

ARTICLE OPEN



A population study of clinically actionable genetic variation affecting drug response from the Middle East

Puthen Veettil Jithesh¹✉, Mohammed Abuhaliqa², Najeeb Syed², Ikhlaq Ahmed², Mohammed El Anbari², Kholoud Bastaki^{1,3}, Shima Sherif¹, Umm-Kulthum Umlai¹, Zainab Jan¹, Geethanjali Gandhi^{1,2}, Chidambaram Manickam², Senthil Selvaraj², Chinnu George², Dhinoth Bangarusamy¹, Rania Abdel-latif⁴, Mashael Al-Shafai⁵, Zohreh Tatari-Calderone², Xavier Estivill⁶, Munir Pirmohamed⁷ and The Qatar Genome Program Research Consortium*

Clinical implementation of pharmacogenomics will help in personalizing drug prescriptions and alleviate the personal and financial burden due to inefficacy and adverse reactions to drugs. However, such implementation is lagging in many parts of the world, including the Middle East, mainly due to the lack of data on the distribution of actionable pharmacogenomic variation in these ethnicities. We analyzed 6,045 whole genomes from the Qatari population for the distribution of allele frequencies of 2,629 variants in 1,026 genes known to affect 559 drugs or classes of drugs. We also performed a focused analysis of genotypes or diplotypes of 15 genes affecting 46 drugs, which have guidelines for clinical implementation and predicted their phenotypic impact. The allele frequencies of 1,320 variants in 703 genes affecting 299 drugs or class of drugs were significantly different between the Qatari population and other world populations. On average, Qataris carry 3.6 actionable genotypes/diplotypes, affecting 13 drugs with guidelines for clinical implementation, and 99.5% of the individuals had at least one clinically actionable genotype/diplotype. Increased risk of simvastatin-induced myopathy could be predicted in ~32% of Qataris from the diplotypes of *SLCO1B1*, which is higher compared to many other populations, while fewer Qataris may need tacrolimus dosage adjustments for achieving immunosuppression based on the *CYP3A5* diplotypes compared to other world populations. Distinct distribution of actionable pharmacogenomic variation was also observed among the Qatari subpopulations. Our comprehensive study of the distribution of actionable genetic variation affecting drugs in a Middle Eastern population has potential implications for preemptive pharmacogenomic implementation in the region and beyond.

npj Genomic Medicine (2022)7:10; <https://doi.org/10.1038/s41525-022-00281-5>

INTRODUCTION

Genetic variation plays an important role in the inter-individual differences in response to medications, and pharmacogenomic (PGx) testing has the potential to provide an informed decision on the appropriate choice and dosage of medications¹. The current progress in next-generation sequencing (NGS) technologies provides several avenues for PGx profiling. Although many studies have promoted exome sequencing or targeted NGS panels for PGx testing at population scale^{2,3}, the benefits of these approaches are mostly limited by their inability to sequence the non-coding regions⁴. Whole-genome sequencing (WGS) can overcome this limitation and hence provide the most comprehensive sequencing approach for more accurate PGx profiling⁵. Furthermore, WGS provides more accurate PGx profiling through the ability to identify potential rare variants/private mutations that may affect drug disposition and response.

Resources such as the PharmVar^{6,7} and Pharmacogenomics KnowledgeBase (PharmGKB)⁸ and guidelines produced by the Clinical Pharmacogenetic Implementation Consortium (CPIC)⁹ and the Dutch Pharmacogenetics Working Group (DPWG)¹⁰ are helping in the clinical implementation of pharmacogenomic testing for a select number of drug-gene combinations with a high level evidence. However, prioritization and implementation of drug-gene combinations for clinical testing in different

ethnic populations require the knowledge of the distribution of genetic variants affecting the drugs and prescription patterns in that population^{11–13}. In addition, guidelines developed by CPIC and DPWG primarily focus on common variants, and a WGS approach would help in the identification of novel variants in the population of interest that are currently not covered by CPIC or DPWG. Although pharmacogenomic screening is established in many medical institutions in the US and Europe^{14,15}, such implementation is lagging in many other parts of the world, including the Middle East, due to the lack of such data¹⁶.

Here we present the first comprehensive characterization of clinically actionable genotypes and diplotypes and their predicted phenotypic effect on efficacy, dosing and the risk of adverse events for several medications with CPIC clinical implementation guidelines in the Qatari population from the analysis of 6045 whole genomes. We also compared the distribution of these frequencies with that of other world populations represented in the 1000 genomes dataset to understand the similarities and distinctiveness of the Qatari population in their predicted response to these medications. As far as we are aware, this is the first such comprehensive study in any Middle Eastern population, with potential implications for pre-emptive pharmacogenomic implementation in the region and beyond.

¹College of Health & Life Sciences, Hamad Bin Khalifa University, Doha, Qatar. ²Research Branch, Sidra Medicine, Doha, Qatar. ³Hamad Medical Corporation, Doha, Qatar. ⁴Qatar Genome Program, Qatar Foundation Research Development and Innovation, Doha, Qatar. ⁵Department of Biomedical Sciences, College of Health Sciences, Qatar University, Doha, Qatar. ⁶Quantitative Genomics Laboratories, Barcelona, Catalonia, Spain. ⁷Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK. *A list of authors and their affiliations appears at the end of the paper. ✉email: jveettil@hbku.edu.qa

RESULTS

Pharmacogenetic variation in the Qatari population

We performed a comprehensive analysis of the variants annotated by PharmGKB to be associated with drugs and based on adjusted *p*-values from two proportions *z*-test, the allele frequencies of 1320 variants in 703 genes affecting 299 drugs or class of drugs were significantly different between the Qatari population (6,045 whole genomes) and other world populations represented in the gnomAD v3 dataset (76,156 whole genomes) (Supplementary Data S15). Of these, 615 variants had higher frequencies in the Qatari population. Some examples of variants with differing frequencies in the Qatari population included rs1137101 in the *LEPR* gene, which was lower in the Qataris, rs2289669 in *SLC47A1* and rs11212617 in *ATM*, both higher in the Qatari population. These differing allele frequencies in the Qatari population compared to other populations provides an avenue for further work in the future to determine whether the clinical outcomes are different based on reported drug-gene associations in other ethnic groups^{17–19}.

We further performed a focused analysis of the distribution of variants in 17 pharmacogenes affecting 48 drugs, which have CPIC Level A annotation and guidelines for clinical implementation. *UGT1A1* diplotypes were not called with confidence and hence not reported here. Results from *CYP4F2* haplotypes are presented along with warfarin dosage calculations in a later section. Our analysis of the remaining 15 genes affecting 46 drugs identified that, on average, individuals carry 3.6 actionable genotypes/diplotypes. Furthermore, 99.5% of the individuals had at least one clinically actionable genotype/diplotype. Qataris, on average, carried pharmacogenetic variations that predict actionable phenotypes affecting 12.9 (28.8%) drugs.

Clinically actionable pharmacogenetic variation in the Qatari population

Among the gene-drug pairs with guidelines for clinical implementation, the greatest number of actionable genotypes were present in *VKORC1* (rs9923231; –1639G > A), with 72.7% of the population predicted to require a lower dosage of the widely used anticoagulant, warfarin²⁰, based on the homozygous alternate (26.4%) or heterozygous genotypes (46.3%) (Table 1). This was followed by the genotypes of *IFNL3*, which are predicted to produce an unfavorable response to treatment for Hepatitis C in 52.5% of the population studied (Hom alt: 10.4%; Het: 42.2%). Both heterozygous and homozygous alternate genotypes of rs12979860 in *IFNL3* lead to the decreased likelihood of response to pegylated interferon- α and ribavirin therapy²¹. At the other extreme, no variant genotypes were present in the population for *CACNA1S* (rs772226819 and rs1800559), a gene known to cause malignant hyperthermia when using potent volatile anesthetic agents such as halothane or the depolarizing muscle relaxant, succinylcholine²². Another related gene, *RYR1* leading to malignant hyperthermia susceptibility²² also had only two individuals (0.003%) with the alternate allele for rs111888148 and rs193922762.

