



The ethnogeographic variability of genetic factors underlying G6PD deficiency

Maria Koromina^{a,b}, Maria Theodora Pandi^{a,c}, Peter J. van der Spek^c, George P. Patrinos^{a,d,e}, Volker M. Lauschke^{f,g,*}

^a University of Patras, School of Health Sciences, Department of Pharmacy, Laboratory of Pharmacogenomics and Individualized Therapy, Patras, Greece

^b The Golden Helix Foundation, London, UK

^c Erasmus University Medical Center, Faculty of Medicine and Health Sciences, Department of Pathology, Bioinformatics Unit, Rotterdam, Netherlands

^d United Arab Emirates University, College of Medicine and Health Sciences, Department of Pathology, Al-Ain, UAE

^e United Arab Emirates University, Zayed Center of Health Sciences, Al-Ain, UAE

^f Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden

^g Dr Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany

ARTICLE INFO

Keywords:

Population genomics
Next generation sequencing
G6PD deficiency
Precision medicine
Genomic medicine

ABSTRACT

Glucose-6-phosphate dehydrogenase (G6PD) deficiency caused by genetic variants in the *G6PD* gene, constitutes the most common enzymopathy worldwide, affecting approximately 5% of the global population. While carriers are mostly asymptomatic, they are at substantial risk of acute hemolytic anemia upon certain infections or exposure to various medications. As such, information about G6PD activity status in a given patient can constitute an important parameter to guide clinical decision-making. Here, we leveraged whole genome sequencing data from 142,069 unrelated individuals across seven human populations to provide a global comprehensive map of *G6PD* variability. By integrating established functional classifications with stringent computational predictions using 13 partly orthogonal algorithms for uncharacterized and novel variants, we reveal the large extent of ethnogeographic variability in G6PD deficiency and highlight its population-specific genetic composition. Overall, estimated disease prevalence in males ranged between 12.2% in Africans, 2.7–3.5% across Asia and 2.1% in Middle Easterners to < 0.3% in Europeans, Finnish and Amish. In Africans, the major deficient alleles were A^{-202A/376 G} (minor allele frequency 11.6%) and A^{-968 C/376 G} (0.5%). In contrast, G6PD deficiency in Middle Easterners was primarily due to the Mediterranean allele (1.3%) and the population-specific Cairo variant (0.4%). In South Asia, the most prevalent deficient alleles were Mediterranean (1.7%), Kerala (1.1%), Gond (0.9%) and Orissa (0.2%), whereas in East Asian populations the Canton (1.1%), Kaiping (0.7%) and Viangchan (0.3%) variants were predominant. Combined, our analyses provide a large dataset of *G6PD* variability across major ethnogeographic groups and can instruct population-specific genotyping strategies to optimize genetically guided therapeutic interventions.

1. Introduction

The glucose-6-phosphate dehydrogenase (*G6PD*) gene encodes the central enzyme of the pentose phosphate pathway that provides NADPH and contributes to glutathione homeostasis and redox balance. G6PD deficiency is a hereditary monogenic defect affecting over 400 million people worldwide that caused by mutations or genetic variants in *G6PD*, leading to reduced enzyme activity and a variety of biochemical and clinical phenotypes [1,2]. In addition to an increased risk of neonatal hyperbilirubinemia, G6PD deficient individuals are at risk for

developing severe acute haemolysis after the intake of certain drugs and 37 drugs contain warnings about their use in G6PD deficient individuals in their labels (<https://www.fda.gov/drugs/science-and-research-drug-s/table-pharmacogenomic-biomarkers-drug-labeling>).

G6PD variants are categorized by the World Health Organization (WHO) into four deficiency classes based on their level of enzyme activity. Variants with < 10% activity of normal are classified as class I or II depending on whether they also cause severe enzyme deficiency with chronic non-spherocytic hemolytic anemia (class I) or not (class II). Variants resulting in 10–60% G6PD activity are classified as class III and

* Corresponding author at: Department of Physiology and Pharmacology, Karolinska Institutet, SE-171 77 Stockholm, Sweden.

E-mail address: volker.lauschke@ki.se (V.M. Lauschke).

<https://doi.org/10.1016/j.phrs.2021.105904>

Received 22 July 2021; Received in revised form 11 September 2021; Accepted 14 September 2021

Available online 20 September 2021

1043-6618/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

variants that do not have major impacts on G6PD function are considered as class IV (60–200% of normal). Notably however, this classification included variants that were characterized using different methodologies and accumulating evidence suggests little functional differences that do not allow for a precise categorization, particularly between class II and class III variants [3]. As *G6PD* is localized on the X-chromosome, the pattern of inheritance differs between the sexes, mostly affecting hemizygous males and homozygous females. In contrast, only around 8–20% of heterozygous females exhibit drug-induced acute hemolysis [4–7].

To date, the vast majority of studies into the genetic basis of G6PD deficiency have utilized genotyping approaches in which individual candidate single nucleotide polymorphism (SNPs) were interrogated. In African populations, population-scale studies are generally restricted to the common A- alleles A^{-202A/376 G} and A^{-968 C/376 G} [8–10]. Similarly, studies across Asia focus only on the few variants that are predominant in the respective regions [11–13]. Such approaches however are selective and do not consider novel variants or known variants that are regarded as too rare in the given population. The increasing availability of Next-Generation Sequencing (NGS) data generated by large scale population sequencing projects enables, for the first time, the comprehensive and systematic analysis of *G6PD* variability and its ethnogeographic profiles beyond commonly interrogated SNPs.

Here, we analyzed whole exome and whole genome sequencing (WES and WGS, respectively) data from 142,069 unrelated individuals and provide consolidated population-specific overviews of clinically important *G6PD* allele frequencies on an unprecedented scale. Furthermore, by integrating variants with clinically established G6PD deficiency with stringent functionality predictions for variants with unknown pathogenicity, we derive comprehensive maps of G6PD genotypes and phenotypes across eight major human populations. We anticipate that the presented data will provide a powerful resource for clinical geneticists, drug developers and public health experts to inform population-specific genotyping strategies and facilitate the implementation of precision public health.

2. Methods

2.1. Data collection and definitions

Single-nucleotide variant (SNV) and indel frequency data across *G6PD* were collected from WES and WGS data from 141,456 unrelated individuals (including 12,487 Africans, 17,720 Latinos, 5185 Ashkenazi Jews, 9977 East Asians, 15,308 South Asians, 12,562 Finns, 64,603 non-Finnish Europeans, and 3614 with unclear or other ethnicity) from the Genome Aggregation Database (gnomAD) v2 [14]. These data were furthermore complemented with 455 Amish and 158 Middle Easterners from gnomAD v3, resulting in a total sample size of 142,069 individuals. An overview of the respective cohorts that are integrated in gnomAD is available on the project website (<https://gnomad.broadinstitute.org/about>) and further ethnogeographic annotations of the cohorts are provided in the [Supplementary Methods](#). The Novel variants were defined as relative to dbSNP version 154. Genomic coordinates are indicated according to GRCh37. For each variant, the transcript that yielded the most severe consequence as defined by Sequence Ontology [15] was retained. PharmGKB and ClinVar were used to extract information about the pathogenicity of *G6PD* variants and haplotypes. Variants with minor allele frequencies (MAFs) above 1% were considered as common, while variants below 1% or below 0.1% were considered as rare and very rare, respectively.

