In silico prediction of HBD gene variants in the Iranian population

Keivan Moradi 1*, Aboozar Mohammadi 2 and Mohsen Kazeminia 3

Abstract

Background: The quantification of hemoglobin A2 (Hb A2; α262) is used as a valuable test to differentiate α- and β-thal carriers in clinical laboratories. Therefore, the HBD (δ-globin) gene variants could result in reduced levels of Hb A2 and have implications for thalassemia screening programs. The aim of the present study was to predict the consequences of HBD gene variants identified in the Iranome project.

Results: The highest number of variants was in the Persian Gulf Islanders. The variants of p.Gln132Glu (HBD: c.394C>G), p.Gly17Arg (HBD: c.14C>T), and p.Ala28Ser (HBD: c.82G>T) presented damage results in three or more prediction tools. In addition, it seems that the p.Gly30= (HBD: c.90C>T) decreases the use of authentic splice and, instead, creates a new donor splice site (DSS) or leads to the use of a cryptic DSS.

Conclusions: Most of these variants have been associated with a decrease in Hb A2 levels. Due to the high mutational diversity in the HBB gene in the Iranian population and the use of Hb A2 quantification to differentiate α- and β-thal carriers among Iranian clinical laboratories, some attention should be taken to a possible co-inheritance of HBD gene variants to avoid the misdiagnosis of β-thal carriers.

Keywords: Hemoglobin A2, δ-globin gene, Variant, Thalassemia, Iran

Background

Hemoglobin A (Hb A; α2β2, 96–98%), hemoglobin A2 (Hb A2; α2δ2, less than 3%), and hemoglobin F (Hb F; α2γ2, less than 1%) are three main components of the total hemoglobin observed in normal adults. Alpha and beta thalassemia (α- and β-thal) are due to mutations in the HBA1/HBA2 and HBB genes, respectively [1–3].

As a member of the β-globin family, the HBD or δ-globin gene is located on chromosome 11. This gene is positioned on the 5′ side of the HBB gene and encodes a 147-amino acid protein that differs from β-globin in only 10 amino acids [3]. The HbVar database (http://globin.bx.psu.edu/hbvar/) is known as a database of information about Hb variants and mutations that cause thalassemia. Although more than 120 variants have been identified in the HBD gene so far, they are far fewer in number than the number of variants reported in the HBB gene [4]. This is because the HBD gene variants are clinically “silent”. Most of the HBD gene variants are missense and result in reduced levels of Hb A2 [5].

On the other hand, Hb A2 accounts for only a small fraction of total hemoglobin and has no known physiological role [3]. However, Hb A2 quantification is used as a valuable test to differentiate α- and β-thal carriers in clinical laboratories; its level is normal or slightly reduced in α-thal carriers and is increased to more than 4% in β-thal carriers. Therefore, in the populations such as Iran in which β-thal is a serious problem in the health system [6–8], any factors affecting the level of Hb A2 could have implications for thalassemia screening programs.

The Iranome database (http://www.iranome.ir/) has recorded the genomic variants found in 800 healthy individuals from eight major ethnic groups in Iran, including Arabs, Azeris, Balochs, Kurds, Lurs, Persians, Persian Gulf Islanders, and Turkmen, with 100 individuals per ethnic group [9]. In this study, we used 14 in silico prediction tools.
tools to identify the deleterious possibility of the HBD variants reported in the Iranome database.

Methods

All studies related to the identification of HBD gene variants in recent years in the Iranian population as well as all HBD gene variants identified in the Iranome project were extracted. To screen the rare variants with frequencies residing around or under 1%, the allele frequency for each variant discovered in 1000 Genomes Project and Genome Aggregation Database (gnomAD) (Available at http://grch37ensembl.org/Homo_sapiens/Info/Index) was used as the reference. In addition, variants were checked for previously reported in the Single Nucleotide Polymorphism database (dbSNP) (https://www.ncbi.nlm.nih.gov/snp/), ITHANET web portal (https://www.ithanet.eu/), HbVar database, as well as in the literature. The ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) was used to search for known variants along with their clinical significances.

