in Iran, the Mediterranean, Chatham, and Cosenza variants, were evaluated. Each of these is located in functionally significant domains of G6PD enzyme and predicted to affect its structure and activity. The Mediterranean variant is situated close to the C-terminal region of the $NADP^+$-binding domain, which is crucial for coenzyme interaction and stabilizes the protein's overall structure (31). The substitution of serine, which is a small polar residue, with phenylalanine, a bulky hydrophobic amino acid, likely introduces steric hindrance and disrupts hydrogen bonding, potentially destabilizing the $NADP^+$ binding pocket. Studies suggest that mutations at or near this region can reduce enzymatic activity and lead to G6PD deficiency (32). The Chatham variant is located in the substrate-binding domain of G6PD enzyme. The replacement of alanine, a small nonpolar amino acid, with threonine, a larger polar amino acid, in the hydrophobic core of the protein may disturb local folding or interfere with glucose-6-phosphate binding (33). Likewise, the Cosenza variant resides in the substrate-binding domain. The positive charge of arginine facilitates salt bridge formation and hydrogen bonding, so its substitution with proline, a helix breaker, can induce structural kinks, reduce protein flexibility, and impair substrate access or overall enzyme conformation (Figure 7). Previous reports have declared similar substitutions at this site with severe enzyme deficiency (34).

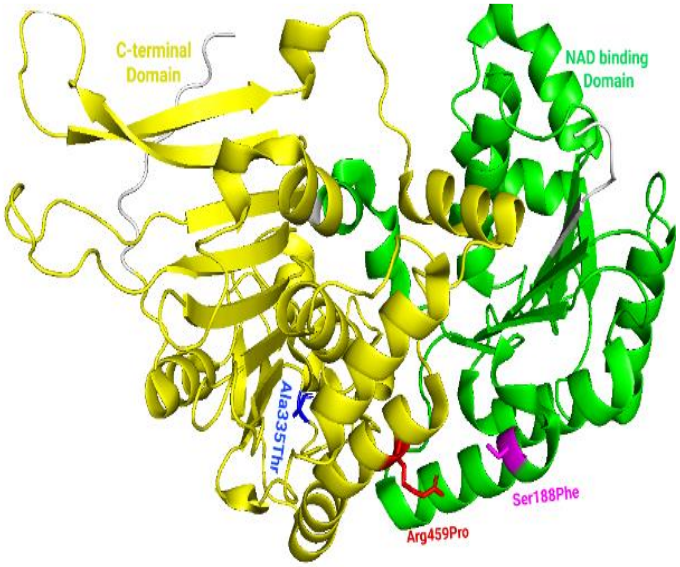

Figure 7. Top three common G6PD variants in Iran: domains and mutation sites. The Mediterranean variant is located within the NAD-binding domain, while the Chatham and Cosenza variants are positioned in the C-terminal domain.

The Mediterranean variant generally shows residual enzyme activity around 10%, placing it in a borderline category between class II and class III depending on the study population and methodology. In contrast, the Chatham and Cosenza variants consistently exhibit enzyme activity values below 10%, a characteristic of class II variants, which are typically associated with more severe clinical manifestations due to the markedly reduced enzymatic activity (35). Previous studies confirm our findings. For example, Sukumar et al. reported that the Mediterranean variant was associated with significantly reduced enzyme activity (26). Similarly, a study conducted by Sirdah et al. demonstrated that a Mediterranean variant associated with clinical manifestations caused structural changes in the protein by substituting serine with phenylalanine in alpha-helix f (At the surface of the protein), disrupting the protein's structure. This structural perturbation reduced the side-chain flexibility of the alpha-helix and induced a conformational shift at the NADP+ pyridine amide position (36).

Despite the comprehensive in silico approach used here, functional assays could not be performed due to resource limitations. Nevertheless, our multi-tool filtering strategy provided a reliable prediction of variant pathogenicity, which could guide future functional experiments. Our study supports the idea that in silico predictions, when backed by multiple tools and conservation scores, can reliably identify deleterious *G6PD* mutations. The bioinformatics tools used in this study were proved to be highly effective in predicting the pathogenicity of the identified variants. Bioinformatic tools offer rapid and cost-effective alternatives for analyzing large-scale biological data, helping to predict variants' biological outcomes, molecular structures, and evolutionary patterns, and to prioritize mutations for further studies. Despite these strengths, bioinformatic analyses have inherent limitations, like relying on in silico predictions that may not fully capture biological complexity. Moreover, outcomes often depend on the quality of underlying algorithms, assumptions, and reference databases. Also, different tools can produce conflicting results; most are optimized for coding regions with limited applicability to regulatory variants, and population-specific data, such as variants in underrepresented groups, may be insufficiently documented.

