

Exploration of interethnic variation in the ibuprofen metabolizing enzyme CYP2C9: a cautionary guide for treatment of COVID-19 symptoms

Ammar Almarzooq (✉ ammar@galoreconsultancy.com)

Galore Consultancy

Research Article

Keywords: COVID-19, ibuprofen, NSAIDs, pharmacogenomics, population genetics

Posted Date: January 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-144279/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Coronavirus disease 2019 (COVID-19), is a rapidly spreading infectious illness that causes a debilitating respiratory syndrome. Supportive therapy remains the standard for mild-to-moderate cases, including treatments with non-steroidal anti-inflammatory drugs (NSAIDs) e.g. ibuprofen, such medications may increase COVID-19 complications when used in patients with acute viral respiratory infections.

Methods: P450 enzyme CYP2C9 are known to be involved in the metabolism of NSAIDs, however, their pharmacogenetic data are limited. This study aims to better understand the genetic landscape of CYP2C9 sequence variation across different ethnic and geographic groups, in correlation with ibuprofen dosing guidelines. A cohort of 101 Jordanian Arab samples were retrospectively recruited and genotyped using Affymetrix DMET Plus Premier Package.

Results: This study identified 18 single nucleotide polymorphisms (SNP) within CYP2C9 in 101 Jordanian Arabs, within the context of over 100,000 global subjects in 417 published reports. Genetic structure analysis across populations revealed that Jordanian Arabs share the closest CYP2C9 sequence homology to Near East and European populations. The Jordanian Arab population was also shown to be genetically heterogeneous for CYP2C9, with approximately 33% of individuals sampled exhibiting an 'at risk' SNP requiring non-standard ibuprofen dosing that is different from the established guidelines.

Conclusions: This is the most comprehensive and up-to-date analysis for CYP2C9 allele frequencies across multi-ethnic populations world-wide. The use of modern genomic tools coupled with a proactive assessment of the most likely gene-drug candidates will lead to a better understanding of the role of pharmacogenetics for COVID-19 and more effective treatments.

Trial registration: : The Institutional Review Board (IRB) of the Jordan University of Science and Technology approved this study on 4/7/2013 with approval number 67/2/2013.

Background

A global pandemic of the novel coronavirus disease 2019 (COVID-19) has caused a global healthcare crisis resulting from high infection and mortality rates [1]. The National Health Commission of China released key guidelines for the of diagnosis and treatment of pneumonitis, and related diseases, caused by COVID-19 [2]. Such guidelines included oxygen therapy, mechanical ventilation and drug therapy, as well as provided recommendations for simultaneous drug treatments in extreme cases. For suspected or confirmed cases of COVID-19, requiring urgent care for conditions such as fever and/or sore throat, pharmacological management, in the form of antibiotics and/or analgesics, is an alternative [2]. In addition, non-steroidal anti-inflammatory drugs (NSAIDs) e.g. ibuprofen, may be prescribed for the management of pain and fever. However, uncertainty related to infection etiology and efficacy, and emerging concerns related to the use of common NSAIDs such as ibuprofen, have presented additional challenges in the treatment of COVID-19 [3, 4]. Patients in France and Europe showing symptoms of

COVID-19 were recommended paracetamol (acetaminophen) rather than ibuprofen, as treatment with ibuprofen could exacerbate the condition [3]. Ibuprofen may offer symptomatic relief, and could provide healthcare professionals additional time to deliver customized care and prevent the spread of infection. Although this strategy is not always reliable since individuals may respond differently to similar treatments.

Genetic factors are one of the major contributors to individual or ethnic differences in drug therapeutic efficacy and toxicity [5, 6]. Consequently, host genetics and demography associated with COVID-19 are crucial aspects of infection and prognosis, hence, integral medication dosing might need to be altered based on a patient's genetic information [7, 8]. There are several gene variants that alter how an individual's body metabolizes and processes COVID-19 therapies, potentially increasing the risk of undesirable adverse effects. The cytochrome P450 enzyme CYP2C9, facilitates metabolism of several NSAIDs including ibuprofen, and CYP2C9 allele frequencies have been shown to vary substantially across diverse ethnic groups [9, 10].

