Blood Reviews xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Blood Reviews

journal homepage: www.elsevier.com/locate/blre

REVIEW

Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East

Nejat Mahdieh [a], Bahareh Rabbani [a,b,*]

[a] Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran
[b] Growth and Development Research Center, Tehran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Available online xxxx

Keywords:
Beta-thalassemia
In silico analysis
HBB mutations

ABSTRACT

Thalassemia is one of the most common single gene disorders worldwide. Nearly 80 to 90 million with minor beta thalassemia and 60–70 thousand affected infants are born annually worldwide. A comprehensive search on several databases including PubMed, InterScience, British Library Direct, and Science Direct was performed extracting papers about mutation detection and frequency of beta thalassemia. All papers reporting on the mutation frequency of beta thalassemia patients were selected to analyze the frequency of mutations in different regions and various ethnicities. Mutations of 31,734 individuals were identified. Twenty common mutations were selected for further analysis. Genotype–phenotype correlation, interactome, and *in silico* analyses of the mutations were performed using available bioinformatics tools. Secondary structure prediction was achieved for two common mutations with online tools. The mutations were also common among the countries neighboring Iran, which are responsible for 71% to 98% of mutations. Computational analyses could be used in addition to segregation and expression analysis to assess the extent of pathogenicity of the variant. The genetics of beta thalassemia in Iran is more extensively heterogeneous than in neighboring countries. Some common mutations have arisen historically from Iran and moved to other populations due to population migrations. Also, due to genetic drift, the frequencies of some mutations have increased in small populations.

1. Introduction

Thalassemias are the most common monogenic disorders [1] in many ethnic groups, due to mutations of beta/alpha hemoglobin chain encoded by *HBB*/*HBA* genes. Thalassemia is not uniformly distributed; the highest frequency is seen in certain geographic regions spanning countries bordering the Mediterranean, parts of North and West Africa, the Middle East, the Indian subcontinent, southern Far East, and southeastern Asia, the so-called thalassemia belt [2,3]. Based on WHO estimations, nearly 1.5% of the world population (i.e. 80 to 90 million) are beta thalassemia minors (heterozygous state) and 60,000–70,000 affected infants are born annually worldwide [3–6]. The frequency of alpha thalassemia carriers is also high around the world, distributed majorly in tropical and subtropical regions. In 1925, Cooley and Lee defined thalassemia as a severe form of anemia with splenomegaly and bone changes [7]. This group of inherited hematological disorders is characterized as early onset of anemia due to reduced synthesis of one or more globin chains [8]. Normal hemoglobins including Hb Portland ($\zeta2\gamma2$), Hb Gower 1 ($\zeta2\varepsilon2$), Hb Gower 2 ($\alpha2\varepsilon2$), HbF ($\alpha2\gamma2$), HbA ($\alpha2\beta2$) and HbA_2 ($\alpha2\delta2$) consist of tetramers of two subunits synthesized during different developmental stages; adult hemoglobins are HbA_1 ($\alpha2\beta2$) and HbA_2 ($\alpha2\delta2$) [9]. Major hemoglobin contains two pairs of polypeptide chains, α and β [10]; genes regulating the synthesis and structure of these globins are organized in two clusters on 16p and 11p, respectively [11–14]. The beta globin gene contains three exons, which are separated by two introns or intervening sequences (IVS). Hemoglobin synthesis is controlled by the locus control region (LCR) which consists of five DNase-hypersensitive sites that lie upstream of β-globin genes [15,16].

Up to now, more than 800 variants have been described in the beta globin gene (*HBB*) worldwide (MIM#141900; GenBank genomic reference sequence NG_000007.3) to cause beta thalassemia [17–19] (http://globin.bx.psu.edu/hbvar/menu.html). The *HBB* mutations are particularly frequent in Sardinia (11%–34%), Sicily (10%), Greece (5%–15%), and Iran (4%–10%) [20–23]. Two types of beta globin gene mutations have been characterized including β^0 (no β globin chain is synthesized) and β^+ (β-globin chains are partly produced) [24].

The majority of *HBB* mutations affect a single nucleotide (point mutations). Mutations of different parts of *HBB* (e.g. promoter, intron,

* Corresponding author at: Genetic Research Lab, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Niayesh-Vali asr Intersection, Tehran, Iran. Tel.: +9821 23922294.
E-mail address: baharehrabbani@yahoo.com (B. Rabbani).

http://dx.doi.org/10.1016/j.blre.2016.07.001
0268-960X/

splice sites, and exons) can affect transcription, splicing, and translation of the gene. A common set of mutations has been reported in various ethnic groups [25–29]; i.e. different mutations may arise from various cohorts, separately. For example, resistance against malaria may play a role in the distribution of *HBB* mutations; natural selection has operated in heterozygotes of *HBB* mutations particularly HbS against *Plasmodium falciparum* malaria. Furthermore, the migration of people may have a main role in the genetic diversity of thalassemia. Studies have shown that many of the mutations found in Lebanon have their origin from Turkish, Iranian, Kurdish, Bulgarian, and Asian Indian cohorts [30,31]. A few deletions have been reported in the *HBB* gene; among 14 reported deletions of *HBB* gene, the 619 bp deletion is the most common.

One of the most effective methods for prevention of beta thalassemia is carrier screening. For example, the Iranian national thalassemia screening program (Supplementary Fig. S1, Supplementary Table S1) has been established for more than ten years [32] in this region of the world. This kind of preventive approach will improve health systems in the Middle East. Here, we base a cohort study on the surveys performed in Middle Eastern countries to find out about the distribution of *HBB* gene mutations and their prevalence in the region. We extend our study into *in silico* analysis of the frequent mutations to unravel the pathogenicity of the variants. The confidence of the methods for predicting the pathogenicity was evaluated based on the analysis of the known mutations. This study also includes genotype–phenotype correlation of frequent mutations to help physicians in their patients' management and prenatal diagnostic studies. This data will be useful for genetic counseling of beta thalassemia families.

2. Materials and methods

2.1. Search strategies

A PubMed database search was conducted to find published reports of beta thalassemia mutations in Middle Eastern countries focusing on the neighboring countries of Iran. The keywords are as follows: “mutation”, “gene”, “beta globin”, and/or “beta thalassemia” and “population's name such as Iran, Azerbaijan, Turkey, Iraq, Kuwait, Saudi Arabia, Qatar, United Arab Emirate, Bahrain, Oman, Jordan, Pakistan”. Among these countries, Iran is the largest country which has a national Thalassemia Network Screening and Genetic Diagnosis Program in the region. Therefore, a comprehensive search on PubMed database was performed, extracting all papers on beta thalassemia in Iran; also, Scientific Information Database (SID), Medlib, and Magiran were searched for Persian papers. All studies were categorized based on health status of the studied subjects, carriers, and cases (thalassemic). Three criteria were considered to include the publications in the study: (1) status of patients and/or carrier, (2) mutation frequencies in each group, and (3) comparable molecular detection methods. Studies were excluded if the statistics of affected patients and those with asymptomatic genetic traits were mixed and were reported together. The following information was collected from each relevant study: number of patients/carriers, type and number of mutations, geographical location, ethnicity, and year of publication.

2.2. Mutation data abstraction

A literature search of *HBB* gene mutations was performed on the above-mentioned populations, to identify all the reported mutations in the database.

2.3. Geographical and ethnical distribution of causal variants in Iran

Research data from Iran were sorted to seven regions in the country: Central, Southern, Eastern, Northern, North West, Western, and South West regions.

2.4. Mutation selection

The number of traits/ patients and carriers was extracted from relevant studies to calculate the mutation frequencies of patients and carriers in different populations. The most frequent mutations were selected for further analysis. The frequently observed mutations in Iranian subjects were rechecked in neighboring countries. Twenty common mutations among these geographical regions and populations were chosen for *in silico* structural and functional analyses.

2.5. In silico analysis

The selected variants were categorized into intronic, exonic, and regulatory mutations based on location of the mutations. Then, the exonic mutations were categorized into missense, nonsense, and frameshift (insertion/deletion [indels]) mutations based on functional effect of the mutation. In this way, different software and servers were used for analysis depending on the type of mutation (see following sections).

2.5.1. Multiple sequence alignment

A multiple sequence alignment was performed by a server named Mutation at a Glance [33], to compare UniProt protein family members (UniProtKB/Swiss-Prot P68871). In this way, conserved domains of protein are distinguished among paralogs.

2.5.2. Structural and functional analysis

The protein sequence of beta globin (UniProtKB/Swiss-Prot P68871) was compared using protein homology/analogy recognition engine V2.0 (Phyre2) [34] to determine the structure and function of the variants in protein. The functional analysis was completed using CombFunc tool [35]. Iterative threading assembly refinement (I-TASSER) server was also applied for protein structure and function predictions [36]. Modeling was based on using LOMETS threading program. The functional analysis was performed by COACH server [37].

2.5.3. Protein interaction network

STRING database version 10.0, a database of known and predicted protein–protein interactions, was used to predict functional association of HBB protein in network of proteins. The interaction includes physical and functional associations derived from genomic context, high-throughput experiments, co-expression, and previous knowledge [38]. The molecular function of proteins within a cell describes the way a system works. A dysfunctional gene in this system may interact with other genes with new phenotypic expression.

2.5.4. SNP annotations

2.5.4.1. Missense, nonsense, and indel mutations. A series of available bioinformatic tools were applied to assess the pathogenic effect of the common functional mutations. Sorting intolerant from tolerant (SIFT) [39], polymorphism phenotyping (PolyPhen-2 v2.1) [40], nonsynonymous single nucleotide polymorphism (nsSNPAnalyzer) [41], protein annotation through evolutionary relationship (PANTHER) [42], screening for non-acceptable polymorphisms (SNAP) [43], single nucleotide polymorphism and gene ontology (SNP&GO) [44], and MutPred [45] were applied to predict the effects of SNPs changing residues in proteins which may cause differences in function. Combined annotation dependent depletion (CADD) is a tool for checking the deleteriousness of single nucleotide variations (SNV) and indel variants in the human genome [46]. MutationTaster [47] and Variation Viewer (http://www.ncbi.nlm.nih.gov/variation/view/) also predict the effect of nucleotide changes in a sequence.

2.5.4.2. Regulatory, splice, and intronic mutations. The effects of intronic SNPs were determined using human splicing finder (HSF) (www.umd.be/HSF/) [48]. HSF, a tool for predicting the effects of SNPs on splicing signals, contains available matrices for auxiliary sequence

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001

prediction and binding sites of some ribonucleoproteins. It includes different algorithms such as RESCUE-ESE,ESE-Finder, MaxEntScan, and FAS-ESS [49]. CADD was used for functional analysis SNV of intronic and regulatory positions. MutationTaster and Variation Viewers were used to assess the pathogenicity.

3. Results

3.1. Selected studies

Data from 43 articles including minor, major, and intermedia thalassemia cases were gathered for analysis from Iran. A total of 14,293 cases were enrolled in the study. A majority of the cases were among high-risk families of B-thalassemia all around Iran. There was no specific characterization criterion of the intermedia cases in the studied samples and, subsequently, they were included in different groups depending on their genotypes.

In addition, 29 articles were included from other ethnicities and populations in the region. Studies in Iranian populations were categorized into affected cases (2104 individuals) and carriers/heterozygous individuals (12,189) (Table 1, Supplementary Table S2); of these, 152 major cases from the southern region, 312 major cases from the northern part, and 481 major patients from the central region. Studies on the western part of Iran showed 527 affected individual, and 632 affected from the south-west regions. No affected patient was observed in the studies from the east (Table 1, Supplementary Table S2). The major and minor individuals were also investigated in other neighboring populations (data not shown). Nearly, 16,583 cases were enrolled in the reported studies.

3.2. Causal genetic variations among regional populations and subpopulations of Iran

A total of ninety mutations were reported from the published studies in Iranian population from different geographical and ethnic groups. In total, 16,397 alleles have been reported and 16% alleles had no mutation. A total of 4208 alleles (25.66%) were studied from major and intermedia subjects. Among the mutations in the majors, 46% had a frequency of <0.05%, i.e. they are considered as rare variants. A total of 12,189 alleles (of 16,397) (74.34%) were studied in minors, and 69 mutations were determined in the minor group; 27 of them (39.13%) had a frequency less than 0.05% (Supplementary Table S2). With this in mind, 90 mutations may be distributed differently among minors and majors.

A total of 30 of 90 mutations were frameshift, 25 splice and intronic, 15 UTR and regulatory, 14 missense, and 6 nonsense mutations. The types and frequencies of mutations in each geographic region are presented. The first twenty frequent mutations including eight intronic, six frameshift, two missense, two nonsense, and two regulatory mutations were chosen for further investigation (Supplementary Table S2). These common mutations were examined in different neighboring populations (Table 2). The frequency values in our patient cohort are ranked from most to least frequent. They are compared to other reported studies in neighboring countries. The three most frequent mutations are IVSI-5(G > C) (27.5%), IVSII-1(G > A) (24.5%), and IVSI-110G > A (5.6%), respectively, in the Iranian population (Fig. 1, Table 2). They account for more than 55% of the reported beta thalassemia mutations.

IVSI-5(G > C) is responsible for 33.6% of mutations in minor patients (Supplementary Table S2; Fig. 1). The major phenotype of beta thalassemia is prominent in intronic mutations, namely, IVSII-1(G > A).

To specify the frequent mutations in seven geographical regions, three common mutations are presented in neighboring countries (Fig. 2, Table 2). IVS-II-1(G > A) was determined in 28.2% alleles as the frequent mutation in homozygotes (Fig. 1). This mutation showed the highest frequency in the northern and central regions of Iran (Table 1). Interestingly, the specified mutations in subpopulations of Iran are also common among the neighboring counties (Table 1; Fig. 2). For example, IVSII-1(G > A) (57.4%) is common in northern regions of Iran as well as being one of the top frequent mutations in Azerbaijan; although the highest frequent mutation is Fsc8 (−AA) (29.4%) which accounts for 4% of the northern mutations in Iran (Table 2 for comparison).

IVSI-110(G > A) accounts for about 40% of beta globin mutations in Turkey, which is also common in Azerbaijan (12%). The most frequent mutations in neighboring countries are as follows: IVS II-1(G > A) in Iraq (19%), Kuwait (31%), and Saudi Arabia (16%); IVSI-5(G > C) in Qatar (35%), UAE (56%), Oman (44%), and Pakistan (38%); and codon 39 in Saudi Arabia (17%), whereas IVSI-25 bp accounts for 35% of the mutations in Bahrain (Table 2). Common mutations in Iraq and Kuwait are more likely similar to the mutations in the western and southern parts of Iran (Fig. 2). Moreover, IVSI-110(G > A) mutation is also high in the south-west parts of Iran as in Saudi Arabia. Also, IVSI-5(G > C) is high in this region as in Qatar and UAE. The most common mutation (IVSI-25 bp) in Bahrain seems to be common in the south-west region of Iran (Supplementary Table S2).