Conclusion

In conclusion, understanding the structural and functional impact of G6PD variants, especially those with variations in highly conserved regions, is essential for both scientific research and clinical practice. These insights help design more accurate diagnostic tools, improve variant classification, and provide effective genetic counseling. Considering the high prevalence of G6PD deficiency in the studied population, there is a clear need for rapid, reliable, and cost-effective diagnostic methods. The strip array technology is proposed as a promising tool for routine screening, enabling simultaneous detection of multiple common variants with high sensitivity and specificity (37). In Iran, where three G6PD variants are the most frequent, prioritizing these mutations in initial screening is both practical and efficient. If these primary variants are not detected, further analysis should target rarer mutations. This stepwise approach maximizes resource use, allows early diagnosis, and enhances the quality of genetic counseling and patient care.

Acknowledgement

The authors would like to thank all those who contributed to this study and provided support and guidance.

Funding sources

Funding for this study was granted by Golestan University of Medical Science (Grant Numbers: 113352).

Ethical statement

Approval for the study was obtained from the Ethics Committee of Golestan University of Medical Sciences (Approval code: IR.GOUMS.REC.1401.523).

Conflicts of interest

No conflict of interest has been reported by authors.

Author contributions

RJK and SST were responsible for investigative procedures, formal analytical tasks, and writing the original draft. MO contributed to concept development, project supervision, manuscript editing and reviewing, validation of procedures, and data organization.

Data availability statement

No additional data were created or used in this study beyond what is presented in the manuscript.

References

1. Arai Y. G6PD Deficiency: A Possible Cardiovascular Risk Factor in Older People. J Atheroscler Thromb. 2021;28(6):586-7. [View at Publisher] [DOI] [PMID] [Google Scholar]
2. Beretta A, Manuelli M, Cena H. Favism: Clinical Features at Different Ages. Nutrients. 2023;15(2):343. [View at Publisher] [DOI] [PMID] [Google Scholar]