In March 2020, the *Clinical Pharmacogenetics Implementation Consortium* (CPIC) published a pharmacogenetic guideline on NSAIDs, with specific therapeutic recommendations for celecoxib, flurbiprofen, ibuprofen, firoxicam, feloxicam, firoxicam, and fenoxicam based on CYP2C9 phenotype [11]. The phenotype was derived from an activity score, obtained by the sum of two individual allele scores. Recent meta-analysis of CYP2C9 alleles and ibuprofen concentrations using the Pharmacogenomics Knowledge Base (PharmGKB), showed strong correlations between the CYP2C9*2 allele (p.R144C; rs1799853) and CYP2C9*3 allele (p.I359L; rs1057910) with plasma levels of ibuprofen [12-16]. Additional meta-analysis of 55 annotations supported by a number of clinical trials, showed a very strong correlation between CYP2C9 genotypes and plasma levels of ibuprofen. Based on these results, PharmGKB assigned the highest level of evidence (level 1A) to these associations [9], indicating strong evidence of pharmacokinetics (PK) or pharmacodynamics (PD) alteration [7]. Furthermore, these have been published as markers by the CPIC dosing guidelines [11], where CPIC level A and B gene/drug pairs have sufficient evidence for at least one prescribing action to be recommended. In contrast, CPIC level C and D gene/drug pairs are considered as having inadequate evidence for actionability to have prescribing recommendations. For example, CYP2C8*3 with level C, has insufficient evidence for recommendations in ibuprofen prescription. However, this study shows that current pharmacogenomic databases can be leveraged to enhance the identification of CYP2C9 alleles, and to determine population differences in drug response/toxicity events. Results from this study have wide ranging impacts in the targeted treatment of COVID-19 patients across broad geographic ranges and ethnic backgrounds, and facilitates drug development processes.

Methods

Sample collection

This study retrospectively included 101 unrelated Jordanian participants, of which 56 were male and 45 were female. After a signed informed consent, 3ml venous blood samples were collected in 3ml EDTA tubes from each participant at the Princess Haya Biotechnology Centre between May 2010 and December 2011. Blood samples were stored at 4°C until DNA extraction. The Institutional Review Board (IRB) of the Jordan University of Science and Technology approved this study on 4/7/2013 under registration number 67/2/2013.

DNA extraction and genotyping

Genomic DNA was extracted from each blood sample using the QIAamp DNA Micro Kit (Qiagen, Germany) according to the manufacturer's instructions. The quality of the purified DNA was determined using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

Genotyping was accomplished using the Affymetrix DMET (Drug Metabolizing Enzymes and Transporters) Plus Premier microarray assay (Santa Clara, CA, USA) to test for drug metabolism associations. The DMET array contains 1936 drug metabolism markers consisting of 1931 single nucleotide polymorphisms (SNPs) and five copy number variations (CNVs) in 225 genes, including 47 phase I enzymes, 80 phase II enzymes, 52 transporters and 46 other genes. These genetic variants were multiplex genotyped using molecular inversion probe (MIP) technology [17]. The profiles for the genotyping call rates and concordance comparisons, were generated by the DMET console software v1.3 (Thermo Fisher Scientific Inc., Waltham, MA, USA), based on the Bayesian robust linear model with Mahalanobis (BRLMM) distance classifier algorithm [18, 19]. Genotypes were determined for each SNP site and reported as homozygous wild-type, heterozygous, homozygous variant, and 'no call' in the case of a lack of genotype call. SNPs with call rate of less than 99% were excluded from subsequent analysis (Figure 1A). Statistical and genetic analyses were performed for selection and validation using Microsoft Excel and SPSS v16. Linkage disequilibrium (LD) analysis was performed to identify non-random SNP associations between populations. LD was in concordance with all worldwide-distributed 1kG-p3

populations [20]. The genotype and allele frequencies were calculated and tested using the chi-square (χ^2) test and the Hardy-Weinberg equilibrium formula ($p > 0.05$).