Table 1
Geographical and ethnical distribution of studied minor and major beta thalassemia cases in the Iranian population.

Regions	No. of carriers	No. of majors	Ethnicity	Common mutations	Refs.
Central	1332 (including 642 couples)	481	Fars	IVS-II-1 (G > A) 36.5%; IVS I-110 (G > C) 9.1%; FSC 8/9(+G) 8.8%	[50]; [27]; [51]; [52]; [53]; [54]; [55]; [56]; [57]; [28]
Southern region	760 (including 365 couples)	152	Fars	IVS II-1 (G > A) 21.1%; IVSI-5(G > C) 18.8%; IVSI-110(G > A) 8.6%	[58]; [59]; [28]; [60]; [61]; [62]
Eastern regions	4731 (including 2190 couples)	0	Balooch and Sistani	IVSI-5 (G > C) 76.1%; FSC8/9(+G) 3.9%; IVSII-1(G > A) 2.5%	[63]; [64]; [65]; [66]; [67]; [29]; [28]
Northern	2313 (including 354 couples)	312	Gilak, Fars	IVS-II-I (G > A) 57.4%; IVSI-5(G > A) 9.4%; CD30(G > A) 6.3%	[68]; [69]; [70]; [71]; [72]; [68]
North West	0	179	Azeri	FSC8/9 21.2%; IVSI-110 21.2%; IVSII-1 (G > A) 19.6%	[73]; [74];
Western	303 (including 100 couples)	348	Kurd and Lur	IVS-II-1 (G > A) 33.4%; FSC8/9(+G) 14.3%; FSC 36/37 (−T)8.3%	[75]; [76]; [77]; [78]; [79]; [80]; [81]
South West	2750 (including 819 couples)	632	Lur and Arab	IVS-II-I (G < A)18.6%; FSC36/37 (−T) 15.7%; IVSI-110 11.1%	[82]; [26]; [83]; [84]; [85]; [86]; [87]; [88]; [28]
Total	12189 (including 4470 couples)	2104			

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001

Table 2
Distribution of twenty common mutations in Iran and neighboring countries.

Mutation	Origin	Iran (%)	Azerbaijan (%)	Turkey (%)	Iraq (%)	Kuwait (%)	Saudi Arabia (%)	Qatar (%)	UAE (%)	Bahrain (%)	Oman (%)	Pakistan (%)	Highest frequency based on HbVar
IVSI-5(G > C)	Mediterranean	27.43	1.76	0.92	4.97	14.75	12.75	35.38	56.44	16.42	44.41	38.02	Bangladesh (60%); Indonesian (54.24%); UAE (53%)
IVS II-1 (G > A)	Mediterranean	24.49	17.06	4.85	19.69	31.15	16.09	9.23	2.96	8.96	2.78	0.58	Iranian (26.47%); Kuwaiti (29%); Yemenite (26.67%)
IVS I-110(G > A)	Middle East (Mediterranean)	5.67	12.35	39.69	11.66	0.82	9.04	6.15	1.65	1.49	0.12		Greek Cypriot (79.86%); Turkish Cypriots (72.19%); Macedonian (47.31%); Greek (42.97%); Albania (42.36%)
Cd36/37(−T)	Kurdish/Iranian	5.52	1.18	0.23	0.19		0.36				0.23		Azerbaijan (2%); Oman (1%); UAE (0.4%)
Fsc 8/9	Asian Indian	5.01	6.18	1.91	4.59	2.46	2.71	26.15	6.27	1.49	0.12	25.53	Pathan (49.05%); Pakistani (25.9%); Punjabi (12.88%); Iranian (11.03%)
IVS I-1 (G > A)	Middle East	2.68	2.06	4.28	4.02	5.74	3.16	1.54		2.98	1.4	0.4	Spanish (31.79%); Hungarian (28.13%); Syria (17%); Lebanese (15%); Greek (13.65%)
IVSI −25 bp	Middle East	2.23			0.96	5.74	9.13	6.15	9.57	35.82	3.61		Bahrain (36%); UAE (8%); Kuwaiti (7.3%)
Fsc8 (−AA)	Mediterranean	2.09	29.41	5.95	4.21	1.64	1.45	3.08			0.23		Russian (38.71%); Azerbaijan (19.2%); Croatian (11.36%)
Codon30 G > C	Black; Asian Indian	1.92	0.59		0.38				0.66		0.47	2.8	Tunisian (2.6%); Pathan (0.95%); Algerian (0.9%); Pakistani (0.9%)
FSC44(−C)	Kurdish	1.73	1.76	1.79	8.41	0.82	1.27	3.08	2.97	4.48	7.23		Oman (9.6%); Bahrain (4.5%); Tunisian (4.4%)
IVS I-6 (T > C)	Mediterranean	1.54	4.18	7.16	12.43	7.38	3.34				0.12		Macedonian (18.56%); Azerbaijan (17.3%); Sicilian (16%); Turkish (14.75%)
Codon39 (C > T)	Middle East	1.34	1.76	4.16	5.54	5.74	17.18	1.54	3.3	23.88	1.17	0.1	Sardinian (95.73%); Italian (66.84%); Argentine (47.06%); French (41.9%)
CD5(−CT)	Mediterranean	1.26	0.29	2.89	7.46		1.54		0.99		1.28	2.1	English (8.7%); Syria (8.5%); Pathan (7.62%); Bulgarian (7.11%)
IVS-I-5 (G > A)	Asian India	1.22	0.29										Spanish (0.93%); Tunisian (0.9%); Algerian (0.9%)
Cd 15 (G > A) (TGG-TGA)	Asian Indian	1.15	1.76	0.06				3.08	0.99	1.49	0.12	1.44	Portuguese (11.79%); Russian (6.45%); Japanese (0.95%)
IVS-II-745, G > C	Mediterranean	1	0.88	3.87	1.34		0.36						Jordan (12%); Sicilian (6.16%); Turkish Cypriots (6.07%)
FSC-22-24 (−AAGT TGG)	Turkish	0.95	0.29	0.29	0.19								Turk
−88 (C > A)	African/ Indian	0.8	0.59	0.06						1.49	0.23	0.1	Bahrain (1.5%); Egyptian (0.6%); Iranian (0.3%)
−28 A > C	Kurdish	0.5	0.59	0.12	0.57	0.82							Azerbaijan (1%); Turkish (0.24%)
CD6 (A > T)(HbS)	Africa, India, the Middle East	0.5	0.88	2.43			4.16				25.29	0.23	American Indian Black
Total mutations		20	19	17	16	11	14	10	10	10	16	10	
Total		88.92	83.86	80.66	86.61	77.06	82.54	95.38	85.8	98.5	88.81	71.3	
Refs		Present study	[89] [90] [91]	[92] [93] [94] [95] [96]	[97] [98] [99] [100] [101]	[102] [103]	[104] [105] [106] [107] [108]	[109]	[110] [111]	[112]	[113] [114] [115]	[116] [117] [118] [119] [120]	

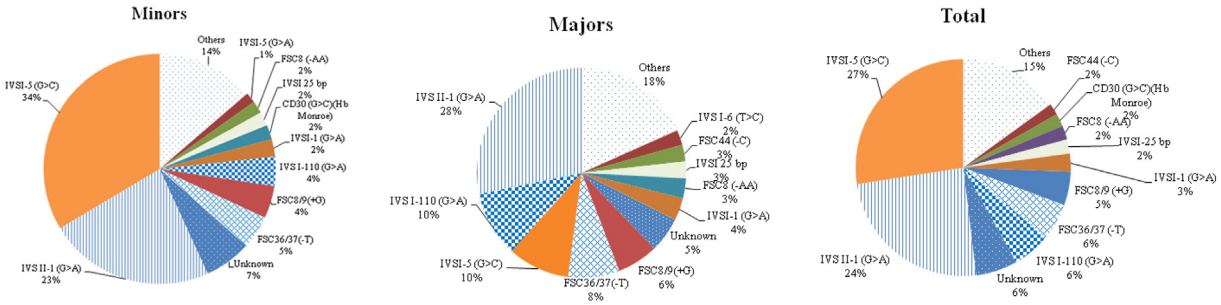

Fig. 1. Frequencies of twenty common mutations among thalassemia minors and majors in Iranian population. A total of 12,189 alleles account for minor thalassemia and 4208 alleles include major thalassemia.

The twenty common mutations were responsible for 89% of all mutant alleles in Iran (Table 2); 10 of these mutations were responsible for 98.5% and 71.3% of mutant alleles in Bahrain and Pakistan, respectively. A panel of common mutations in this region of the world would be helpful for screening programs.

3.3. Genotype–phenotype correlation

Genotype–phenotype analysis of patients and heterozygous/carriers of the common mutation was performed based on the available data (data from Iran). The top ranked mutations consist of intronic mutations. Among the homozygous mutations, IVSII-1 was common (28%) among the major cases in Iran. Among homozygous frameshift mutations, c.112delT (cd36/37(−T)) showed the highest ranked value (8%) among the major thalassemias. The heterozygous cases show minor thalassemia with highest value (34%) in IVSI-5. Also, IVSI-5 is about 10% among majors (Fig. 1).

3.4. In silico analysis

We focus on missense and nonsense variants among the common variants, p.E7V (cd 6 A > T), p.W16X (cd15G > A), p.R31T (cd30G > C), and p.Q40X (cd39C > T). Two nonsense mutations lead to truncated protein. The structural and functional analyses of two missense mutations were investigated in further sections. The frameshift (including small insertions and small deletions) mutations cause truncated protein and impaired functional product. Other mutations were analyzed with different tools depending on the ability of servers (discussed below).

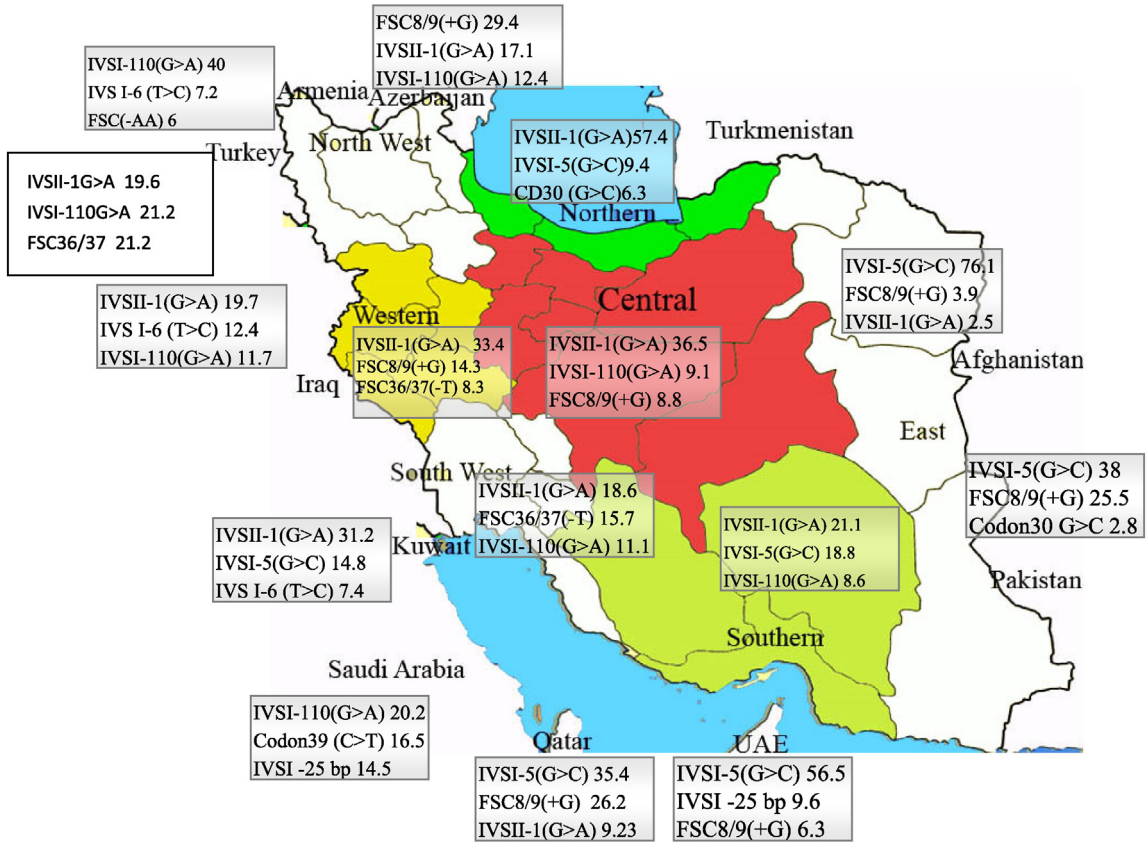

Fig. 2. Distribution of three common mutations of beta globin in different regions of Iran. Central: Isfahan, Tehran, Yazd, Semnan, Markazi, Qazvin; Southern region: Kerman, Fars, Hormozgan; Eastern regions: Sistan va Balouchestan, Khorasan; Northern: Gilan, Mazandaran, Golestan; Northwest: Azarbayjans, Ardabil, Zanjan; Western: Kurdestan, Kermanshah, Lorestan, Hamadan, Ilam; Southwest: Khuzestan, KohkiluyehVaBuyer ahmadi, Bushehr. The numbers in front of the mutations indicate the % of mutation in that region.

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001

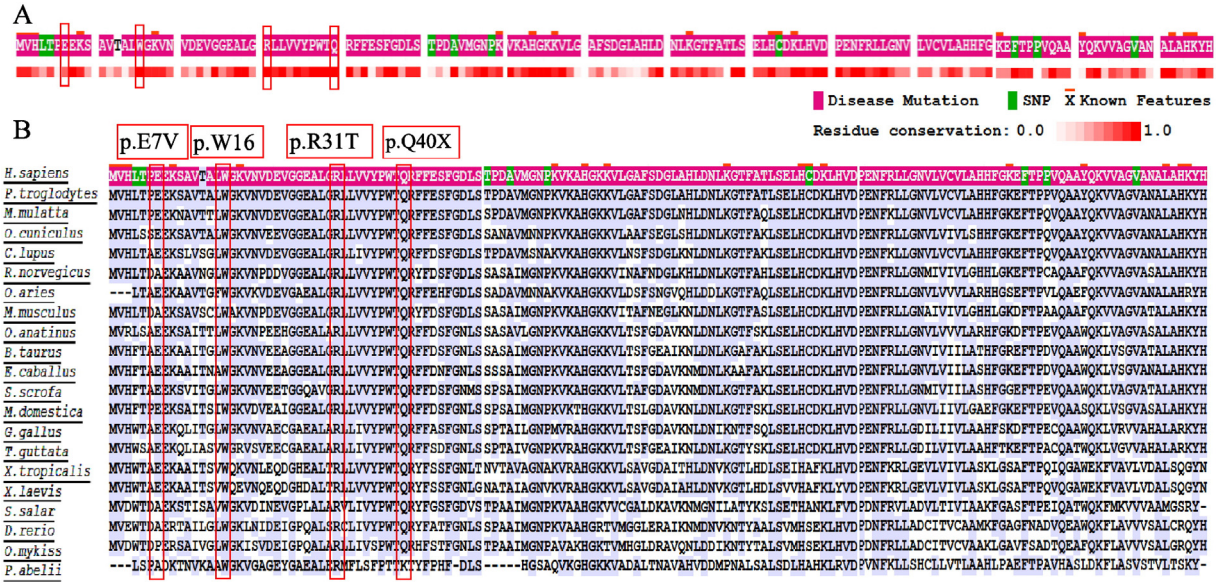

Fig. 3. Protein alignment of HBB amino acid sequence among vertebrates. (A) Distribution of mutations and neutral substitutions along the HBB protein. The degree of conservation is also determined (shown in color range from 0 to 1.0). R31 is highly conserved for residual conservation which is close to 1 as in the diagram. (B) Multiple amino acid alignment of HBB family adapted from Mutation at a Glance homepage (http://harrier.nagahama-i-bio.ac.jp/mutation/ataglance.cgi) and UniProt protein family members. R31 is a highly conserved residue among other organisms; different orthologous and paralogous members are specified in the left side of the protein sequences. Missense (p.R31T) and nonsense (p.W16X and p.Q40X) mutations as indicated in the red box show high conservation among different species but not at p.E7V position.

