

PRIMARY RESEARCH

Open Access



Seven novel glucose-6-phosphate dehydrogenase (G6PD) deficiency variants identified in the Qatari population

Shaza Malik^{1†}, Roan Zaied^{1,2†}, Najeeb Syed³, Puthen Jithesh^{3,4} and Mashaal Al-Shafai^{1*}

Abstract

Background: Glucose-6-phosphate dehydrogenase deficiency (G6PDD) is the most common red cell enzymopathy in the world. In Qatar, the incidence of G6PDD is estimated at around 5%; however, no study has investigated the genetic basis of G6PDD in the Qatari population yet.

Methods: In this study, we analyzed whole-genome sequencing data generated by the Qatar Genome Programme for 6045 Qatar Biobank participants, to identify G6PDD variants in the Qatari population. In addition, we assessed the impact of the novel variants identified on protein function both in silico and by measuring G6PD enzymatic activity in the subjects carrying them.

Results: We identified 375 variants in/near *G6PD* gene, of which 20 were high-impact and 16 were moderate-impact variants. Of these, 14 were known G6PDD-causing variants. The most frequent G6PD-causing variants found in the Qatari population were p.Ser188Phe (*G6PD* Mediterranean), p.Asn126Asp (*G6PD* A+), p.Val68Met (*G6PD* Asahi), p.Ala335Thr (*G6PD* Chatham), and p.Ile48Thr (*G6PD* Aures) with allele frequencies of 0.0563, 0.0194, 0.00785, 0.0050, and 0.00380, respectively. Furthermore, we have identified seven novel *G6PD* variants, all of which were confirmed as G6PD-causing variants and classified as class III variants based on the World Health Organization's classification scheme.

Conclusions: This is the first study investigating the molecular basis of G6PDD in Qatar, and it provides novel insights about G6PDD pathogenesis and highlights the importance of studying such understudied population.

Keywords: *G6PD* deficiency, Whole-genome sequencing (WGS), Novel variants, Qatar Biobank (QBB), Qatar Genome Programme (QGP)

Background

Glucose-6-phosphate dehydrogenase (G6PD) is an omnipresent cytosolic enzyme that has an important housekeeping role in all cells. In red blood cells (RBCs), nicotinamide adenine dinucleotide phosphate (NADPH) is produced mainly by the action of G6PD in the first

step of the pentose phosphate pathway [1]. NADPH, among other cellular functions, is particularly important in preventing the buildup of reactive oxygen species [2]. Normal activity of G6PD thus helps protect RBCs from oxygen-derived oxidative stress [3]. Glucose-6-phosphate dehydrogenase deficiency (G6PDD) patients might develop symptoms after exposure to compounds that trigger oxidative stress in RBCs (favism and drug-induced hemolytic anemia), and it is inherited as X-linked recessive phenotype [4]. The WHO classifies G6PDD-causing variants into five classes: class I is the most severe causing chronic non-spherocytic hemolytic anemia (G6PD

*Correspondence: malshafai@qu.edu.qa

†Shaza D. Malik and Roan E. Zaied are joint first authors

¹ Department of Biomedical Sciences, College of Health Sciences, QU Health, Qatar University, Doha, Qatar

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

activity < 1%), class II is also considered severe and associates with acute hemolytic anemia (G6PD activity < 10%), class III is considered as moderate deficiency and it is associated with occasional acute hemolytic anemia (G6PD activity of 10–60%), while classes IV and V are asymptomatic (G6PD activity of 60–150% and > 150%, respectively) [5].

G6PDD affects around 400 million people globally making it the most common human enzymopathy [6]; it is particularly common in the Middle East with prevalence rates reaching up to 39.8% and 30% in Saudi Arabia and Syria, respectively [7, 8]. Over 217 variants of *G6PD* have been reported worldwide, the vast majority of which are point mutations [9]. The most common variants among Arabs are p.Ser188Phe (*G6PD* Mediterranean), p.Ile48Thr (*G6PD* Aures), p.Asn126Asp (rs1050829), and p.Val68Met (*G6PD* Asahi) [10]. However, no studies addressed the molecular basis of G6PDD in Qatar. In this study, we investigated G6PDD-causing variants in the Qatari population using whole-genome sequencing (WGS) data from Qatar Genome Programme (QGP) for 6,045 Qatar Biobank (QBB) participants [11].

Materials and methods

Study participants

The study subjects consisted of 6045 QBB participants, and these subjects are Qatari nationals or long-term residents of Qatar (at least 15 years). Participants are aged 18 years or above, and they appeared phenotypically healthy. This includes 3403 females and 2642 males. Detailed phenotypic and lifestyle data were available for the study participants including lifestyle and food intake questionnaires as well as biochemical tests. However, no data were available on the G6PD activity levels. All participants provided informed consent.

Genetic data

WGS was conducted as part of the Qatar Genome Programme using the Illumina HiSeq TenX platform to an average coverage of 30X. Raw sequencing reads were converted to paired FASTQ format using the bcl2fastq software from Illumina [12], and fastq files were aligned against the reference genome sequence (GRCh37) using bwakit (v. 0.7.11). Variant calling was performed using GATK 3.4, and then, individual vcf's were undergone for joint calling step to transform into multi-sample VCF file for all participants. After performing the VQSR step, only PASS variants were further used for downstream analysis. Variant calling and filtering steps were performed following GATK best practices [13]. All variants were described in relation to coding DNA reference sequence NM_001042351.3, NM_000402.4, or NM_001042351.1 (specified where relevant). Principal component analysis

plot for the genomes used in this study is given in supplementary material, Additional file 1: Figure S1.