3. Luzzatto L, Ally M, Notaro R. Glucose-6-phosphate dehydrogenase deficiency. Blood. 2020;136(11):1225-40. [View at Publisher] [DOI] [PMID] [Google Scholar]
4. Liu Z, Yu C, Li Q, Cai R, Qu Y, Wang W, et al. Chinese newborn screening for the incidence of G6PD deficiency and variant of G6PD gene from 2013 to 2017. Hum Mutat. 2020;41(1):212-21. [View at Publisher] [DOI] [PMID] [Google Scholar]
5. Mondal A, Mukherjee S, Dar W, Singh S, Pati S. Role of glucose 6-phosphate dehydrogenase (G6PD) deficiency and its association to Autism Spectrum Disorders. Biochim Biophys Acta Mol Basis Dis. 2021;1867(10):166185. [View at Publisher] [DOI] [PMID] [Google Scholar]
6. Boonyuen U, Jacob BAC, Wongwigkan J, Chamchoy K, Singha-Art N, Pengsuk N, et al. Genetic analysis and molecular basis of G6PD deficiency among malaria patients in Thailand: implications for safe use of 8-aminoquinolines. Malar J. 2024;23(1):38. [View at Publisher] [DOI] [PMID] [Google Scholar]
7. Ge T, Yang J, Zhou SH, Wang Y, Li Y, Tong X. The Role of the Pentose Phosphate Pathway in Diabetes and Cancer. Front Endocrinol (Lausanne). 2020:11:365. [View at Publisher] [DOI] [PMID] [Google Scholar]
8. Belfield KD, Tichy EM. Review and drug therapy implications of glucose-6-phosphate dehydrogenase deficiency. Am J Health Syst Pharm. 2018;75(3):97-104. [View at Publisher] [DOI] [PMID] [Google Scholar]
9. Yen W-C, Wu Y-H, Wu C-C, Lin H-R, Stern A, Chen S-H, et al. Impaired inflammasome activation and bacterial clearance in G6PD deficiency due to defective NOX/p38 MAPK/AP-1 redox signaling. Redox Biol. 2020;28:101363. [View at Publisher] [DOI] [PMID] [Google Scholar]
10. Stone SN, Reisig KV, Saffel HL, Miles CM. Management of Athletes With G6PD Deficiency: Does Missing an Enzyme Mean Missing More Games? Sports Health. 2019;12(2):149-53. [View at Publisher] [DOI] [PMID] [Google Scholar]
11. Tabatabaei SM, Salimi Khorashad A, Sakeni M, Raeisi A, Metanat Z. Prevalence of glucose-6-phosphate dehydrogenase (G6PD) deficiency in southeast Iran: implications for malaria elimination. J Infect Dev Ctries. 2015;9(3):289-97. [View at Publisher] [DOI] [PMID] [Google Scholar]
12. Kosaryan M, Nasehi MM, Karami H, Parsaii MR, Mahdavi MR, Zakizadeh R,et al. Neonatal Screening for G6PD Deficiency in Mazandaran Province, Iran 2007-2010. IJBC. 2011;2(4):113-6. [View at Publisher] [Google Scholar]
13. Moosazadeh M, Amiresmaili M, Aliramezany M. Prevalence of G6PD Deficiency in Iran, a Mata-analysis. Acta Med Iran. 2014;52(4):256-64. [View at Publisher] [PMID] [Google Scholar]
14. Ansari-Moghaddam A, Adineh HA, Mohammadi M, Tabatabae SM, Zareban I, Ranjba M,et al. Prevalence of glucose-6-phosphate dehydrogenase (G6PD) deficiency in malaria endemic region of Iran (Sistan and Baluchestan Province): Epidemiological profile and trends over time. Asian Pac J Trop Dis. 2017;7(10):587-91. [View at Publisher] [DOI]
15. Mehrjoo Z, Fattahi Z, Beheshtian M, Mohseni M, Poustchi H, Ardalani F, et al. Distinct genetic variation and heterogeneity of the Iranian population. PLoS Genet. 2019;15(9):e1008385. [View at Publisher] [DOI] [PMID] [Google Scholar]
16. Fattahi Z, Beheshtian M, Mohseni M, Poustchi H, Sellars E, Nezhadi SH, et al. Iranome: A catalog of genomic variations in the Iranian population. Hum Mutat. 2019;40(11):1968-84. [View at Publisher] [DOI] [PMID] [Google Scholar]
17. Saadat M, Maryam Ansari-Lari M, Farhud D. Short Report Consanguineous marriage in Iran. Ann Hum Biol. 2004;31(2)263-9. [View at Publisher] [DOI] [PMID] [Google Scholar]
18. Kane M, Pratt VM, Scott SA, Pirmohamed M, Esquivel B, Kattman BL, et al. Primaquine therapy and G6PD and CYP2D6 genotype. In: Pratt VM, Scott SA, Pirmohamed M, Esquivel B, Kane MS, editors. Medical Genetics Summaries . Bethesda (MD): National Center for Biotechnology Information (US); 2012. [PMID] [Google Scholar]