3.4.1. Protein sequence alignment

p.E7V and p.R31T were the subjects of sequence alignment among other species for conserved residues and the degree of conservation for each amino acid was investigated (Fig. 3, A and B). Arginine at position 31 (R31) is highly conserved (Fig. 3A); in addition, multiple alignment shows that Arg31 is also conserved among different organisms (Fig. 3B). The degree of conservation at glutamic acid 7 (E7) is not as conserved as R31 (Fig. 3A); multiple alignments of protein sequences of different organisms also show lesser conservation in this position. The nonsense mutations as p.W16X and p.Q40X are also positioned at highly conserved regions (~1.0); therefore, a change in amino acid sequence may cause pathogenicity.

3.4.2. Structure prediction

Secondary structure prediction was achieved for p.E7V and p.R31T with online tools. I-TASSER prediction is basically based on threading model [36]. Secondary structure was determined as a helix structure by I-TASSER server with high confidence value for p.E7V and p.R31T which shows these two positions were not structurally changed due to amino acid modification (Fig. 4); solvent accessibility of the given amino acids was also predicted as illustrated, Arg at position 31 is less buried in protein (more exposed) and the accessibility to solvent is higher; though amino acid change to threonine at position 31 had modified the solubility to 3 value (2 is the value in normal structure) which means the amino acid is positioned more within the protein. The accessibility of protein was changed in p.R31T (Fig. 4B).

B-factor profile (BFP) value of E7 and R31 were 0.31 and −0.87, respectively. BFP for position p.E7V has higher value (0.34) than the BFP value of p.R31T (−0.83); p.R31T value was changed compared to normal amino acid sequence which defines that p.R31T is more stable in the structure than p.E7V and this change may be more effective for this position (Fig. 4B).

Threading template of the query protein was performed using LOMETS threading program [121]. The top threading template prediction for p.R31T was based on PDB 1fhjB (Aquomet hemoglobin-I of the maned wolf—*Chrysocyon brachyuru*—identity 0.89; normalized *Z*-score of the threading alignments = 2.48 and coverage = 0.99 [36, 122,123]. The top template for p.E7V is also 1fhjB (identity 0.89; norm. *Z*-score = 2.49, coverage 0.99) (data not shown). The predicted threading template of a normal sequence also based on 1fhjB—identity 0.90; normalized *Z*-score = 2.50 and coverage = 0.99—showed higher *Z*-score in comparison with other sequence substitutions. The difference between normal structure and p.E7V structure are less than p.R31T.

Top five final models of predicted secondary structures were listed on I-TASSER with 3-D predicted tertiary structures (Fig. 4A). To note, decoys from a large structural conformation models were generated to simulate the final model based on pair-wise structure similarity of the decoys (using SPICKER program). The confidence is measured by *C*-score. For p.R31T the *C*-score = 1.25, estimated TM score = 0.89 ± 0.07 and RMSD = 2.3 ± 1.8 Å. For p.E7V the first model was indicated with *C*-score = 1.26. *C*-score of higher value signifies a model with a high confidence and vice versa [36,122,123]. TM-score value indicates the similarity of the predicted structures to the native structure. Structural analogs of the predicted protein are investigated with structural alignment program, TM-align. TM-align aligns the first model to the PDB library models. The proposed model for p.E7V matched with PDB 1fhjB and 1dxtB with TM-score = 0.981. TM-score > 0.5 determined the structure class/protein family of the predicted query protein structure. p.R31T model also matches the PDB 1fhjB with TM-score = 0.980 (data not shown). Therefore, because of structural similarity, the model protein has similar functions; slight changes in the sequence may slightly affect the function due to the position and binding sites of the protein.

Phyre2 server predicted the structure based on template-based homology modeling and fold recognition [34]. The globin protein with p.E7V and p.R31T was modeled based on the members of globin-like superfamily, named d2d5xb1 (crystal structure of carbonmonoxy horse hemoglobin complexed with L35) with confidence score 100; identity 83% and coverage of 99%. As shown, there is a high secondary structure prediction confidence score (red) at both positions (Fig. 5). The second structure has structurally changed the order of amino acid sequences. The alpha helix-strand structure was predicted with mid-disordered confidence at position p.E7 and with low confidence at position p.R31. This means that low disordered regions are lower in flexibility, dynamicity and lower extension in solution and sensitive to a change. To explain, the total disordered score of second structure changes was

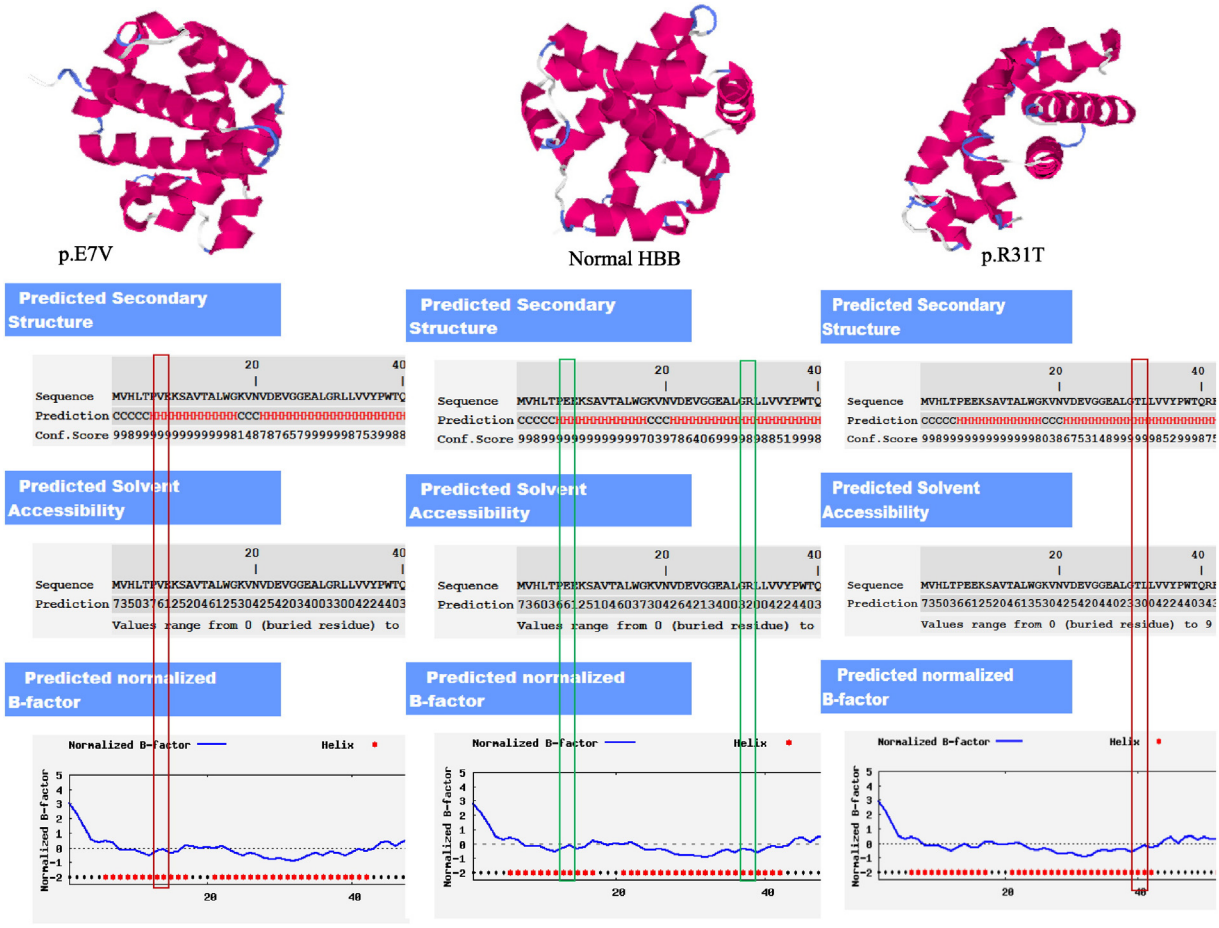

Fig. 4. The data provided by I-TASSER server. (A) Cartoon structural models of human HBB protein and mutant HBB. (B) Predicted secondary structure solvent accessibility and normalized B-factor are shown. The first line indicates the sequence, second line (C: random coil; H: alpha-helix; S: beta-strand) shows the secondary structure which is determined to be helix at positions 7 and 31with a confidence score of 9 (third line). The range of confidence is 0–9, wherein a higher score indicates a prediction with higher confidence. The solvent accessibility of the sequence is predicted as buried amino acids which is determined to be 6 and 2 for amino acids at positions 7 and 31, respectively (range 0–9 wherein a higher value means higher accessibility, i.e. 0—buried to 9—exposed). p.E7V and p.R31T accessibility value were 6 and 3, respectively. The accessibility of protein was changed in p.R31T comparing to p.E7V. The predicted normalized B-factor tells about the stability of the predicted secondary structure. Negative value shows that residue is more stable in the structure. B-factor profile (BFP) value of E7 and R31 were 0.31 and 0.87, respectively. p.E7V shows the BFP of 0.34 (>0) means less stable in experimental structure than −0.83 for p.R31T. BFP is predicted using a combination of both template-based assignment and profile-based prediction. Based on the distributions and predictions of the BFP, residues with BFP values >2 are less stable in experimental structures. p.R31T value was changed compared to normal amino acid sequence which defines that p.R31T is more stable in the structure than p.E7V and this change may be more effective for this position. Threading template of the query protein was performed using LOMETS threading program. The highest significant alignment regions of the templates are chosen with the *Z*-score measurement. Alignment is based on the blocks and spatial positions in the assembly. The best selected templates are chosen from each program which is comparable with normalized *Z*-score. The top threading template prediction for p.R31T and p.E7V was based on PDB 1fhjB (Aquomet hemoglobin-I of the maned wolf—*Chrysocyon brachyuru*—*Z*-score > 1 showed a good alignment (R31T: identity 0.89; normalized *Z*-score of the threading alignments = 2.48 and coverage = 0.99; p.E7V: Identity 0.89; Norm. *Z*-score = 2.49, coverage 0.99). PDB 1v4wA the second template has higher *Z*-score (3.24) but has lower identity = 0.47 for p.R31T. The cartoon structures of the modeled temples are shown here (A). Modeled templates for p.R31T had the *C*-score = 1.25, estimated TM score = 0.89 ± 0.07 and RMSD = 2.3 ± 1.8 Å and model for p.E7V had *C*-score = 1.26, *C*-score is typically in the range of [−5, 2], where a *C*-score of higher value signifies a model with a high confidence and vice versa; *C*-score >−1.5 shows correct model of topology. The estimated TM-score = 0.89 ± 0.07-A TM-score > 0.5 indicates a model of correct topology (global fold similarity) and <0.17 means random fold similarity) and root-mean-square deviation (RMSD) of 2.3 ± 1.8 Å [36,122,123]. These two values indicate the similarity of the predicted structures to the native structure.

higher in p.R31T (16%) than p.E7V (15%). R31 showed that it has a tetramer interface; consequently, changes affect tetramer interference. p.R31T is more sensitive to changes than p.E7V (Fig. 5).

3.4.3. Functional analysis

Amino acid positions 2, 3, 83, and 144 of the beta globin sequence are binding sites of 2,3-bisphosphoglycerate, 64 and 93 are metal binding (iron), 60 and 144 are sites of modification (UniProtKB P68871). Depending on the position and kind of variant change, the structure is disturbed and consequently affects other binding sites modifying the function of the protein.

CombFunc analysis relating to Phyre2 server predicted the function of sequence based on gene ontology; the predicted molecular functions are as follows: metal ion binding, hemoglobin alpha binding, oxygen transporter activity, oxygen binding, haptoglobin binding, and peroxidase activity; similar to normal hemoglobin. The probability of these functions slightly differed in p.E7V in comparison to normal globin. Biological prediction showed oxygen transport and blood coagulation as the top listed functions. ConFunc predicted molecular functions as protein binding, metal ion binding, cation binding, ion binding, binding, substrate-specific transporter activity, oxygen transporter activity, transporter activity, and, to lesser extent, haptoglobin binding, oxidoreductase activity, acting on peroxide as acceptor, oxidoreductase activity, antioxidant activity, peroxidase activity, catalytic activity, hemoglobin binding, hemoglobin alpha binding, and oxygen binding functions. These functions were predicted for p.R31T but with lower probability, that means p.R31T is more potent than p.E7V to alteration (data not shown) [35].