2.2. Variant annotation

Variant annotation was conducted by using Variant Effect Predictor (VEP). To predict the functional consequences of missense variants, we used a panel of 13 partly orthogonal algorithms that consider sequence

conservation, genomic features, physicochemical amino acid properties, solvent accessibility, and structural features. Specifically, we selected SIFT [16], Polyphen2 [17], MutationAssessor [18], BayesDel [19], CADD [20], DANN [21], fathmm_mkl [22], Condel [23], MetaLR and MetaSVM [24], PROVEAN [25], DEOGEN2 [26] and ClinPred [27]. Frameshift and stop-gain variants were evaluated using LOFTEE (<https://github.com/konradjk/loftee>).

2.3. Evaluation of variant pathogenicity

Variant functionality was assessed using a tiered model that considered different levels of evidence (Fig. 1). Variants were first matched against WHO classification data. Specifically, variants classified per WHO as class I, II or III were considered as deficient, while class IV variants were assigned normal activity. Only those variants for which no WHO classification was established, were in a second step matched against ClinVar annotations. Variants annotated as pathogenic or likely pathogenic by ClinVar were classified as G6PD deficient, whereas benign or likely benign variants were considered as neutral. Only for those variants for which no classification data was available, computational algorithms were used. To establish the optimal classification threshold, we evaluated the performance of different combinations of algorithms (Supplementary Figure 1). Based on these results, variants were classified as G6PD deficient if they were flagged as deleterious by at least 10 out of 13 algorithms (for missense variants), a threshold which achieved a sensitivity of 56% and a specificity of 100% on a set of 43 and 21 variants with known pathogenic and benign effects, respectively. Furthermore, variants were considered deficient if were annotated as loss-of-function by LOFTEE (for frameshift and stop-gain variants). Otherwise, variants were annotated as functionally neutral.

To estimate the distribution of G6PD activity across populations, we calculated the aggregated frequency of G6PD deficient alleles in a given population (q) based on the independence probability theory as previously reported [28] as $q^2 = [1 - \prod_{i=1}^n (1 - f_i)]^2$, with f_i being the MAF of every pathogenic allele $i = \{1, \dots, n\}$ in the given population. Males hemizygous for a reduced function variant (class I to III) were assumed to be G6PD deficient. Females were considered G6PD deficient if they were homozygous or compound heterozygous for a reduced function variant (class I to III). Furthermore, we considered in our calculations a fraction of heterozygous females as G6PD deficient based on previous

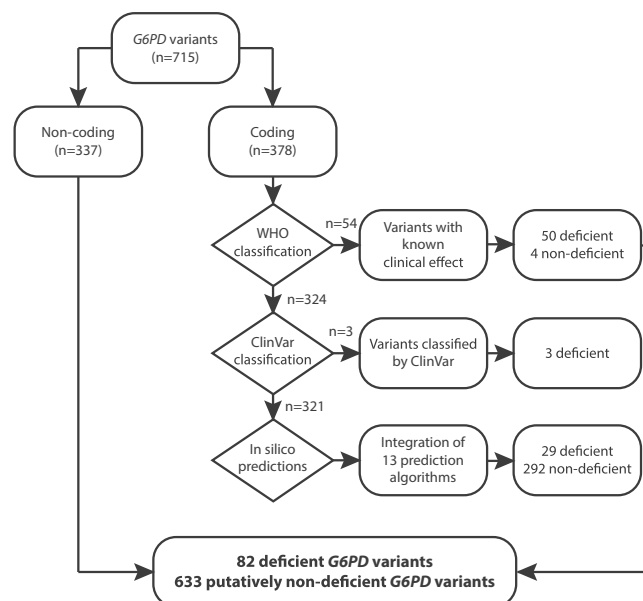


Fig. 1. Flowchart for variant pathogenicity assessment.

genotype-phenotype association studies. As estimates differ, we provide ranges using the two scenarios of assuming 8% [5–7] and 20% [4] G6PD deficiency among heterozygous females.

3. Results

3.1. The genetic landscape of the human G6PD locus

Across 142,069 individuals, we identified a total of 715 variants within the *G6PD* locus. A high percentage of those (337 variants; 47.1%) affected introns, or non-coding regions upstream or downstream of transcriptional start and stop sites, respectively (Fig. 2A). While these can have potentially relevant regulatory functions, current algorithms do not allow to accurately predict their functional impacts and these non-coding variants were thus omitted from further analyses. Of the remaining 378 coding variants, the majority caused amino acid exchanges (205 variants), followed by synonymous variants (129 variants) and variations in the untranslated regions (UTRs) of the transcript (35 variants; Fig. 2A, inset). Furthermore, few variants were identified that introduced stop codons (3 variants), caused frameshifts (2 variants) or affected splice sites (1 variant).

Notably, only 8 variants were common with a MAF > 1%, whereas approximately 40% were singletons only identified in a single allele (Fig. 2B). Moreover, we observed a large number of population-specific *G6PD* variants. Specifically, 225 variants were exclusively found in Europeans, while Africans and Asian populations harbored 72 and 124 population-specific variants, respectively (Fig. 2C). Latinos harbored 56 population-specific variants, while only 5, 2 and 1 unique variants were identified in Ashkenazi Jews, Middle Easterners and Amish, respectively.

3.2. Global frequencies of functionally relevant G6PD polymorphisms

To understand the distribution of genetic risk factors for G6PD deficiency, we first analyzed the frequencies of G6PD alleles with established functional consequences (Table 1). Within the analyzed dataset, rs1050829 was the most common variant that, if occurring in isolation, encodes the “A” allele, which does not impact G6PD functionality (class IV). Frequency of “A” was highest in Africans (MAF=19.5%) and Latino populations (MAF=1%), whereas it was absent in individuals of European descent and East Asians. Among the reduced function variants, A^{-202A/376 G} and A^{-968 C/376 G} were most common, particularly in Africans (MAFs of 11.6% and 0.5%, respectively) and, to a lesser extent, in admixed Latinos (MAFs of 0.4% and 0.08%, respectively). In Europeans, the highest frequencies were observed for the Seattle variant (MAF=0.11%) and the Mediterranean allele (MAF=0.07%). As expected, in genetically more homogeneous

populations, the pattern of variants that form the molecular basis for G6PD deficiency was less complex. In Finnish and Ashkenazim, the only alleles with known deficiency were Seattle (MAF=0.09%) and Mediterranean (MAF=0.6%), respectively, whereas no deficient alleles were identified in Amish.

In Asia, the molecular composition of G6PD deficiency is diverse and a multitude of alleles have been identified. Here, we found that the Canton (MAF=1.1%) and Kaiping (MAF=0.7%) variants were the most common causes of G6PD deficiency in East Asia, whereas the Mediterranean (MAF=1.74%), Kerala (MAF=1.14%) and Gond (MAF=0.87%) alleles were most frequent in South Asia. G6PD deficiency in the Middle Eastern population was attributed to both European (Mediterranean; MAF=1.26%) and African (A^{-202A/376 G}; MAF=0.42%) alleles, as well as unique variants exclusively found in this ethnogeographic group (Cairo; MAF=0.42%). Combined, these results consolidate information about the prevalence of well-described G6PD risk alleles and illustrate the remarkable diversity of *G6PD* molecular genetics.

3.3. Ethnogeographic distribution of genetically encoded G6PD deficiency

Next, we integrated information about the previously known deficient *G6PD* alleles described above, with frequency data of variants that were identified as putatively deficient using 13 partly orthogonal *in silico* tools. Notably, computational predictions identified 29 variants (23 missense, 3 stop-gain, 2 frameshift and 1 splice variant) with putative functional impacts that had not been characterized previously.