Among the highly polymorphic genes with actionable diplotypes, predicted CYP2C19 rapid (29.8%), ultrarapid (6.5%), poor (1.9%) and intermediate (19.8%) metabolizers were present in the Qatari population (58%), which are known to affect the metabolism and necessitate prescribing alternate drugs or a change in dosage of clopidogrel²³, voriconazole²⁴, several antidepressants^{25,26} and proton pump inhibitors²⁷. For example, CYP2C19 intermediate and poor metabolizers (21.7%) would require the prescription of an alternate antiplatelet therapy such as prasugrel or ticagrelor instead of clopidogrel, especially when treating patients with acute coronary syndromes (ACS) undergoing percutaneous coronary intervention (PCI)²³. Similarly, CYP2C19 ultrarapid, rapid and poor metabolizers

(38.3%) would need prescribing an alternate antifungal agent instead of voriconazole²⁴.

Poor (8.2%) and intermediate (37.8%) metabolizers were predicted from CYP2B6 diplotypes (46%), who may have increased risk of adverse events affecting the central nervous system when treated with the HIV type-1 nonnucleoside reverse transcriptase inhibitor, efavirenz. Both CYP2B6 rapid and ultrarapid metabolizers were also predicted to be present in the population, but there is no evidence of these affecting the plasma concentration of efavirenz, though other drugs may be affected²⁸. Increased risk of simvastatin-induced myopathy²⁹ could be predicted in 32.4% of Qataris from the diplotypes of *SLCO1B1* (Poor function diplotypes: 5.6%; Decreased function diplotypes: 26.7%). Increased risk of toxicity when treated with the anti-epileptic drug, phenytoin, and several nonsteroidal anti-inflammatory drugs (NSAIDs) metabolized by CYP2C9^{30,31} could be predicted from poor (1.6%) and intermediate (30.3%) metabolizer status in 31.9% of the population.

Actionable diplotypes of one of the most important pharmacogenes, *CYP2D6* were observed in 33.7% of the sampled Qatari population (Poor: 1.9%; Ultrarapid: 8.6%; Intermediate: 23.3%). *CYP2D6* metabolizes approximately 25% of drugs, including several antidepressants, codeine, atomoxetine, ondansetron, tropisetron, and tamoxifen^{25,26,32–35}. CYP3A5 expresser phenotype was predicted in 18% of the population (Extensive metabolizer: 1.4%; intermediate: 16.5%), which has the CPIC recommendation for an increased starting dose of tacrolimus compared to the standard dose for achieving immunosuppression³⁶. Only a small proportion of the population may have an increased risk of thiopurine-related leukopenia, neutropenia or myelosuppression as predicted from the diplotypes of *TPMT* (2%) or *NUDT15* (4.2%)³⁷. Similarly, *DPYD* diplotypes predicted only 0.1% of the population may be at increased risk of severe or fatal toxicity when treated with fluoropyrimidines³⁸.

A high risk of abacavir hypersensitivity was predicted in 2.6% of the population due to the presence of *HLA-B*57:01* homozygous or heterozygous genotypes³⁹. *HLA-B*58:01* diplotypes (homozygous or heterozygous) were present in 5.9% of the population studied, indicating significantly increased risk of allopurinol-induced severe cutaneous adverse reaction (SCAR), which is manifested by Stevens–Johnson syndrome (SJS), toxic epidermal necrolysis (TEN), or drug reaction with eosinophilia and systemic symptoms (DRESS)⁴⁰. An increased risk of SJS/TEN is also predicted due to the presence of heterozygous or homozygous *HLA-B*15:02* in 0.4% of the population if treated with phenytoin/fosphenytoin³⁰ or carbamazepine/oxcarbazepine⁴¹. Furthermore, 5.4% of the Qatari population studied have the *HLA-A*31:01* heterozygous or homozygous genotypes, indicating further risk of SJS/TEN and other hypersensitivity phenotypes with carbamazepine treatment⁴¹.

The clinically actionable genotype/diplotype distribution of 10 genes were found to be statistically significantly different in the Qatari population compared to other world populations represented in the 1000 genomes dataset, with only three genes having higher frequencies in the Qatari population (Table 1; Fig. 1a). The actionable diplotype frequency of *SLCO1B1* was more than double in the Qatari population compared to other world populations (QGP: 32% vs. 1000 genomes: 15%, $p = 3.2 \times 10^{-59}$). Other genes with higher frequencies were *CYP2C9* (32% vs 23%, $p = 9.2 \times 10^{-14}$) and *VKORC1* (73% vs 49%, $p = 7.3 \times 10^{-96}$). Furthermore, nine genes had distinctly different actionable frequency distributions between the Qatari population and the European superpopulation from the 1000 genomes dataset, with *CYP2B6*, *CYP3A5*, *NUDT15*, *SLCO1B1*, *VKORC1* and *HLA-B*58:01* having higher frequencies in the Qatari population (Table 1; Fig. 1b).

Table 1. Actionable genotype/diplotype frequencies of clinically important pharmacogenes in the Qatari population.

Gene	Examples of Affected drugs/category of drugs with clinical guidelines	Number of Variants analysed, Number of star alleles analysed	Major Actionable Genotypes/ Diplotypes in the population, present in at least 5 individuals in the dataset	Phenotypic effect warranting change in drug, drug dose or drug monitoring	Number of individuals with actionable genotypes/diplotypes in the Qatari population (%)	Total Number of individuals with actionable genotypes/diplotypes in the Qatari genome populations (%)
CACNA1S	Potent Volatile Anesthetic Agents, Succinylcholine	2	(rs772226819 TT (c.520 C > T), rs1800559 AA (c.3257 G > A))	Malignant Hyperthermia Susceptibility	0 (0)	0
CYP2B6	Efavirenz	63, 38	*6/*6, *6/*18, *6/*36 *1/*6, *2/*6, *6/*22, *4/*6, *1/*18, *1/*36, *1/*9, *2/*9	Poor metabolizer Intermediate metabolizer	497 (8.22) 2284 (37.78)	1295 (51.72)
CYP2C9	Phenytoin, NSAIDs	94, 71	*1/*2, *1/*3, *2/*2, *1/*11, *2/*11, *2/*9, *2/*3, *3/*3	Intermediate metabolizer	1832 (30.31)	588 (23.48)
CYP2C19	Clopidogrel, Voriconazole, Antidepressants, Proton Pump Inhibitors	71, 39	*1/*17 *17/*17 *2/*2, *2/*35 *1/*2, *2/*17, *1/*35, *2/*13, *1/*3, *17/*35	Poor metabolizer Rapid metabolizer Ultrarapid metabolizer Poor metabolizer Intermediate metabolizer	99 (1.64) 1804 (29.84) 395 (6.53) 113 (1.87) 1197 (19.8)	1483 (59.23)
CYP2D6	Atomoxetine, Codeine, Ondansetron, Tropicisetron, Tamoxifen, Antidepressants	355, 145	*4/*4, *4/*68 + *4, *68 + *4/ *68 + *4, *4/*5, *5/*68 + *4/ *1/*2x2, *2/*2x2, *2x2/*41, *1x2/*2, *1/*1x2, *1x2/*41, *1x2/*1x2, *2x2/*2x2, *1x2/ *2x2, *17/*2x2, *2x2/*35, *2x2/*27x2, *2x2/*33, *1x2/ *17	Poor metabolizer Ultrarapid metabolizer	114 (1.89) 517 (8.55)	982 (39.22)
CYP3A5	Tacrolimus	25, 9	*1/*4, *1/*68 + *4, *41/*41, *2/*4, *1/*5, *4/*41, *2/*68 + *4, *41/*68 + *4, *10/*41, *2/*5, *17/*41, *41/*5, *17/*4, *4/*10, *35/*4, *1/*3, *10/*68 + *4, *1/ *13, *17/*68 + *4, *10/*10, *1/ *40, *1/*6, *5/*10, *10/*17, *1/ *8, *29/*4, *1/*36 + *10, *1/*7, *1_2_68, *17/*17, *17/*29, *41/*9, *9/*68 + *4 *1/*1	Intermediate metabolizer Extensive metabolizer (CYP3A5 expressor)	1407 (23.28) 86 (1.42)	1191 (47.56)
DPYD	Fluoropyrimidines		*1/*3, *1/*6, *1/*7	Intermediate metabolizer (CYP3A5 expressor)	996 (16.48)	10 (0.4)
HLA-A	Carbamazepine		HLA-A*31:01 Hom HLA-A*31:01 Het	Intermediate metabolizer Risk of SJS/TEN	9 (0.15) 10 (0.16)	10 (0.4) 125 (4.99)
HLA-B	Phenytoin, Carbamazepine, Oxcarbazepine		(HLA-B*15:02 Hom) HLA-B*15:02 Het	Risk of SJS/TEN	323 (5.27) 0 (0) 25 (0.41)	88 (3.51)