Variant annotation

The variants identified were annotated using SnpEff (SnpEff/SnpSift (v4.3t)), which classified variants as high-, moderate-, low-, or modifier-impact variants based on their potential impact at the protein level [14]. High-impact variants include structural, nonsense, frameshift, loss of start codon variants as well as splice site donors/acceptors variants, while moderate-impact variants include 3' and 5' UTR variants, exon loss variants, missense variants, conservative in-frame deletions, and conservative in-frame insertions. Low-impact variants include synonymous variants, stop-retain variants, 5'-UTR-premature-start codon gain variants, and splice region variants, while the modifier-impact variants include variants in the flanking regions of the genes, intronic regions, and non-coding regions [14]. Here, we focused on studying the high- and moderate-impact variants, since they are likely to affect the protein function.

Assessment of the high- and moderate-impact variants

High- and moderate-impact variants were annotated using different databases such as Human Gene Mutation Database (HGMD) and ClinVar. They were also assessed through pathogenicity prediction tools including sorting intolerant from tolerant (SIFT), Polyphen, combined annotation-dependent depletion (CADD), and dbNSFP-Polyphen2-HDIV. Allele frequencies (AF) of those variants were determined using the QGP dataset, the genome aggregation database (gnomAD), Greater Middle East (GME), and the 1000 Genomes (1 KG) Project database.

Molecular structure analysis using PyMol

The crystallized partial G6PD structure was retrieved from PDB (PDB ID: 2BHL [15]). Only non-synonymous variants that were not previously investigated in Doss et al. [10] and whose native residues were crystallized in 2BHL were analyzed. PyMol version 2.4.1 was used to model the 18 selected variants in order to visualize the potential structural changes they introduce to the native protein structure [16].

Sanger sequencing

The novel variants identified in this study were confirmed by PCR and Sanger sequencing. The primers used are given in Additional 2: Table SA1 using the ABI 373 automated sequencer, and the data were analyzed using the

Mutation Surveyor software (<http://www.softgenetics.com/mutationSurveyor.html>).

G6PD deficiency assay

To investigate the functional impact of the novel *G6PD* variants identified, we used a quantitative assay to measure G6PD enzymatic activity in frozen red blood cells (RBCs) obtained from the variant carriers. The samples were tested for their enzyme activity at the diagnostic laboratories of Hamad Medical Cooperation (HMC), using G6PDH assay (RANDOX, cat. no. PD 410), following the manufacturer’s protocol. G6PD activity measured in units of enzyme activity was normalized for RBC

count as per the following formula. The reference range used to determine deficiency is as per HMCs protocol (224.1–516.9 mU/10⁹ RBC), and this is based on the normal activity of the enzyme observed in hemizygous males from the Qatari population.

$$G6PDH\ mU/10^9RBCs = \frac{G6PDH\ activity\ in\ mU/mL}{/RBC's\ count\ 10^6/\mu l.}$$

Results

G6PD variants in the Qatari population

We identified from the WGS data of 6,045 QGP participants 375 variants in/nearby the *G6PD* gene. These include 20 high-impact, 16 moderate-impact, 19 low-impact, and 320 modifier-impact variants (Fig. 1). We focused on the high-impact and moderate-impact variants (36 variants) as previously indicated in Sect. 2.3. The five most common variants seen in the Qatari population and their corresponding frequencies in other databases are indicated in Fig. 2. The high-impact and moderate-impact variants are indicated in a liner protein diagram in Fig. 3, constructed using DOG V 2.0.1 [17] based on the annotation mentioned in the literature [6, 15].

Analysis of the high- and moderate-impact variants

We investigated 36 variants, including 20 high-impact and 16 moderate-impact variants. Of those, seven variants were novel (four high-impact and three moderate-impact variants). The frequencies of those variants and their CADD scores are given in Tables 1 and 2. Their predicted pathogenicity scores using SIFT, Polyphen, and dbNSFP-Polyphen2-HDIV are also presented as additional material in Additional file 2: Tables SA2 and SA3. The WHO classification for the well-established

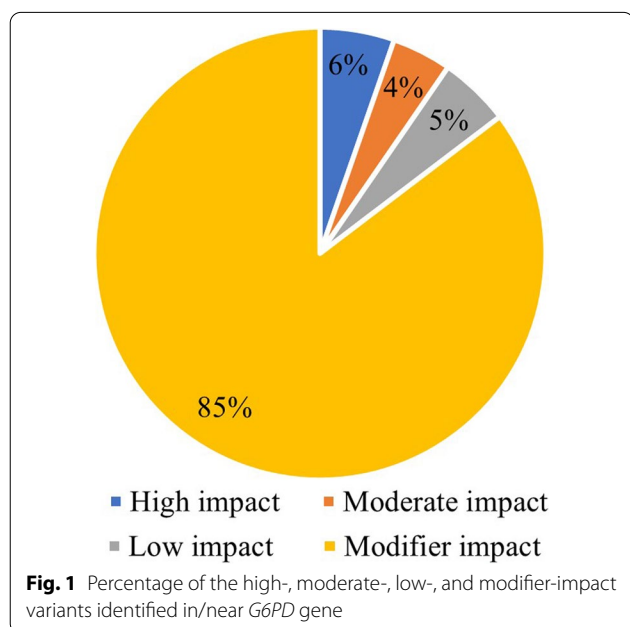


Fig. 1 Percentage of the high-, moderate-, low-, and modifier-impact variants identified in/near *G6PD* gene