Selection of CYP2C9 variants

LD analysis was performed using the LDlink tool to generate D' and r^2 values [21], and a matrix was generated for visualization. These allele frequencies were compared to 139 different CPIC reports from European and Near Eastern population groups. The statistical comparison of allele frequencies on experimental data and reference populations were performed by Pearson's chi-squared test with Bonferroni correction and the negative logarithm of the adjusted significance values $[-\log_{10}(\text{adj. p.val})]$ using the R statistical package v3.6.2 with ggplot2 and visualized using Rstudio v1.3.1056 (Boston, MA). The assembled DNA flanking sequence for each of the SNP loci was also subjected to BLAT [22] to determine the specificity of the array's probes matched to the target sequence set. A probe alignment was considered to be specific if 40 consecutive base pairs of the probe were fully aligned with the target sequence [17]. In order to detect the homologous sequences which likely result in false-positive or false-negative variant calls, sequence similarity searches were performed using the Ensembl BLAST/BLAT search programs with default parameters.

Population structure analyses

To identify cryptic relatedness from the genomic data, principal component analysis (PCA) was performed using the base R function "prcomp" within the R software package and multidimensional scaling (MDS) plots were generated from the PCA results using ggfortify and ggplot2 [23]. Cryptic population structure was inferred using CYP2C9 SNP data to identify the ancestral relatedness between the Jordanian Arab population and three defined datasets:

Ethnic defined populations by the 1000 Genomes Project Phase III (1kG-p3) dataset [20], consisting of 1810 individuals from 22 populations within four defined ancestral groups (Additional file 2). The admixed populations were excluded to simplify the ethnic identification analyses.

1. African (AFR): Americans of African Ancestry in SW, USA (ASW), Esan in Nigeria (ESN), Gambian in Western Divisions in the Gambia (GWD), Mende in Sierra Leone (MSL), Luhya in Webuye, Kenya (LWK), Yoruba in Ibadan, Nigeria (YRI) and African Caribbeans in Barbados (ACB).
2. European (EUR): Utah Residents (CEPH) with Northern and Western European Ancestry (CEU), Finnish in Finland (FIN), British in England and Scotland (GBR), Iberian Population in Spain (IBS) and Toscani in Italia (TSI).
3. East Asian (EAS): Han Chinese in Beijing, China (CHB), Chinese Dai in Xishuangbanna, China (CDX), Southern Han Chinese (CHS), Japanese in Tokyo, Japan (JPT) and Kinh in Ho Chi Minh City, Vietnam (KHV).
4. South Asian (SAS): Bengali from Bangladesh (BEB), Indian Telugu from the UK (ITU), Punjabi from Lahore, Pakistan (PJL), Tamil from the UK (STU) and Gujarati Indian from Houston, Texas (GIH).

Two more higher-resolution groups of populations consisting of 3413 individuals from 18 global reports were analyzed (Table 4). The findings were subsequently validated by increasing the coverage of this structure analysis to include more populations from the CPIC group across two geographically-defined groups European (EUR) and Near Eastern (NEA). The genotypic data of individuals from 118 reports consisting of 31880 individuals (Additional file 3), were downloaded from the CPIC updated report in March 2020 [11].

Fixation Index (F_{st}) was used to quantify population differentiation from genetic structure using SNP allele frequencies. The R package “BEDASSLE” function was used to assess genetic similarity between ethnic populations by generating pairwise F_{st} values (0 indicates no divergence, 1 indicates complete separation) between the Jordanian Arab population and the other populations listed above [23].

Pharmacogenetic analyses

Frequencies of the 14 actionable pharmacogenomics biomarkers were assessed cumulatively for the Jordanian Arabs against nine biogeographical groups, consisting of 101407 individuals from 412 global populations [11, 24].