Table 1 continued

Gene	Examples of Affected drugs/category of drugs with clinical guidelines	Number of Variants analysed, Number of star alleles analysed	Major Actionable Genotypes/ Diplotypes in the population, present in at least 5 individuals in the dataset	Phenotypic effect warranting change in drug, drug dose or drug monitoring	Number of individuals with actionable genotypes/diplotypes in the Qatari population (%)	Total Number of individuals with actionable genotypes/ diplotypes in the Qatari population (%)	Total Number of individuals with actionable genotypes/ diplotypes in the thousand genome populations (%)
HLA-B	Abacavir		(HLA-B*57:01 Hom)	Hypersensitivity risk	2 (0.03)	161 (2.62)	151 (6.03)
HLA-B	Allopurinol		HLA-B*57:01 Het (HLA-B*58:01 Hom) HLA-B*58:01 Het	Risk of SCAR	159 (2.59) 4 (0.07) 359 (5.85)	363 (5.92)	165 (6.59)
<i>IFNL3</i>	Pegylated Interferon alpha, Ribavirin	1	rs12979860 Hom Alt	Unfavourable response	626 (10.35)	3175 (52.51)	1353 (54.03)
<i>NUDT15</i>	Thiopurines	19, 20	rs12979860 Het *1/*3 *3/*3	Unfavourable response Intermediate metabolizer Poor metabolizer Indeterminate	2549 (42.15) 245 (4.05) 5 (0.08) 2 (0.03)	252 (4.17)	185 (7.39)
<i>RYR1</i>	Potent Volatile Anesthetic Agents, Succinylcholine	2	(rs111888148 c.1589 G > A, rs193922762 c.982 C > T)	Malignant Hyperthermia Susceptibility	2 (0.03)	2 (0.03)	0
<i>SLCO1B1</i>	Simvastatin	29, 36	*1/*15, *1/*5, *1/*17, *1/*31 *15/*15, *5/*15, *15/*17, *5/ *17, *5/*5, *17/*17 *1/*3, *1/*2	Decreased function (Increased risk of myopathy) Poor function (High risk of myopathy) Intermediate metabolizer Poor metabolizer	1616 (26.73) 341 (5.64) 120 (1.98) 1 (0.02)	1957 (32.37)	376 (15.02)
<i>TPMT</i>	Thiopurines	43, 44		Intermediate metabolizer Poor metabolizer	1596 (26.39)	4395 (72.68)	1230 (49.12)
<i>VKORC1</i>	Warfarin	1	rs9923231 (-1639G > A) Hom (AA) rs9923231 (-1639G > A) Het (GA)	Lower dosage requirement Lower dosage requirement	2799 (46.29)		

NSAIDs: Nonsteroidal anti-inflammatory drugs, *SJS/TEN*: Stevens–Johnson syndrome, toxic epidermal necrolysis, *SCAR*: Severe cutaneous adverse reaction.

Comparison of the frequencies of actionable genotypes/diplotypes in the Qatari population (6045 genomes) with that of the thousand genome populations (2,504 genomes). Examples of drugs predicted to have an effect based on CPIC guidelines are also provided.

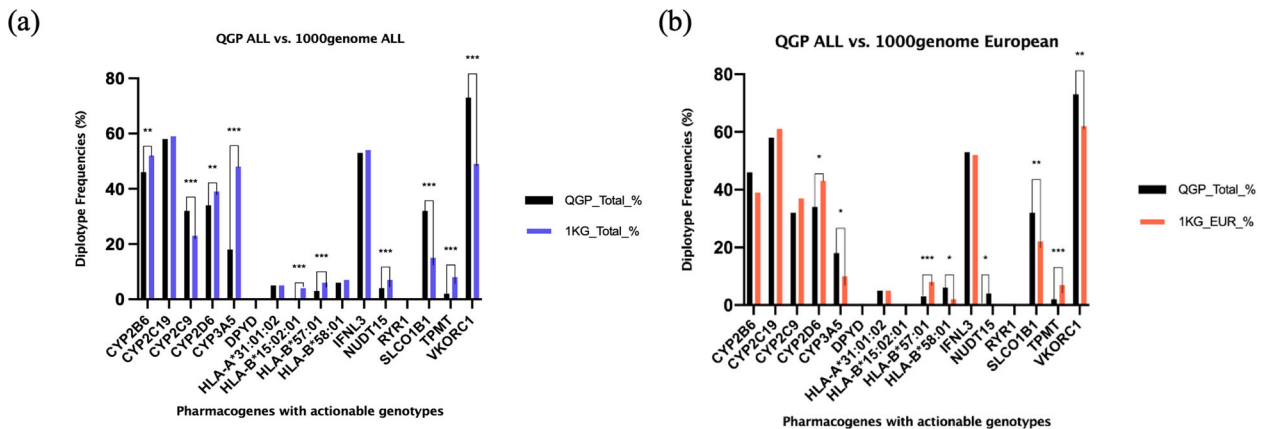


Fig. 1 Comparison of actionable genotype or diplotype frequencies in the Qatari population. Actionable frequencies are compared between the Qatari cohort ($n = 6045$) and (a) the overall 1000 genomes cohort ($n = 2504$), or (b) the European superpopulation in the 1000 genomes cohort ($n = 503$). p values of significantly differing frequencies with Bonferroni adjustment indicated as follows: * = $0.0001 \leq p$ value < 0.05 , ** = $0.000001 < p$ value < 0.0001 , *** = p value ≤ 0.000001 .

Clinically actionable pharmacogenetic variation in the Qatari subpopulations

Diverse distribution of actionable genotype/diplotype frequencies was observed among the Qatari subpopulations (Fig. 2, Supplementary Data S1–S13). For example, the actionable diplotype frequencies of *CYP3A5* varied from around 12% in the Peninsular Arabs and General Arabs to around 40–50% in South Asian and African subpopulations, with the Persian and Admixed subpopulations having frequencies in the middle of the spectrum (~25%). The Peninsular Arabs had the lowest frequency of actionable diplotypes among the Qatari subpopulations, and the highest in African subpopulation for *TPMT* (PAR: 0.6%, AFR:10.9%), and *CYP2B6* (PAR: 31.3%, AFR: 55.4%). The actionable diplotypes for *DPYD* were totally absent in the Peninsular Arabs, Africans, and the South Asian subpopulations.