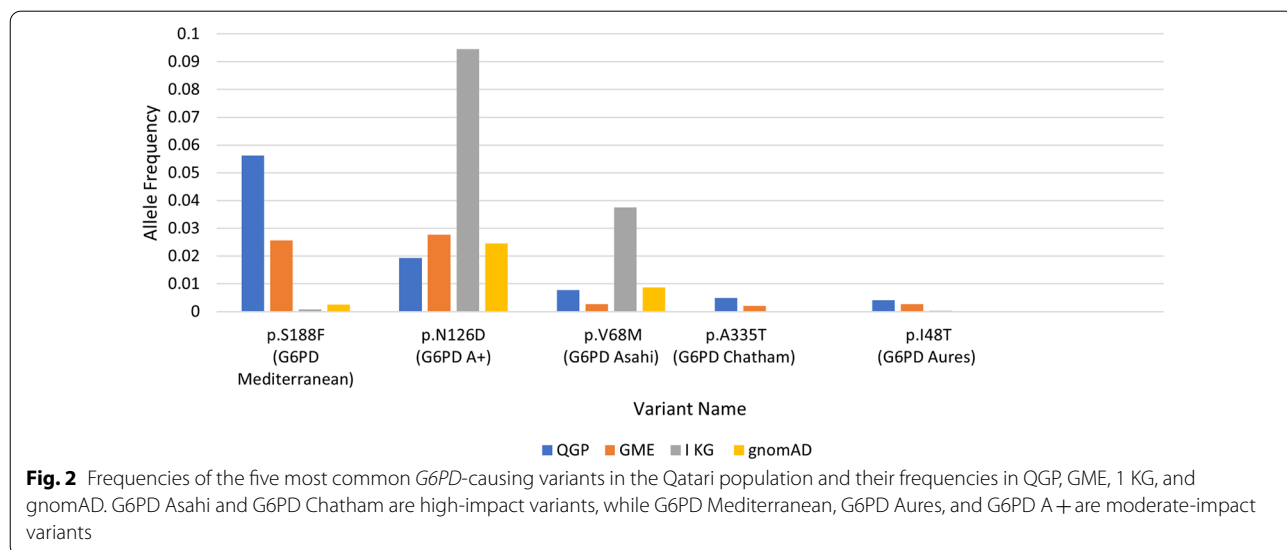


Fig. 2 Frequencies of the five most common *G6PD*-causing variants in the Qatari population and their frequencies in QGP, GME, 1 KG, and gnomAD. G6PD Asahi and G6PD Chatham are high-impact variants, while G6PD Mediterranean, G6PD Aures, and G6PD A+ are moderate-impact variants

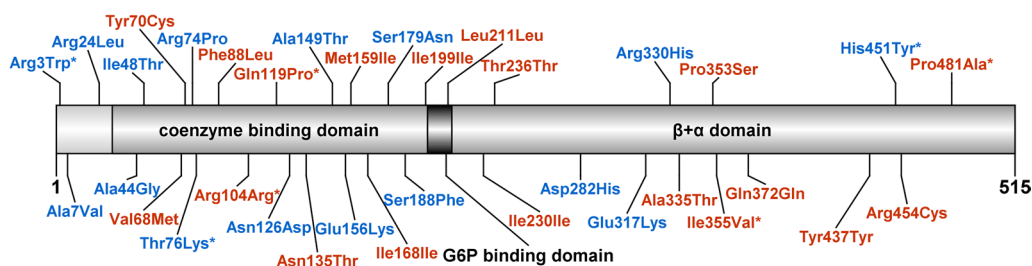


Fig. 3 Linear G6PD structure showing the positions of the 36 reported G6PD mutations. High-impact variants are shown in red, while moderate-impact variants are shown in blue. Novel variants are denoted with an asterisk (*). For p.Met159Ile, two different genetic variants resulted in the the same amino acid change

G6PDD-causing variants were also included when available (Additional file 2: Tables SA2 and SA3). The subpopulation frequencies of the 36 variants were also collected from QGP data for the six subpopulations in Qatar which are Peninsular Arabs (PAR) (17.4%), General Arabs (GAR) (38.2%), West Eurasian/Persians (WEP) (22.7%), Africans (AFR) (1.5%), South Asians (SAS) (0.6%), and Admixed (ADM) (19.5%) (Additional file 2: Table SA4).

Molecular visualization

In silico analysis using PyMol was preformed to reveal potential structural changes introduced by the identified variants. Modeled variants include six previously reported high-impact variants (p.Arg454Cys, p.Met159Ile, p.Phe88Leu, p.Pro353Ser, p.Ala335Thr, and p.Tyr70Cys), seven moderate-impact variants (p.Ala149Thr, p.Glu156Lys, p.Arg330His, p.Glu317Lys, p.Ala44Gly, p.Arg74Pro, and p.Asp282His), and five novel variants (p.Pro481Ala, p.Ile355Val, p.Gln119Pro, p.His451Tyr, and p.Thr76Lys). The variants p.Gln119Pro, p.Thr76Lys, p. Tyr70Cys, and p.Asp282His are predicted to cause loss of polar contacts, which could lead to the destabilization of the protein structure (Fig. 4c, d, i, o). The variants p.Pro481Ala, p.Ala149Thr, p.Asp282His, p.His451Tyr, and p.Pro353Ser resulted in gain of polar contacts, which could conversely over-stabilize the protein structure (Fig. 4a, k, o, p, r). Rotamers of p.Arg454Cys, p.Asp282His, and p.Ala335Thr were predicted to clash with surrounding residues, which could lead to steric hindrance in G6PD as shown in Fig. 4e, o, q. Finally, substitutions that greatly change the size of the native amino acid were also observed such as p.Glu156Lys and p.Arg74Pro shown in Fig. 4j and l, respectively.

Enzyme activity assay

The seven novel variants were seen in a total of 14 participants (12 females and 2 males). Of those, all were G6PD

deficient, except for two female participants who had G6PD levels that fall in the lower normal range (Table 3).