These nine groups were defined by global autosomal genetic structure and based on data from large-scale sequencing initiatives [24] and are used to illustrate the broad diversity of global allele frequencies in this study [24-27]. Furthermore, this biogeographic grouping system meets a key need in pharmacogenetics research by enabling consistent communication of the scale of variability in global allele frequencies and are now used by PharmGKB and CPIC [10, 11 & 24]. The genotypic data of individuals from 412 global populations were downloaded from the CPIC updated report in March 2020.

1. American (AME): The American genetic ancestry group includes populations from both North and South America with ancestors predating European colonization, including American Indian, Alaska Native, First Nations, Inuit, and Métis in Canada, and Indigenous peoples of Central and South America.
2. Central/South Asian (SAS): The Central and South Asian genetic ancestry group includes populations from Pakistan, Sri Lanka, Bangladesh, India, and ranges from Afghanistan to the western border of China.
3. East Asian (EAS): The East Asian genetic ancestry group includes populations from Japan, Korea, and China, and stretches from mainland Southeast Asia through the islands of Southeast Asia. In addition, it includes portions of central Asia and Russia east of the Ural Mountains.
4. European (EUR): The European genetic ancestry group includes populations of primarily European descent, including European Americans. We define the European region as extending west from the Ural Mountains and south to the Turkish and Bulgarian border.

5. Near Eastern (NEA): The Near Eastern genetic ancestry group encompasses populations from northern Africa, the Middle East, and the Caucasus. It includes Turkey and African nations north of the Saharan Desert.
6. Oceanian (OCE): The Oceanian genetic ancestry group includes pre-colonial populations of the Pacific Islands, including Hawaii, Australia, and Papua New Guinea.
7. Sub-Saharan African (SSA): The Sub-Saharan African genetic ancestry group includes individuals from all regions in Sub-Saharan Africa, including Madagascar [8].
8. African American/Afro-Caribbean (AAC): Individuals in the African American/Afro-Caribbean genetic ancestry group reflect the extensive admixture between African, European, and Indigenous ancestries and, as such, display a unique genetic profile compared to individuals from each of those lineages alone. Examples within this cluster include the Coriell Institute's African Caribbean in Barbados (ACB) population, the African Americans from the Southwest US (ASW) population [9], and individuals from Jamaica and the US Virgin Islands.
9. Latino (LAT): The Latino genetic ancestry group is not defined by an exclusive geographic region, but includes individuals of Mestizo descent, individuals from Latin America, and self-identified Latino individuals in the United States. Like the African American/Afro-Caribbean group, the admixture in this population creates a unique genetic pattern compared to any of the discrete geographic regions, with individuals reflecting mixed native and indigenous American, European, and African ancestry.

The total frequency of two SNPs with a Level 1A for the Jordanian Arab population within nine geographically-defined groups were mapped for global impact visualization of allele frequency on ibuprofen response (Table 1, Figure 1). Inferred frequency for CYP2C9*1 was excluded from our biogeographical analyses as no population studies have tested for all known variant alleles, and *1 was not genotyped directly in many studies [11].

Table 1. Sample size accumulated per biogeographical group.

Population Group	Subjects Genotyped (n)
African-American/Afro-Caribbean	4277
American	6791
Central/South Asian	13519
East Asian	26906
European	26739
Latino	6352
Near Eastern	11821
Oceanian	1132
Sub-Saharan Africa	3870
Total	101407

Results

Selection and analysis of CYP2C9 Variations

Eighteen CYP2C9 variants across 101 Jordanian individuals of Arab descent associated with reduced enzyme function were selected (Table 2 & Additional file 4). The defective allele *2 (rs1799853) was the most abundant variant (0.094), followed by allele *3 (rs1057910) (0.084). In addition, two rare variants, c.1425A>T (rs1057911) and 50196C>T (rs2017319), were also detected. These two SNPs had frequencies of less than 0.005. CYP2C9 *2 and *3 together accounted for 17.8% of the allele frequency and about 32.7% of the reduced or non-functional genotype/phenotype associations. The four genotype frequencies CYP2C9 *1/*1, *1/*2, *1/*3 and *2/*3 were 0.673, 0.158, 0.139 and 0.03, respectively. Moreover, the genotype frequencies showed no deviation from HWE ($p > 0.05$; Table 3).