Another bioinformatic tool, 3DLigandSite, was used to predict binding sites at the modified positions. It seems that there is no effective

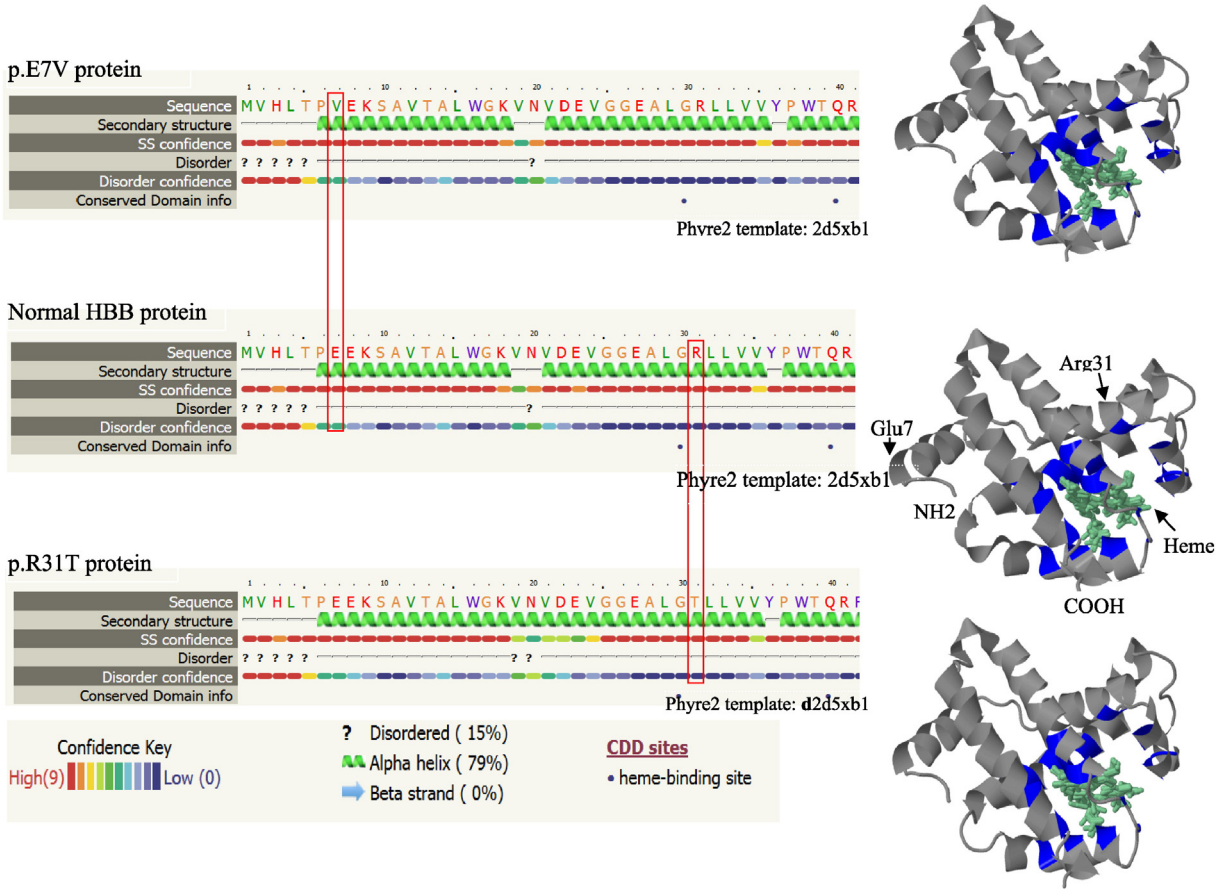

Fig. 5. Prediction of the secondary structure based on template/homology modeling by Phyre2 server. The first line indicates the amino acid sequence and the second line is the secondary structure prediction which is determined as alpha helix (H) extended or B strand and coiled structure with the confidence value of low to high average (red, depicted in third line). The fourth line calculates the structurally ordered sequence using disoPred program indicating two states of ordered and disordered. The homology modeling of p.E7V was based on d2d5xb1, members of globin-like superfamily model. A change p.E7V has 79% alpha helix structure changes with 100% confidence. The structure disorder is about 15% but at E7 change shows average confidence. In addition, p.R31T was also modeled based on d2d5xb1model with 100% confidence and 83% identity; also, p.R31T position is disordered, which means it is not flexible and dynamic with low value (blue). p.E7V and p.R31T are the most common missense mutations. The chemical properties of the amino acids determine the biological activity of the protein. Valine (V), an essential amino acid, is hydrophobic (aliphatic and nonpolar) with the chemical formula [$HO_2CCH(NH_2)CH(CH_3)_2$]. Glutamic acid (E, formula: [$C_5H_9NO_4$]) is an acidic and polar (charged) amino acid. Threonine (T, formula: [$HO_2CCH(NH_2)CH(OH)CH_3$]), bearing an alcohol group, is polar amino acid. Structural view of prediction of normal, p.E7V and p.R31T mutant protein was constructed by Phyre2 server. The cartoon format of the amino acid sequence illustrated the changes in heme binding and tetramer interference positions. Conserved domain information gives the changes in heme binding site and tetramer interface. As the second structure is changed, the sites are changed. The predicted binding sites (blue) and other residues (grey). Heme binding sites are: L32, F42, F43, H64, K67, V68, A71, L89, H93, L97, V99, N103, F104, L107, L142; tetramer interface: T31, V34, V35, W38, R41, H98, D100, N103, N109, V112, C113, A116, H117, G120, F123, T124, P125, Q128, A129, Q132 in normal model which are changed with amino acid changes.

change in binding sites when p.E7V occurs (Fig. 6). Although, if Arg at position 31 changes to Thr there is a change in binding sites as compared to normal HBB protein (Fig. 6, right columns). Binding sites are more influenced in p.R31T than p.E7V in comparison with normal amino acid sequence of beta globin protein (Fig. 6). No change was observed in predicted binding site in p.E7V but contacts, number of contacts, and the distance residues were altered in p.R31T. As change in amino acid R31 modifies heme binding site partiality at other positions.

I-TASSER predicts function using COACH server based on the predicted structure. Functional homologous templates are used to determine the ligand binding sites, enzyme commission, and gene ontology. Based on the ligand binding site analysis, p.R31T matched to PDB 3gdjD (crystal structure camel—*Camelus dromedarius*—hemoglobin) with *C*-score = 1.00 as hem for its binding factor. Model p.E7V functional analysis matched PDB 2h35B (solution structure of human normal adult hemoglobin) with HEC (Heme C) binding site, *C*-score = 1.00 with different ligand binding sites. Therefore, we conclude that these two changes may influence other hem binding sites but are not directly affected sites.

Enzyme commission for R31 is low (*C*-score = 0.374 and TM-score = 0.762, identity = 0.145) with oxidoreductase activity based on PDB 1gvhA (X-ray structure of ferric *Escherichia coli* flavohemoglobin reveals an unexpected geometry of the distal heme Pockettm) with two active sites. The second comparison to the PDB 1cqxA shows (*C*-score = 0.358, TM-score = 0.711 and identity = 0.144) with no active site. This shows that this protein has low enzyme activity. Position p.E7V also has low enzyme commission based on PDB 1gvhA (*C*-score = 0.378, TM-score = 0.763, identity = 0.145). Comparing the two amino acid position changes,p.R31T is more effective in enzyme commission analysis (*C*-score = 0.374 is lower than 0.378 for p.E7V).

The functional analysis is also investigated using gene ontology. Based on gene ontology p.R31T matched to PDB 1dxtB with coverage = 1.00 and *C*-score = 0.91, TM-score = 0.9767, identity = 0.99 and the second ranked template is PDB 1fhjB *C*-score = 0.77, TM-score = 0.9795, coverage = 0.99, and identity = 0.89. p.E7V based on gene ontology matched to PDB 1dxtB with coverage of 1.00 and *C*-score = 0.91, TM-score = 0.9813, identity 0.99; the second-ranked PDB 1a9wE (crystal structure of a human embryonic Gower II carbonmonoxy hemoglobin) had a *C*-score = 0.76 TM-score = 0.9667 coverage 0.99 and identity 0.74, which shows that gene ontology of both amino acid changes has not typically changed the protein functions and that alterations are more damaging to the protein's quality of function and not

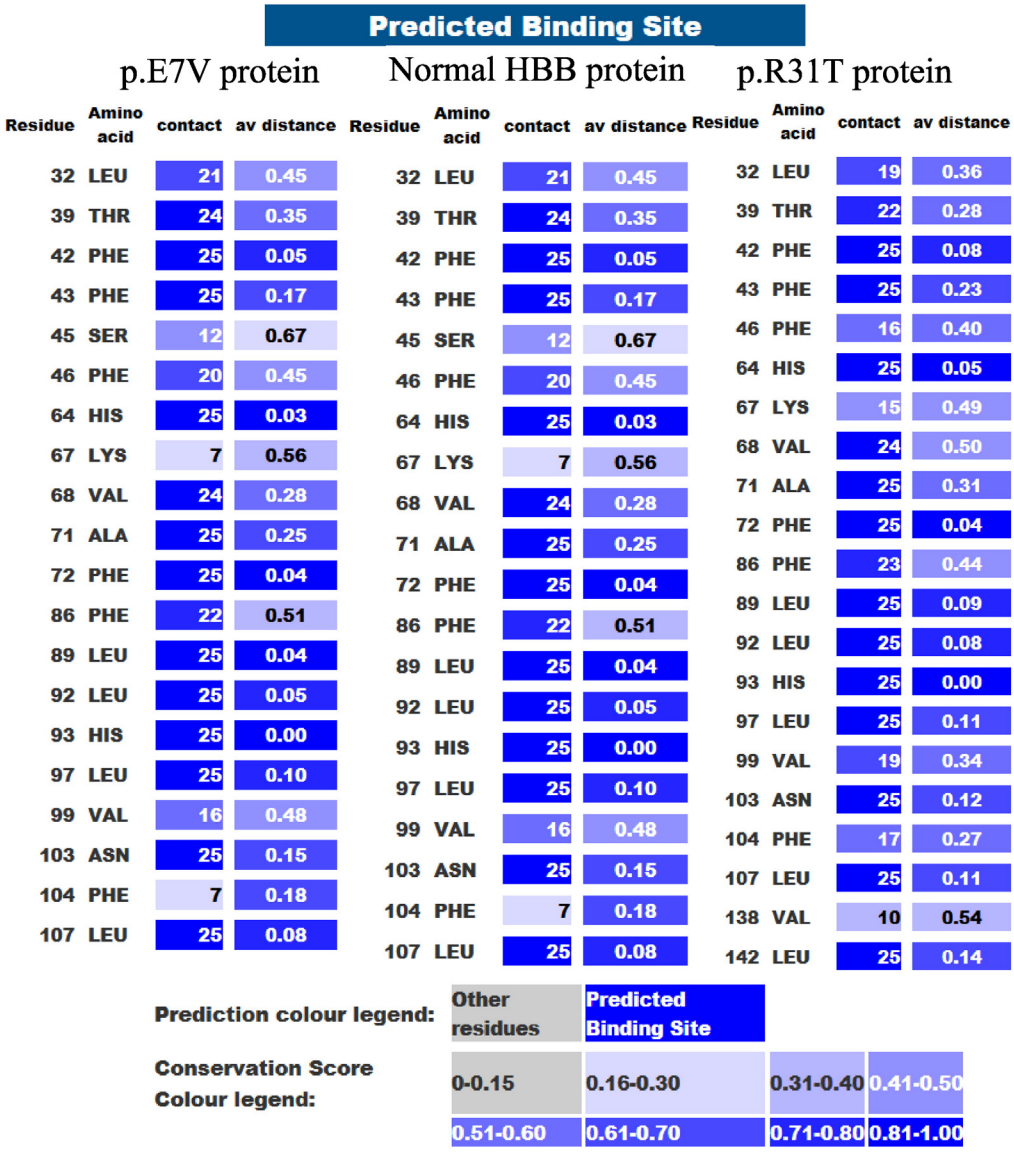

Fig. 6. Phyre2 predicting binding site submits the data to 3DLigandSite server to predict potential binding sites (cluster) in comparison to normal amino acid sequence. These tables list all of the predicted binding-site residues with details of the number of ligands that they contact, the average distance between the residue and the residue conservation score. As illustrated the residues' contact slightly change in p.R31T (right table) in comparison with p.E7V (left table). Number of contacts for each residue may be changed; the average distance from conservation score of each residue (range: 0–1.00) is defined for each residue. The color shows the binding site range. Low-distance shows high accuracy and lower coverage and as the distance increases the accuracy lowers while the coverage increases. The maximum range is 0.68 is obtained at a 0.8 A° distance. The correlation decreases at lower and higher cut offs.

the ontology. Additionally, *in vitro* analysis is needed to investigate the function of each mutation but *in silico* analysis would reveal a fast functional analysis of the new variants.

3.4.4. Interactome analysis

STRING v10.0 server was used to investigate the interaction of HBB with other genes in network system. Though a change in other genes associated within a network may cause phenotypic variability in the function of the HBB protein or vice versa. The following proteins were predicted to interact somehow with HBB as follows: HBA1 (hemoglobinA1), HBA2 (hemoglobin alpha2), AHSP (alpha hemoglobin stabilizing protein), KLF1 (Kruppel-like factor 1 (erythroid)), HP (haptoglobin), HBZ (hemoglobin zeta), HBG2 (hemoglobin, gamma G), NFE2 (nuclear factor (erythroid-derived 2)), AQP1 (aquaporin 1 (Colton blood group)), and HPX (hemopexin) (Fig. 7). A change in B globin structure modifies the function of its protein; therefore, its interaction with other proteins may evolve the phenotype. B globin with unusual or truncated structure lessens the tetramer formation and alters the biological functions of hemoglobin.

3.4.5. Pathogenic analysis of causal variants

Depending on the type of mutation, the probability of pathogenicity of variants was determined using different bioinformatics online software.

3.4.5.1. Missense, nonsense, and indel mutations. In this study, SIFT (≤0.05 pathogen) and PolyPhen2 (0.5–1.5 possibly/probably damaging) were calculated for the p.E7V (cd 6 A > T), and p.R31T (cd30G > C). The p.R31T was damaging by both tools, though p.E7V was benign by polyphen2. SNAP value was exemplified as non-neutral for p.R31T. This suggested that the p.R31T could be deleterious but p.E7V is neutral. Also, SNP&Go predicted the pathogenicity of the p.R31T as diseased with a reliability index of 7 (unreliable, 0–reliable, 10). Pathogenic annotation of amino acid change p.R31T was predicted to be deleterious by all software tools (Table 3A); but the pathogenicity of the p.E7V was predicted just by SIFT, Variation Viewer and MutPred. CADD, MutationTester, and Variation Viewer showed the pathogenic effect of p.R31T mutation (Table 3A). CADD indicates that p.R31T is more

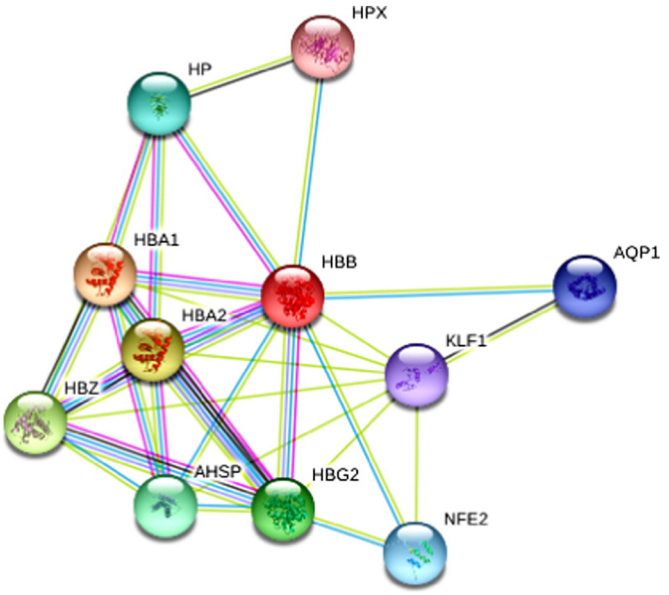

Fig. 7. Protein–protein interaction network of HBB. This protein has a crucial role in oxygen transport from the lung to the various peripheral tissues. HBA1: hemoglobin alpha1; HBG2: hemoglobin, gamma G; NFE2: nuclear factor (erythroid-derived 2); AQP1: aquaporin 1 (Colton blood group); HBA2: hemoglobin alpha2; AHSP: alpha hemoglobin stabilizing protein; KLF1: Kruppel-like factor 1 (erythroid); HP: haptoglobin; HPX: hemopexin; HBZ: hemoglobin zeta. This figure was generated by STRING (V10.0).

deleterious than p.E7V but with a lower scaled score compared to other SNVs in the human genome indicating a rare variant at the top end list of all SNVs (Table 3A).