Discussion

This study is the first to investigate the genetic basis of G6PDD in Qatar. We used WGS data from 6045 phenotypically healthy QBB participants and identified p.Ser188Phe (G6PD Mediterranean), p.Asn126Asp (G6PD A+), p.Val68Met (G6PD Asahi), p.Ala335Thr (G6PD Chatham), and p.Ile48Thr (G6PD Aures) as the five most common disease-causing G6PD variants in our QBB cohort with AFs of 0.0563, 0.0194, 0.00785, 0.0050, and 0.00380, respectively. This is consistent with findings from the Arab world, in which p.Ser188Phe (G6PD Mediterranean), p.Asn126Asp (G6PD A+), p.Val68Met (G6PD Asahi), and p.Ile48Thr (G6PD Aures) were reported to be the most common G6PD variants in Arab G6PDD patients, with prevalence of 56.80%, 10.03%, 10%, and 5.51%, respectively [10]. The p.Tyr437Tyr (rs2230037) variant was the most common variant seen in this study with AF of 0.288. It has been described as a polymorphism in various studies [18, 19], a “benign” variant in ClinVar, and as potentially causative in some studies [20].

Studies conducted in the UAE, Jordan, Iraq, and Oman also reported relatively low prevalence (0.5–8.7%) of G6PD Chatham cases among G6PDD patients [21–24]. These similarities in G6PD variations observed across different Arab populations reflect the common ancestry and similar genetic background of Arabs. Moreover, Doss and Alasmar [10] report four variants to be exclusively found in the Arab world, namely p.Asp135Thr, p.Ser179Asn, p.Arg246Leu, and p.Glu307Pro. The variant p.Asp135Thr was observed in Egypt, Palestine, and Jordan [10, 22, 25], and we report it also here in our cohort with AF of 0.00033074. The variant p.Ser179Asn was reported in Palestine and was also seen in our study (AF = 0.00016537). On the other hand, the other two variants, p.Arg246Leu and p.Glu307Pro, which were initially seen in Tunisia, were not detected in our study.

Table 1 High-impact variants identified in the *G6PD* gene

Position	rs ID	Variant name	CDS substitution	Amino acid substitution	Exon	CADD score	HGMD	Frequency in QGP	Frequency in GME	Frequency in gnomAD
+ 153760419			c.1441C>G	p.Pro481Ala	12	7.925		0.00008268	0.00139	
153760605	rs398123546	<i>G6PD</i> Union	c.1360C>T	p.Arg454Cys	11	34	DM	0.00008269		0.0001071
# 153760654	rs2230037	-	c.1311C>T	p.Tyr437Tyr	11	6.988	DM?	0.284687		0.169
153760953	rs2230036	-	c.1116G>A	p.Gln372Gln	10	4.741	-	0.00529188	0.00691	0.009705
+ 153761006			c.1063A>G	p.Ile355Val	10	6.273		0.00041342		
153761012	rs137852333	<i>G6PD</i> lerapetra	c.1057C>T	p.Pro353Ser	10	24.8	DM	0.00033074		0.0002306
153761205	rs5030869	<i>G6PD</i> Chatham	c.1003G>A	p.Ala335Thr	9	27.3	DM	0.00504382	0.00207	0.00001647
153762312	rs782118135	-	c.708C>T	p.Thr236Thr	7	8.209	-	0.00024805		0.0007578
153762330	rs781917123	-	c.690C>T	p.Ile230Ile	7	13.01	-	0.00016537		-
153762564	rs781866029	-	c.633C>T	p.Leu211Leu	6	10.35	-	0.00008268		
153762600	rs369516039	-	c.597C>T	p.Ile199Ile	6	10.89	-	0.00008268		
153762693	rs782678149	-	c.504C>T	p.Ile168Ile	6	12.96	-	0.00016537		0.00001647
153763391	rs370918918		c.477G>C	p.Met159Ile	5	18.24	DM?	0.00223251		0.001219
153763391	rs370918918		c.477G>A	p.Met159Ile	5	20.2		0.00016537		
153763464	rs782322505		c.404A>C	p.Asn135Thr	5	24.5	DM	0.00033074		0.000008237
+ 153763512			c.356A>C	p.Gln119Pro	5	2.238		0.00008268		
+ 153763556			c.312C>T	p.Arg104Arg	5	2.941	-	0.00016537		
153764155	rs781794862	-	c.264C>G	p.Phe88Leu	4	2.678		0.00008268		0.0001565
153764210	rs782090947		c.209A>G	p.Tyr70Cys	4	22.9	DM	0.00016537		0.000008237
153764217	rs1050828	<i>G6PD</i> Asahi	c.202G>A	p.Val68Met	4	27.2	DM	0.00785513	0.00275	0.0009974

Novel variants are denoted with a dagger (†). Variants are described in relation to NM_001042351.1; those denoted with (#) are described in relation to NM_001042351.1. CDS: coding sequence