Table 2. Details of the 18 allele frequencies of CYP2C9 gene.

SNP ID	Physical Position	Common Name	CYP2C9 Allele	Marker Call Rate	Minor Allele Frequency
rs1799853	96702047	CYP2C9*2_3608C>T(R144C)	*2	100	0.094
rs1057910	96741053	CYP2C9*3_42614A>C(I359L)	*3	99	0.084
rs56165452	96741054	CYP2C9*4_42615T>C(I359T)	*4	99	0
rs28371686	96741058	CYP2C9*5_42619C>G(D360E)	*5	100	0
rs9332131	96709039	CYP2C9*6_10601delA(K273X)	*6	100	0
rs67807361	96698494	CYP2C9*7_5080C>A(L19I)	*7	99	0.198
rs2256871	96708974	CYP2C9*9_10535A>G(H251R)	*9	100	0
rs9332130	96709037	CYP2C9*10_10598A>G(E272G)	*10	100	0
rs28371685	96740981	CYP2C9*11_42542C>T(R335W)	*11	100	0
rs9332239	96748777	CYP2C9*12_50338C>T(P489S)	*12	100	0
rs72558187	96701715	CYP2C9*13_3276T>C(L90P)	*13	100	0
rs72558189	96701991	CYP2C9*14_3552G>A(R125H)	*14	100	0
rs72558190	96707539	CYP2C9*15_9100C>A(S162x)	*15	100	0
rs72558192	96731936	CYP2C9*16_33497A>G(T299A)	*16	100	0
rs869277704 (rs72558188)	96701970	CYP2C9*25_3531_3540del10	*25	100	0
rs2017319	96748635	CYP2C9_55221C>T(A441A)	NA	99	0.045
rs1057909	96741051	CYP2C9_42612A>G(Y358C)	NA	100	0
rs1057911	96748737	CYP2C9_55323A>T(G475G)	NA	100	0.045

Table 3. Distribution of CYP2C9 alleles and genotypes in Jordanian Arabs.

SNP ID	Allele	Allele Count	Allelic Frequency	Genotype	Genotypic Frequency	Observed Genotypes	Expected Genotypes	χ^2	p -value
rs1799853	C	183	0.91	CC	0.81	82	82.89	1.09	0.58
	T	19	0.09	CT	0.19	19	17.21		
				TT	0.00	0	0.89		
rs1057910	A	185	0.92	AA	0.83	84	84.72	0.85	0.65
	C	17	0.08	AC	0.17	17	15.57		
				CC	0.00	0	0.00		

Significant D' values were observed spanning the entire genomic region, following LD measurements for pairs of SNPs distributed across the 52-kb region. Most allele pairs of CYP2C9 have a D' value equal to 1.0 (indicates complete LD), whereas, r^2 values across the same region, show a LD block between the *7 allele (rs67807361) at exon 1 and the *14 allele (rs72558189) at exon 3. A clear LD block was also observed between CYP2C9*3 (rs1057910) at exon 3 and between c.1425A>T (rs1057911) at exon 9, crossing an approximately 8-kb region (Figure 2A and Additional file 5). The experimental group rs67807361 was significantly different from the other sampled populations ($p=4.9 \times 10^{-22}$; Table 5). However, the nucleotide BLAT search showed that the DNA sequence obtained from the flanking region of this SNP (124bp) had 100% sequence identity with the CYP2C19 and AL583836.1 genes at the region of 10:94762716-94762839. Therefore, this variant was excluded from the analyses since the individual probes in the MIP assay only bind to a genomic footprint of ~40bp. Thus, the homologous sequences would likely result in false-positive or false-negative variant calls [17].