In addition, the pathogenicity and diseased effect of frameshift mutations were predicted by CADD and MutationTaster, and confirmed in Variation Viewer (Table 3B). All frameshift mutations led to premature protein. Obviously, W16X and Q40X generate stop codons which lead to a truncated protein.

3.4.5.2. Regulatory and intronic mutations. The pathogenic effect of regulatory and intronic mutations was predicted by HSF, Mutation Taster, Variation Viewer, and CADD. To exemplify, *in silico* analysis of all the intronic changes were predicted to create new sites or disrupt the sites as ESE, ESS, intron-identifying elements (IIE), enhancer, and silencer, which affect the splicing process (Table 3C).

Mutation Taster and Variation Viewer showed the pathogenic effect of the mutations except for the regulatory regions that could not be analyzed by MutationTaster (Table 3C). CADD also reveals a high Phred score for intronic positions but lower than the score of the exonic and frameshift changes. As a matter of fact, the calculated *C*-scoreof CADD shows that these SNVs are among the top deleterious mutations (Table 3C, PHRED score). CADD has the ability to evaluate the small insertions, deletions, and regulatory variants throughout the genome as displayed in Table 3C.

4. Discussion

For simplicity, better understanding, and analysis, we have classified the studies into populations and subpopulations. Iran, in the center of the Middle East, is positioned on the ancient Silk Road and has acted as a bridge or meeting place between the Eastern and Western civilizations. The existence of various ethnicities with different cultures, invasions, and historical wars as well as being bordered by more than ten countries are factors responsible for mutation heterogeneity of beta thalassemia and other diseases [124–128]. Nevertheless, specific customs and traditions within some ethnicities such as intragroup marriages could decrease the heterogeneity of the mutations. A total of

Table 3A
Pathogenic prediction of the missense mutations of the *HBB* gene.

Mutation	HGVS nomenclature	dbSNP	Type	CADD		MutationTaster prediction	Variation Viewer	PolyPhen2	nsSNP Analyzer	SIFT	PANTHER	SNAP	SNP &GO	MutPred
				Raw score	PHRED									
Codon30 G > C	c.92G > C	rs33960103	B^0	4.991742	29.3	Disease causing p.R31T	Pathogenic	probably damaging (score 0.988)	Disease	Damaging (0)	Disease	Non-neutral (effect)	Disease	0.906
CD6 (A > T)(HbS)	c.20 A > T	rs334		1.02061	9.169	Polymorphism p.E7V	Pathogenic	Benign (Score 0.001)	Neutral	Damaging (0)	Neutral	Neutral	Neutral	0.844

Note: SNP&GO predicts the mutation relating disease with an accuracy of 82%. SNAP could identify over 80% of the non-neutral mutations at 77% accuracy and over 76% of the neutral mutations at 80% accuracy at its default threshold [43]. MutPred predicts the probability of deleterious mutation by range of (0–10); CADD provides raw and scaled scores. The higher raw values indicate a variant is more likely to be not observed and therefore more possibly to have deleterious consequences. Using the rank of each variant relative to all 8.6 billion SNVs in the reference genome, these researchers determined phred-like scores (−10*log10 (rank/total)) ("scaled *C*-scores") ranging from 1 to 99. A scaled *C*-score ≥ 10 indicates that the variants are predicted to be the 10% most deleterious that you can do to the human genome, a score of greater or equal to 20 indicates the 1% most deleterious and so on.

Table 3B
Pathogenicity of nonsense and frameshift mutations of HBB gene by CADD, MutationTaster, and Variation Viewer. Clinical presentation of the mutations in heterozygous form is also shown.

Mutation	HGVS nomenclature	dbSNP	Type	CADD		Mutation Taster prediction	Variation Viewer	Hematology and clinical presentation (heter)
				Raw score	PHRED			
Cd36/37(−T)	c.112delT	rs63750532	B^0	5.080768	31	Disease causing (W38GfsX24)	Pathogenic	
Fsc 8/9	c.27-28insG	rs35699606	B^0	1.135077	9.626	Disease causing (S10VfsX14)	Pathogenic	Hb A_2 5.9% MCH 20.5 pg MCV 64 fL
Fsc8 (−AA)	c.25-26delAA	rs35497102	B^0	2.205739	13.33	Disease causing (K9VfsX14)	Pathogenic	Hb A_2 4.1–5.1% MCH 16.3–18.8 pg MCV 68–74 fL
FSC44(−C)	c.135delC	rs80356820	B^0	4.806348	27.2	Disease causing (F46LfsX16)	Pathogenic	
Codon39 (C > T)	c.118C > T	rs11549407	B^0	3.365565	17.34	Disease causing (Q40X)	Pathogenic	Hb A_2 4.35–5.35% MCH 19.1–21.3 pg MCV 64.1–81.1 fL
CD5(−CT)	c.17-18delCT	rs34889882	B^0	1.583009	11.25	Disease causing (P6RfsX17)	Pathogenic	Hb A_2 5.2–5.5% MCH 18.8–20.7 pg MCV 68–81 fL
Cd15 (G > A) (TGG-TGA)	c.48G > A	rs34716011	B^0	4.784389	26.9	Disease causing (W16X)	ND	Hb A_2 5.1%); MCH 21 pg MCV 66.5 fL
FSC-22-24 (−AAGTTGG)	c.68-74delAAGTTGG	rs281864898	B^0	5.041626	29.9	Disease causing (E23VfsX37)	ND	

14,293 affected carriers and their families were investigated in Iran. Also, a total of 16,583 cases were studied from neighboring countries of Iran. Over the past decades, several studies have been performed on B thalassemia patients and all the reported mutations were included in this cohort.

HBB gene mutations arise from different functional mutations of the gene. Of these, we investigated the 20 most frequently reported mutations. This study provides a large number of beta thalassemia patients in the region with complete evaluation of pathogenic variations and their correlation to phenotypes (Fig. 1). The clinical evaluation was assigned by clinical presentation, hematological analysis, familial studies, globin chain synthesis ratios, and physicians' confirmations.

Mutation diversity in Iranian cohorts of border regions is similar to neighboring countries of that area, suggesting historical immigration and emigration of these populations. Overall, the ranking of mutations in subpopulations are similar in the neighboring countries as shown (Fig. 2) in Turkey, Iraq, Oman, and other countries.

Different pathogenic *HBB* variations and their frequencies were gathered. This kind of data is valuable for the analysis of heterogenous and homogenous populations and subpopulations and their comparison. The other reason is that this data may be used as a panel of mutations for this region of the world to choose the most cost-effective strategy for screening patients. The prevalent form of thalassemia with a given genotype could help health providers and physicians in

Table 3C
The bioinformatic analyses of the pathogenicity of regulatory, splicing, and intronic mutations of HBB gene. Clinical presentation of the mutations in heterozygous form is also shown.

Mutation	HGVS nomenclature	dbSNP	Type	CADD		HSF	Mutation Taster prediction	Variation Viewer	Hematology and clinical presentation
				Raw score	PHRED				
IVSI-5(G > C)	c.92 + 5G > C	rs33915217	B^+	0.62489	7.361	IIEs and ESS site broken Silencer motif new site	Disease causing	Pathogenic	Hb A 5%–8%
IVS II-1 (G > A)	c.315 + 1G > A	rs33945777	β^0	2.32385	13.73	IIEs, and ESE new site ESS and Silencer Site broken	Disease causing	Pathogenic	Hb A_2 4.2%–5.6% MCH 18.5–22.1 MCV 63.4–80.4 fL
IVS I-110(G > A)	c.93-21G > A	rs35004220	B^+	0.327489	5.776	IIEs, ESE, silencer and enhancer site broken hnRNP protein new site	Disease causing	Pathogenic	
IVS I-1 (G > A)	c.92 + 1G > A	rs33971440	β^0	4.161918	21.5	ESS and ESE site broken IIEs and silencer new site	Disease causing	Pathogenic	Hb A_2 4.75%–5.55% MCH 18.7–20.5 pg MCV 62.1–77.9 fL
IVSI −25 bp	c.93-21_96del	rs63750223	B^0	2.407984	14.01			ND	Hb A_2 5.2% MCH 19.5 pg MCV 74.5 fL
IVS I-6 (T > C)	c.92 + 6 T > C	rs35724775	B^+	1.033218	9.22	ESE and enhancer new site IIEs site broken	Disease causing	Pathogenic	Hb A_2 3.35%–4.45% MCH 20.6–24 pg MCV 64.7–77.3 fL
IVS-I-5 (G > A)	c.92 + 5G > A	rs33915217	B^+	0.62489	7.361	ESE new site ESS site broken	Disease causing	Pathogenic	Hb A_2 2%–5% MCH 20–22 pg MCV 71–79 fL (homo)
IVS-II-745, G > C	c.316-106C > G	rs34690599	B^+	−0.445053	1.935	IIEs and ESE site broken hnRNP protein and silencer new site	Disease causing	Pathogenic	Hb A_2 4.4%–5.4% MCH 19.1–21.7 pg MCV 64.9–76.5 fL
−88 (C > A)	c.-138C > A	rs33944208	β^+	0.820753	8.305			Pathogenic	
−28 A > C	c.-78 A > C	rs33931746	β^+	4.316974	22.6			Pathogenic	

Exonic splicing enhancers (ESEs); exonic splicing silencers (ESSs); intronic splicing enhancers (ISEs); intronic splicing silencers (ISSs); intronic identifying elements (IIE).
In HSF server, the exonic splicing enhancers (ESEs) are specific short nucleotide sequences targeted usually by serine/arginine-rich (SR) proteins. The spliceosome can ignore pseudo-exons and decoy splice sites using exonic splicing silencers (ESSs). ESSs provide binding sites for proteins promoting exon exclusion (mainly hnRNP proteins). Intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs) are also intronic *cis*-elements playing similar roles of ESEs and ESSs.

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001

building a framework for the country. This will assist in premarital and carrier counseling.

It seems that there may be some bias due to incomplete analysis of some reports; this may be due to methodology used in each study. Most studies used ARMS-PCR and RFLPs for specific analysis of patients and other mutations could have been missed though sequencing and complete analysis of the gene in some studies were needed for complete evaluation. Complete Sanger sequencing of the gene is helpful for phenotype genotype analysis of patients. The critical distinction between our data and neighboring studies may be due to subpopulational analysis and detailed survey of the affected individuals.

The clinical significance of mutations with a given phenotype was explored to compare the variability of the reported data. One of the common mutation is IVSI-5(G > C) [HBB:c.92 + 5G > C], located at the 5′ spice site and because of the high degree of conservation of the surrounding region, +5G > C leads to the abolition of mRNA splicing and is clinically important [129]. This mutation is the most common in the eastern part of Iran (Fig. 2), accounting for 76.1% of mutations. IVSI-5(G > C) is also the most common mutation in Pakistan (>38%), Oman (>44%), UAE (>56%), and Qatar (>35%) (Table 2; Fig. 2) [109,111,113–118,130]. Sinha et al. studied 8505 alleles from India and they found that a total of 52 mutations account for 97.5% of all beta-thalassemia alleles and IVSI-5(G > C) is the most common mutation (responsible for 54.7% of mutations) [131]. The frequency of this mutation is low in western Iran. In Turkey and Iraq, its prevalence is decreased by less than 1% and 5%, respectively [92,93,97,100,132]. Its distribution, thus, shows an east-to-west gradient. Also, the southern part of Iran has a high frequency of this allele, which may be due to migrations from Persian Gulf countries. The origin of the mutation is probably one of these neighboring countries and then distributed to others during migration; several studies have described high frequency (58%–72%) of the mutation in southern and eastern India, suggesting its ancient origin from these regions [133]. One study on UAE haplotypes showed an independent origin of this mutation from this population [134]. Iran's eastern border with Pakistan through land and with Oman through the sea; Sistan va baloochestan of Iran and Khorasan is bordered by Pakistan and Afghanistan. This mutation probably occurred for the first time in one of these countries and then spread to others. After that, intramarriages, genetic drift, and other phenomena may affect it so that its frequency has been increased. One important point is that this migration has been done through southeastern Iran. Furthermore, up to now, no study has been published about the HBB mutations in Afghanistan, our study suggests that the type and frequency of HBB mutation in this country might be similar to those mutations observed in northeastern Iran. A total of 4094 carriers had minor phenotypes though 403 alleles were found in major/intermedia patients. It causes B^+ (reduced protein synthesis) thalassemia.

IVS II-1 (G > A) [HBB:c.315 + 1G > A] as beta mutation abolishes the 5′ splicing site. When homozygous, it causes transfusion dependent beta-thalassemia major and in heterozygous form changes hematological factors are as follow Hb 10.95 ± 1.65; MCV 71.9 ± 8.5 fL; MCH 20.3 ± 1.8; Hb A_2 4.9 ± 0.7%; Hb F 5.75 ± 4.55% [135]. Our study shows that this mutation has the highest frequency (57.44%) in northern Iran (Table 1, Fig. 2) [71]. The frequency of the mutation is decreased from north to the south of Iran [71,85,136]. It also has high frequency in the west and southwest of Iran whereas it decreases to the east. The lowest frequency (2.54%) is observed in the eastern region, but the lowest frequency (6%) of IVSII-1 has been reported in Kerman Province (southern region) [61]. A total of 1186 alleles of homozygous (major) group had IVSII-1(G > A). IVS II-1 (G > A) is the most common mutation in Iraq and Kuwait and is relatively high in Azerbaijan. It seems that this mutation may have arisen from northern Iran then moved to other populations. This mutation is observed in some small countries like Kuwait [102,103] with high prevalence; most likely intramarriage in addition to effects of genetic drifts lead to the high prevalence of the mutation. Further studies are required to determine the ancient origin of IVS II-1 (G > A) and an accurate estimate of the time when it arose.