Analysis of the fixation index based on the pharmacogenes revealed varying levels of population differentiation between the Qatari subpopulations, with the Peninsular Arabs farther away from the South Asians and Africans, and closer to the West Eurasian/Persians, while being closest to the General Arabs (Fig. 2b). However, the Peninsular Arab subpopulation showed a distinct frequency distribution compared to the General Arabs for some of the genes studied. For example, the Peninsular Arabs had a lower percentage of actionable diplotypes and hence a potentially lower risk of adverse reactions or inefficacy of corresponding drugs when compared with the General Arab subpopulation for *CYP2B6* (31.3% vs. 45.5%), *CYP2C19* (52.9% vs. 61.2%), *CYP2C9* (28.3% vs. 34.2%), and *CYP2D6* (27.38% vs. 31.76%) while the converse was true for certain other genes: *NUDT15* (5.2% vs. 4.3%), *SLCO1B1* (40% vs. 36.4%) (Fig. 2a, c). Notably, the *SLCO1B1* actionable diplotype frequency for the Peninsular Arabs was the highest among all the subpopulations, which was also almost double that seen in the European, East Asian and American populations (~22%), the highest observed among the populations studied in the 1000 genomes program (Fig. 2a).

Among the HLA alleles, *HLA-B*15:02* was completely absent in the Peninsular Arabs, while *HLA-A*31:01* was seen at a lower frequency (4.1%) compared to most of the other subpopulations, especially the General Arabs which had the highest frequency in the population (7.2%). The frequency of *HLA-B*58:01* was the lowest in the Peninsular Arabs (3.8%) compared to the other subpopulations, with the General Arabs having higher (6.5%) and the African subpopulation having the highest (9%). By contrast, Peninsular Arabs had one of the highest frequencies for *HLA-B*57:01* (4.9%) compared to the other subpopulations, including

General Arabs (2.1%), and this genotype was absent in the African subpopulation (Fig. 2).

Warfarin dosing and the potential pre-emptive pharmacogenomic implementation

A major concern in the use of warfarin, one of the most commonly used anticoagulants worldwide, is its narrow therapeutic index, which necessitates accurate dosage calculation and therapeutic drug monitoring. We assessed the potential dosage requirements of individuals in the cohort by a combinatorial calculation of the effects of clinical and genetic factors following the IWPC algorithm⁴² that included age, sex, height, weight, ethnicity, and the concurrent use of drugs that alter warfarin requirements, in addition to the genotypes of *VKORC1* and diplotypes of *CYP2C9* as described in the Methods section. The distribution of the predicted weekly dose of warfarin in the population ranged from 5.4 mg to 66.4 mg, with 593 (10%) requiring a lower dose (≤ 21 mg per week) and 313 (5%) requiring a higher dose (≥ 49 mg per week). In comparison, in the patients of European ethnicity from the EU-PACT trial ($n = 325$ patients with stable doses), 80 (25%) needed a lower dose and 42 (13%) needed a higher dose (Fig. 3).

In addition to the alleles used for the calculation of warfarin dosage, other *CYP2C9* haplotypes were also present in the population. For example, *CYP2C9*11* known to affect dosage in the African population⁴³ was seen in the Qatari population (0.5%), with the African subpopulation having the highest frequency (2.7%), close to that seen in the African population in the 1000 genomes dataset (2.3%). *CYP4F2*3* haplotype associated with warfarin dosing was present in ~24% of the population, with the lowest frequency in the African subpopulation (8.69%), and the highest in Peninsular Arabs (38.2%). The overall frequency of the *CYP2C* rs12777823 variant, which is known to lead to a lower dose requirement of warfarin, was ~12%, with the lowest frequency in the Peninsular Arabs (7.8%) and higher frequencies in the African (21.7%) and South Asian (34.2%) subpopulations.

Medications with actionable genetic variation in the Qatari population

Further to deciphering the landscape of pharmacogenomic variation affecting the CPIC Level A drugs, we also examined the availability of these drugs through the major healthcare provider in Qatar. Out of the 50 drugs, 13 (26%), including abacavir, atazanavir, and ivacaftor were unavailable, while efavirenz was available as non-formulary (Supplementary Data S16).

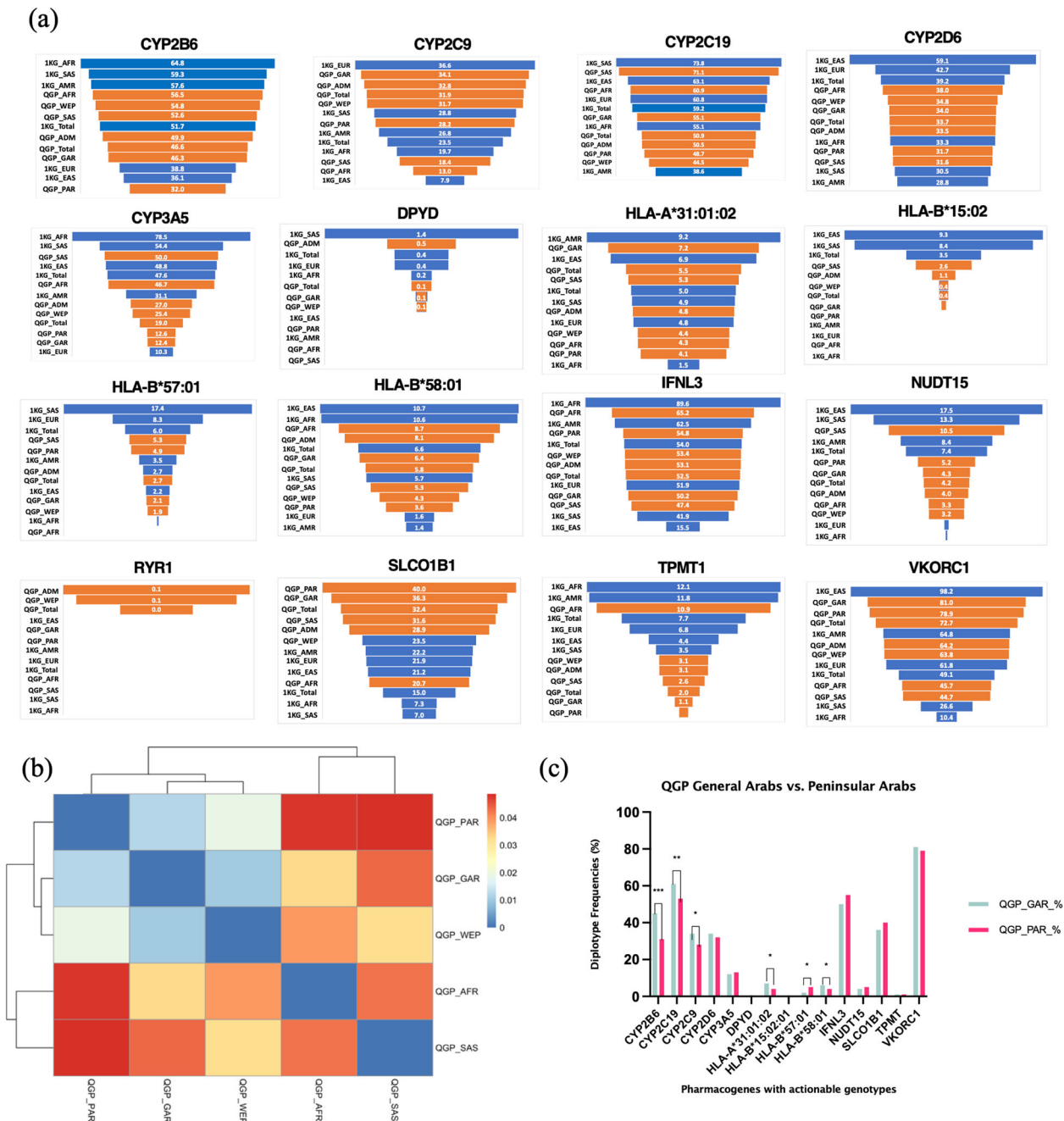


Fig. 2 Actionable genotype or diplotype frequencies in the Qatari subpopulations. (a) Actionable genotype/diplotype frequencies of clinically important pharmacogenes in the Qatari population and subpopulations (shown as orange bars) along with that of the overall 1000 genomes and the superpopulations (shown as blue bars). (b) Clustering of QGP subpopulations based on F_{ST} calculated from the pharmacogenes. (c) Comparison of the actionable frequencies of QGP General Arab and Peninsular Arab subgroups. QGP subpopulations PAR, Peninsular Arabs; GAR, General Arabs; ADM, Admixed; WEP, West Eurasian/Persian; AFR, African; SAS, South Asian; 1KG, thousand genomes superpopulations: EUR, European; AMR, American; AFR, African; SAS, South Asian; EAS, East Asian.