Table 2 Moderate-impact variants identified in the *G6PD* gene

Position	rs ID	Variant name	CDS substitution	Amino acid substitution	Exon	CADD score	HGMD*	Frequency in QGP	Frequency in GME	Frequency in gnomAD
†153760614	-	-	c.1351C>T	p.His451Tyr	11	25.6	-	0.00016537	-	-
153761219	rs868950643	-	c.989G>A	p.Arg330His	9	7.666	-	0.0116587	-	-
153761259	rs137852339	G6PD Kalyan	c.949G>A	p.Glu317Lys	9	22.6	DM	0.00090954	-	0.001664
153761811	rs137852318	G6PD Modena	c.844G>C	p.Asp282His	8	24.8	DM	0.00008268	-	0.000865
153762634	rs5030868	G6PD Mediterranean	c.563C>T	p.Ser188Phe	6	24.2	DM	0.0563089	0.02569	0.003097
153762661	rs281860640	-	c.536G>A	p.Ser179Asn	6	25.9	DM	0.00016537	-	-
153763402	rs137852313	G6PD Ilesha	c.466G>A	p.Glu156Lys	5	8.019	DM	0.0010749	-	0.000313
153763423	rs782669677	-	c.445G>A	p.Ala149Thr	5	10.11	DM	0.00024805	-	0.00001647
153763492	rs1050829	G6PD A +	c.376A>G	p.Asn126Asp	5	0.037	DM	0.0194311	0.027759	0.029
†153764192	-	-	c.227C>A	p.Thr76Lys	4	18.76	-	0.00008268	-	-
153764198	-	-	c.221G>C	p.Arg74Pro	4	10.28	-	0.00107491	-	-
153764371	rs76645461	G6PD Aures	c.143 T>C	p.Ile48Thr	3	18.09	DM	0.0038035	0.00276	0.000008237
153764383	rs78478128	G6PD Orissa	c.131C>G	p.Ala44Gly	3	28.9	-	0.00016537	-	0.000272
#153775015	rs782006658	-	c.71G>T	p.Arg24Leu	1	12.05	-	0.00041377	-	0.001074
153775066	rs782150651	-	c.20C>T	p.Ala7Val	1	22.2	-	0.00049718	-	0.00003676
#†153775079	-	-	c.7C>T	p.Arg31Trp	1	19.35	-	0.00024888	-	-

Novel variants are denoted with a dagger (†). Variants are described in relation to NM_001042351.3; those denoted with (#) are described in relation to NM_000402.4. CDS: coding sequence

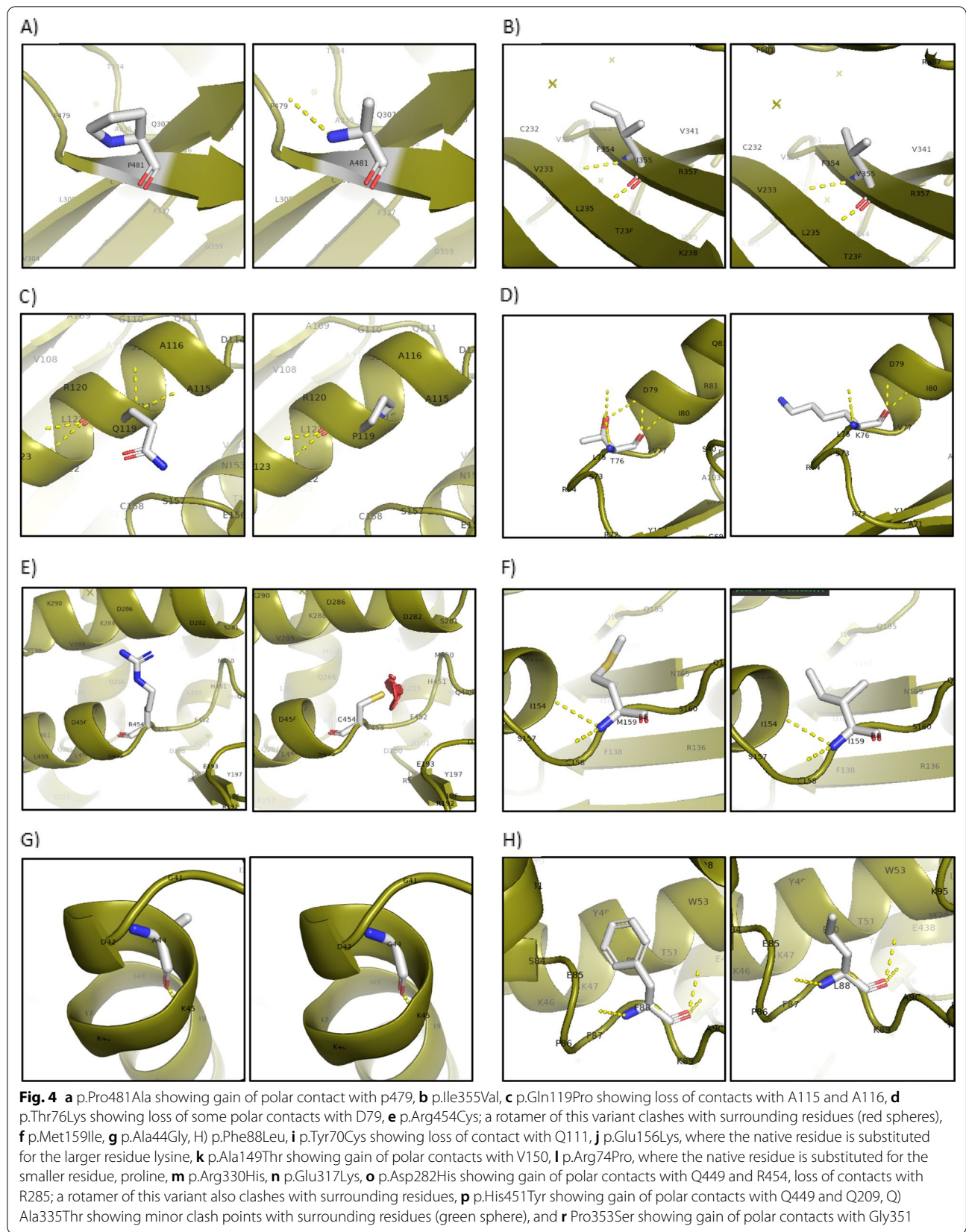


Table 3 Enzyme activity assay results for the carriers of the seven novel variants identified in the Qatari population