IVS-I-110 (G > A) as a $beta^+$ mutation and one of the most common mutations in Mediterranean countries, occurred in 21 nucleotides 5′ to the acceptor splice site (AG…GC); it produced a new splice site resulting in an 80% abnormal spliced mRNA and 20% normal mRNA [135,137,138]. When homozygous, it leads to almost invariably transfusion-dependent beta-thalassemia major and, if heterozygous, hematological factors are as follows: Hb 11.85 ± 1.8 g/dl; MCV 69.5 ± 6.9 fL; MCH 20.9 ± 2.0 fL; Hb A_2 4.65 ± 0.5%; Hb F 1.2 ± 1.15% [139]. IVS-I-110 (G > A) is the most common mutation in Cyprus, Turkey, Greece, and Albania [92–94,96,135,139–143]. Zahed et al. studied this mutation in Lebanese subjects and concluded that IVS-I-110 (G > A), following its emergence in Turkey, was probably later introduced into Lebanon by migration or settlements [31]. This mutation shows a high frequency in northwestern Iran in the vicinity of Turkey and Azerbaijan, suggesting population immigrations among these populations during old times. It seems therefore that this mutation flowed to central parts of Iran with the migration of peoples.

Each of the Cd36/37(−T) and Fsc 8/9 mutations with a frequency of about 5% are the next commonest mutations in Iran. Cd36/37(−T) is more frequent in the southwest (15.7%) and west (8.3%) of the country. While the presumed origin of Cd36/37(−T) mutation is from Kurdish population, the mutation showing an east-to-west gradient is the most common mutation in Khozestan and Lorestan provinces [80,85] but is less common in east Iran [66]. If so, consequences of gene flow and genetic drift may increase its frequency in subpopulations of the latter provinces. FSC 8/9 [HBB:c.27_28insG] mutation is distributed in different parts of Iran, with a high frequency in the northwest and western regions, 21.23% and 14.3%, respectively (Table 2) [73,74,77,78]. FSC 8/9 is one of three common mutations in Pakistan and India [130,144]. It is also high in UAE and Qatar; that may be a gene flow from the south and southeast to the northwest regions of Iran.

The following mutations have been reported in all areas: IVSI-1 (G > A), IVSI-5 (G > C), IVSI-25 bp, IVS I-6(T > C), IVS I-110(G > A), IVS II-1(G > A), IVS II-745(G- > C), FSC5(−CT), FSC8(−AA), FSC8/9(+G), CD15(G > A)(TGG-TGA), FSC36/37(−T), and FSC44(−C). The frequency of some of them, including FSC8/9(+G), FSC44 (−C), and IVSI-1(G > A), has a usually constant rate among different populations; these points might be considered as hotspots.

CD6 [HBB:c.20 A > T] mutation lead to E7 amino acid substitution with valine (p.E7V); HbS occurs at a high frequency in different areas such as sub-Saharan Africa, parts of the Mediterranean region, the Middle East, and certain regions of India. It has very low frequency (0.5%) in Iranian populations but is high (25%) in Oman. CD30 [HBB:c.92G > C] or p.R31T, CD15G > A or c.48G > A, and CD40C > T or c.118C > T (nonsense) mutations are very rare in Iran and neighboring countries.

Our goal is to improve management and counseling of families. To achieve this, we need to evaluate the clinical status of affected individuals and correlate this with their genotype to build a comprehensive knowledge base in the relevant communities of Iran and the Middle East.

4.1. Clinical presentations of common mutations

The $beta^+$ and $beta^0$ mutations lead to the reduced or absent synthesis of the beta chain, respectively. Impaired B-globin synthesis leads to imbalanced proportion of beta and alpha globin chains which the excess of a-chain aggregates in RBC precursors, and causes abnormal cell maturation and their premature destruction in the bone marrow. Anemia thus occurs as a definitive consequences of this abnormal process which leads to splenomegaly, bone disease, and endocrine and cardiac implications. Based on the severity and the mutation type, beta thalassemia generally may be observed in one of three clinical and hematological forms including the beta-thalassemia carrier state (trait or minor or heterozygote), thalassemia intermedia, and thalassemia major (a severe

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001

transfusion-dependent anemia) [145]. Heterozygotes are clinically asymptomatic and have specific hematological features (92%–95% HbA; >3.8 HbA_2; and variable amounts of HbF (0.5%–4%)). Thalassemia intermedia is categorized as a clinically and genotypically heterogeneous group of thalassemia, ranging from the asymptomatic carrier state to the severe transfusion-dependent type; it is due to the coinheritance of homozygosity or compound heterozygosity for mild beta-thalassemia alleles. The rate of imbalance between the globin chains determines the clinical severity of beta-thalassemia.

RBC indices in normal individuals are MCV (male: 89.1 ± 5.01, female: 87.6 ± 5.5 fL), MCH (male: 30.9 ± 1.9, female: 30.2 ± 2.1 pg) and Hb (male: 15.9 ± 1.0, female: 14.0 ± 0.9 g/dL). The quantities of these indices are changed in beta-thalassemia as follows: MCV (50–70 fL for major and <79 fL for carrier state), MCH (12-20 pg for major and <27 for minor) and Hb (<7 g/dL for major and 9.1–15.3 for minor).

All common nonsense and frameshift mutations of the *HBB* gene are categorized as B^0 type. More severe clinical and hematological consequences are expected to be observed in individuals carrying these mutations. Hematological and RBC indices in heterozygotes of these mutations are Hb A_2 = 4.1–5.9%, MCH = 16.3–21.3 pg, and MCV = 64–81 fL.

Depending on the severity of intronic mutation, these indices may show variable quantities. Analysis by MutationTaster2 predicted that the common intronic mutations are disease causing [47]. IVS II-1 (G > A), IVS I-1 (G > A), and IVS I-25 bp are B^0 type mutations and affect these variables more than other intronic mutations. IVS I-25 bp disrupts RNA processing and in turn leads to a clinical phenotype [146]. CADD values for these positions show higher Phred scores (>10) (Table 3C).

Direct genotype–phenotype correlation of the available data (RBC indices and hemoglobin patterns) for the selected 20 mutations came to the close relation to the phenotype of major and minors; there was some difficulty with intermedia individuals which were not specified in different studies for detailed analysis.

Genotype–phenotype analysis expands our knowledge in predicting phenotypic traits and to investigate important genetic effects of the heritability of a trait. However, the pathogenicity of the traits remains unexplained due to the single focus of the study designs. A system biology approach would model based on different levels of genetic, genomic, proteomic, transcriptome, methylomic, and metalobomic data to underpin the phenotype [147,148].

Genotypes and phenotypes were not specifically distinguished in each study and therefore it was difficult to conclude the genotype correlation in all studies. Also, we suggest that in research programs for patients in whom there is a discordance of phenotypes and genotypes, there may be benefit of whole exome sequencing to enroll the genes involved. Since there may be other effective genes modifiers, etc., for causing disease with different severity.

4.2. In silico analyses

With advances in sequencing technology and improvement of variant calling, it is very difficult to deal with structural and functional analysis in parallel. With more than 800 mutations for the *HBB* gene, the biochemical analysis of all these mutations is demanding; however, not all of them have been characterized and the clinical consequences remain unsolved. *In silico* analysis made it straightforward [149]. Mutations that disrupt oxygen transport activity, heme binding, hemoglobin binding, iron–iron binding, and oxygen binding sites cause loss of function of globin and subsequently hemoglobin protein. Mutations causing leaky amounts of beta globin production cause minor or B^+ type effect.

The secondary structure of B globin protein is helix with main molecular heme, hemoglobin, iron, oxygen binding sites, and oxygen transporter activity. This protein functions in tetramer. Any variant in the gene sequence may affect the amino acid sequence, expression, and affect protein function. The stability may be changed due to conformational and folding positions at mRNA and consequently at protein level. The *structural* studies, at the DNA level, were investigated for missense mutations by I-TASSER and Phyre2. The servers could be used for the prediction of amino acid changes in the protein as it specifies residual amino acid changes. The predicted structural data could help practitioners and geneticists to assign phenotypes to novel variants.

We document structural and pathogenic aberrations computationally in previously characterized and uncharacterized mutations on the basis of changes that each brings to the protein to evaluate the confidence of these tools on the novel mutations. Our data demonstrates that we could predict the pathogenicity of a variant in coding regions based on structural changes. Prediction of protein stability changes upon single point mutations was investigated throughout structural and functional analysis of p.R31T and p.E7V. The frameshift mutations led to premature protein termination with no function, predicted by CADD and MutationTaster, and confirmed with the phenotypic consequences at Variation Viewer. The native and altered globin amino acid sequences were used to determine the secondary structures to construct models. The solvent accessibility was determined to investigate the protein stability. The dynamic models were compared before and after mutation to evaluate the altered and native protein models. The consequences of these structural alterations were explored in the function of protein for biological functions (COACH), ligand binding sites, enzyme activity, gene ontology, and binding sites to evaluate the changes. In fact, slight changes were based on gene ontology with missense mutations which could not be directly reliable. The functional study in this investigation may be a good model for additional future studies.

Generally, point mutations affecting the b-globin transcription (e.g. promoter and 5′UTR mutations) are mild beta + mutations and were analyzed with bioinformatic software to calculate the risk of pathogenicity. Those affecting splicing process including splice junctions, polyadenylation, consensus sequences, and other 3′UTR mutations cause both silent and beta + mutations. Mutations in the coding regions harbor nonsense, frameshift, and initiation codon cause mild or B^0 mutations. Mostly, deletion, insertion, initiation codon, frameshift, and splicing junction mutations cause B^0 mutations. Interestingly, frameshift and intronic mutations have higher frequencies than other functional mutations in this cohort, respectively.

We found that the predictability of *in silico* analysis for amino acid substitutions was easier to investigate than the intronic and noncoding variants since more software were available. The predication of pathogenicity of intronic sites leads to detailed investigation regarding the phenotypic diversity in beta thalassemia. Data analysis by mutation taster and CADD could demonstrate the pathogenicity of the variants. CADD and MutationTaster showed a high confidence value for the probability to predict the extent of pathogenic effect of unknown intronic variants. However, CADD could define the SNV in the regulatory and downstream gene alterations. Introns (intervening sequences, IVS) are removed during the process of RNA splicing. Splice sites located at the 5′ and 3′ ends of introns and branch site partly define the specificity of splicing. ESSs, ESEs, and cis-regulatory elements regulate the use of adjacent splice sites. These elements recruit protein factors such as the serine/arginine-rich (SR) protein family interacting favorably or unfavorably with components of the core splicing machinery [150]. Any nucleotide change in these sequences, therefore, may affect the efficiency of RNA processing. Bioinformatic analyses showed that the given intronic mutations of *HBB* gene are harmful and pathogenic sequences or create such a sequence (HSF in Table 3C). IVSI-25 bp is a 25-nucleotide deletion involving the IVS-1 acceptor splice site; it disrupts processing of RNA sequences and leads to the formation of low levels of RNA containing an intact IVS-1 [146].

A simple *in silico* evaluation of uncharacterized gene mutations could thus potentially help predict the pathogenicity of a variant. This makes it easier and faster to predict the effect of new and novel variants and could inform the families at risk of beta thalassemia. Our data suggests that the mutations in the noncoding regions of *HBB* gene may be responsible for some of the phenotypes which could be scrutinized

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001

with bioinformatic servers (Table 3C). Likewise, the mutations not reported in SNP database should be investigated if found in thalassemic individuals. However, this prediction should be confirmed by segregation or expression analysis. Identification of the pathogenic mutations allows confirmation of a defect in individuals and genetic testing and counseling of other high-risk family members. Also, finding a pathogenic variant in a population makes it easier to relate the mutation in others clinically affected with variable phenotypic expression. Of course, it is notable that there is a wide diversity of the phenotypic presentation and exact prediction of phenotype with such analysis may not be effective. Though we could only predict the pathogenicity of the mutations with *in silico* analysis.

4.3. Interactome network of proteins

A wide phenotypic diversity has been observed even among individuals with the same beta-thalassemia genotype. Modifier variants modulate and modify the clinical phenotype. There are a cluster of β-chain variants of hemoglobin at the mouse β-globin region; this cluster is developmentally regulated by multiple elements spanning a region of 100 kb including the locus control region (LCR). LCR acts as a super enhancer during the different ontological periods. Furthermore, specific transcription factors including erythroid transcription factor (GATA1) and Kruppel-like factor 1 (KLF1) are needed for the expression of *HBB* gene. An extensive transcription interactome exists in erythroid cells. As the interactome analysis showed, the interaction is seen in ten proteins including HBA_1 and HBA_2, which are mainly involved in oxygen transport [38]. AHSP protein as a chaperone prevents the damaging aggregation of alpha-hemoglobin during normal erythroid cell development. It specifically protects free alpha-hemoglobin from precipitation [151]. The KLF1 mediates preferential aggregation of its own regulated genes by binding to enhancer and promoter elements [152]. Notably, a change in the beta globin would have diverse effects on the interacted protein and factors, i.e. transcription, translation, and development and function of hemoglobin and consequently the phenotype.

5. Conclusion

In particular, we have analyzed all reported mutations defined in this region of the world that have high incidence. This is valuable for the analysis of panel-based studies in the region that may have no access to advanced genetic technologies. To identify a specific panel of the most frequent mutations, such investigations are needed. Difficulty of defining the clinical significance of a new mutation is also important. Comparison of the previously defined pathogenic value could increase our dependence on different servers to define the pathogenicity for interim reporting and decision making for family members. This article will be of special interest for performing a prenatal diagnosis of beta thalassemia. It is used to identify the fetal phenotype but will not predict it's phenotype with high accuracy for unknown variants. In case of novel variants, computational analysis could be used in addition to segregation and expression analysis to assess the extent of pathogenicity of the variant. The software could be used to predict the phenotype but for prenatal diagnosis, careful analysis is necessary for at-risk patients. Those mutations with known functional consequences and clinical phenotype give high confidence for genetic counseling of at-risk families to know the probability of the affected child. The families have the right to undergo different decisions, e.g. termination, treatment of patients, or preimplantation genetic diagnosis (PGD).

5.1. Practice points

- Genetic screening tests for common mutations in the region
- The use of bioinformatic tools for assessment of pathogenicity

5.2. Research agenda

- Regional distribution of *HBB* mutations
- Frequency of *HBB* gene mutations
- *In silico* analysis of *HBB* common mutations
- Efficacy and reliability of bioinformatic software for fast analysis of the pathogenicity

Conflict of interest

The authors have no conflicts of interest to declare in relation to this manuscript.

Acknowledgments

We would like to thank Professor Edward Tuddenham (University College London) for helpful comments on the draft manuscript. We are thankful to Dr. Hirofumi Nakaoka for *in silico* analysis using CADD server.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.blre.2016.07.001.