DISCUSSION

In this study, we assessed the landscape of clinically actionable pharmacogenomic variation associated with drug response in the Qatari population. One important point to highlight about our study is that, unlike previous studies in the population⁴⁴, we conducted a fine-grained analysis that led to results with a high resolution. For example, rather than just using single variants as proxy to define the star alleles, we used all the variants known to contribute to the haplotype, wherever possible. Furthermore, we analyzed whole-genome sequencing data from ~2% of the

Qatari population, capturing broader variant and population landscape than exome studies in small cohorts. Thus, the genotypes and diplotypes we report in the Qatari population can be compared with results from future studies of other populations with confidence. Another aspect of our study is the use of multiple types of variants, for example, structural variants to define the star alleles of *CYP2D6*, unlike some of the recent large-scale studies⁴⁵. This has helped us to capture and present a more accurate account of the distribution of the pharmacogenomic variation in the population. More importantly, we provide a more

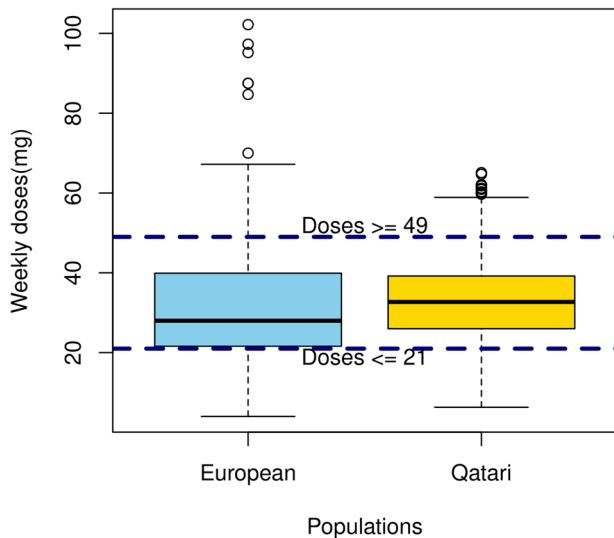


Fig. 3 Warfarin dosing prediction. Distribution of the predicted weekly dose (mg) of warfarin in the Qatari population (yellow) using the IWPC algorithm based on age, sex, height, weight, ethnicity, the concurrent use of drugs that decrease warfarin requirements, and the genotypes/diplotypes of *VKORC1* and *CYP2C9*. Also shown is the distribution of dosage in patients of European ethnicity from the EU-PACT trial (blue). Weekly doses in mg are plotted in the Y-axis for the two populations. The box is drawn with the interquartile range and the central horizontal line showing the median, while values above the range shown as whiskers. Individuals with values ≤ 21 mg per week (below the bottom horizontal line) would be predicted to need a lower dose, and those with values ≥ 49 mg per week (above the top horizontal line) would be predicted to need a higher dose.

useful and clinically translatable measure of genetic variability in terms of combined genotype or diplotype frequencies that are actionable based on available guidelines, rather than just reporting allele or haplotype frequencies as done in most studies.

Based on the results from the current study and considering the disease epidemiology and prescription pattern, it is possible to recommend which drug-gene combinations need to be prioritized for pharmacogenomic testing in the hospitals in Qatar and potentially other Middle East countries. For example, the frequency of the actionable *SLCO1B1* diplotypes is high in the population, especially in the Arab subpopulations, suggesting higher risk of muscle toxicity when treated with commonly prescribed lipid lowering HMG CoA reductase inhibitors (statins) such as simvastatin and potentially atorvastatin. Dyslipidemia is the topmost noncommunicable disease (~30%) identified in the QBB cohort and statins are among the highly prescribed medications in Qatar⁴⁶. Thus, for implementation, genotyping of *SLCO1B1* before statins are prescribed or undertaking genotyping in individuals intolerant of statins, should be considered, especially since there are other statins, which have a lower substrate affinity for OATP1B1 could be considered as alternate hypolipidemic drugs²⁹.

By contrast, the frequencies of actionable diplotypes of certain other genes were lower in the Qatari population compared to other world populations. A striking example is *CYP3A5* (QGP: 18% vs 1000 genomes: 48%), the lower frequency of which suggests that tacrolimus dosage adjustments may be necessary only in fewer patients for achieving immunosuppression³⁶. Of the two genes which can lead to adverse effects when treated with thiopurines, including azathioprine, mercaptopurine and thioguanine, *TPMT* had a lower frequency of actionable diplotypes in the Qatari population (2% vs. 8%), while *NUDT15* had a higher frequency when compared to other populations represented in the thousand genomes dataset (4% vs. 0.2%). *NUDT15* loss-of-function alleles are known to be

more prevalent in Asians and Hispanics leading to myelosuppression, rather than *TPMT* deficiency, which is the primary genetic cause of thiopurine intolerance in Europeans and Africans³⁷. Our results suggest that *NUDT15* testing may be more important in the Qatari population than *TPMT* before initiating thiopurine treatment, but in order to improve the safety of thiopurines, we would suggest genotyping for allelic variation in both genes, in keeping with the CPIC guidance³⁷.

Genetic variation in the non-coding region also contributes significantly in determining the efficacy or dosage requirements of certain treatments. For example, a non-coding variant in the promoter region of *VKORC1*, c.-1639G > A (rs9923231) alters the transcription factor binding site, leading to decreased expression of vitamin K epoxide reductase enzyme, a target of warfarin. Patients carrying at least one A allele at the -1639 locus require lower initial and maintenance doses of warfarin compared to the patients carrying a G/G genotype⁴⁷. Our study indicates a high frequency of the alternate genotypes in the Qatari population. This example stresses the importance of WGS to identify genetic variation in the non-coding regions, which will be missed by exome sequencing.

Furthermore, *VKORC1* c.-1639G > A and *CYP2C9**2 and *3 were found to be the major genetic variants to predict warfarin dosage in a Qatari patient cohort, explaining 39.2% of warfarin dose variability, along with hypertension, smoking and heart failure, while *CYP4F2**3 was not associated with warfarin dose⁴⁸. Studies in other populations in the region also showed the significant contribution of these genetic variants in explaining 30–63% of warfarin dose variability, highlighting the clinical importance of PGx-guided warfarin dose recommendations^{49–51}. Although many of those populations share the same ancestry and are similar in their warfarin dose predictors, a population-specific dosing algorithm may be better suited for the prospective estimation of warfarin dose. A multivariate model, which included demographic, clinical and pharmacogenetic variables together explained 63% of the overall inter-patient variability in warfarin dose requirement in an admixed Omani patient cohort. This locally developed model performed better than the IWPC model as the latter could only explain 34% of the inter-patient variability in Omani patients⁵⁰.

In addition to the distribution of actionable genotypes, it is also important to understand the prescription pattern of medications to prioritize the gene-drug pairs for clinical implementation. We observed that some of the CPIC Level A drugs were not prescribed by the major healthcare provider in Qatar. Thus, it may be recommended that genes, which affect the response to drugs, such as atazanavir and ivacaftor (*UGT1A1* and *CFTR*) should be given the lowest priority when implementing comprehensive pharmacogenomic testing in Qatar. This also points towards the need for concerted efforts in Qatar and other countries to develop guidelines for clinical implementation of pharmacogenes that affect drugs, which are highly prescribed in these countries and have lower efficacy and safety profiles. Such efforts will complement and contribute to the activities of CPIC, DPWG and other international consortia.