Pathogenicity of HBD variants was predicted by using in silico tools such as MutationTaster (http://www.mutationtaster.org/), SIFT (https://sift.bii.a-star.edu.sg/), PROVEAN (http://provean.jcvi.org/index.php), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/), FATHMM-XF (http://fathmm.biocompute.org.uk/fathmm-xf/), I-Mutant disease (http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi), SNPs&GO (https://snps.biofold.org/snps-and-go/), PhD-SNP (https://https://snps.biofold.org/phd-snp/index.html), PMut (http://mbir.barcelona.org/PMut) and CADD (https://cadd.gs.washington.edu/).

Deleterious thresholds used in the above tools were as follows: SIFT < 0.05 [10], PolyPhen-2 > 0.5 [11], PROVEAN < -2.5 [12], FATHMM > 0.5 [13], SNPs&GO > 0.5 [14], PhD-SNP > 0.5 [15], and PMut > 0.5 [16]. For CADD, we used the highest phred-like score cutoff recommended by the authors, i.e., 20 [17]. MutationTaster predicts an alteration as one of four possible types: disease causing (i.e., probably deleterious), disease-causing automatic (i.e., known to be deleterious), polymorphism (i.e., probably harmless), and polymorphism automatic (i.e., known to be harmless). In addition, we used VarSEAK (https://varseak.bio/), MaxEntScan (http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html), NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/), and NNSplice (https://www.fruityly.org/seq_tools/splice.html), as some splice site prediction tools to predict the effect of variants on splicing events. The output of these tools are based on score (%), maximum entropy score, confidence score (0-1), and score (0-1), respectively. Higher score implies a higher probability/confidence of the sequence being a true splice site [18].

Here, we used VarSome database (http://varsome.com) for the interpretation of sequence variants [19]. This database uses ACMG standards and guidelines for interpretation [20]. The gene reference sequence was NG_063112.2. The NM_000519.4 was used to determine the variant position. Position of the variants in protein was determined based on UniProtKB/SwissProt P02042.

Results

Among the 800 healthy individuals studied in the Iranome project, HBD gene variants have been reported in 46 individuals from different ethnicities. Accordingly, the highest number of variants was in the Persian Gulf Islanders, followed by Balochs, Lurs, Kurds, Persians, Turkmen, Azeris, and Arabs, respectively (Table 1). According to NM_000519.4 reference transcript, a total of 16 different single nucleotide variations (SNVs), including seven exonic and nine intronic variants, were identified in the HBD gene. All variants were single nucleotide substitutions, and no insertion or deletion variants were observed. In addition, all variants were detected in heterozygous states (Table 1).

Except for HBD: c.315+55G>T and HBD: c.92+43A>G, the other intronic variants were previously recorded in dbSNP. However, none of the intronic variants were reported in the ITHANET, HbVar, and ClinVar databases (Table 1). Also, analysis on the MutationTaster, FATHMM-XF, PhD-SNP® (except for HBD: c.315+199A>G), and CADD tools showed that all intronic variants were in the category of benign/polymorphisms. In addition, based on the VarSEAK, MaxEntScan, NetGene2, and NNSplice tools, none of these variants had effect on splicing events (Table 2).

Exonic variants were divided into two groups, synonymous SNVs and non-synonymous SNVs, which accounted for two and five variants, respectively (Fig. 1). The non-synonymous variants were p.Gln132Glu (HBD: c.394C>G), p.Asps53Glu (HBD: c.159 T>G), p.Gly17Arg (HBD: c.49G>C), p.Thr5Ile (HBD: c.14C>T), and p.Alac28Ser (HBD: c.82G>T). Although p.Asps53Glu (HBD: c.159 T>G) showed neutral results in all 14 tools, the other four non-synonymous SNVs presented damage results in three or more prediction methods (Table 3). The analysis of HBD: c.82G>T variant on the splice site prediction tools revealed that the replacement of guanine with thymine at position c.82 activates a cryptic donor splice site (DSS) at c.78. This new splice site was much stronger than the authentic splice site at c.92+1 position (Fig. 2a).

None of the two synonymous SNVs, p.Gly30= (HBD: c.90C>T) and p.His98= (HBD: c.294C>T), were found in both ITHANET and HbVar databases (Table 1). Unlike the HBD: c.294C>T variant, HBD: c.90C>T had deleterious results on MutationTaster, CADD, and PhD-SNP® web tools (Table 3). In addition, based on the splice site prediction tools, the replacement of cytosine with thymine at position 90 decreases the score for the use of
authentic DSS at c.92+1 and, instead, creates a new DSS at c.89 position or leads to the use of a cryptic splice site located at 16 nt upstream of authentic DSS (Fig. 2b).