Aggregation of all deficient alleles and variants revealed that genetically encoded G6PD deficiency was highest among African males with a prevalence of 12.4%, followed by African females with 3.3–5.9% (Fig. 3A; Table 2). Notably, the vast majority of G6PD deficiency in Africa (93.5%) was attributed to the A^{-202A/376 G} allele (Fig. 3B). Besides African populations, G6PD deficiency was common across Asia with prevalence between 4.4% in South Asian males and 0.5–1.2% in East Asian females. The genetic composition differed drastically between East and South Asian populations and was considerably more complex than in Africa. In South Asia, the three most common G6PD deficient alleles (Mediterranean, Kerala and Gond) combined explained around 85% of G6PD deficiencies. Similarly, interrogation of five variant alleles (Canton, Kaiping, Viangchan, Gaohe and Chinese-5) was required in East Asia to capture 85% of G6PD deficiencies. Furthermore, there was only minimal overlap between the common G6PD deficient alleles in South and East Asia.

In the Middle East, overall G6PD deficiency pivoted between 2.1% in males and 0.4–0.9% in females and was allotted to the Mediterranean (60.1%), Cairo (20%) and A^{-202A/376 G} (19.9%) alleles. While the prevalence of G6PD deficient alleles in Latinos was overall considerably lower than in African populations (0.9% in males and 0.2–0.4% in

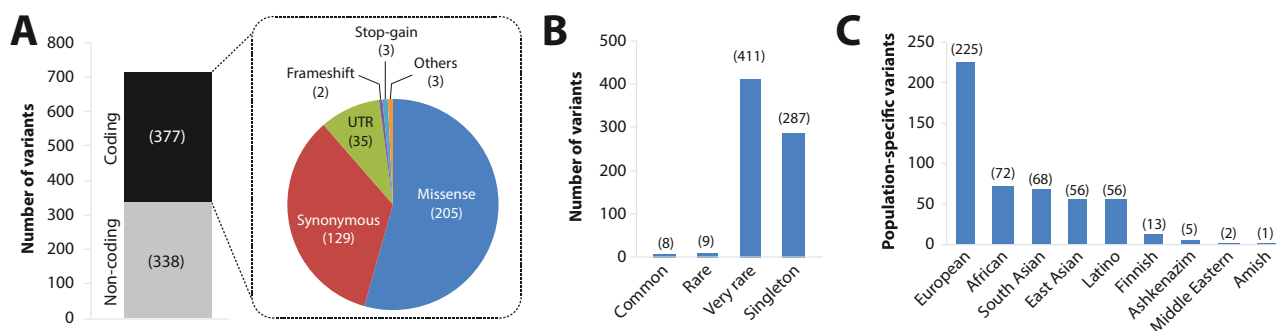


Fig. 2. Overview of the *G6PD* genetic inventory. A, Across 142,069 individuals, a total of 338 different non-coding (grey) and 377 coding (black) variants were identified. Inset: pie chart showing the composition of coding *G6PD* variations by variant class. B, Overview of the frequency distribution of all identified *G6PD* variants. Note that the vast majority of variants (97.6%) were very rare with minor allele frequencies < 0.01% or singletons. C, Of the total 715 variants, 498 (69%) were only identified in a single population. Most population-specific variants were identified in Europeans, likely due to their relative overrepresentation in the available genomic dataset.

Table 1

Variant frequencies of well characterized *G6PD* alleles across ethnogeographic groups. Functionally relevant alleles with frequencies above 0.1% in a given population are indicated in bold. Genomic coordinates provided according to GRCh38. *N* indicates the number of analyzed alleles for a given population. AMR = Latino; AFR = African; ASJ = Ashkenazi Jewish; EAS = East Asian; FIN = Finnish; NFE = Non-Finnish Europeans = ME = Middle Eastern; SAS = South Asian.

Trivial allele name	rsID or genomic coordinate	Protein change	Variant allele frequency (in %)									
			Global	AFR (<i>N</i> = 24,974)	NFE (<i>N</i> = 129,206)	AMI (<i>N</i> = 912)	ASJ (<i>N</i> = 10,370)	FIN (<i>N</i> = 25,124)	AMR (<i>N</i> = 35,440)	ME (<i>N</i> = 316)	EAS (<i>N</i> = 19,954)	SAS (<i>N</i> = 30,616)
Severe												
Montpellier	rs371489738	p.G378S	0.001	0	0.003	0	0	0	0	0	0	0
Mediterranean Canton	rs5030868 rs72554665	p.S188F p.R459L	0.23 0.08	0.02 0	0.07 0	0 0	0.6 0	0 0	0 0	0 0.004	0 0	1.74 0
Kaiping	rs72554664	p.R463H	0.05	0	0	0	0	0	0	0	0	0.7
Viangchan	rs137852327	p.V291M	0.02	0	0	0	0	0	0	0	0	0.25
Chatham	rs5030869	p.A335T	0.02	0	0.007	0	0	0	0	0	0	0.13
Union	rs398123546	p.R454C	0.01	0.04	0.005	0	0	0	0.01	0	0	0.06
Taipei, Chinese-3	rs137852331	p.N165D	0.006	0	0	0	0	0	0	0	0	0.08
Coimbra	rs137852330	p.R198C	0.003	0	0.001	0	0	0	0	0	0.007	0.02
Shunde												
Cosenza	rs72554665	p.R459P	0.002	0	0.002	0	0	0	0	0	0.03	0
Kamogawa	chrX:154536035	p.R57W	0.001	0	0	0	0	0	0.009	0	0	0
Liuzhou	rs119197862	p.E148K	0.001	0.003	0	0	0	0	0	0	0	0
Ludhiana	chrX:154533064	p.G310E	0.001	0	0	0	0	0	0	0	0	0.04
Radlowo	rs1557230213	p.R227W	0.001	0.008	0	0	0	0	0	0	0	0
Vanua Lava	rs78365220	p.L128P	0.001	0	0	0	0	0	0	0	0.03	0
Moderate												
A ^{-202A/376 G}	rs1050828 and rs1050829	p.V68M and p.N126D	3.11	11.64	0.02	0	0	0	0	0.41	0.42	0
Kerala	rs137852339	p.E317K	0.11	0	0.005	0	0	0	0	0	0	1.14
Gond	rs370918918	p.M159I	0.08	0	0.005	0	0	0	0.004	0	0	0.87
Seattle	rs137852318	p.D282H	0.07	0.01	0.11	0	0	0.09	0.03	0	0	0.07
A ^{-968 C/376 G}	rs1050829 and rs76723693	p.N126D and p.L323P	0.06	0.5	0	0	0	0	0.08	0	0	0
Ilesha	rs137852313	p.E156K	0.03	0.05	0.001	0	0	0	0.12	0	0	0.04
Sierra Leone	rs181277621 and rs1050829	p.R104H and p.N126D	0.02	0.02	0.02	0	0	0	0	0	0	0.04
Orissa	rs78478128	p.A44G	0.02	0	0	0	0	0	0	0	0	0.16
Chinese-5	rs137852342	p.L342F	0.01	0.005	0	0	0	0	0	0	0.17	0
Malaga	rs5030872	p.D181V	0.01	0.11	0.002	0	0	0	0.03	0	0	0
Gaohe	rs137852340	p.H32R	0.01	0	0	0	0	0	0	0	0.2	0
Quing Yan	rs137852341	p.G131V	0.007	0	0	0	0	0	0	0	0.1	0
Mexico City	rs137852328	p.R227Q	0.005	0	0.005	0	0	0	0.02	0	0	0
Mahidol	rs137852314	p.G163S	0.004	0	0	0	0	0	0	0	0.007	0.04
Ube, Konan	rs138687036	p.R81C	0.002	0.01	0.004	0	0	0	0.007	0	0.007	0
Nanning	rs782757170	p.L235F	0.002	0	0	0	0	0	0	0	0.03	0
Cairo	rs782322505	p.N135T	0.001	0	0	0	0	0	0	0.42	0	0
Montalbano	rs74575103	p.R285H	0.001	0.008	0	0	0	0	0	0	0	0
Sibari	rs782754619	p.M212V	0.001	0.008	0.001	0	0	0	0	0	0	0
Kamiube, Keelung	rs1557229502	p.R463C	0.001	0	0.002	0	0	0	0	0	0	0
Lagosanto	rs782308266	p.R81H	0.001	0	0.001	0	0	0	0	0	0	0
Aures	rs76645461	p.I48T	0.001	0	0	0	0	0	0	0	0.01	0
Palestina	rs1000937138	p.R57Q	0.001	0.003	0	0	0	0	0	0	0	0
Sinnai	rs797043472	p.V12L	0.001	0	0.002	0	0	0	0	0	0	0
Normal												
A	rs1050829	p.N126D	3.44	19.52	0	0	0.01	0	1	0.002	0	0.04
Mira d' Aire	rs34193178	p.D350H	0.05	0.39	0.001	0	0	0	0.06	0	0	0
Sao Borja	rs5030870	p.D113N	0.003	0.003	0	0	0	0	0.02	0	0	0