Even though the analysis of the whole-genome sequencing data from more than 6000 Qataris was intended to understand the landscape of clinically actionable pharmacogenomic variants in the population, this analysis has also generated the pharmacogenomic profile of each individual for all these drug-gene combinations. Thus, our study also provides an interesting opportunity, following further discussions with clinicians and healthcare authorities, to implement the first step toward pre-emptive pharmacogenomics in the country by returning the results to the participants or inclusion of the pharmacogenomics report of each individual in their electronic health records for use anytime during their visit to the hospitals.

A limitation of this study is the use of translation tables for genotype/diplotype generation, and their prediction of phenotypes

developed based on the literature, which is dominated by studies from European or other populations, and not from the Middle East. Although it can be safely assumed that the effects of the diplotypes should be similar in different populations, the role of population-specific rare variants in the inter-individual variability of drug response is being recognized widely⁵². We have also observed novel combinations of variants in some of the genes in certain individuals, which are not currently mapped into 'star alleles', and hence phenotype prediction is not possible unless further functional characterization of these variants is performed.

In conclusion, we presented the first comprehensive analysis of the distribution of actionable variation in clinically important genes affecting drug efficacy or safety in the Qatari population and subpopulations. We hope that this will not only fill the gap in the literature, but also help in the implementation of precision medicine in Qatar and beyond.

METHODS

Study samples

The study population consisted of an observational longitudinal cohort of 6218 apparently healthy adult Qatari individuals, consented and recruited by the Qatar Biobank (QBB)⁴⁶, and whose genomes were sequenced as part of the first phase of the Qatar Genome Program (QGP). The cohort included 3528 women (56.7%), and the mean age of the participants at the start of the study was 40.2 years (SD: ±12.66). A recent publication on the genome-wide association of several traits in the same cohort provides more details of the dataset⁵³. In the present study, only anonymized datasets were accessed and used for the analysis after obtaining approval from the QBB Institutional Review Board (E/2017/QGP-RES-PUB-008/0014).

Whole-genome sequencing data

Methods used for whole-genome sequencing (WGS) and bioinformatics data processing to identify genetic variants in the cohort are described in detail elsewhere⁵³ and briefly described here. As part of the QGP, the blood samples collected by the QBB were sequenced using Illumina short-read technology on HiSeq X machines to an average coverage of 30x. The raw WGS data were processed through the standard Burrows-Wheeler (BWA) algorithm (v0.7.12)⁵⁴ and Genome Analysis Toolkit (GATK, v3.4)⁵⁵ pipelines, following the GATK best practices for processing the data. A single multisample VCF file with all the samples sequenced and generated using the GRCh37 human genome reference was used for the analysis. After quality control, 8 samples were removed for excess heterozygosity, one for low-call rates, 65 for gender mismatch, 87 for population outliers, and 10 for identical matching⁵³. Therefore, WGS data from 6045 participants were taken forward for analysis.

Copy number variation (CNV) analysis was performed using Canvas (v1.11.0)⁵⁶ and structural variant (SV) analysis was performed using Manta (v0.29.6)⁵⁷ and Delly (v0.7.8)⁵⁸.

Phenotypic data

Phenotypic data available for the QBB participants included age, sex, anthropometric measurements, self-reported questionnaire data on health conditions, lifestyle, diet, and medications.

Selection of pharmacogenes

First, we selected all the variants known to be associated with response to drugs as annotated by PharmGKB (using `clinical_ann_metadata.tsv` file in `clinicalAnnotations.zip` downloaded from <https://www.pharmgkb.org/downloads/> on 15 Dec 2020) and extracted the allele frequencies in the QGP dataset. Only simple variants (SNVs and indels) with a dbSNP rsID were analyzed in this stage and included 2629 variants in 1026 genes known to affect 559 drugs or class of drugs as annotated by PharmGKB. We also performed a detailed analysis of clinically important pharmacogenes for which annotations were made available by CPIC for their interaction with specific drugs. We analyzed 17 genes affecting 48 drugs with CPIC Level A annotation (Accessed from <https://cpicpgx.org/genes-drugs/> on 15 Dec 2020).

Pharmacogenomic analysis

Pharmacogenes with CPIC Level A annotations and one or a few variants influencing response to drugs (*CACNA1S*, *IFNL3*, *RYR1*, and *VKORC1*), we calculated the allele and genotype frequencies in the QGP population from the allele counts and allele numbers directly. For highly polymorphic genes (*CYP2B6*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A5*, *CYP4F2*, *DPYD*, *NUDT15*, *SLCO1B1*, *TPMT*, and *UGT1A1*), we developed custom Python scripts that took the multisample VCF as input and converted the combination of variants to haplotypes as 'star alleles' based on the gene-specific allele definition tables for each pharmacogene created by PharmVar, CPIC and PharmGKB. Computational phasing was performed to assign the variants to maternally or paternally derived chromosomes before the star allele assignment. The individual VCF files for each of the genes of interest were extracted using `bcftools` (<http://samtools.github.io/bcftools/bcftools.html>) from the WGS multisample VCF file. The reference-based phasing algorithm, Eagle2 (v 2.4)⁵⁹ was used to phase the haplotypes for each of the individual genes. The genetic map file (`genetic_map_hg19_withX.txt.gz`) with recombination frequencies was used as a reference for the phasing algorithm.

All possible diplotypes were calculated based on the star alleles (haplotypes) detected on both the homologous chromosomes. Suballeles were considered together with the main alleles for the diplotype calculation. The CPIC diplotype-phenotype translation tables for each gene were used for diplotype to phenotype mapping and assessing their priority when implemented in electronic health records (EHR), which was used to deem the phenotype as 'actionable'. Thus, diplotypes with the EHR priority notation 'Abnormal/Priority/High Risk' were considered as actionable diplotypes.

Genotype frequencies were combined when the homozygous alternate and heterozygous genotypes were predicted to lead to actionable phenotypes. Similarly, actionable diplotype frequencies were derived from the combination of multiple diplotypes predicted to lead to actionable phenotypes as described above. For example, *CYP2B6* poor and intermediate metabolizers have an increased risk of adverse events when treated with efavirenz. Hence actionable frequencies were reported as combined frequencies for poor and intermediate metabolizer diplotypes for *CYP2B6*. The metabolizer phenotype can itself be derived from multiple diplotypes. For example, the diplotypes *6/*6, *6/*9, *6/*18, *6/*36, and *18/*20 present in our dataset were considered as *CYP2B6* poor metabolizers based on the CPIC annotation. Details of the actionable diplotypes in each gene present in the population are provided in Supplementary Data S1–S13.

Since the CPIC star allele definition tables do not cover all the combinations of the included variants, some individuals were predicted to have 'novel' star alleles for certain genes, in which case, we assigned them with an 'uncertain' phenotype.

For the identification of the haplotypes and diplotypes of *CYP2D6* in the dataset, Cyrius v1.1 was also employed since our standard analysis was unable to call all the star alleles. Cyrius is a recent tool for specifically genotyping *CYP2D6* from short-read genome sequencing data, which is capable of alleviating the problematic alignment with *CYP2D7*, as well as utilizing information from structural variants⁶⁰.

For the analysis of HLA genes, we used the Population Reference Graph (PRG) framework⁶¹. We identified HLA alleles at 6-digit allelic resolution which takes into account both nonsynonymous and synonymous single-nucleotide variants in the protein-coding region of the HLA genes.

The distribution of frequencies of the actionable variants in the overall Qatari population as well as in the Qatari subpopulations were calculated. As described elsewhere, population stratification using principal components and admixture analysis have identified six distinct subpopulation clusters within the Qatari population with different ethnicity backgrounds: Peninsular Arabs (PAR: 1,052; 17.4%), General Arabs (GAR: 2,311; 38.2%), West Eurasian/Persians (WEP: 1,372; 22.7%), Africans (AFR: 92; 1.5%), South Asians (SAS: 38; 0.6%) and Admixed (ADM: 1,180; 19.5%)⁵³.