References

[1] Thein SL. Dominant thalassaemia molecular basis and pathophysiology. Br J Haematol 1992;80:273–7.
[2] Birgens H, Ljung R. The thalassaemia syndromes. Scand J Clin Lab Invest 2007;67:11–25.
[3] Abolghasemi H, Amid A, Zeinali S, Radfar MH, Eshghi P, Rahiminejad MS, et al. Thalassemia in Iran: epidemiology, prevention, and management. J Pediatr Hematol Oncol 2007;29:233–8.
[4] Galanello R, Origa R. Beta-thalassemia. Orphanet J Rare Dis 2010;5:11.
[5] Modell B, Khan M, Darlison M, King A, Layton M, Old J, et al. A national register for surveillance of inherited disorders: beta thalassaemia in the United Kingdom. Bull World Health Organ 2001;79:1006–13.
[6] Cousens NE, Gaff CL, Metcalfe SA, Delatycki MB. Carrier screening for beta-thalassaemia: a review of international practice. Eur J Hum Genet 2010;18:1077–83.
[7] Cooley TB, Lee P. A series of cases of splenomegaly in children with anemia and peculiar bone changes. Trans Am Pediatr Soc 1925;37:29–30.
[8] Low LC. Growth of children with beta-thalassemia major. Indian J Pediatr 2005;72:159–64.
[9] Peschle C, Mavilio F, Migliaccio G, Migliaccio AR, Russo G, Mastroberardino G, et al. Erythropoietic development and hemoglobin switching in human embryos: cellular and molecular aspects. Prog Clin Biol Res 1985;191:383–96.
[10] Marengo-Rowe AJ. Structure–function relations of human hemoglobins. Proc (Baylor Univ Med Cent) 2006;19:239–45.
[11] Fritsch EF, Lawn RM, Maniatis T. Molecular cloning and characterization of the human beta-like globin gene cluster. Cell 1980;19:959–72.
[12] Spritz RA, DeRiel JK, Forget BG, Weissman SM. Complete nucleotide sequence of the human delta-globin gene. Cell 1980;21:639–46.
[13] Baralle FE, Proudfoot NJ, Clegg JB. The structural analysis of the human epsilon-globin gene and its product. Ann N Y Acad Sci 1980;344:76–82.
[14] Slightom JL, Blechl AE, Smithies O. Human fetal G gamma- and A gamma-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 1980;21:627–38.
[15] Talbot D, Collis P, Antoniou M, Vidal M, Grosveld F, Greaves DR. A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. Nature 1989;338:352–5.
[16] Collins FS. Identifying human disease genes by positional cloning. Harvey Lect 1990;86:149–64.
[17] Giardine B, Borg J, Viennas E, Pavlidis C, Moradkhani K, Joly P, et al. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic Acids Res 2014;42:D1063–9.
[18] Huisman THJ, Carver MFH, Baysal E, Efremov GD. http://globin.bx.psu.edu/hbvar/menu.html. [(A database of human hemoglobin variants and thalassemias)] 2014.
[19] Patrinos GP, Giardine B, Riemer C, Miller W, Chui DH, Anagnou NP, et al. Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. Nucleic Acids Res 2004;32:D537–41.
[20] Guiso L, Frogheri L, Pistidda P, Angioni L, Dore F, Pardini S, et al. Frequency of delta$^+$27-thalassaemia in Sardinians. Clin Lab Haematol 1996;18:241–4.
[21] Lukens JN. The Thalassemias and Related Disorders, Quantitative Disorders of Hemoglobin Synthesis. Philadelphia: Lea & Febiger; 1993.
[22] Haghshenas M, Zamani J. Thalassemia. , 4Shiraz University of Medical Sciences Publishing Center; 1997 50–68[in Persian].

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001

[23] Loukopoulos D, Kaltsoya-Tassiopoulou A, Fessas P. Thalassemia control in Greece. Birth Defects Orig Artic Ser 1988;23:405–16.
[24] Weatherall D. 2003 William Allan Award address. The thalassemias: the role of molecular genetics in an evolving global health problem. Am J Hum Genet 2004;74:385–92.
[25] Parikh P, Cotton M, Boehm C, Kazazian Jr HH. Ethnic distribution of beta-thalassaemia in Indian subcontinent. Lancet 1990;336:1006.
[26] Galehdari H, Salehi B, Pedram M, Oraki Kohshour M. High prevalence of rare mutations in the beta globin gene in an ethnic group in Iran. Iran Red Crescent Med J 2011;13:356–8.
[27] Verma IC, Kleanthous M, Saxena R, Fucharoen S, Winichagoon P, Raizuddin S, et al. Multicenter study of the molecular basis of thalassemia intermedia in different ethnic populations. Hemoglobin 2007;31:439–52.
[28] Rahim F, Abromand M. Spectrum of ß-thalassemia mutations in various ethnic regions of Iran. Pak J Med Sci 2008;24:410–5.
[29] Miri-Moghaddam E, Zadeh-Vakili A, Nikravesh A, Sistani SS, Naroie-Nejad M. Sistani population: a different spectrum of beta-thalassemia mutations from other ethnic groups of Iran. Hemoglobin 2013;37:138–47.
[30] Zahed L, Talhouk R, Saleh M, Abou-Jaoudeh R, Fisher C, Old J. The spectrum of beta-thalassaemia mutations in the Lebanon. Hum Hered 1997;47:241–9.
[31] Zahed L, Demont J, Bouhass R, Trabuchet G, Hanni C, Zalloua P, et al. Origin and history of the IVS-I-110 and codon 39 beta-thalassemia mutations in the Lebanese population. Hum Biol 2002;74:837–47.
[32] Samavat A, Modell B. Iranian national thalassaemia screening programme. BMJ 2004;329:1134–7.
[33] Hijikata A, Raju R, Keerthikumar S, Ramabadran S, Balakrishnan L, Ramadoss SK, et al. Mutation@A Glance: an integrative web application for analysing mutations from human genetic diseases. DNA Res 2010;17:197–208.
[34] Kelley LA, Sternberg MJ. Protein structure prediction on the web: a case study using the Phyre server. Nat Protoc 2009;4:363–71.
[35] Wass MN, Barton G, Sternberg MJ. CombFunc: predicting protein function using heterogeneous data sources. Nucleic Acids Res 2012;40:W466–70.
[36] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010;5:725–38.
[37] Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 2013;29:2588–95.
[38] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein–protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013;41:D808–15.
[39] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073–81.
[40] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9.
[41] Saunders CT, Baker D. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. J Mol Biol 2002;322:891–901.
[42] Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 2013;8:1551–66.
[43] Bromberg Y, Yachdav G, Rost B. SNAP predicts effect of mutations on protein function. Bioinformatics 2008;24:2397–8.
[44] Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 2009;30:1237–44.
[45] Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 2009;25:2744–50.
[46] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.
[47] Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014;11:361–2.
[48] Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009;37, e67.
[49] Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 2004;11:377–94.
[50] Najmabadi H, Pourfathollah AA, Neishabury M, Sahebjam F, Krugluger W, Oberkanins C. Rare and unexpected mutations among Iranian beta-thalassemia patients and prenatal samples discovered by reverse-hybridization and DNA sequencing. Haematologica 2002;87:1113–4.
[51] Behfar M, Ehsani M, Salamati P, Holakouie Naieni K, Jamshidi R, Derakhshandeh-Peykar P. Associations of red blood corpuscle mean volume and hematocrit with severity of beta-globin gene mutations in beta-thalassemia carriers. J Sch Public Health Inst Public Health Res 2011;8:41–9.
[52] Arab A, Karimipoor M, Rajabi A, Hamid M, Arjmandi S, Zeinali S. Molecular characterization of beta-thalassemia intermedia: a report from Iran. Mol Biol Rep 2011;38:4321–6.
[53] Tafazoli M, Nourmohammadi I, Zaker F, Kheradmankia S. Survey of prevalence of beta-globin gene mutations of beta thalassemia in province of Ghazven. Koomesh 2005;6:255–8.
[54] Sarookhani MR, Ahmadi MH, Amirizadeh N. Rare and unexpected beta thalassemic mutations in Qazvin Province of Iran. Razi J Med Sci 2009;16.
[55] Marashi SJ, Ataollahi Eshkoor S, Mirinargesi MS, Sarookhani MR, Rahmat AB, Ismail PB. Detection of eight common β-globin gene mutation in thalassemia major patients using real time polymerase chain reaction (PCR)-high resolution melting and EvaGreen™ dye. Afr J Biotechnol 2012;11:448–59.
[56] Derakhshandeh-Peykar P, Hourfar H, Heidari M, Kheirollahi M, Miryounesi M. The spectrum of β-thalassemia mutations in Isfahan Province of Iran. Iran J Public Health 2008;37:106–11.
[57] Salehi R, Fisher CA, Bignell PA, Eslami G, Old JM. Identification of three novel mutations [−41 (A > C), codon 24 (−G), and IVS-I-109 (−T)], in a study of beta-thalassemia alleles in the Isfahan region of Iran. Hemoglobin 2010;34:115–20.
[58] Rahiminejad MS, Zeinali S, Afrasiabi A, Valeshabad AK. beta-Thalassemia mutations found during 1 year of prenatal diagnoses in Fars Province, Iran. Hemoglobin 2011;35:331–7.
[59] Bordbar M, Haghpanah S, Afrasiabi A, Dehbozorgian J, Karimi M. Genotype–phenotype correlation related to lipid profile in beta-thalassemia major and intermedia in southern Iran. J Clin Lipidol 2012;6:108–13.
[60] Rahimi Z, Vaisi Raygani A, Merat A, Haghshenass M, Gerard N, Nagel RL, et al. Thalassemic mutations in southern Iran. Iran J Med Sci 2006;31:70–3.
[61] Saleh-Gohari N, Bazrafshani M. Distribution of beta-globin gene mutations in thalassemia minor population of Kerman Province, Iran. Iran J Public Health 2010;39:69–76.
[62] Nikuei P, Hadavi V, Rajaei M, Saberi M, Hajizade F, Najmabadi H. Prenatal diagnosis for beta-thalassemia major in the Iranian province of Hormozgan. Hemoglobin 2008;32:539–45.
[63] Hamzehloei T, Mohajer Tehran F. The spectrum of mutations in 100 thalassemic carriers referred to Ghaem Hospital of Mashhad. Iran J Ped Hematol Oncol 2012;2:49–53.
[64] Miri-Moghaddam E, Zadeh-Vakili A. Profile of beta-thalassemia and its prenatal diagnosis in Khorasan-e-Jonobi Province, Iran. Hemoglobin 2012;36:456–63.
[65] Miri Moghadam E, Taroovi Nejad M, Eshghi P, Zeinali S, Savadkoohi F. Molecular basis and prenatal diagnosis of B-thalassemia in southeast if Iran. J Mazandaran Univ Med Sci 2005;15:105–11.
[66] Eshghi P, Zadeh-Vakili A, Rashidi A, Miri-Moghaddam E. An unusually frequent beta-thalassemia mutation in an Iranian province. Hemoglobin 2008;32:387–92.
[67] Miri-Moghaddam E, Zadeh-Vakili A, Rouhani Z, Naderi M, Eshghi P, Khazaei Feizabad A. Molecular basis and prenatal diagnosis of beta-thalassemia among Balouch population in Iran. Prenat Diagn 2011;31:788–91.
[68] Derakhshandeh-Peykar P, Akhavan-Niaki H, Tamaddoni A, Ghawidel-Parsa S, Naieni KH, Rahmani M, et al. Distribution of beta-thalassemia mutations in the northern provinces of Iran. Hemoglobin 2007;31:351–6.
[69] Modjtahed zadeh F. Beta thalassemia gene mutations in thalassemic patients referred to Boo Ali Sina Hospital o Sari in the year 1373. J Mazandaran Univ Med Sci 1999;9:32–7.
[70] Hashemi Soteh M, Akhavan Niaki H, Kowsarian M, Aliasgharian A, Banihashemi A. Frequency of beta-globin gene mutations in beta-thalassemia patients from east of Mazandaran. J Mazandaran Univ Med Sci 2008;18:17–25.
[71] Akhavan-Niaki H, Derakhshandeh-Peykar P, Banihashemi A, Mostafazadeh A, Asghari B, Ahmadifard MR, et al. A comprehensive molecular characterization of beta thalassemia in a highly heterogeneous population. Blood Cells Mol Dis 2011;47:29–32.
[72] Valizadeh F. Frequency of beta-globin gene mutations in beta-carrier couples in Babolsar, Iran, 2001–2011. J Mazandaran Univ Med Sci 2014;23:17–23.
[73] Hossein Pour Feizi MA, Hossein Pour Feizi AA, Asghar Zadeh M, Amin Bakhsh M, Azarfam P, Pouladi N. Common B-thalassemia mutations in northwestern Iran. J Shahid Sadoughi Univ Med Sci 2005;12 [44–39].
[74] Hosseinpour Feizi MA, Hosseinpour Feizi AA, Pouladi N, Haghi M, Azarfam P. Molecular spectrum of beta-thalassemia mutations in northwestern Iran. Hemoglobin 2008;32:255–61.
[75] Haghi M, Khorshidi S, Hosseinpour Feizi MA, Pouladi N, Hosseinpour Feizi AA. Beta-thalassemia mutations in the Iranian Kurdish population of Kurdistan and West Azerbaijan provinces. Hemoglobin 2009;33:109–14.
[76] Yousefi MH. Beta thalassemia trait in Sanandaj. Sci J Kurdistan Univ Med Sci 1997;4:11–3.
[77] Fathollah-pour A, Naghshi-Zadeian R. Common beta-thalassemia mutations in the city of Sanandaj. Sci J Kurdistan Univ Med Sci 2003;3:21–6.
[78] Rahimi Z, Muniz A, Parsian A. Detection of responsible mutations for beta thalassemia in the Kermanshah Province of Iran using PCR-based techniques. Mol Biol Rep 2010;37:149–54.
[79] Mehrabi M, Alibakhshi R, Fathollahi S, Farshchi MR. The spectrum of beta-thalassemia mutations in Kermanshah Province in west Iran and its association with hematological parameters. Hemoglobin 2013;37:544–52.
[80] Kiani AA, Mortazavi Y, Zeinali S, Shirkhani Y. The molecular analysis of beta-thalassemia mutations in Lorestan Province, Iran. Hemoglobin 2007;31:343–9.
[81] Sharifi A, Aminzadeh M, Pourmoghaddam Z, Mahdieh N. The frequency of common beta-thalassemia mutations among couples referred to health centers of Ilam during a five years period. J Ilam Univ Med Sci 2014;22:17–23.
[82] Khodaei H, Zeinali S, Delmaghani S. Molecular studies on beta-thalassemia mutations in Bushehr Province. Iran South Med J 2001;2:89–93.
[83] Rahim F. Microcytic hypochromic anemia patients with thalassemia: genotyping approach. Indian J Med Sci 2009;63:101–8.
[84] Dehghanifard A, Shahjahani M, Galehdari H, Rahim F, Hamid F, Jaseb K, et al. Prenatal diagnosis of different polymorphisms of beta-globin gene in Ahvaz. Int J Hematol Oncol Stem Cell Res 2013;7:17–22.
[85] Galehdari H, Salehi B, Azmoun S, Keikhaei B, Zandian KM, Pedram M. Comprehensive spectrum of the beta-thalassemia mutations in Khuzestan, southwest Iran. Hemoglobin 2010;34:461–8.
[86] Galehdari H, Pedram M, Salehi B, Andashti B. Wide spectrum of mutations in the beta globin gene cause beta thalassemia major in southwest Iran. Iran J Pediatr Soc 2010;2:4–8.
[87] Mohammadi-Asl J, Samafbafzadeh A, Makvandi M, Zandian KM, Pedram M. Investigation of common mutations of beta globin gene among thalassemia