Discussion

More than 1.5 million variants have been identified in the genomes of individuals studied in the Iranome project [9]. Using 14 prediction tools, we evaluated a number of 16 HBD gene variants reported in the Iranome database (Table 1). Based on the ACMG guidelines, none of these variants were categorized as pathogenic or likely pathogenic (Tables 2 and 3).

Zhang et al. [21] reported p.Gln132Glu (HBD: c.394C>G) as a novel δ-globin variant in a healthy Chinese 35-year-old man in 2019 and named it Hb A2-Puer. The hematological and electrophoretic data related to this Hb variant in heterozygous state were as follows: Hb (g/dL) 16.1, MCV (fL) 85.2, MCH (pg) 29.0, Hb A (%) 97.4, Hb A2 (%) 1.3, and Hb X (%) 1.4. Our analysis showed that HBD: c.394C>G is a variant with deleterious effects on MutationTaster, FATHMM-XF, I-Mutant disease, and PhD-SNP, and neutral/benign on SIFT, PROVEAN, PolyPhen-2, SNPs&GO, VarSEAK, and Pmut prediction tools. In addition, with a score of 18.06 in the CADD web tool, this variant could not get the phred-like score cutoff at 20 to locate in the top 1% probability of being deleterious. On the other hand, the same variant has been reported on the HBB gene (Hb Camden: HBB: c.394C>G) with conflicting interpretations of pathogenicity, from silent or likely benign/uncertain significance in the literature [22] or in the ClinVar database, respectively, to causative in the ITHANET database. Finally, based on the ACMG guidelines, the HBD: c.394C>G variant was classified as a variant of uncertain significance [20]. According to the Iranome database, the HBD: c.394C>G variant has been observed in a Kurdish healthy individual in heterozygous form. Since the only report of this variant in the literature is related to Zhang et al.’s [21] study, it can be assumed that the present study is the second one to report this variant in the world and the first study to annotate it in Iran.

The HBD: c.49G>C variant, also known as Hb A2′ or Hb B2, is one of the most common δ-globin gene variants, mainly found in Black families and occurs in nearly 1% of African-Americans [1]. The same substitution was found on the HBB and HBG2 genes [Hb D-Bushman (HBB: c.49G>C) and Hb F-Melbourne (HBG2: c.49G>C), respectively]. The Hb A2-Yialousa (HBD: c.82G>T) and HBD: c.14C>T had also been reported in the Iranome database. The Hb A2-Yialousa has been shown to be...
Table 2: In silico prediction of intronic variants observed in the HBD gene in the Iranian population, based on the Iranome project (http://www.iranome.ir/)

| c.316-76T>C | c.315+55G>T | c.93-34C>T | c.92+43A>G | c.316-80G>A | c.316-102A>G | c.315+199A>G | c.315+146T>G | c.316-32T>C |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Mutation Taster | P | P | P | P | P | P | P | P |
| FATHMM-XF | Benign | Benign | Benign | Benign | Benign | Benign | Benign | Benign |
| CADD | 8948 | 0.201 | 2.825 | 5.735 | 5.194 | 3.008 | 3.105 | 6.307 |
| VarSEAK | No splicing effect | No splicing effect | No splicing effect | No splicing effect | No splicing effect | No splicing effect | No splicing effect | No splicing effect |
| MaxEntScan | No change | No change | No change | No change | No change | No change | No change | No change |
| NetGene2 | No change | No change | No change | No change | No change | No change | No change | No change |
| NNSplice | No change | No change | No change | No change | No change | No change | No change | No change |
| PhD-SNPg | Benign | Benign | Benign | Benign | Benign | Pathogenic | Benign | Benign |
| Frequency in 1000 genomes Project | 0.001 | 0.0002 | NA | 0.0002 | 0.014 | 0.002 | 0.06 | 0.019 |
| Frequency in gnomAD | 0.00006 | 0.000008 | NA | 0.0013 | 0.003 | 0.055 | 0.018 |
| ACMG | VUS | VUS | VUS | VUS | VUS | Likely benign | Likely benign | Benign |