Genetic structure of CYP2C9 across populations

The two leading principal components from the 13 variants shared between the Jordanian Arab population and the 22 global populations from the 1000 Genomes Project Phase III (1kG-p3) dataset (Figure 3A), captured 60.63% and 21.38% of the variance respectively, showing a well-defined separation between Jordanian Arabs and AFR, EAS, and SAS super populations. Jordanian Arabs had a close affinity with EUR (comprised of GBR, FIN, TSI, CEPH, and IBS), and validated by pairwise F_{st} analyses (Additional file 6). The lowest level of differentiation was observed between the Jordanian Arab population and GBR ($F_{st}=5.97 \times 10^{-3}$), followed by IBS ($F_{st}=6.39 \times 10^{-3}$), and FIN ($F_{st}=6.69 \times 10^{-3}$), whereas the greatest divergence was observed with GWD ($F_{st}=8.54 \times 10^{-2}$).

Lack of genomic data for additional ethnic groups in the 1000 Genomes Project such Middle Eastern populations (ESN), can reduce robustness and potentially result in biased geographic-based genomic analysis. Therefore, a secondary analysis was performed to include under-represented Arab populations. The two leading principal components shared between the Jordanian Arab population, and the 18 global reports including ESN for *1, *2 (rs1799853) and *3 (rs1057910); (Table 4) captured 98.35% and 1.62% of the variance, respectively, suggesting a well-defined genetic separation between Jordanian Arabs and AFR and EAS populations (Figure 3B). In addition, defined clusters of EUR and ESN populations were found, which were further validated using pairwise F_{st} analyses (Additional file 7). The lowest level of differentiation was observed between the Jordanian Arab population and Saudi Arabian population ($F_{st}=7.4 \times 10^{-4}$), followed by Italian ($F_{st}=1.62 \times 10^{-3}$) and Turkish populations ($F_{st}=1.8 \times 10^{-3}$), whereas the greatest divergence was observed with the Korean population ($F_{st}=1.13 \times 10^{-1}$).

Table 4. Reference populations of 19 public worldwide populations for the MDS analysis (with the Jordanian Arab population).

		CYP2C9 alleles (%)				
Population		*1	*2	*3	N	Reference
Near Eastern	Jordanian*	82.2	9.4	8.4	101	Current Study
	Lebanese	79.2	11.2	9.6	161	[28]
	Saudi Arabian1	80.8	11.0	8.0	112	[29]
	Egyptian	82.0	12.0	6.0	247	[30]
	Tunisia	78.0	13.9	8.1	258	[31]
	Turkish	79.4	10.6	10.0	499	[32]
	Iranian	79.3	11.0	9.7	160	[33]
European	British	79.0	12.5	8.5	100	[34]
	Italian	79.6	11.2	9.2	157	[35]
	Spanish	74.5	15.6	9.8	102	[36]
African	Afro-American1	98.5	1.0	0.5	100	[37]
	Afro-American2	94.6	2.5	1.3	120	[38]
	Ethiopian	93.3	4.3	2.3	150	[35]
East Asian	Chinese1	98.3	0.0	1.7	115	[39]
	Chinese2	95.1	0.0	4.9	102	[40]
	Japanese1	98.2	0.0	1.8	140	[41]
	Japanese2	97.9	0.0	2.1	218	[42]
	Korean	98.9	0.0	1.1	574	[43]
	Taiwanese	97.4	0.0	2.6	98	[37]

These findings were investigated further by increasing the coverage of the variant analysis to include populations from 118 reports across EUR and ESN (Additional File 3). MDS analysis showed that the two leading principal components of CYP2C9 alleles *2 and *3, shared between the Jordanian Arab population and the 118 reports from EUR and ESN captured 64.45% and 35.55% of the variance, respectively. Jordanian Arabs cluster with

Turkish, Israeli, Caucasian, Italian, Romanian, Iranian and Lebanese populations (Figure 3C).