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001

majors using RDB in Khuzestan Province. Jundishapur Sci M J 2007;6:398–403.
[88] Hosseini H, Jalali Far MA, Saki N, Galehdari H, Nasir shalal M, Saki A. Study of beta-globin chain mutations in patients with beta thalassemia trait among different racial groups in Khouzestan Province. Genet 3rd millennium, 10; 2013. p. 2868–73.
[89] Kuliev AM, Rasulov IM, Dadasheva T, Schwarz EI, Rosatelli C, Saba L, et al. Thalassaemia in Azerbaijan. J Med Genet 1994;31:209–12.
[90] Tagiev AF, Surin VL, Gol'tsov AA, Lukianenko AV, Solovyev G, Gulieva EA, et al. The spectrum of beta-thalassemia mutations in Azerbaijan. Hum Mutat 1993;2:152–4.
[91] Curuk MA, Yuregir GT, Asadov CD, Dadasova T, Gu LH, Baysal E, et al. Molecular characterization of beta-thalassemia in Azerbaijan. Hum Genet 1992;90:417–9.
[92] Basak AN, Ozcelik H, Ozer A, Tolun A, Aksoy M, Agaoglu L, et al. The molecular basis of beta-thalassemia in Turkey. Hum Genet 1992;89:315–8.
[93] Tadmouri GO, Tuzmen S, Ozcelik H, Ozer A, Baig SM, Senga EB, et al. Molecular and population genetic analyses of beta-thalassemia in Turkey. Am J Hematol 1998;57:215–20.
[94] Guvenc B, Canataroglu A, Unsal C, Yildiz SM, Turhan FT, Bozdogan ST, et al. Beta-thalassemia mutations and hemoglobinopathies in Adana, Turkey: results from a single center study. Arch Med Sci 2012;8:411–4.
[95] Keser I, ManguoÚlu E, Guzeloglu Kayisli O, Kurt F, Bagci G, Luleci G, et al. Prenatal diagnosis of β-thalassemia in the Antalya Province. Turk J Med Sci 2005;354.
[96] Fettah A, Bayram C, Yarali N, Isik P, Kara A, Culha V, et al. Beta-globin gene mutations in Turkish children with beta-thalassemia: results from a single center study. Mediterr J Hematol Infect Dis 2013;5, e2013055.
[97] Al-Allawi NA, Jalal SD, Mohammad AM, Omer SQ, Markous RS. beta -thalassemia intermedia in Northern Iraq: a single center experience. Biomed Res Int 2014;2014:262853.
[98] Abdelaziz Omer W. Molecular characterization of beta-thalassemia mutations in Baghdad. Iraqi J Comm Med 2010;23:90–5.
[99] Abdelaziz Omer W. Molecular characterization of beta-thalassemia mutations in Ninawa governorate. Ann Coll Med 2009;35:124–33.
[100] Al-Allawi NA, Al-Mousawi BM, Badi AI, Jalal SD. The spectrum of beta-thalassemia mutations in Baghdad, Central Iraq. Hemoglobin 2013;37:444–53.
[101] Al-Allawi NA, Jubrael JM, Hughson M. Molecular characterization of beta-thalassemia in the Dohuk region of Iraq. Hemoglobin 2006;30:479–86.
[102] Adekile A, Haider M, Kutlar F. Mutations associated with beta-thalassemia intermedia in Kuwait. Med Princ Pract 2005;14(Suppl. 1):69–72.
[103] Adekile AD, Gu LH, Baysal E, Haider MZ, al-Fuzae L, Aboobacker KC, et al. Molecular characterization of alpha-thalassemia determinants, beta-thalassemia alleles, and beta S haplotypes among Kuwaiti Arabs. Acta Haematol 1994;92:176–81.
[104] el-Hazmi MA, al-Swailem AR, Warsy AS. Molecular defects in beta-thalassaemias in the population of Saudi Arabia. Hum Hered 1995;45:278–85.
[105] El-Harth EH, Kuhnau W, Schmidtke J, Stuhrmann M, Nasserallah Z, Al-Shahiri A. Identification and clinical presentation of beta thalassaemia mutations in the eastern region of Saudi Arabia. J Med Genet 1999;36:935–7.
[106] Abuzenadah AM, Hussein IM, Damanhouri GA, FM AS, Gari MA, Chaudhary AG, et al. Molecular basis of beta-thalassemia in the western province of Saudi Arabia: identification of rare beta-thalassemia mutations. Hemoglobin 2011;35:346–57.
[107] Al-Sultan A, Phanasgaonkar S, Suliman A, Al-Baqushi M, Nasrullah Z, Al-Ali A. Spectrum of beta-thalassemia mutations in the eastern province of Saudi Arabia. Hemoglobin 2011;35:125–34.
[108] Al-Ali AK, Al-Ateeq S, Imamwerdi BW, Al-Sowayan S, Al-Madan M, Al-Muhanna F, et al. Molecular bases of beta-thalassemia in the eastern province of Saudi Arabia. J Biomed Biotechnol 2005;2005:322–5.
[109] Al-Obaidli A, Hamodat M, Fawzi Z, Abu-Laban M, Gerard N, Krishnamoorthy R. Molecular basis of thalassemia in Qatar. Hemoglobin 2007;31:121–7.
[110] el-Kalla S, Mathews AR. A significant beta-thalassemia heterogeneity in the United Arab Emirates. Hemoglobin 1997;21:237–47.
[111] Quaife R, al-Gazali L, Abbes S, Fitzgerald P, Fitches A, Valler D, et al. The spectrum of beta thalassaemia mutations in the UAE national population. J Med Genet 1994;31:59–61.
[112] Jassim N, Merghoub T, Pascaud O, al Mukharraq H, Ducrocq R, Labie D, et al. Molecular basis of beta-thalassemia in Bahrain: an epicenter for a Middle East specific mutation. Ann N Y Acad Sci 1998;850:407–9.
[113] Daar S, Hussein HM, Merghoub T, Krishnamoorthy R. Spectrum of beta-thalassemia mutations in Oman. Ann N Y Acad Sci 1998;850:404–6.
[114] Daar S, Gravell D. Diagnosis of beta-thalassaemia carriers in the Sultanate of Oman. Sultan Qaboos Univ Med J 2006;6:27–31.
[115] Hassan SM, Vossen RH, Chessa R, den Dunnen JT, Bakker E, Giordano PC, et al. Molecular diagnostics of the HBB gene in an Omani cohort using bench-top DNA ion torrent PGM technology. Blood Cells Mol Dis 2014.
[116] Ansari SH, Shamsi TS, Ashraf M, Bohray M, Farzana T, Khan MT, et al. Molecular epidemiology of beta-thalassemia in Pakistan: far reaching implications. Int J Mol Epidemiol Genet 2011;2:403–8.
[117] Baig SM, Azhar A, Hassan H, Baig JM, Kiyani A, Hameed U, et al. Spectrum of beta-thalassemia mutations in various regions of Punjab and Islamabad, Pakistan: establishment of prenatal diagnosis. Haematologica 2006;91:ELT02.
[118] Mahmood Baig S, Sabih D, Rahim MK, Azhar A, Tariq M, Sajid Hussain M, et al. Beta-thalassemia in Pakistan: a pilot program on prenatal diagnosis in Multan. J Pediatr Hematol Oncol 2012;34:90–2.
[119] Moatter T, Kausar T, Aban M, Ghani S, Pal JA. Prenatal screening for beta-thalassemia major reveals new and rare mutations in the Pakistani population. Int J Hematol 2012;95:394–8.
[120] Usman M, Moinuddin M, Ghani R. Molecular genetics of beta-thalassaemia syndrome in Pakistan. East Mediterr Health J 2010;16:972–6.
[121] Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res 2007;35:3375–82.
[122] Roy A, Yang J, Zhang Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res 2012;40:W471–7.
[123] Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinf 2008;9:40.
[124] Mahdieh N, Shirkavand A, Raeisi M, Akbari MT, Tekin M, Zeinali S. Unexpected heterogeneity due to recessive and de novo dominant mutations of GJB2 in an Iranian family with nonsyndromic hearing loss: implication for genetic counseling. Biochem Biophys Res Commun 2010;402:305–7.
[125] Mahdieh N, Bagherian H, Shirkavand A, Sharafi M, Zeinali S. High level of intrafamilial phenotypic variability of non-syndromic hearing loss in a Lur family due to delE120 mutation in GJB2 gene. Int J Pediatr Otorhinolaryngol 2010;74:1089–91.
[126] Mahdieh N, Rabbani B, Shirkavand A, Bagherian H, Movahed ZS, Fouladi P, et al. Impact of consanguineous marriages in GJB2-related hearing loss in the Iranian population: a report of a novel variant. Genet Test Mol Biomarkers 2011;15:489–93.
[127] Ramazani A, Kahrizi K, Razaghiazar M, Mahdieh N, Koppens P. The frequency of eight common point mutations in CYP21 gene in Iranian patients with congenital adrenal hyperplasia. Iran Biomed J 2008;12:49–53.
[128] Mahdieh N, Shirkavand A, Rabbani B, Tekin M, Akbari B, Akbari MT, et al. Screening of OTOF mutations in Iran: a novel mutation and review. Int J Pediatr Otorhinolaryngol 2012;76:1610–5.
[129] Treisman R, Orkin SH, Maniatis T. Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature 1983;302:591–6.
[130] Ahmed S, Petrou M, Saleem M. Molecular genetics of beta-thalassaemia in Pakistan: a basis for prenatal diagnosis. Br J Haematol 1996;94:476–82.
[131] Sinha S, Black ML, Agarwal S, Colah R, Das R, Ryan K, et al. Profiling beta-thalassaemia mutations in India at state and regional levels: implications for genetic education, screening and counselling programmes. HUGO J 2009;3:51–62.
[132] Jalal SD, Al-Allawi NA, Bayat N, Imanian H, Najmabadi H, Faraj A. Beta-thalassemia mutations in the Kurdish population of northeastern Iraq. Hemoglobin 2010;34:469–76.
[133] Black ML, Sinha S, Agarwal S, Colah R, Das R, Bellgard M, et al. A descriptive profile of beta-thalassaemia mutations in India, Pakistan and Sri Lanka. J Community Genet 2010;1:149–57.
[134] De Leo R, Deidda G, Novelletto A, el-Kalla S, Mathews AR, Felicetti L. Analysis of beta-thalassemia mutations in the United Arab Emirates provides evidence for recurrent origin of the IVSI nt 5 (G-C) mutation. Hum Mutat 1995;5:327–8.
[135] Oner R, Altay C, Gurgey A, Aksoy M, Kilinc Y, Stoming TA, et al. Beta-thalassemia in Turkey. Hemoglobin 1990;14:1–13.
[136] Eshghi P, Rashidi A, Zadeh-Vakili A, Miri-Moghadam E. Hematological phenotype of the IVS-I-5 (G > C) beta-thalassemia mutation and assessment of Iran's national screening criteria. Hemoglobin 2008;32:440–5.
[137] Spritz RA, Jagadeeswaran P, Choudary PV, Biro PA, Elder JT, deRiel JK, et al. Base substitution in an intervening sequence of a beta$^+$-thalassemic human globin gene. Proc Natl Acad Sci U S A 1981;78:2455–9.
[138] Westaway D, Williamson R. An intron nucleotide sequence variant in a cloned beta$^+$-thalassaemia globin gene. Nucleic Acids Res 1981;9:1777–88.
[139] Baysal E, Indrak K, Bozkurt G, Berkalp A, Aritkan E, Old JM, et al. The beta-thalassaemia mutations in the population of Cyprus. Br J Haematol 1992;81:607–9.
[140] Efremov GD. Thalassemias and other hemoglobinopathies in the Republic of Macedonia. Hemoglobin 2007;31:1–15.
[141] Koliakos GG, Dimitriadou-Vaphiadou A, Christakis J, Trakatellis A. Beta zero and beta$^+$ thalassemia genes in northern Greece. Gene Geogr 1991;5:55–60.
[142] Stefanis L, Kanavakis E, Traeger-Synodinos J, Tzetis M, Metaxotou-Mavromati A, Kattamis C. Hematologic phenotype of the mutations IVS1-n6 (T→C), IVS1-n110 (G→A), and CD39 (C→T) in carriers of beta-thalassemia in Greece. Pediatr Hematol Oncol 1994;11:509–17.
[143] Boletini E, Svobodova M, Divoky V, Baysal E, Curuk MA, Dimovski AJ, et al. Sickle cell anemia, sickle cell beta-thalassemia, and thalassemia major in Albania: characterization of mutations. Hum Genet 1994;93:182–7.
[144] Chakrabarti P, Gupta R, Mishra A, Rai M, Singh VP, Dash D. Spectrum of beta-thalassemia mutations in North Indian states: a beta-thalassemia trait with two mutations in cis. Clin Biochem 2005;38:576–8.
[145] Weatherall DJ. Phenotype–genotype relationships in monogenic disease: lessons from the thalassaemias. Nat Rev Genet 2001;2:245–55.
[146] Orkin SH, Sexton JP, Goff SC, Kazazian Jr HH. Inactivation of an acceptor RNA splice site by a short deletion in beta-thalassemia. J Biol Chem 1983;258:7249–51.
[147] Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. Mamm Genome 2007;18:389–401.
[148] Ritchie MD. Finding the epistasis needles in the genome-wide haystack. Methods Mol Biol 2015;1253:19–33.
[149] Alanazi M, Abduljaleel Z, Khan W, Warsy AS, Elrobh M, Khan Z, et al. In silico analysis of single nucleotide polymorphism (SNPs) in human beta-globin gene. PLoS One 2011;6(10), e25876.
[150] Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. Cell 2004;119:831–45.
[151] Favero ME, Costa FF. Alpha-hemoglobin-stabilizing protein: an erythroid molecular chaperone. Biochem Res Int 2011;2011:373859.
[152] Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet 2010;42:53–61.

Please cite this article as: Mahdieh N, Rabbani B, Beta thalassemia in 31,734 cases with *HBB* gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East..., Blood Rev (2016), http://dx.doi.org/10.1016/j.blre.2016.07.001