Variant position	Amnio acid change	Sample ID	Subject gender	Subject zygosity status	RBC count ($\times 10^6/\mu\text{l}$)	G6PDH (mU/mL)	G6PDH (mU/ 10^9 RBC)	Interpretation*
153775079	p.Arg3Trp	QU000105000014	M	Hemizygous	6	1010	168	Deficient
		QU000105000006	F	Heterozygous	4.4	997	227	Normal
		QU000105000007	F	Heterozygous	5	1013	203	Deficient
153760614	p.His451Tyr	QU000105000011	F	Heterozygous	4.5	602	134	Deficient
		QU000105000010	F	Heterozygous	5.8	878	151	Deficient
153764192	p.Thr76Lys	QU000105000003	F	Heterozygous	4.4	700	159	Deficient
153760419	p.Pro481Ala	QU000105000001	F	Heterozygous	5.6	1075	192	Deficient
153761006	p.Ile355Val	QU000105000013	M	Hemizygous	4.2	733	175	Deficient
		QU000105000012	F	Heterozygous	4	1004	251	Normal
		QU000105000009	F	Heterozygous	4.2	725	173	Deficient
		QU000105000004	F	Heterozygous	4	526	132	Deficient
153763512	p.Gln119Pro	QU000105000002	F	Heterozygous	4.2	341	81	Deficient
153763556	P.Arg104Arg	QU000105000005	F	Heterozygous	5.3	668	126	Deficient
		QU000105000008	F	Heterozygous	3.9	678	174	Deficient

M: male, F: female

*G6PDH of 244 mU/ 10^9 RBC or below is interpreted as deficient

We also investigated the frequencies of the 36 variants within the six distinct subpopulations that constitute the Qatari population. The most common variant identified in this study, G6PD Mediterranean, was predominantly seen in West Eurasian/Persians (WEP) subpopulation with AF of 0.134475, while in gnomAD it was mostly seen in South Asians subpopulation with AF of 0.0173. Three variants G6PD A+, G6PD Asahi, and G6PD Aures were mainly identified in Africans (AFR) with AFs of 0.173913, 0.0706522, and 0.0108696, respectively, while G6PD Chatham was mainly seen in WEP (AF=0.0193149). In gnomAD, the frequencies and distributions of those variants among subpopulations are different; G6PD A+ is mainly identified in Latinos (AF=0.001130), G6PD Chatham in Europeans (AF=0.00001232), G6PD Aures in East Asians (AF=0.0001443), and G6PD Asahi in Africans (AF=0.1164). This can be due to the fact that Arabs and Middle Eastern populations are underrepresented in gnomAD and other publicly available databases.

Conservation studies report three motifs in the G6PD protein to show particularly high conservation: these being the 198-RIDHYLGKE-206 motif which is required for the binding and catalysis of glucose-6-phosphate (G6P) [26], the 38-GASGDLA-44 motif (termed the nucleotide-binding fingerprint) which constitutes a region where the coenzyme binds, and finally 170-EKPxG-174 which is needed for the correct positioning of both the coenzyme and the substrate [15, 27, 28]. Only two variants were seen within these motifs in our study: the synonymous variant p.Ile199Ile with AF of

0.000082685 and the G6PD Orissa variant (p.Ala44Gly) with AF of 0.000165371. Their low frequencies further suggest that variations within these regions might not be tolerated. Finally, most of the analyzed variants (22%) were found to be within exon 5, suggesting that this exon is a hot spot for missense variants. We also identified variants in exons 6, 10, and 13, which are reported to be within the substrate-binding domain of G6PD (Additional file 2: Tables SA2 and SA3) [29].

This study identified seven novel G6PD variants (not seen in other variants databases) and investigated their impact on G6PD activity. These were classified as class III variants since they resulted in enzymatic activity between the range of 10 and 60% [5]. Table 3 shows a summary of the G6PD activity assay findings for the novel variants. Among the high-impact novel variants, the p.Pro481Ala variant, which substitutes the polar amino acid proline with the non-polar amino acid alanine, is predicted to introduce additional polar contacts with the nearby Pro479 (Fig. 4a). The variant has a relatively low CADD score of 7.925 and is predicted to be “benign” and “tolerated” in Polyphen and SIFT, respectively. This variant was identified in one heterozygous female which was G6PD deficient based on the enzymatic activity test, which contradicts the predication and might also cause deficiency due to X-inactivation skewness in favor of the deficient allele in the female carrier. In the second novel variant, p.Ile355Val, both the native and mutant amino acids are non-polar and are of similar size (Fig. 4b). No loss or gain of polar contacts is observed after molecular visualization

which correlates with their annotation as “benign” and “tolerated” in Polyphen and SIFT, respectively. The biochemical activity test showed that three participants (one hemizygous male and two heterozygous females) carrying this variant were found to be G6PD deficient. However, another heterozygous female appeared to have G6PD activity in the lower normal range (251 mU/10⁹ RBC). The variation in the G6PD status observed among the heterozygous females from deficient to normal is likely to reflect variations in the ratios of G6PD deficient to G6PD normal RBCs as a result of variations in X-inactivation patterns during early embryogenesis [30, 31]. The p.Gln119Pro is the third high-impact novel variant. Here, the polar-uncharged amino acid glycine is substituted for the non-polar amino acid, proline. This was predicted to cause loss of polar contacts with two of the nearby amino acids, Ala115 and Ala116, indicating loss of stability (Fig. 4e, f). The variant was, however, predicted to be “benign” and “tolerated” by Polyphen and SIFT, respectively. One heterozygous female carried this variant, and she was tested to have deficient G6PD activity. Finally, p.Arg104Arg is a synonymous variant and it has a low CADD score of 2.941, but was classified as high-impact variant by SnpEff which takes into account the location of the variant as well as its role in interaction [14]. In agreement with SnpEff prediction, the female carrier of this variant was G6PD deficient.