19. Harcke SJ, Rizzolo D, Harcke HT. G6PD deficiency: An update. JAAPA Off J Am Acad Physician Assist. 2019;32(11):21-6. [View at Publisher] [DOI] [PMID] [Google Scholar]
20. Shenkutie TT, Nega D, Hailu A, Kepple D, Witherspoon L, Lo E, et al. Prevalence of G6PD deficiency and distribution of its genetic variants among malaria-suspected patients visiting Metehara health centre, Eastern Ethiopia. Malar J. 2022;21(1):260. [View at Publisher] [DOI] [PMID] [Google Scholar]
21. Li H, Ch'ih Y, Li M, Luo Y, Liu H, Xu J,et al. Newborn screening for G6PD deficiency in HeFei, FuYang and AnQing, China: Prevalence, cut-off value, variant spectrum. J Med Biochem. 2024;43(1):86-96. [View at Publisher] [DOI] [PMID] [Google Scholar]
22. Bernardo J, Nock M. Pediatric Provider Insight into Newborn Screening for G6PD Deficiency. Clin Pediatr (Phila). 2015;54(6):575-8. [View at Publisher] [DOI] [PMID] [Google Scholar]
23. Malik S, Zaied R, Syed N, Jithesh P, Al-Shafai M. Seven novel glucose-6-phosphate dehydrogenase (G6PD) deficiency variants identified in the Qatari population. Hum Genomics. 2021;15(1):61. [View at Publisher] [DOI] [PMID] [Google Scholar]
24. Nannelli C, Bosman A, Cunningham J, Dugué P-A, Luzzatto L. Genetic variants causing G6PD deficiency: Clinical and biochemical data support new WHO classification. Br J Haematol. 2023;202(5):1024-32. [View at Publisher] [DOI] [PMID] [Google Scholar]
25. Moosazadeh M, Nekoei-Moghadam M, Aliram-Zany M, Amiresmaili M. Identification of Mutation of Glucose-6-Phosphate Dehy-drogenase (G6PD) in Iran: Meta- analysis Study. Iran J Public Health. 2013;42(9):1007-15. [View at Publisher] [PMID] [Google Scholar]
26. Sukumar S, Mukherjee MB, Colah RB, Mohanty D. Molecular basis of G6PD deficiency in India. Blood Cells Mol Dis. 2004;33(2):141-5. [View at Publisher] [DOI] [PMID] [Google Scholar]
27. Al-Jaouni SK, Jarullah J, Azhar E, Moradkhani K. Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in Jeddah, Kingdom of Saudi Arabia. BMC Res Notes. 2011;4:436. [View at Publisher] [DOI] [PMID] [Google Scholar]
28. Shahjahani M, Mortazavi Y, Heli B, Dehghanifard A. Prevalence of G6PD Deficiency in Iran. Int J Hematol-Oncol Stem Cell Res. 2013;7(1):48-9. [View at Publisher] [PMID] [Google Scholar]
29. Noori-Daloii MR, Soltanian S, Gangi SHM, Yousefi A, Hejazi S, Bani-hashem A, et al. Molecular Identification of the Most Prevalent Mutations of Glucose-6-Posphate Dehydrogenase (G6PD) Gene in Deficient Patients in Khorasan Province of Iran. 2006;17(2):103-6. [View at Publisher] [Google Scholar]
30. Au SW, Gover S, Lam VM, Adams MJ. Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency. Structure. 2000;8(3):293-303. [View at Publisher] [DOI] [PMID] [Google Scholar]
31. Beutler E. G6PD deficiency. Blood. 1994;84(11):3613-36. [View at Publisher] [DOI] [PMID]
32. Minucci A, Moradkhani K, Hwang MJ, Zuppi C, Giardina B, Capoluongo E. Glucose-6-phosphate dehydrogenase (G6PD) mutations database: Review of the "old" and update of the new mutations. Blood Cells Mol Dis. 2012;48(3):154-65. [View at Publisher] [DOI] [PMID] [Google Scholar]
33. Luzzatto L, Nannelli C, Notaro R. Glucose-6-Phosphate Dehydrogenase Deficiency.Hematol Oncol Clin North Am. 2016;30(2):373-93. [View at Publisher] [DOI] [PMID] [Google Scholar]
34. di Montemuros F, Dotti C, Tavazzi D, Fiorelli G, Cappellini MD. Molecular heterogeneity of glucose-6-phosphate dehydrogenase (G6PD) variants in Italy. Haematologica. 1997;82(4):440-5. [View at Publisher] [PMID] [Google Scholar]
35. World Health Organization. MPAG March 2022 Session 2 Technical Consultation on G6PD Classification [Internet]. Geneva: WHO; 2022 [cited 2025 Oct 3]. Available from: https://cdn.who.int/media/docs/default-ource/malaria/mpac-documentation/mpag-mar2022-session2-technical-consultation-g6pd-classification.pdf. [View at Publisher]