females), the molecular genetics were similar with A^{-202A/376 G} being most common, accounting for 45.1% of cases. In addition, Latinos prominently feature population-specific variants (Ilesha and Mexico City). In populations of European origin (Ashkenazi Jews, Finnish and non-Finnish Europeans), the Seattle, Mediterranean and the not characterized putatively deleterious stop-gain variant rs202181505 were most common and jointly accounted for 65–99% of cases.

4. Discussion

G6PD deficiency is considered the most common genetic disorder,

affecting approximately 5% of the world's population [29] and its enzymatic basis and X-chromosomal mode of inheritance are known since almost 70 years [30]. Importantly, *G6PD* deficiency exhibits conspicuous variability across ethnic groups. Frequencies are highest in Sub-Saharan Africa and Southeast Asia with reported prevalence rates ranging from 1% to 30%, whereas frequencies in Europe and Japan are considerably lower (0–10%). Besides various histological and enzymatic assessments, genetic testing of risk variants constitutes a common strategy to predict *G6PD* deficiency.

Over the last decades, a multitude of studies have evaluated the prevalence of *G6PD* variants across many different regions and

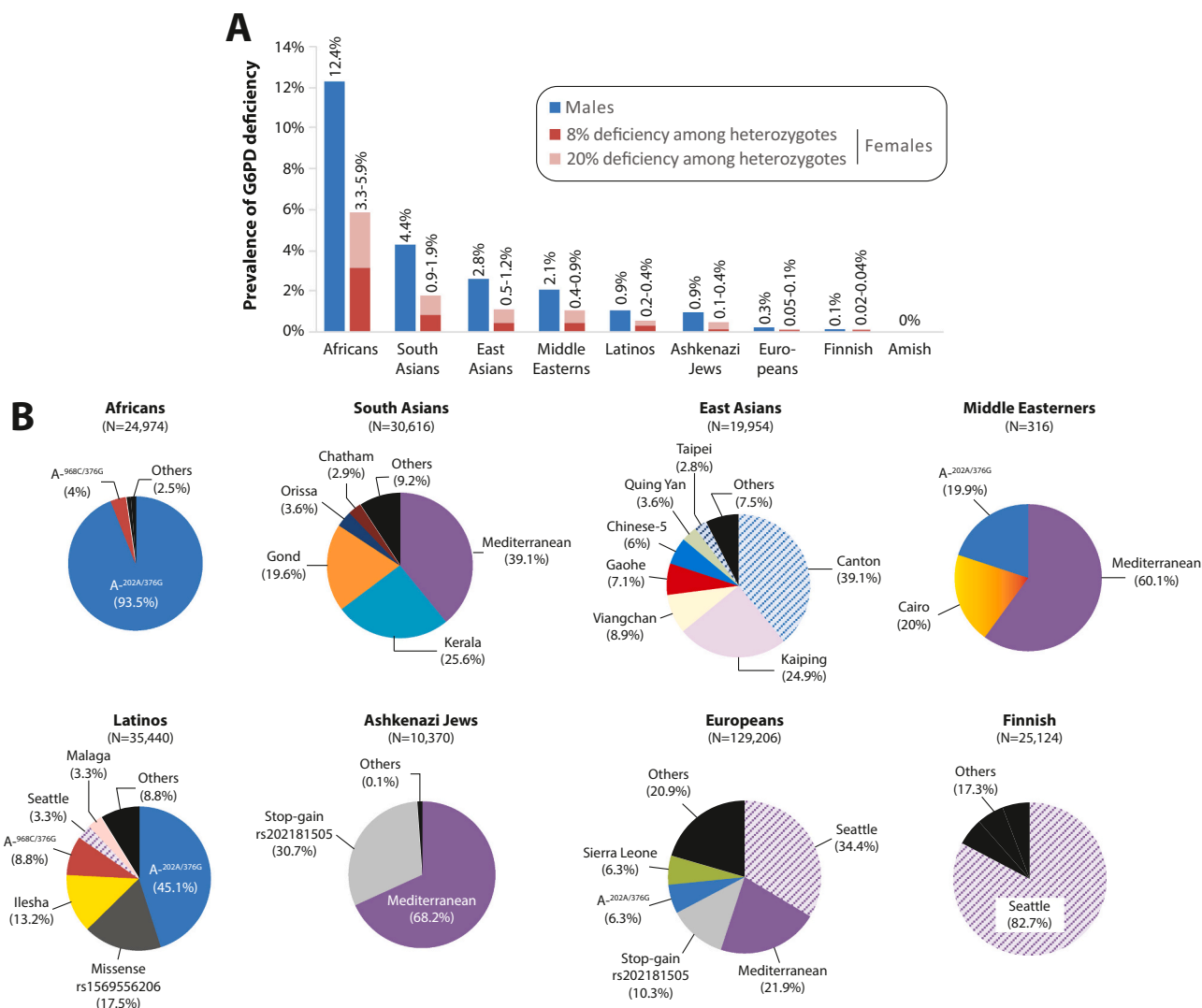


Fig. 3. The distribution and molecular genetics of G6PD deficiency differs drastically between populations. A, The overall prevalence of G6PD deficiency is shown for males and females for each population analyzed based on both well-characterized deficient variants and the variants predicted as reduced function using stringent computational assessments (see methods section for details). The confidence intervals for the respective estimates are provided in Table 2. Frequencies of G6PD deficiency range between 12.4% in African males to 0% in Amish. B, Pie charts showing the genetic composition of population-specific G6PD deficiency. Note the drastic differences in genetic complexity between populations. While two reduced function alleles explain >97% of G6PD deficiencies in Africa, disease risk is distributed across a multitude of different variants in Asian populations and Latinos. The total number of analyzed alleles (*N*) is shown for each population.