Among the moderate-impact novel variants, the polar uncharged threonine is replaced with the polar, positively charged, and much larger lysine in the p.Thr76Lys variant. This is predicted to cause loss of some polar contacts with the neighboring Asp79, potentially destabilizing the enzyme (Fig. 4g, h). The variant is classified as “damaging” and “possibly damaging” in Polyphen and SIFT, respectively. This variant was seen in one female participant (heterozygous) within our study who was reported as G6PD deficient. In the second moderate-impact novel variant, p.Arg3Trp, the positively charged, polar amino acid, arginine is substituted for the aromatic amino acid tryptophan. This substitution was predicted to be damaging by SIFT, and it had a relatively high CADD score of 19.35. Of the three participants tested for this variant, two were found to be deficient, while one heterozygous female showed low normal G6PD activity (227 mU/10⁹ RBC). In the final moderate-impact novel variant, p.His451Tyr, the polar amino acid histidine is substituted for the aromatic amino acid tyrosine. It was predicted to be “probably damaging” and “tolerated” by Polyphen and SIFT, respectively, and has the highest CADD score of 25 among the novel variants identified. The two participants carrying this variant (heterozygous females) were found to be G6PD deficient, supporting their causality.

Conclusion

In summary, this study revealed novel G6PDD variants and elucidated their genotype–phenotype correlation. In addition, we determined the frequencies of some common G6PD variants in the Qatari population. Our work highlights the importance of investigating understudied populations in providing novel insights about disease pathogenesis.

Abbreviations

ADM: Admixed; AF: Allele frequency; AFR: Africans; CADD: Combined annotation-dependent depletion; ExAC: Exome Aggregation Consortium; GME: The Greater Middle East database; GAR: General Arab; G6PD: Glucose-6-phosphate dehydrogenase; G6PDD: Glucose-6-phosphate dehydrogenase deficiency; gnomAD: Genome aggregation database; HMC: Hamad Medical Cooperation; HGMD: Human Gene Mutation Database; NADPH: Nicotinamide adenine dinucleotide phosphate; QBB: Qatar Biobank; QGP: Qatar Genome Programme; RBC: Red blood cell; RS ID: Reference SNP cluster ID; ASA: South Asians; SIFT: Sorting intolerant from tolerant; WEP: West Eurasian/Persians; WGS: Whole-genome sequencing; WHO: World Health Organization; VCF: Variant call format; PAR: Peninsular Arab; Polyphen: Prediction of functional effects of human nsSNPs; 1 Kg: 1000 Genomes Project database.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-021-00358-9>.

Additional file 1: Figure S1. Principal component analysis plot for the genomes used in this study.

Additional file 2: Table SA1. Primers used to confirm the novel G6PD variants using Sanger sequencing. Table A2. High-impact variants identified in the G6PD gene. Table A3. Moderate-impact variants identified in the G6PD gene. Table A4. Subpopulation frequencies of the identified high- and moderate-impact variants.

Acknowledgements

We thank QBB and QGP for providing the phenotypic and whole-genome sequencing data used in this study, with special thanks to Dr. Nahla Afifi, the Director of QBB, and Mrs. Fatima Qafoud, QBB Manager of laboratory and clinic, Dr. Said Ismail, the Director of QGP, and Prof. Asma Al-Thani, the Chair of the National Genome Committee, for supporting this study. We also thank the QBB participants for donating their samples and data for research purposes and Ms. Sumbul Bushra for her help with the molecular visualization analysis.

Authors' contributions

MA designed the study; SM and RZ performed the data analysis under the guidance and supervision of MA; PJ and NS helped with the WGS data extraction and annotation; and SM, RZ, and MA wrote the manuscript. All authors read and approved the final version.

Funding

Open access was supported by Qatar University Grant (QUST-1-CHS-2021–12).

Availability of data and materials

The informed consent given by the study participants does not cover posting of participant level phenotype and genotype data of Qatar Biobank/Qatar Genome Project in public databases. However, access to QBB/QGP data can be obtained through an established ISO-certified process by submitting a project request at <https://www.qatarbiobank.org.qa/research/how-to-apply> which is subject to approval by the QBB IRB Committee. The datasets supporting the conclusions of this article are included within the article and its additional file.

Declarations

Ethics approval and consent to participate

Ethical approval for this study was obtained from Qatar Biobank's (QBB) (approval number: E -2019-RES-ACC-0165-0088) and Qatar University's (approval number: QU-IRB 1288-EA/20) institutional review boards, in concordance with the Helsinki Declaration of ethical principles for medical research.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interest.

Author details

¹Department of Biomedical Sciences, College of Health Sciences, QU Health, Qatar University, Doha, Qatar. ²Liggins Institute, The University of Auckland, Auckland, New Zealand. ³Applied Bioinformatics Core, Sidra Medicine, Doha, Qatar. ⁴Hamad Bin Khalifa University, Doha, Qatar.