36. Sirdah M, Reading NS, Vankayalapati H, Prchal JT. A computational study of structural differences of binding of NADP+ and G6P substrates to G6PD Mediterraneanc.563T, G6PD A−c.202A/c.376G, G6PD Cairoc.404C and G6PD Gazac.536A mutations. Blood Cells Mol Dis. 2021;89:102572. [View at Publisher] [DOI] [PMID] [Google Scholar]

37. Rezaie N, Ghazanfari SS, Khosravi T, Vaghefi F, Oladnabi M. A comprehensive in silico analysis of mutation spectrum of maple syrup urine disease (MSUD) genes in Iranian population. Mol Biol Res Commun. 2024;13(4):235-46. [View at Publisher] [DOI] [PMID] [Google Scholar]

Cite this article as:

Jafari Kalokan R, Sharifi Tabar S, Oladnabi M. Bioinformatic analysis of *G6PD* variants with a focus on key Iranian mutations. *JCBR.* 2025;9(X):X. http://dx.doi.org/10.29252/JCBR.9.X.X

Supplementary Table 1. Bioinformatic filters applied on G6PD variants. Initially, 215 missense variants were identified. Based on the first filter (ACMG), 115 variants were classified as pathogenic. Among these, 108 variants had a CADD score greater than 20 and were chosen for further investigations. In the next step, PANTHER, SIFT, FATHMM, and PolyOhen-2 were used, according to which 82 variants were retained as harmful or probably harmful. Finally, based on predictions from I-Mutant and MUpro, 69 variants were predicted to decrease protein stability.