We also analyzed 2,504 genomes from the thousand genomes phase 3 consisting of 661 African, 347 American, 504 East Asian, 503 European, 489 South Asian samples⁶² through the same pipeline for all the genes except *CYP2D6* and HLA genes, for which we calculated the actionable genotypes/diplotypes from previously published studies^{60,63}.

Statistical analysis

For each variant known to be associated with drug response as annotated by PharmGKB, we applied a two-proportions z-test to identify statistically significant variants between QGP and gnomAD genomes v 3.1.1. For each

gene in the study of CPIC genes, we applied a two-proportions z-test to compare the observed proportions in the two groups either QGP vs thousand genomes or within the Qatari subpopulations. In all cases, we corrected for multiple hypotheses testing using a Bonferroni procedure. All genes with an adjusted p value < 0.05 were considered as having different proportions in the two groups. Analyses were run using R software version 4.0.4 and graphs plotted using Prism 9 for Mac OS.

The fixation index, F_{ST} , was calculated for understanding the genetic distance between the subpopulations based on the pharmacogenes. We extracted variants found in CPIC genes from original VCF files based on location coordinates of CPIC genes. Hail (v 0.2.45) was used for the extraction of the VCF file. Then the VCF file was filtered for variant missingness of 0.01 and those failed the Hardy Weinberg equilibrium test. Plink 2.0 was used for calculating F_{ST} based on the predefined subpopulation clusters.

Warfarin dosing

We predicted weekly warfarin dosage (mg/week) requirements for all the individuals in the cohort using the International Warfarin Pharmacogenetic Consortium (IWPC) algorithm as shown in Supplementary Data S14⁴². The race was considered as Asian for individuals in the QGP South Asian subpopulation, Black or African American for individuals in the QGP African subpopulation, and the rest were considered Missing or Mixed race. Enzyme inducer status (whether taking carbamazepine, phenytoin, rifampin, or rifampicin) and amiodarone status were obtained from questionnaire data on self-reported medications.

To compare the distribution of the predicted warfarin dosages in the Qatari population with the European population, we analysed the dosages calculated for patients recruited in the EU-PACT trial⁶⁴. Stable dosages available for 325 patients from both arms in this trial were plotted along with the Qatari population data.

Formulary drugs

The availability of CPIC Level A drugs for prescription in the Hamad Medical Corporation (HMC), the primary provider of health services and treatment in Qatar, was checked in the drug formulary database, which was accessed on the 7th January 2021.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

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-new> which is subject to approval by the QBB IRB committee.

Received: 19 July 2021; Accepted: 22 December 2021;

Published online: 15 February 2022

REFERENCES

- Pirmohamed, M. Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. *Annu. Rev. Genomics Hum. Genet.* **15**, 349–370 (2014).
- Gordon, A. S. et al. PGRNseq: a targeted capture sequencing panel for pharmacogenetic research and implementation. *Pharmacogenomics Genom* **26**, 161–168 (2016).
- Al-Mahayri, Z. N. et al. Variation in 100 relevant pharmacogenes among emiratis with insights from understudied populations. *Sci. Rep.* **10**, 21310 (2020).
- Londin, E. R. et al. Performance of exome sequencing for pharmacogenomics. *Personalized Med.* **12**, 109–115 (2014).
- Caspar, S. M., Schneider, T., Stoll, P., Meienberg, J. & Matyas, G. Potential of whole-genome sequencing-based pharmacogenetic profiling. *Pharmacogenomics* **22**, 177–190 (2021).
- Gaedigk, A. et al. The Pharmacogene Variation (PharmVar) Consortium: incorporation of the human cytochrome P450 (CYP) allele nomenclature database. *Clin. Pharmacol. Therapeutics* **103**, 399–401 (2018).

- Gaedigk, A. et al. The evolution of PharmVar. *Clin. Pharmacol. Therapeutics* **105**, 29–32 (2019).
- Whirl-Carrillo, M. et al. Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Therapeutics* **92**, 414–417 (2012).
- Relling, M. V. & Klein, T. E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin. Pharmacol. Therapeutics* **89**, 464–467 (2011).
- Swen, J. et al. Pharmacogenetics: from bench to byte. *Clin. Pharmacol. Therapeutics* **83**, 781–787 (2008).
- Mette, L., Mitropoulos, K., Vozikis, A. & Patrinos, G. P. Pharmacogenomics and public health: implementing 'populationalized' medicine. *Pharmacogenomics* **13**, 803–813 (2012).
- Yasuda, S., Zhang, L. & Huang, S. M. The role of ethnicity in variability in response to drugs: focus on clinical pharmacology studies. *Clin. Pharmacol. Therapeutics* **84**, 417–423 (2008).
- Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).
- Dunnenberger, H. M. et al. Preemptive clinical pharmacogenetics implementation: current programs in five US medical centers. *Annu. Rev. Pharmacol. Toxicol.* **55**, 89–106 (2015).
- van der Wouden, C. H. et al. Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium. *Clin. Pharmacol. Therapeutics* **101**, 341–358 (2017).
- Jithesh, P. V. & Scaria, V. From genomes to genomic medicine: enabling personalized and precision medicine in the Middle East. *Personalized Med.* **14**, 377–382 (2017).
- Sun, Y.-M., Wang, L.-F., Li, J., Li, Z.-Q. & Pan, W. The 223A>G polymorphism of the leptin receptor gene and lipid-lowering efficacy of simvastatin in Chinese patients with coronary heart disease. *Eur. J. Clin. Pharmacol.* **65**, 157–161 (2009).
- Tkáč, I. et al. Pharmacogenomic association between a variant in SLC47A1 gene and therapeutic response to metformin in type 2 diabetes. *Diabetes, Obes. Metab.* **15**, 189–191 (2013).
- Pedersen, A. J., Stage, T. B., Glintborg, D., Andersen, M. & Christensen, M. M. H. The pharmacogenetics of metformin in women with polycystic ovary syndrome: A randomized trial. *Basic Clin. Pharmacol. Toxicol.* **122**, 239–244 (2018).
- Johnson, J. A. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. *Clin. Pharmacol. Therapeutics* **102**, 397–404 (2017).
- Muir, A. et al. Clinical pharmacogenetics implementation consortium (CPIC) guidelines for IFNL3 (IL28B) genotype and PEG interferon- α -based regimens. *Clin. Pharmacol. Therapeutics* **95**, 141–146 (2014).
- Gonsalves, S. G. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for the use of potent volatile anesthetic agents and succinylcholine in the context of RYR 1 or CACNA 1S genotypes. *Clin. Pharmacol. Therapeutics* **105**, 1338–1344 (2019).
- Scott, S. et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin. Pharmacol. Therapeutics* **94**, 317–323 (2013).
- Moriyama, B. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for CYP2C19 and voriconazole therapy. *Clin. Pharmacol. Therapeutics* **102**, 45–51 (2017).
- Hicks, J. K. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2D6 and CYP2C19 genotypes and dosing of selective serotonin reuptake inhibitors. *Clin. Pharmacol. Therapeutics* **98**, 127–134 (2015).
- Hicks, J. K. et al. Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. *Clin. Pharmacol. Therapeutics* **102**, 37–44 (2017).
- Lima, J. J. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2C19 and proton pump inhibitor dosing. *Clin. Pharmacol. Therap.* (2020).
- Desta, Z. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2B6 and efavirenz-containing antiretroviral therapy. *Clin. Pharmacol. Therapeutics* **106**, 726–733 (2019).
- Ramsey, L. B. et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. *Clin. Pharmacol. Therapeutics* **96**, 423–428 (2014).
- Karnes, J. H. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2C9 and HLA-B Genotypes and Phenytoin Dosing: 2020 Update. *Clin. Pharmacol. Therapeutics* **109**, 302–309 (2021).
- Theken, K. N. et al. Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2C9 and nonsteroidal anti-inflammatory drugs. *Clin. Pharmacol. Therapeutics* **108**, 191–200 (2020).
- Crews, K. R. et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. *Clin. Pharmacol. Therapeutics* **95**, 376–382 (2014).