ethnogeographic groups, naturally with a focus on populations in which deficiency is most prevalent, such as Sub Saharan Africa [31,32], West Africa [33], Kenya [34], Sudan [35], Sierra Leone [36], Gabon [37], Mauritania [38], Botswana [39], Cameroon [40], Saudi Arabia [41], Mexico [42], Colombia [43], the Dominican Republic [44], Iraqi Kurds [45], Pashtun in Afghanistan [46], India [47,48], China [49–52], Laos [53] and Thailand [54]. Yet, the molecular genetics of *G6PD* is complex, and most studies evaluate only one or few candidate variants that are most common in the population under study. To consolidate the available data, various literature surveys and meta-analyses have been conducted [55–57]. However, these are limited to variants for which a sufficient number of analyses are available and study integration can be problematic due to differences in genotyping strategies and assay panels. Furthermore, current studies focused on specific populations and often analyzed cohorts with differing severity of malaria infections. Given that G6PD deficiency has been associated with protective effects against malaria [58,59], analyses of *Plasmodium* infected cohorts might not accurately represent the frequency of *G6PD* variability in the population at large.

The increased amount of available population-scale sequencing data allows for the first time to systematically derive frequency estimates of all *G6PD* variants. Using fully consistent and compatible data from 142,069 unrelated individuals across seven ethnogeographic groups, we map the prevalence of WHO-classified *G6PD* risk alleles and pinpoint variants for population-adjusted genotyping strategies. Furthermore, by using stringent computational assessments based on 13 partly orthogonal algorithms, we provide estimates for the relative contributions of previously uncharacterized or novel *G6PD* variants. The resulting dataset provides, to the best of our knowledge, the most comprehensive overview of *G6PD* variability presented to date with potentially important implications for the optimization of precision public health strategies. Of note, while very rare, class I deficiencies that are characterized by chronic transfusion-dependent anemia, are likely underrepresented in the analyzed data set, as patients with clinically manifest signs of genetic disease were excluded during recruitment.

To generate representative allelic frequency data for different ethnogeographic groups multiple strategies can be pursued. Firstly, the composition of the analyzed cohort can be balanced based on the

Table 2

Estimated frequencies of population-specific G6PD deficiency. Two-sided 90% confidence intervals were calculated based on the binomial proportion. Asterisks indicate one-sided confidence intervals due to a measured frequency of zero. AMR = Latino; AFR = African; ASJ = Ashkenazi Jewish; EAS = East Asian; FIN = Finnish; NFE = Non-Finnish Europeans = ME = Middle Eastern; SAS = South Asian.

Population	Allele number (N)	G6PD Deficiency Frequency (90% confidence interval)		
		Males	Females (assuming 8% of het females to be G6PD deficient)	Females (assuming 20% of het females to be G6PD deficient)
AFR	24,974	12.45% (12.11%; 12.8%)	3.29% (3.11%; 3.49%)	5.91% (5.67%; 6.16%)
NFE	129,206	0.32% (0.04%; 0.06%)	0.05% (0.04%; 0.06%)	0.13% (0.11%; 0.15%)
AMI	912	0% (0%; 0.33%)*	0% (0%; 0.33%)*	0% (0%; 0.33%)*
ASJ	10,370	0.88% (0.73%; 1.04%)	0.15% (0.09%; 0.22%)	0.36% (0.27%; 0.47%)
FIN	25,124	0.11% (0.08%; 0.15%)	0.02% (0.01%; 0.04%)	0.04% (0.02%; 0.07%)
AMR	35,440	0.91% (0.83%; 1%)	0.15% (0.12%; 0.19%)	0.37% (0.32%; 0.43%)
ME	316	2.1% (1.04%; 4.12%)	0.37% (0.02%; 1.49%)	0.86% (0.26%; 1.31%)
EAS	19,954	2.82% (2.63%; 3.02%)	0.52% (0.44%; 0.61%)	1.17% (1.05%; 1.31%)
SAS	30,616	4.42% (4.23%; 4.62%)	0.87% (0.79%; 0.96%)	1.89% (1.76%; 2.02%)

absolute numbers of individuals belonging to its subpopulations. This approach attempts to approximate the likelihood of identifying a carrier of a given variant if a random individual from the superpopulation is blindly sampled. However, the disadvantage of this approach is that very populous populations dominate the overall result whereas minority populations have little impact. Alternatively, data can be averaged across populations irrespective of the size of the respective group. This strategy is very sensitive to population definitions and requires a substantial number of groups to not skew frequencies in the direction of outlier populations. Furthermore, both strategies have problems in accounting for individuals with admixed ancestry, which account for a major fraction of many global populations [60]. Based on these limitations, we here followed a third pragmatic approach that utilizes all available data of sufficient quality. This strategy allows for the incorporation of admixed individuals and individuals for which the geographic sampling location has not been reported, thus making maximal use of the data; however, its downsides are that frequencies are sensitive to the underlying sampling strategies and that ethnogeographic resolution is mostly limited to aggregated superpopulations. Based on these considerations, we think that large population-scale data, as presented here, should complement rather than replace smaller studies into more specific populations to provide increasing information about variant frequencies in different populations across multiple levels of granularity.

Genotype-to-phenotype concordance of G6PD deficiency is limited, particularly for heterozygous females. While G6PD activity shows a bimodal distribution in males that clearly separates hemizygotously deficient and enzymatically normal individuals, these differences are blurred in females in which the activity pattern is continuous and normally distributed [61]. The cause of this phenomenon is random

X-inactivation as the mechanism for dosage compensation in females. Consequently, heterozygous females do not have approximately 50% of G6PD activity in each cell but each erythrocyte is rather enzymatically normal or G6PD deficient, resulting in two distinct cell populations. This entails that upon exposure to a trigger of acute hemolytic anemia, only the deficient population will be broken down. The proportion of normal vs. deficient erythrocytes shows substantial inter-individual variability, giving rise to variable severity of hemolytic anemia in heterozygous females. Here, we accounted for these events by considering 8–20% of heterozygous women to be clinically G6PD deficient based on previous genotype-phenotype correlations [4–7].

The presented findings have important implications for pharmacogenomics and personalized medicine. Due to the geographic overlap between G6PD deficiency and malaria endemic regions, low G6PD activity is particularly relevant for the treatment with antimalarials of the 8-aminoquinolines class, such as primaquine, that can induce hemolytic anemia. In contrast, there is overwhelming evidence that 4-aminoquinolines, such as chloroquine and hydroxychloroquine, are safe in G6PD deficient individuals [62]. Examples of other drugs for which G6PD deficiency is clinically relevant include the rasburicase, used for the treatment of acute hyperuricemia, the nonsteroidal antiandrogen flutamide, as well as sulfonamide antibiotics. Current quantitative assessment methods for G6PD deficiency are inaccurate, particularly for heterozygous females with intermediate phenotypes, or are difficult to conduct in a low-resource setting [63]. Thus, further improvements in diagnostic tests are required to improve the predictive accuracy. From a public health perspective, it is furthermore important to establish whether newborn screening for G6PD deficiency constitutes a cost-effective allocation of health care resources and, if so, to reevaluate whether screening should be conducted irrespective of ethnicity and which genotyping strategy should be selected. Currently, the WHO recommends screening of all newborns in populations with a prevalence of more than 3% in males and the presented data might provide updated guidance for the selection of such ethnogeographic groups.