Variant	Protein	ACMG	Panther	CADD	FATHM	SIFT	I-Mutant 2.0	Mupro	Poly Phen 2
c.130G>A	p.Ala44Thr	Pathogenic	Probably damaging 0.95	26.7	Damaging (-5.07)	Damaging	Decrease (DDG=-0.55)	Decrease (DDG=-1.07)	Probably damaging (0.999)
c.148C>T	p.Pro50Ser	Pathogenic	Probably damaging 0.95	26.7	Damaging (-5.97)	Damaging	Decrease (DDG=-0.85)	Decrease (DDG=-0.84)	Probably damaging (1.000)
c.152C>T	p.Thr51Ile	Pathogenic	Probably damaging 0.85	26.1	Damaging (-4.86)	Damaging	Decrease (DDG=-1.52)	Decrease (DDG=-0.25)	Probably damaging (0.998)
c179T>C	p.Leu60Pro	Pathogenic	Probably damaging 0.85	26.5	Damaging (-6.06)	Damaging	Decrease (DDG=-2.25)	Decrease (DDG=-2.33)	Probably damaging (1.000)
c.185C>A	p.Pro62His	Pathogenic	Probably damaging 0.95	24.7	Damaging (-6.30)	Damaging	Decrease (DDG=-1.90)	Decrease (DDG=-0.69)	Probably damaging (1.000)
C.224T>C	p.Leu75Pro	Pathogenic	Probably damaging 0.74	27	Damaging (-4.98)	Damaging	Decrease (DDG=-1.66)	Decrease (DDG=-2.31)	Probably damaging (1.000)
c.404A>C	p.Asn135Thr	Pathogenic	Probably damaging 0.95	23.7	Damaging (-6.76)	Damaging	Decrease (DDG=-2.18)	Decrease (DDG=-1.43)	Probably damaging (0.987)
c.409C>T	p.Leu137Phe	Pathogenic	Probably damaging 0.85	25.7	Damaging (-6.32)	Damaging	Decrease (DDG=-1.22)	Decrease (DDG=-1.56)	Probably damaging (0.981)
c.473G>A	p.Cys158Tyr	Pathogenic	Probably damaging 0.95	25.1	Damaging (-6.17)	Damaging	Decrease (DDG=-0.55)	Decrease (DDG=-0.87)	Probably damaging(1.000)
c.488G>A	p.Gly163Asp	Pathogenic	Probably damaging 0.95	28.5	Damaging (-6.62)	Damaging	Decrease (DDG=-0.46)	Decrease (DDG=-0.68)	Probably damaging (0.993)
c.497G>A	p.Arg166His	Pathogenic	Probably damaging 0.95	26.3	Damaging (-6.73)	Damaging	Decrease (DDG=-1.25)	Decrease (DDG=-0.89)	Probably damaging (1.000)
c.514C>T	p.Pro172Ser	Pathogenic	Probably damaging 0.95	27.2	Damaging (-7.09)	Damaging	Decrease (DDG=-1.36)	Decrease (DDG=-0.65)	Probably damaging (1.000)
c.517T>C	p.Phe173Leu	Pathogenic	Probably damaging 0.95	25.7	Damaging (-6.67)	Damaging	Decrease (DDG=-2.26)	Decrease (DDG=-0.59)	Probably damaging (1.000)
c.527A>G	p.Asp176Gly	Pathogenic	Probably damaging 0.95	27.3	Damaging (-5.31)	Damaging	Decrease (DDG=-2.55)	Decrease (DDG=-1.76)	Probably damaging (0.999)
c.536G>A	p.Ser179Asn	Pathogenic	Probably damaging 0.95	24.9	Damaging (-5.46)	Damaging	Decrease (DDG=-1.32)	Decrease (DDG=-0.3)	Probably damaging (0.838)
c.592C>T	p.Arg198Cys	Pathogenic	Probably damaging 0.95	31	Damaging (-5.84)	Damaging	Decrease (DDG=-1.15)	Decrease (DDG=-0.82)	Probably damaging (1.000)
c.592C>A	p.Arg198Ser	Pathogenic	Probably damaging 0.95	28.6	Damaging (-5.81)	Damaging	Decrease (DDG=-2.44)	Decrease (DDG=-1.19)	Probably damaging (1.000)
c.593G>C	p.Arg198Pro	Pathogenic	Probably damaging 0.95	26.9	Damaging (-5.83)	Damaging	Decrease (DDG=-1.77)	Decrease (DDG=-1.44)	Probably damaging (1.000)
c.593G>A	p.Arg198His	Pathogenic	Probably damaging 0.95	25.8	Damaging (-5.81)	Damaging	Decrease (DDG=-1.97)	Decrease (DDG=-1.3)	Probably damaging (1.000)
c.679C>T	p.Arg227Trp	Pathogenic	Probably damaging 0.86	29.6	Damaging (-7.13)	Damaging	Decrease (DDG=-1.29)	Decrease (DDG=-1.39)	Probably damaging (1.000)
c.737G>T	p.Arg246Leu	Pathogenic	Probably damaging 0.95	25.9	Damaging (-5.93)	Damaging	Decrease (DDG=-0.13)	Decrease (DDG=-0.27)	Probably damaging (1.000)
c.769C>G	p.Arg257Gly	Pathogenic	Probably damaging 0.95	27.8	Damaging (-5.55)	Damaging	Decrease (DDG=-0.99)	Decrease (DDG=-1.6)	Probably damaging (0.999)
c.770G>T	p.Arg257Leu	Pathogenic	Probably damaging 0.95	35	Damaging (-5.51)	Damaging	Decrease (DDG=-1.25)	Decrease (DDG=-0.51)	Probably damaging (1.000)
c.806G>A	p.Cys269Tyr	Pathogenic	Probably damaging 0.85	26.9	Damaging (-6.94)	Damaging	Decrease (DDG=-0.24)	Decrease (DDG=-1.6)	Probably damaging (0.988)
c.820G>A	p.Glu274Lys	Pathogenic	Probably damaging 0.95	28.1	Damaging (-5.29)	Damaging	Decrease (DDG=-0.57)	Decrease (DDG=-1.28)	Probably damaging (0.992)
c.825G>C	p.Lys275Asn	Pathogenic	Probably damaging 0.85	22.3	Damaging (-7.01)	Damaging	Decrease (DDG=-0.76)	Decrease (DDG=-0.55)	Probably damaging (1.000)
c.826C>T	p.Pro276Ser	Pathogenic	Probably damaging 0.95	26.9	Damaging (-6.67)	Damaging	Decrease (DDG=-1.53)	Decrease (DDG=-0.51)	Probably Damaging (1.000)
c.844G>C	p.Asp282His	Pathogenic	Probably damaging 0.74	23.5	Damaging (-7.06)	Damaging	Decrease (DDG=-1.28)	Decrease (DDG=-1.6)	Probably damaging (0.995)
c.844G>T	p.Asp282Tyr	Pathogenic	Probably damaging 0.74	23.8	Damaging (-7.07)	Damaging	Decrease (DDG=-0.63)	Decrease (DDG=-1.24)	Probably damaging (0.999)