33. Brown, J. T. et al. Clinical Pharmacogenetics Implementation Consortium Guideline for Cytochrome P450 (CYP)2D6 Genotype and Atomoxetine Therapy. *Clin. Pharmacol. Therapeutics* **106**, 94–102 (2019).
34. Bell, G. C. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2D6 genotype and use of ondansetron and tropisetron. *Clin. Pharmacol. Therapeutics* **102**, 213–218 (2017).
35. Goetz, M. P. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and Tamoxifen Therapy. *Clin. Pharmacol. Therapeutics* **103**, 770–777 (2018).
36. Birdwell, K. A. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. *Clin. Pharmacol. Therapeutics* **98**, 19–24 (2015).
37. Relling, M. V. et al. Clinical Pharmacogenetics Implementation Consortium Guideline for Thiopurine Dosing Based on TPMT and NUDT15 Genotypes: 2018 Update. *Clin. Pharmacol. Therapeutics* **105**, 1095–1105 (2019).
38. Amstutz, U. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clin. Pharmacol. Therapeutics* **103**, 210–216 (2018).
39. Martin, M. A. et al. Clinical pharmacogenetics implementation consortium guidelines for HLA-B genotype and abacavir dosing. *Clin. Pharmacol. Therapeutics* **91**, 734–738 (2012).
40. Saito, Y. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for human leukocyte antigen B (HLA-B) genotype and allopurinol dosing: 2015 update. *Clin. Pharmacol. Therapeutics* **99**, 36–37 (2016).
41. Phillips, E. J. et al. Clinical Pharmacogenetics Implementation Consortium Guideline for HLA Genotype and Use of Carbamazepine and Oxcarbazepine: 2017 Update. *Clin. Pharmacol. Therapeutics* **103**, 574–581 (2018).
42. Consortium, I. W. P. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N. Engl. J. Med.* **360**, 753–764 (2009).
43. Asimwe, I. G. et al. Genetic Factors Influencing Warfarin Dose in Black-African Patients: A Systematic Review and Meta-Analysis. *Clin. Pharmacol. Therapeutics* **107**, 1420–1433 (2020).
44. Sivadas, A. & Scaria, V. Pharmacogenomic survey of Qatari populations using whole-genome and exome sequences. *Pharmacogenomics J.* **18**, 590–600 (2018).
45. McInnes, G. et al. Pharmacogenetics at Scale: An Analysis of the UK Biobank. *Clin. Pharmacol. Therapeutics* **109**, 1528–1537 (2021).
46. Al Thani, A. et al. Qatar biobank cohort study: study design and first results. *Am. J. Epidemiol.* **188**, 1420–1433 (2019).
47. Johnson, J. A. et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin. Pharmacol. Therapeutics* **102**, 397–404 (2017).
48. Bader, L. et al. The effect of genetic and nongenetic factors on warfarin dose variability in Qatari population. *Pharmacogenomics J.* **20**, 277–284 (2020).
49. Ghozlan, M. F., Foad, D. A., Darwish, Y. W. & Saad, A. A. Impact of CYP2C9 and VKORC1 genetic polymorphisms upon warfarin dose requirements in Egyptian patients with acute coronary syndrome. *Blood Coagul. Fibrinolysis* **26**, 499–504 (2015).
50. Pathare, A. et al. Warfarin pharmacogenetics: development of a dosing algorithm for Omani patients. *J. Hum. Genet.* **57**, 665–669 (2012).
51. Namazi, S. et al. The impact of genetic polymorphisms and patient characteristics on warfarin dose requirements: a cross-sectional study in Iran. *Clin. Therapeutics* **32**, 1050–1060 (2010).
52. Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y. & Lauschke, V. M. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genom.* **12**, 26 (2018).
53. Thareja, G. et al. Whole genome sequencing in the Middle Eastern Qatari population identifies genetic associations with 45 clinically relevant traits. *Nat. Commun.* **12**, 1250 (2021).
54. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
55. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
56. Roller, E., Ivakhno, S., Lee, S., Royce, T. & Tanner, S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* **32**, 2375–2377 (2016).
57. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
58. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
59. Loh, P.-R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443 (2016).
60. Chen, X. et al. Cyrius: accurate CYP2D6 genotyping using whole-genome sequencing data. *Pharmacogenomics J.* **21**, 251–261 (2021).
61. Dilthey, A. T. et al. High-accuracy HLA type inference from whole-genome sequencing data using population reference graphs. *PLoS computational Biol.* **12**, e1005151 (2016).
62. Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
63. Abi-Rached, L. et al. Immune diversity sheds light on missing variation in worldwide genetic diversity panels. *PLoS ONE* **13**, e0206512 (2018).
64. Pirmohamed, M. et al. A randomized trial of genotype-guided dosing of warfarin. *N. Engl. J. Med.* **369**, 2294–2303 (2013).

ACKNOWLEDGEMENTS

PVJ is supported by faculty funding from the College of Health & Life Sciences, HBKU. Qatar Biobank and Qatar Genome Program are Research, Development & Innovation's entities within Qatar Foundation for Education, Science and Community Development. Funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

AUTHOR CONTRIBUTIONS

M.A., N.S., I.A., M.E., K.B., S.S., U.K.U., Z.J., G.G., C.M., S.S., D.B. C.G., and R.A. analyzed data and contributed to writing. M.A.S., Z.T.C., X.E., and M.P. supervised the work and contributed to writing. P.V.J. conceptualized, implemented, supervised, analyzed and wrote the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

MP receives research funding from various organisations including the UK MRC and NIHR. He has also received partnership funding for the following: MRC Clinical Pharmacology Training Scheme (co-funded by MRC and Roche, UCB, Eli Lilly and Novartis); a PhD studentship jointly funded by EPSRC and Astra Zeneca; and grant funding from Vistagen Therapeutics. He has also unrestricted educational grant support for the UK Pharmacogenetics and Stratified Medicine Network from Bristol-Myers Squibb and UCB. He has developed an HLA genotyping panel with MC Diagnostics, but does not benefit financially from this. He is part of the IML Consortium ARDAT (www.ardat.org). MP is also Vice Chair of the Qatar Precision Medicine Initiative International Scientific Advisory Committee. The remaining authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-022-00281-5>.

Correspondence and requests for materials should be addressed to Puthen Veettil Jithesh.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

THE QATAR GENOME PROGRAM RESEARCH CONSORTIUM

Rania Abdel-latif⁴, Tariq Abu Saqri², Tariq Abu Zaid², Nahla Afifi⁸, Rashid Al-Ali², Souhaila Al-Khodori², Wadha Al-Muftah⁴, Yasser Al-Sarraj⁴, Omar Albagha^{1,9}, Eiman Alkhayat⁸, Fatima Alkuwari⁴, Hakeem Almabrazi², Mashael Alshafai⁵, Asmaa Althani^{4,8}, Muhammad Alvi⁴, Ramin Badii³, Radja Badji⁴, Lotfi Chouchane¹⁰, Dima Darwish⁴, Ahmed El Khouly², Maryem Ennaifar⁴, Xavier Estivill⁶, Tasnim Fadl⁴, Khalid Fakhro^{2,10}, Eleni Fethnou⁸, Mehshad Hamza², Said I. Ismail⁴, Puthen V. Jithesh¹, Mohammedhusen Khatib², Wei Liu², Stephan Lorenz², Hamdi Mbarek⁴, Younes Mokrab^{2,10}, Tushar Pathare², Shafeeq Poolat², Fatima Qafoud⁸, Fazulur Rehaman Vempalli², Chadi Saad⁴, Karsten Suhre^{10,11}, Najeeb Syed², Zohreh Tatari², Ramzi Temanni², Sara Tomei² and Heba Yasin⁴

⁸Qatar Biobank for Medical Research, Qatar Foundation, Doha, Qatar. ⁹Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. ¹⁰Weill Cornell Medicine-Qatar, Doha, Qatar. ¹¹Weill Cornell Medicine, New York, NY, USA.