In conclusion, our analyses comprehensively mapped the genetic variability of the *G6PD* locus across major aggregated populations. These data provide a valuable resource for the ethnogeographic distribution of *G6PD* deficient alleles, promote the integration of comprehensive NGS-based genotyping and highlight key variants for population-adjusted *G6PD* genotyping strategies.

Conflict of interest

V.M.L. is CEO and shareholder of HepaPredict AB, co-founders and shareholder of PersoMedix AB, and discloses consultancy work for Enginzyme AB. G.P.P. is Full Member and National Representative at the European Medicines Agency, Committee for Human Medicinal Products (CHMP) – Pharmacogenomics Working Party, Amsterdam, the Netherlands. The other authors declare no conflict of interest.

Acknowledgments

V.M.L. is funded by the Swedish Research Council [grant agreement numbers: 2016–01153, 2016–01154 and 2019–01837], by the EU/EFPIA/OICR/McGill/KTH/Diamond Innovative Medicines Initiative 2 Joint Undertaking (EUBOPEN grant number 875510) and by the Robert Bosch Foundation, Stuttgart, Germany. G.P.P. is funded by the European Union's Horizon 2020 research and innovation program Ubiquitous Pharmacogenomics (grant agreement number 668353). The authors also wish to acknowledge Mrs Zoi Gkizarioti and Mr Konstantinos Petropoulos for providing useful literature material regarding the assessed G6PD haplotypes.

CRediT authorship contribution statement

Maria Koromina: Data curation, Formal analysis, Methodology,

Software, Writing - original draft; Writing - review & editing, **Maria Theodora Pandi**: Methodology, Software, **Peter J. van der Spek**: Supervision, **George P. Patrinos**: Conceptualization, Funding acquisition, Supervision, Writing - review & editing, **Volker M. Lauschke**: Conceptualization, Funding acquisition, Supervision, Visualization, Writing - review & editing.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.phrs.2021.105904](https://doi.org/10.1016/j.phrs.2021.105904).