* P: polymorphism, NA: not available, VUS: variant of uncertain significance
most common HBD gene variant in the Mediterranean area [1] as well as in Iran [23], and a rare variant in China [24]. It seems that the replacement of guanine with thymine at position c.82 activates a cryptic DSS at the upstream (Fig. 2a). The HBD: c.14C>T variant was first identified in Greek Cypriots [25]. The frequency of this variant is low and has been reported in other populations such as Oman [26] and China [24]. All of these variants have been associated with a decrease in Hb A2 levels [23–26]. Based on our analyses, these variants presented damage results in three or more prediction methods (Table 3). In addition, according to ACMG guidelines, they were classified as likely benign. Therefore, the pathogenicity or neutrality of them remains unknown.

Another missense variant, HBD: c.159 T>G, was previously recorded in the dbSNP (rs757106601). However, there was no information associated to this variant in the literature, as well as in the ClinVar, ITHANET, and HbVar databases (Table 1). In addition, the same variant has not been reported on the HBB gene. Our analysis showed neutral results related to this variant in all 14 Table 3 In silico prediction of exonic variants observed in the HBD gene in the Iranian population, based on the Iranome project (http://www.iranome.ir/).

| Tool               | c.394C>G/p.Gln132Glu | c.159 T>G/p.Asp53Glu | c.90C>T/p.Gly30= | c.49G>C/p.Gly17Arg | c.14C>T/p.Ala28Ser | c.294C>T/p.His98= | c.82G>T/p.Ala28Ser |
|--------------------|----------------------|----------------------|-----------------|-------------------|-------------------|-----------------|-------------------|
| MutationTaster      | Disease causing      | Polymorphism         | Disease causing | Polymorphism      | Disease causing   | Disease causing  |
| SIFT               | Tolerated            | Tolerated            | Tolerated       | Deleterious       | Tolerated         | Tolerated        |
| PROVEAN            | Neutral              | Neutral              | Neutral         | Neutral           | Neutral           | Neutral          |
| PolyPhen-2         | Benign               | Benign               | NA              | Benign            | Possibly damaging| NA               |
| FATHMM-XF          | Pathogenic           | Benign               | Benign          | Benign            | Benign            | Benign           |
| I-Mutant Disease   | Disease              | Neutral              | NA              | Disease           | Neutral           | Neutral          |
| SNPsGO             | Neutral              | Neutral              | NA              | Disease           | Neutral           | Neutral          |
| CADD               | 18.06                | 11.91                | 23.5            | 0.904             | 15.64             | 7.897           |
| VarSEAK            | No splicing effect   | No splicing effect   | New DSS         | No splicing effect| No splicing effect| New DSS         |
| MaxEntScan         | No change            | No change            | New DSS         | No change         | No change         | New DSS         |
| NetGene2           | No change            | No change            | New DSS         | No change         | No change         | New DSS         |
| NNSplice           | No change            | No change            | New DSS         | No change         | No change         | New DSS         |
| PhD-SNPg           | Pathogenic           | Benign               | Pathogenic      | Benign            | Benign            | Benign          |
| Pmut               | Neutral              | Neutral              | NA              | Neutral           | Neutral           | Neutral          |
| Frequency in 1000 Genomes Project | NA | NA | NA | 0.0003 | 0.001 | 0.008 | 0.001 |
| Frequency in gnomAD | NA | 0.000004 | NA | 0.003 | 0.001 | 0.008 | 0.001 |
| ACMG               | VUS                  | Likely benign        | VUS             | Likely benign     | Likely benign     | Likely benign    |

NA not available, DSS donor splice site, VUS variant of uncertain significance.
prediction tools used in the present study and classified as VUS (Table 3).

Two synonymous $HBD$ gene variants had been reported in the Iranome database (Tables 1 and 3). The p.Gly30= (p.Gly30= (HBD: c.90C>T) variant had deleterious results on MutationTaster, CADD, and PhD-SNP web tools (Table 3). It seems that the replacement of cytosine with thymine at position 90 decreases the score for the use of authentic DSS and, instead, creates a new DSS or leads to the use of a cryptic DSS (Fig. 2b). The same variant was found on the $HBB$ gene, p.Gly30= (HBB: c.90C>T), which has been categorized as a pathogenic variant in the ClinVar database. According to the ACMG guidelines, HBD: c.90C>T was classified as VUS (Table 3). This variant was reported for the first time in the Iranome database in a healthy individual with Arab ethnicity. To the best of our knowledge, there is no report of this variant in the literature so far.