Supplementary Table 1 (Continued)

c.848A>G	p.Asp283Gly	Pathogenic	Probably damaging 0.89	24.4	Damaging (-7.02)	Damaging	Decrease (DDG=-1.99)	Decrease (DDG=-1.61)	Probably damaging (1.000)
c.851T>C	p.Val284Ala	Pathogenic	Probably damaging 0.57	24.9	Damaging (-5.15)	Damaging	Decrease (DDG=-2.36)	Decrease (DDG=-2.64)	Probably Damaging (0.995)
c.853C>T	p.Arg285Cys	Pathogenic	Probably damaging 0.95	27.5	Damaging (-7.88)	Damaging	Decrease (DDG=-1.17)	Decrease (DDG=-0.71)	Probably damaging(1.000)
c.910G>T	p.Val304Phe	Pathogenic	Probably damaging 0.86	23.8	Damaging (-7.44)	Damaging	Decrease (DDG=-1.86)	Decrease (DDG=-0.89)	Probably damaging (0.995)
c.920A>C	p.Gln307Pro	Pathogenic	Probably damaging 0.95	24.6	Damaging (-7.61)	Damaging	Decrease (DDG=-1.74)	Decrease (DDG=-1.26)	Probably damaging (0.995)
c.921G>C	p.Gln307His	Pathogenic	Probably damaging 0.95	22.1	Damaging (-7.61)	Damaging	Decrease (DDG=-1.28)	Decrease (DDG=-1.1)	Probably damaging (1.000)
c.962G>T	p.Gly321Val	Pathogenic	-	24.8	Damaging (-7.23)	Damaging	Decrease (DDG=-1.77)	Decrease (DDG=0.32)	Probably damaging (1.000)
c.962G>A	p.Gly321Glu	Pathogenic	Probably damaging 0.95	24.8	Damaging (-7.21)	Damaging	Decrease (DDG=-0.72)	Decrease (DDG=-0.60)	Probably damaging (1.000)
c.1006A>G	p.Thr336Ala	Pathogenic	Probably damaging 0.95	23.8	Damaging (-5.90)	Damaging	Decrease (DDG=-0.38)	Decrease (DDG=-0.69)	Probably damaging (1.000)
c.1052G>T	p.Gly351Val	Pathogenic	Probably damaging 0.95	27.7	Damaging (-8.32)	Damaging	Decrease (DDG=-1.24)	Decrease (DDG=-0.64)	Probably damaging (1.000)
c.1057C>T	p.Pro353Ser	Pathogenic	Probably damaging 0.95	23.6	Damaging (-6.47)	Damaging	Decrease (DDG=-1.48)	Decrease (DDG=-0.36)	Probably damaging (0.999)
c.1081G>A	p.Ala361Thr	Pathogenic	Probably damaging 0.95	23.9	Damaging (-7.08)	Damaging	Decrease (DDG=-0.68)	Decrease (DDG=-0.79)	Probably damaging (1.000)
c.1138A>G	p.Ile380Val	Pathogenic	Probably damaging 0.86	23.5	Damaging (-6.94)	Damaging	Decrease (DDG=-1.49)	Decrease (DDG=-0.53)	Probably damaging (0.972)
c.1139T>C	p.Ile380Thr	Pathogenic	Probably damaging 0.86	23.5	Damaging (-7.17)	Damaging	Decrease (DDG=-3.21)	Decrease (DDG=-1.33)	Probably damaging (1.000)
c.1141T>C	p.Phe381Leu	Pathogenic	Probably damaging 0.95	25.9	Damaging (-7.34)	Damaging	Decrease (DDG=-2.23)	Decrease (DDG=-0.70)	Probably damaging (1.000)
c.1159C>T	p.Arg387Cys	Pathogenic	Probably damaging 0.95	26.8	Damaging (-7.04)	Damaging	Decrease (DDG=-0.77)	Decrease (DDG=-0.53)	Probably damaging (1.000)
c.1160G>A	p.Arg387His	Pathogenic	Probably damaging 0.95	27.5	Damaging (-7.01)	Damaging	Decrease (DDG=-1.91)	Decrease (DDG=-0.95)	Probably damaging (1.000)
c.1166A>G	p.Glu389Gly	Pathogenic	Probably damaging 0.95	28.7	Damaging (-7.07)	Damaging	Decrease (DDG=-0.89)	Decrease (DDG=-2.02)	Probably damaging (1.000)
c.1175T>C	p.Ile392Thr	Pathogenic	Probably damaging 0.57	25.9	Damaging (-7.63)	Damaging	Decrease (DDG=-3.98)	Decrease (DDG=-2.09)	Probably damaging (1.000)
c.1177C>G	p.Arg393Gly	Pathogenic	Probably damaging 0.95	24.6	Damaging (-7.12)	Damaging	Decrease (DDG=-1.84)	Decrease (DDG=-1.77)	Probably damaging (1.000)
c.1178G>A	p.Arg393His	Pathogenic	Probably damaging 0.95	24.3	Damaging (-7.12)	Damaging	Decrease (DDG=-2.09)	Decrease (DDG=-1.36)	Probably damaging (1.000)
c.1186C>G	p.Pro396Ala	Pathogenic	Probably damaging 0.95	24.8	Damaging (-8.18)	Damaging	Decrease (DDG=-1.83)	Decrease (DDG=-1.14)	Probably damaging (0.998)
c.1187C>G	p.Pro396Arg	Pathogenic	Probably damaging 0.95	26.9	Damaging (-8.21)	Damaging	Decrease (DDG=-0.72)	Decrease (DDG=-0.62)	Probably damaging (0.999)
c.1192G>A	p.Glu398Lys	Pathogenic	Probably damaging 0.95	25.6	Damaging (-7.21)	Damaging	Decrease (DDG=-0.55)	Decrease (DDG=-1.45)	Probably damaging (1.000)
c.1220A>C	p.Lys407Thr	Pathogenic	Probably damaging 0.95	24.9	Damaging (-7.72)	Damaging	Decrease (DDG=-0.48)	Decrease (DDG=-1.1)	Probably damaging (0.999)
c.1225C>T	p.Pro409Ser	Pathogenic	Probably damaging 0.95	24.3	Damaging (-7.22)	Damaging	Decrease (DDG=-0.48)	Decrease (DDG=-0.83)	Probably damaging (1.000)
c.1226C>A	p.Pro409Gln	Pathogenic	Probably damaging 0.95	25.7	Damaging (-7.27)	Damaging	Decrease (DDG=-0.4)	Decrease (DDG=-0.55)	Probably damaging (1.000)
c.1229G>A	p.Gly410Asp	Pathogenic	Probably damaging 0.95	23.7	Damaging (-7.30)	Damaging	Decrease (DDG=-0.84)	Decrease (DDG=-0.38)	Probably damaging (0.997)
c.1316G>C	p.Arg439Pro	Pathogenic	Probably damaging 0.95	26.9	Damaging (-7.16)	Damaging	Decrease (DDG=-1.71)	Decrease (DDG=-1.17)	Probably damaging (1.000)
c.1318C>T	p.Leu440phe	Pathogenic	Probably damaging 0.95	27	Damaging (-5.91)	Damaging	Decrease (DDG=-0.59)	Decrease (DDG=-1.74)	Probably damaging (0.999)
c.1339G>A	p.Gly447Arg	Pathogenic	Probably damaging 0.95	27.2	Damaging (-8.32)	Damaging	Decrease (DDG=-1.00)	Decrease (DDG=-0.32)	Probably damaging (1.000)

Supplementary Table 1 (Continued)