References

- [1] K.D. Belfield, E.M. Tichy, Review and drug therapy implications of glucose-6-phosphate dehydrogenase deficiency, *Am. J. Health-Syst. Pharm.* 75 (3) (2018) 97–104.
- [2] L. Luzzatto, M. Ally, R. Notaro, Glucose-6-phosphate dehydrogenase deficiency, *Blood* 136 (11) (2020) 1225–1240.
- [3] G. Bancone, C.S. Chu, G6PD variants and haemolytic sensitivity to primaquine and other drugs, *Front. Pharmacol.* 12 (2021), 638885.
- [4] M.K. Johnson, T.D. Clark, D. Njama-Meya, P.J. Rosenthal, S. Parikh, Impact of the method of G6PD deficiency assessment on genetic association studies of malaria susceptibility, *PLoS One* 4 (9) (2009) 7246.
- [5] C.S. Chu, G. Bancone, F. Nosten, N.J. White, L. Luzzatto, Primaquine-induced haemolysis in females heterozygous for G6PD deficiency, *Malar. J.* 17 (1) (2018) 101.
- [6] S. Dechyotin, K. Sakunthai, N. Khemtonglang, S. Yamsri, K. Sanchaisuriya, K. Kitcharoen, S. Kitcharoen, Prevalence and molecular characterization of glucose-6-phosphate dehydrogenase (G6PD) deficiency in females from previously malaria endemic regions in northeastern Thailand and identification of a novel G6PD variant, *Mediterr. J. Hematol. Infect. Dis.* 13 (1) (2021), 2021029.
- [7] A.W. Satyagraha, A. Sadhewa, L.V. Panggalo, D. Subekti, I. Elyazar, S. Soebianto, N. Mahpud, A.R. Harahap, J.K. Baird, Genotypes and phenotypes of G6PD deficiency among Indonesian females across diagnostic thresholds of G6PD activity guiding safe primaquine therapy of latent malaria, *PLoS Negl. Trop. Dis.* 15 (7) (2021), 0009610.
- [8] C. De Araujo, F. Migot-Nabias, J. Guitard, S. Pelleau, T. Vulliamy, R. Ducrocq, The role of the G6PD Aeth376G/968C allele in glucose-6-phosphate dehydrogenase deficiency in the Seerer population of Senegal, *Haematologica* 91 (2) (2006) 262–263.
- [9] S.S. Awandu, J. Raman, T.I. Makhathisa, P. Kruger, J. Frean, T. Bousema, J. Niemand, L.M. Birkholtz, Understanding human genetic factors influencing primaquine safety and efficacy to guide primaquine roll-out in a pre-elimination setting in Southern Africa, *Malar. J.* 17 (1) (2018) 120.
- [10] J. May, C.G. Meyer, L. Grossterlinden, O.G. Ademowo, F.P. Mockenhaupt, P. E. Olumese, A.G. Falusi, L. Luzzatto, U. Bienzle, Red cell glucose-6-phosphate dehydrogenase status and pyruvate kinase activity in a Nigerian population, *Trop. Med. Int. Health* 5 (2) (2000) 119–123.
- [11] H. Matsuoka, J. Wang, M. Hirai, M. Arai, S. Yoshida, T. Kobayashi, A. Jalloh, K. Lin, F. Kawamoto, Glucose-6-phosphate dehydrogenase (G6PD) mutations in Myanmar: G6PD Mahidol (487G>A) is the most common variant in the Myanmar population, *J. Hum. Genet.* 49 (10) (2004) 544–547.
- [12] H. Matsuoka, C. Nguon, T. Kanbe, A. Jalloh, H. Sato, S. Yoshida, M. Hirai, M. Arai, D. Socheat, F. Kawamoto, Glucose-6-phosphate dehydrogenase (G6PD) mutations in Cambodia: G6PD Viangchan (871G>A) is the most common variant in the Cambodian population, *J. Hum. Genet.* 50 (9) (2005) 468–472.
- [13] P. Phompradit, J. Kuesap, W. Chaijaroenkul, R. Rueangweerayut, Y. Hongkaew, R. Yamnuan, K. Na-Bangchang, Prevalence and distribution of glucose-6-phosphate dehydrogenase (G6PD) variants in Thai and Burmese populations in malaria endemic areas of Thailand, *Malar. J.* 10 (2011) 368.
- [14] K.J. Karczewski, L.C. Francioli, G. Tiao, B.B. Cummings, J. Alfoldi, Q. Wang, R. L. Collins, K.M. Laricchia, A. Ganna, D.P. Birnbaum, L.D. Gauthier, H. Brand, M. Solomonson, N.A. Watts, D. Rhodes, M. Singer-Berk, E.M. England, E.G. Seaby, J.A. Kosmicki, R.K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J.X. Chong, K.E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E.V. Minikel, B. Weisburd, M. Lek, J.S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K.M. Connolly, M. Covarrubias, S. Donnelly, S. Ferreira, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M.E. Talkowski, Genome aggregation database consortium, B.M. Neale, M.J. Daly, D.G. MacArthur, The mutational constraint spectrum quantified from variation in 141,456 humans, *Nature* 581 (7809) (2020) 434–443.
- [15] K. Eilbeck, S.E. Lewis, C.J. Mungall, M. Yandell, L. Stein, R. Durbin, M. Ashburner, The sequence ontology: a tool for the unification of genome annotations, *Genome Biol.* 6 (5) (2005) 44.
- [16] P.C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function, *Nucleic Acids Res.* 31 (13) (2003) 3812–3814.
- [17] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, *Nat. Methods* 7 (4) (2010) 248–249.
- [18] B. Reva, Y. Antipin, C. Sander, Predicting the functional impact of protein mutations: application to cancer genomics, *Nucleic Acids Res.* 39 (17) (2011) 118.
- [19] B.J. Feng, PERCH: a unified framework for disease gene prioritization, *Hum. Mutat.* 38 (3) (2017) 243–251.
- [20] P. Rentzsch, M. Schubach, J. Shendure, M. Kircher, CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores, *Genome Med.* 13 (1) (2021) 31.
- [21] D. Quang, Y. Chen, X. Xie, DANN: a deep learning approach for annotating the pathogenicity of genetic variants, *Bioinformatics* 31 (5) (2015) 761–763.
- [22] H.A. Shihab, M.F. Rogers, J. Gough, M. Mort, D.N. Cooper, I.N. Day, T.R. Gaunt, C. Campbell, An integrative approach to predicting the functional effects of non-coding and coding sequence variation, *Bioinformatics* 31 (10) (2015) 1536–1543.
- [23] A. González-Pérez, N. López-Bigas, Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, *Condel, Am. J. Hum. Genet.* 88 (4) (2011) 440–449.
- [24] C. Dong, P. Wei, X. Jian, R. Gibbs, E. Boerwinkle, K. Wang, X. Liu, Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies, *Hum. Mol. Genet.* 24 (8) (2015) 2125–2137.
- [25] Y. Choi, A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels, *Bioinformatics* 31 (16) (2015) 2745–2747.
- [26] D. Raimondi, I. Tanyalcin, J. Ferté, A. Gazzo, G. Orlando, T. Lenaerts, M. Rooman, W. Vranken, DEOGEN2: prediction and interactive visualization of single amino acid variant deleteriousness in human proteins, *Nucleic Acids Res.* 45 (W1) (2017) W201–w206.
- [27] N. Alirezai, K.D. Kernohan, T. Hartley, J. Majewski, T.D. Hocking, ClinPred: prediction tool to identify disease-relevant nonsynonymous single-nucleotide variants, *Am. J. Hum. Genet.* 103 (4) (2018) 474–483.
- [28] Q. Xiao, V.M. Lauschke, The prevalence, genetic complexity and population-specific founder effects of human autosomal recessive disorders, *NPJ Genomic Medicine* 6 (1) (2021) 41.
- [29] M.D. Cappellini, G. Fiorelli, Glucose-6-phosphate dehydrogenase deficiency, *Lancet* 371 (9606) (2008) 64–74.
- [30] A.S. Alving, P.E. Carson, C.L. Flanagan, C.E. Ickes, Enzymatic deficiency in primaquine-sensitive erythrocytes, *Science* 124 (3220) (1956) 484–485.
- [31] J.E.B. da Rocha, H. Othman, C.T. Tiemessen, G. Botha, M. Ramsay, C. Masimirembwa, C. Adebamowo, A. Choudhury, J.T. Brandenburg, M. Matshaba, G. Simo, F.J. Gamo, S. Hazelhurst, G6PD distribution in sub-Saharan Africa and potential risks of using chloroquine/hydroxychloroquine based treatments for COVID-19! Abstract, *Pharm. J.* (2021), <https://doi.org/10.1038/s41397-021-00242-8>.
- [32] Y.A. Aynalem, G.B. Mulu, T.Y. Akalu, W.S. Shiferaw, Prevalence of neonatal hyperbilirubinaemia and its association with glucose-6-phosphate dehydrogenase deficiency and blood-type incompatibility in sub-Saharan Africa: a systematic review and meta-analysis, *BMJ Paediatr. Open* 4 (1) (2020), 000750.
- [33] S.S. Shah, K.A. Rockett, M. Jallow, F. Sisay-Joof, K.A. Bojang, M. Pinder, A. Jeffreys, R. Craik, C. Hubbard, T.E. Welles, D.P. Kwiatkowski, Heterogeneous alleles comprising G6PD deficiency trait in West Africa exert contrasting effects on two major clinical presentations of severe malaria, *Malar. J.* 15 (2016) 13.
- [34] J.J. Gilchrist, S. Uyoga, M. Pirinen, A. Rautanen, S. Mwarumba, P. Njuguna, N. Mturi, A.V.S. Kenyan Bacteraemia Study Group, J.A.G. Hill, T.N. Scott, Williams, Risk of pneumococcal bacteremia in Kenyan children with glucose-6-phosphate dehydrogenase deficiency, *BMC Med.* 18 (1) (2020) 148.
- [35] M.M.A. Hamid, M. Albsheer, M. Muneer, L. Altinae, A.A. Lover, PO 8609 Prevalence and risk factors for glucose-6-phosphate dehydrogenase (G6PD) deficiency in two *P. vivax* malaria-endemic areas in Sudan, *BMJ Glob. Health* 4 (2019) A62.2–A62.
- [36] A. Jalloh, M. Jalloh, I. Gamanga, D. Baion, F. Sahr, A. Gbakima, V.R. Willoughby, H. Matsuoka, G6PD deficiency assessment in Freetown, Sierra Leone, reveals further insight into the molecular heterogeneity of G6PD A, *J. Hum. Genet.* 53 (7) (2008) 675–679.
- [37] L. Pernaute-Lau, A.A. Adegnik, Y. Zhou, J.F. Zinsou, J.P. Gil, S. Krishna, P. G. Kremsner, V.M. Lauschke, T.P. Velavan, Pharmacogene sequencing of a gabonese population with severe plasmodium falciparum malaria reveals multiple novel variants with putative relevance for antimalarial treatment, *Antimicrob. Agents Chemother.* 65 (7) (2021), 0027521.
- [38] O.K.M. Djigo, M.S. Salem, S.M. Diallo, M.A. Bollahi, B.M. Boushab, A. Garre, N. P. Mze, L. Basco, S. Briolant, A.O. Boukhary, Molecular epidemiology of G6PD genotypes in different ethnic groups residing in saharan and sahelian zones of mauritania, *Pathogens* 10 (8) (2021) 931.
- [39] T. Motshoge, G. Ababio, L. Aleksenko, S. Souda, C.W. Muthoga, N. Mutukwa, L. Tawe, P. Ramatillo, L. Gabaitiri, S. Chihanga, T. Mosweunyane, S. Hamda, K. Moakofi, D. Ntebela, E. Peloeuwetse, L. Mazhani, J.M. Pernice, J. Read, I. K. Quaye, G.M. Paganotti, Prevalence of G6PD deficiency and associated haematological parameters in children from Botswana, *Infect., Genet. Evol.* 63 (2018) 73–78.
- [40] S.M. Lauden, S. Chongwain, A. Achidi, E. Helm, S.E. Cusick, A. Krug, T.M. Slusher, T.C. Lund, Prevalence of glucose-6-phosphate dehydrogenase deficiency in Cameroonian blood donors, *BMC Res. Notes* 12 (1) (2019) 195.
- [41] M.A. Gari, A.G. Chaudhary, M.H. Al-Qahtani, A.M. Abuzenadah, A. Waseem, H. Banni, F.M. Al-Sayes, A. Al-Harbi, S. Lary, Frequency of Mediterranean mutation among a group of Saudi G6PD patients in Western region-Jeddah, *Int. J. Lab. Hematol.* 32 (1 Pt 2) (2010) 17–21.
- [42] M. Vela-Amieva, M.A. Alcantara-Ortigoza, A. Gonzalez-Del Angel, L. Belmont-Martinez, C. Lopez-Candiani, I. Ibarra-Gonzalez, Genetic spectrum and clinical early natural history of glucose-6-phosphate dehydrogenase deficiency in Mexican