Thalassemia is a serious health problem in the Iranian population. Numerous studies conducted in Iran have shown a high mutational diversity for $\alpha$- and $\beta$-thal diseases [27–32]. However, to our knowledge, only a limited number of studies have been performed in Iran to identify the $HBD$ gene variants. In fact, with the exception of the study by Kordafshari et al. [23], which reported the spectrum of $HBD$ gene variants in 21 individuals, other studies were case reports that identified a specific variant in one or a limited number of individuals [33–35]. Therefore, at least six different types of variants in the $HBD$ gene were reported before in the Iranian population: Hb A2-Yialousa (HBD: c.82G>T), Hb A2-Coburg (HBD: c.350G>A), Hb A2-NYU (HBD: c.39 T>A), Hb A2-Etolia (HBD: c.257 T>C), Hb A2-Fitzroy (HBD: c.428C>A), and HBD: c.92+5G>T. The HBD: c.82G>T and HBD: c.92+5G>T variants were the most frequent ones among Iranian population [23, 33–35].

**Conclusions**

Given the small number of studies performed on the $HBD$ gene in Iran and the fact that the $HBD$ gene variants are clinically “silent,” it can be assumed that the spectrum of variants of this gene in our population is much wider. Evidence of this claim is revealed by the $HBD$ gene variants identified in individuals who participated in the Iranome project. Out of the 16 $HBD$ gene variants reported in the Iranome database, five variants including HBD: c.394C>G, HBD: c.49G>C, HBD: c.82G>T, HBD: c.14C>T, and HBD: c.90C>T showed the potential of deleterious effects in the present study. All of these variants, except for HBD: c.90C>T as a novel variant, have been associated with a decrease in Hb A2 levels. Due to the high mutational diversity in the $HBB$ gene in the Iranian population and the use of Hb A2 quantification to differentiate $\alpha$- and $\beta$-thal carriers among Iranian clinical laboratories, some attention should be taken to a possible co-inheritance of $HBD$ gene variants to avoid the misdiagnosis of $\beta$-thal carriers.
29. Moradi K, Aznab M, Azimi A, Biglari M, Shafieenia S, Alibakhshi R (2020) \(\alpha\)-Thalassemia mutations in Ilam province, West Iran. Hemoglobin. https://doi.org/10.1080/03630269.03632019.01694033

30. Alibakhshi R, Moradi K, Aznab M, Dastafkan Z, Tahrnasebi S, Ahmadi M et al (2020) The spectrum of \(\alpha\)-thalassemia mutations in Kurdistan province, West Iran. Hemoglobin. 44(3):156–161

31. Alibakhshi R, Moradi K, Aznab M, Azimi A, Shafieenia S, Biglari M (2019) The spectrum of \(\beta\)-thalassemia mutations in Hamadan province, West Iran. Hemoglobin. 43(1):18–22

32. Moradi K, Aznab M, Biglari M, Shafieenia S, Azimi A, Bijari N et al (2020) Molecular genetic analysis of \(\alpha\)-thalassemia in Hamadan province, West Iran. Hemoglobin. 4(5):319–324

33. Amirian A, Karimipoor M, Jafarinejad M, Taghavi M, Kordafshari A, Azar SF et al (2011) First report on the co-inheritance of beta-globin IVS-I-5 (G-- > C) thalassemia with delta globin CD12 [Asn-- > Lys (AAT-- > AAA)] HbA2-NYU in Iran. Arch Iran Med 14(1):8–11

34. Amirian A, Jafarinejad M, Kordafshari AR, Mosayyebzadeh M, Karimipoor M, Zeinali S (2010) Identification of a novel \(\delta\)-globin gene mutation in an Iranian family. Hemoglobin. 34(6):594–598

35. Valaei A, Eghbalpour F, Kanimoghaddam Z, Bayat F, Basranj MT, Karimipoor M et al (2012) Co-inheritance of beta & delta-globin gene (HbYialousa) mutations in an Iranian \(\beta\)-thalassemia carrier. Int J Clin Med 3: 633–636

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.