c.1340G>T	p.Gly447Val	Pathogenic	Probably damaging 0.95	26.5	Damaging (-8.32)	Damaging	Decrease (DDG=-1.09)	Decrease (DDG=-0.07)	Probably damaging (1.000)
c.1358T>A	p.Val453Glu	Pathogenic	Probably damaging 0.95	24.6	Damaging (-7.38)	Damaging	Decrease (DDG=-1.49)	Decrease (DDG=-1.21)	Probably damaging (1.000)
c.1361G>A	p.Arg454His	Pathogenic	Probably damaging 0.95	26.5	Damaging (-7.27)	Damaging	Decrease (DDG=-1.99)	Decrease (DDG=-1.07)	Probably damaging (1.000)
c.1366G>C	p.Asp456His	Pathogenic	Probably damaging 0.95	25.2	Damaging (-7.44)	Damaging	Decrease (DDG=-1.39)	Decrease (DDG=-0.70)	Probably damaging (0.967)
c.1367A>T	p.Asp456Val	Pathogenic	Probably damaging 0.95	23.8	Damaging (-7.45)	Damaging	Decrease (DDG=-1.04)	Decrease (DDG=-0.09)	Probably damaging (0.998)
c.1381G>A	p.Ala461Thr	Pathogenic	Probably damaging 0.95	25.7	Damaging (-7.14)	Damaging	Decrease (DDG=-0.82)	Decrease (DDG=-1.52)	Probably damaging (0.986)
c.1387C>T	p.Arg463Cys	Pathogenic	Probably damaging 0.85	24.2	Damaging (-7.01)	Damaging	Decrease (DDG=-0.66)	Decrease (DDG=-0.55)	Probably damaging (1.000)
c.1463G>T	p.Gly488Val	Pathogenic	Probably damaging 0.95	24.8	Damaging (-8.32)	Damaging	Decrease (DDG=-1.08)	Decrease (DDG=-0.60)	Probably damaging (1.000)
c.1465C>T	p.Pro489Ser	Pathogenic	Probably damaging 0.95	24.4	Damaging (-6.20)	Damaging	Decrease (DDG=-1.72)	Decrease (DDG=-1.04)	Probably damaging (1.000)

Supplementary Table 2. Conservation scores (ConSurf) of G6PD variants. Based on the scores obtained from this tool, 28 variants received the highest score of 9, followed by 12 variants with a score of 8 while no variant received the lowest scores of 1 or 2.

Variant	Protein	Consurf
c.130G>A	p.Ala44Thr	8
c.148C>T	p.Pro50Ser	9
c.152C>T	p.Thr51Ile	8
c179T>C	p.Leu60Pro	4
c.185C>A	p.Pro62His	5
C.224T>C	p.Leu75Pro	3
c.404A>C	p.Asn135Thr	7
c.409C>T	p.Leu137Phe	5
c.473G>A	p.Cys158Tyr	4
c.488G>A	p.Gly163Asp	3
c.497G>A	p.Arg166His	9
c.514C>T	p.Pro172Ser	9
c.517T>C	p.Phe173Leu	8
c.527A>G	p.Asp176Gly	9
c.536G>A	p.Ser179Asn	8
c.592C>T	p.Arg198Cys	9
c.592C>A	p.Arg198Ser	9
c.593G>C	p.Arg198Pro	9
c.593G>A	p.Arg198His	9
c.679C>T	p.Arg227Trp	6
c.737G>T	p.Arg246Leu	9
c.769C>G	p.Arg257Gly	9
c.770G>T	p.Arg257Leu	9
c.806G>A	p.Cys269Tyr	6
c.820G>A	p.Glu274Lys	9
c.825G>C	p.Lys275Asn	7
c.826C>T	p.Pro276Ser	9
c.844G>C	p.Asp282His	5
c.844G>T	p.Asp282Tyr	5
c.848A>G	p.Asp283Gly	5
c.851T>C	p.Val284Ala	7
c.853C>T	p.Arg285Cys	9
c.910G>T	p.Val304Phe	8
c.920A>C	p.Gln307Pro	9

Variant	Protein	Consurf
c.921G>C	p.Gln307His	9
c.962G>T	p.Gly321Val	6
c.962G>A	p.Gly321Glu	6
c.1006A>G	p.Thr336Ala	9
c.1052G>T	p.Gly351Val	7
c.1057C>T	p.Pro353Ser	9
c.1081G>A	p.Ala361Thr	7
c.1138A>G	p.Ile380Val	6
c.1139T>C	p.Ile380Thr	6
c.1141T>C	p.Phe381Leu	8
c.1159C>T	p.Arg387Cys	5
c.1160G>A	p.Arg387His	4
c.1166A>G	p.Glu389Gly	5
c.1175T>C	p.Ile392Thr	7
c.1177C>G	p.Arg393Gly	7
c.1178G>A	p.Arg393His	7
c.1186C>G	p.Pro396Ala	9
c.1187C>G	p.Pro396Arg	9
c.1192G>A	p.Glu398Lys	9
c.1220A>C	p.Lys407Thr	9
c.1225C>T	p.Pro409Ser	8
c.1226C>A	p.Pro409Gln	8
c.1229G>A	p.Gly410Asp	8
c.1316G>C	p.Arg439Pro	7
c.1318C>T	p.Leu440phe	9
c.1339G>A	p.Gly447Arg	9
c.1340G>T	p.Gly447Val	9
c.1358T>A	p.Val453Glu	7
c.1361G>A	p.Arg454His	9
c.1366G>C	p.Asp456His	9
c.1367A>T	p.Asp456Val	8
c.1381G>A	p.Ala461Thr	8
c.1387C>T	p.Arg463Cys	3
c.1463G>T	p.Gly488Val	8