- children detected through newborn screening, *Orphanet J. Rare Dis.* 16 (1) (2021) 103.
- [43] S.H. Valencia, I.D. Ocampo, M.I. Arce-Plata, J. Recht, M. Arévalo-Herrera, Glucose-6-phosphate dehydrogenase deficiency prevalence and genetic variants in malaria endemic areas of Colombia, *Malar. J.* 15 (1) (2016) 291.
- [44] J.Z. Xu, R.O. Francis, L.E.L. Nadal, M. Shirazi, V. Jobanputra, E.A. Hod, J.S. Jhang, B.A. Stotler, S.L. Spitalnik, S.W. Nicholas, G6PD deficiency in an HIV clinic setting in the dominican republic, *Am. J. Trop. Med. Hyg.* 93 (4) (2015) 722–729.
- [45] N. Al-Allawi, A.A. Eissa, J.M. Jubrael, S.A. Jamal, H. Hamamy, Prevalence and molecular characterization of glucose-6-phosphate dehydrogenase deficient variants among the Kurdish population of Northern Iraq, *BMC Blood Disord.* 10 (2010) 6.
- [46] N. Jamornthanyawat, G.R. Awab, N. Tanomsing, S. Pukrittayakamee, F. Yamin, A. M. Dondorp, N.P. Day, N.J. White, C.J. Woodrow, M. Imwong, A population survey of the glucose-6-phosphate dehydrogenase (G6PD) 563C>T (Mediterranean) mutation in Afghanistan, *PLoS One* 9 (2) (2014) 88605.
- [47] R. Chalvam, M.B. Mukherjee, R.B. Colah, D. Mohanty, K. Ghosh, G6P.D. Namoru, 208 T-> C) is the major polymorphic variant in the tribal populations in southern India, *Br. J. Haematol.* 136 (3) (2007) 512–513.
- [48] R. Devendra, V. Gupta, R. Shanmugam, M. Singh, P. Patel, N. Valecha, N. Mishra, N. Ahmed, S.L. Hoti, H.V. Hegde, P. Warang, A. Chiddarwar, P. Kedar, P. Mayekar, M.B. Mukherjee, Prevalence and spectrum of mutations causing G6PD deficiency in Indian populations, *Infect., Genet. Evol.* 86 (2020), 104597.
- [49] Y. He, Y. Zhang, X. Chen, Q. Wang, L. Ling, Y. Xu, Glucose-6-phosphate dehydrogenase deficiency in the Han Chinese population: molecular characterization and genotype–phenotype association throughout an activity distribution, *Sci. Rep.* 10 (1) (2020) 17106.
- [50] J. Gao, S. Lin, S. Chen, Q. Wu, K. Zheng, J. Su, Z. Guo, S. Duan, Molecular characterization of glucose-6-phosphate dehydrogenase deficiency in the shenzhen population, *Hum. Hered.* 85 (3–6) (2020) 110–116.
- [51] W. Dai, T. Yang, Y. Wang, Q. Zhao, Y. Zhan, J. Ye, L. Han, W. Qiu, H. Zhang, L. Liang, X. Gu, Y. Yu, Rapid detection of twenty-nine common Chinese glucose-6-phosphate dehydrogenase variants using a matrix-assisted laser desorption/ionization-time of flight mass spectrometry assay on dried blood spots, *Clin. Biochem.* 94 (2021) 27–34.
- [52] Y. Zheng, J. Wang, X. Liang, H. Huang, Y. Ma, L. Lin, C. Wang, X. Zhan, L. Yang, G. Zha, P. Yang, X. Zou, Z. Chen, X. Chen, W. Chen, X. Liu, M. Lin, Epidemiology, evolutionary origin, and malaria-induced positive selection effects of G6PD-deficient alleles in Chinese populations, *Mol. Genet. Genom. Med.* 8 (12) (2020) 1540.
- [53] A. Sanephonasa, C.L. Cheepsunthorn, N. Khaminsou, O. Savongsy, I. Nuchprayoon, K. Leecharoenkiat, Molecular characterization of G6PD mutations reveals the high frequency of G6PD Aures in the Lao Theung population, *Malar. J.* 20 (1) (2021) 30.
- [54] B. Boonyawat, T. Phetthong, N. Suksumek, C. Traivaree, Genotype-phenotype correlation of G6PD mutations among central thai children with G6PD deficiency, *Anemia* 2021 (2021), 6680925.
- [55] E. Beutler, G6PD: population genetics and clinical manifestations, *Blood Rev.* 10 (1) (1996) 45–52.
- [56] R.E. Howes, M. Dewi, F.B. Piel, W.M. Monteiro, K.E. Battle, J.P. Messina, A. Sakuntabhai, A.W. Satyagraha, T.N. Williams, J.K. Baird, S.I. Hay, Spatial distribution of G6PD deficiency variants across malaria-endemic regions, *Malar. J.* 12 (2013) 418.
- [57] W.M. Monteiro, F.F. Val, A.M. Siqueira, G.P. Franca, V.S. Sampaio, G.C. Melo, A. C. Almeida, M.A. Brito, H.M. Peixoto, D. Fuller, Q. Bassat, G.A. Romero, F.O. Maria Regina, G.L. Marcus, Vinícius, G6PD deficiency in Latin America: systematic review on prevalence and variants, *Mem. do Inst. Oswaldo Cruz* 109 (5) (2014) 553–568.
- [58] E.C. Mbanefo, A.M. Ahmed, A. Titouna, A. Elmarazy, N.T. Trang, N. Phuoc Long, N. Hoang Anh, T. Diem Nghi, B. The Hung, M. Van Hieu, N. Ky Anh, N.T. Huy, K. Hirayama, Association of glucose-6-phosphate dehydrogenase deficiency and malaria: a systematic review and meta-analysis, *Sci. Rep.* 7 (2017) 45963.
- [59] G.R. Awab, F. Aaram, N. Jamornthanyawat, K. Suwannasin, W. Pagornrat, J. A. Watson, C.J. Woodrow, A.M. Dondorp, N.P. Day, M. Imwong, N.J. White, Protective effect of Mediterranean-type glucose-6-phosphate dehydrogenase deficiency against *Plasmodium vivax* malaria, *eLife* 10 (2021), e62448.
- [60] G. Hellenthal, G.B. Busby, G. Band, J.F. Wilson, C. Capelli, D. Falush, S. Falush, A genetic atlas of human admixture history, *Science* 343 (6172) (2014) 747–751.
- [61] C.S. Chu, G. Bancone, F. Nosten, N.J. White, L. Luzzatto, Primaquine-induced haemolysis in females heterozygous for G6PD deficiency, *Malar. J.* 17 (1) (2018) 101.
- [62] W.H.K. Schilling, G. Bancone, N.J. White, No evidence that chloroquine or hydroxychloroquine induce hemolysis in G6PD deficiency, *Blood Cells, Mol., Dis.* 85 (2020), 102484.
- [63] G.J. Domingo, N. Advani, A.W. Satyagraha, C.H. Sibley, E. Rowley, M. Kalnok, J. Cohen, M. Parker, M. Kelley, Addressing the gender-knowledge gap in glucose-6-phosphate dehydrogenase deficiency: challenges and opportunities, *Int. Health* 11 (1) (2019) 7–14.