Received: 7 April 2021 Accepted: 6 September 2021

Published online: 07 October 2021

References

- Stincione A, Prigione A, Cramer T, Wamelink MMC, Campbell K, Cheung E, et al. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol Rev Camb Philos Soc.* 2015;90(3):927–63.
- Pollak N, Dolle C, Ziegler M. The power to reduce: pyridine nucleotides—small molecules with a multitude of functions. *Biochem J.* 2007;402(2):205–18.
- Salvemini F, Franze A, Iervolino A, Filosa S, Salzano S, Ursini MV. Enhanced glutathione levels and oxidoresistance mediated by increased glucose-6-phosphate dehydrogenase expression. *J Biol Chem.* 1999;274(5):2750–7.
- Olusanya BO, Osibanjo FB, Slusher TM. Risk factors for severe neonatal hyperbilirubinemia in low and middle-income countries: a systematic review and meta-analysis. *PLoS ONE.* 2015;10(2):e0117229.
- World Health Organization. Glucose-6-phosphate dehydrogenase deficiency. WHO Working Group. *Bulletin of the World Health Organization.* 1989;67(6):601–11.
- Au SW, Gover S, Lam VM, Adams MJ. Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency. *Structure.* 2000;8(3):293–303.
- Alabdullaali MK, Alayed KM, Alshaikh AF, Almashhadani SA. Prevalence of glucose-6-phosphate dehydrogenase deficiency and sickle cell trait among blood donors in Riyadh. *Asian J Transfus Sci.* 2010;4(1):31.
- Usanga EA, Ameen R. Glucose-6-phosphate dehydrogenase deficiency in Kuwait, Syria, Egypt, Iran, Jordan and Lebanon. *Hum Heredity.* 2000;50(3):158–61.
- Gómez-Manzo S, Marcial-Quino J, Vanoye-Carlo A, Serrano-Posada H, Ortega-Cuellar D, González-Valdez A, et al. Glucose-6-phosphate dehydrogenase: update and analysis of new mutations around the world. *Int J Mol Sci.* 2016;17(12):2069.
- Doss CG, Alasmal DR, Bux RI, Sneha P, Bakhsch FD, Al-Azwani I, et al. Genetic epidemiology of glucose-6-phosphate dehydrogenase deficiency in the Arab world. *Sci Rep.* 2016;6:37284.
- Al Thani A, Fthenou E, Pappadopoulos S, Al Marri A, Shi Z, Qafoud F, et al. Qatar biobank cohort study: study design and first results. *Am J Epidemiol.* 2019;188(8):1420–33.
- bcl2fastq software [Available from: <https://github.com/brwnj/bcl2fastq>].
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernyt-sky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303.
- Cingolani P, Cunningham F, McLaren W, Wang K. Variant annotations in VCF format. 2015.
- Kotaka M, Gover S, Vandeputte-Rutten L, Au SW, Lam VM, Adams MJ. Structural studies of glucose-6-phosphate and NADP+ binding to human glucose-6-phosphate dehydrogenase. *Acta Crystallogr D Biol Crystallogr.* 2005;61(Pt 5):495–504.
- Schrodinger L. The PyMOL molecular graphics system. Version. 2010;1(5):0.
- Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. DOG 1.0: illustrator of protein domain structures. *Cell Res.* 2009;19(2):271–3.
- Arnaout HH, El-Gharbawy NM, Shaheen IA, Afifi RA, Abd EL-Dayem OY. Incidence and association of 563 C/T Mediterranean and the silent 1311C/T G6PD mutations in G6PD-deficient Egyptian children. *Lab Med.* 2011;42(6):355–60.
- Wisnumurti DA, Sribudiani Y, Porsch RM, Maskoen AM, Rahayuningsih SE, Asni EK, et al. G6PD genetic variations in neonatal Hyperbilirubinemia in Indonesian Deutromalay population. *BMC Pediatr.* 2019;19(1):506.
- Sirdah MM, Shubair ME, Al-Kahlout MS, Al-Tayeb JM, Prchal JT, Reading NS. Possible association of 3' UTR+ 357 A> G, IVS11-nt 93 T> C, c. 1311 C> T polymorphism with G6PD deficiency. *Hematology.* 2017;22(6):370–4.
- Amro SAB, Zaabi E, Hussain S, Aly AM, Baqir HS, Zaki A-H, et al. Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in Abu Dhabi District, United Arab Emirates. *Trop J Pharm Res.* 2014;13(5):731–7.
- Al-Sweedan SA, Awwad N. Molecular characterization of glucose-6-phosphate dehydrogenase deficiency among Jordanians. *Acta Haematol.* 2012;128(4):195–202.
- Al-Musawi BM, Al-Allawi N, Abdul-Majeed BA, Eissa AA, Jubrael JM, Hamamy H. Molecular characterization of glucose-6-phosphate dehydrogenase deficient variants in Baghdad city-Iraq. *BMC Blood Disord.* 2012;12(1):4.
- Daar S, Vulliamy TJ, Kaeda J, Mason PJ, Luzzatto L. Molecular characterization of G6PD deficiency in Oman. *Hum Hered.* 1996;46(3):172–6.
- Sirdah M, Reading NS, Perkins SL, Shubair M, Aboud L, Prchal JT. Hemolysis and Mediterranean G6PD mutation (c. 563 C> T) and c. 1311 C> T polymorphism among Palestinians at Gaza Strip. *Blood Cells Mol Dis.* 2012;48(4):203–8.
- Bautista J, Mason PJ, Luzzatto L. Human glucose-6-phosphate dehydrogenase Lysine 205 is dispensable for substrate binding but essential for catalysis. *FEBS Lett.* 1995;366(1):61–4.
- Gómez-Manzo S, López-Velázquez G, García-Torres I, Hernández-Alcantara G, Méndez-Cruz ST, Marcial-Quino J, et al. Deficiencia de glucosa-6-fosfato deshidrogenasa: De lo clínico a lo bioquímico. *Acta Bioquímica Clín Latinoamericana.* 2014;48(4):409–20.
- Mason PJ, Bautista JM, Gilsanz F. G6PD deficiency: the genotype-phenotype association. *Blood Rev.* 2007;21(5):267–83.
- Minucci A, Moradkhani K, Hwang MJ, Zuppi C, Giardina B, Capoluongo E. Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the "old" and update of the new mutations. *Blood Cells Mol Dis.* 2012;48(3):154–65.
- Roper D, Layton M, Rees D, Lambert C, Vulliamy T, De la Salle B, et al. Laboratory diagnosis of G6PD deficiency. A British Society for Haematology Guideline. *Br J Haematol.* 2020;189(1):24–38.
- Ley B, Luter N, Espino FE, Devine A, Kalnoky M, Lubell Y, et al. The challenges of introducing routine G6PD testing into radical cure: a workshop report. Springer; 2015.

Publisher's Note

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