Pharmacogenetic analyses by biogeographic grouping system

Across the nine biogeographical groups, 27% of subjects were of East Asian origin, followed by Europeans (26%), South Central Asians (13%), Near Easterns (12%), Americans (7%), Latinos (6%), African Americans/Afro-Caribbeans (4%), Sub-Saharan Africans (4%), and Oceanians (1%; Table 1). Distinct differences were found among these populations, with direct impact on ibuprofen clinical outcomes (Figure 4B). The CYP2C9 *2 (rs1799853) and *3 (rs1057910) allele frequencies were significantly higher in the Central/South Asian origin (0.224), followed by Near Easterns (0.212), Europeans (0.203), Jordanian Arabs (0.178), and Latinos (0.116), indicating a decreased metabolism and clearance of ibuprofen as compared to Americans (0.064), Oceanians (0.045), East Asians (0.04), African Americans/Afro-Caribbeans (0.036) and Sub-Saharan Africans (0.024, Table 7). These significant variant alleles and genotypes were classified as PharmGKB Level 1A evidence with reduced enzyme function, and therefore are associated with recommended changes to ibuprofen dosing [7]. Interpretation of the translation into specific dosing guidelines for individual ibuprofen-diplotype pairs [7, 11] showed that Central/South Asian, Near Eastern, and European populations are 7.9x to 5.9x more likely to show impaired CYP2C9 metabolism than African populations (Sub-Saharan and African American/Afro-Caribbean populations, respectively), and 4.9x more likely than East Asian populations (Table 8 and Additional File 8).

Interestingly, a large number of generally less common alleles were also identified based on this approach (Additional File 9). Allele *9 (rs2256871) was significantly over-represented in Sub-Saharan Africans (0.13), but was not detected in other global populations (Figure 4A). Alleles *5 (rs28371686), *6 (rs9332131), *8 (rs7900194) and *11 (rs28371685) were significantly over-represented in African populations (Sub-Saharan African and African Americans/Afro-Caribbean) and under-represented in other

populations. East Asian populations over-represented alleles *42 (rs12414460) and *55 (42620C>A).

Discussion

Although no direct evidence of pharmacogenomics data in patients with COVID-19 was available at the time of writing this manuscript, there are plausible mechanisms by which genetic determinants may play a role in adverse drug responses (ADRs). Having diverse population genetic information and genetic databases, could help clinicians avoid additional risks for treating COVID-19 patients. While NSAIDs, including ibuprofen, may be prescribed for the management of pain and fever, controversy arose on the use of ibuprofen due to the possibility of a worse COVID-19 prognosis. In this work, several genetic markers were analyzed across diverse ethnic backgrounds to identify population differences in drug responses and toxicity events associated with ibuprofen treatment. Results from this study showed that pharmacogenomics studies can be leveraged to enhance the understanding of adverse reactions to the treatment of COVID-19 symptoms and support advancement of drug development pipelines.

Based on the CPIC updated report of the CYP2C9 guidelines [11], seven geographically-defined groups (American, Central/South Asian, East Asian, European, Near Eastern, Oceanian, and Sub-Saharan African) and two admixed groups (African American/Afro-Caribbean and Latino) are described, defined by global autosomal genetic structure based on data from large-scale sequencing initiatives [11, 24]. Geographic grouping pattern was selected as geography has historically been the greatest predictor of genetic variation between human populations, with genetic distance increasing as a function of geographic distance [7]. We also intended these groups to represent people with a predominance of ancestors who were in the region pre-diaspora and pre-colonization [24]. Pharmacogenomics studies are typically not large, from a single country or ethnic group, therefore difficult to implement or incorporate into broader research goals. Therefore, frequencies from larger groups of subjects from multiple different sampled populations are more likely to result in a more accurate estimate. However, it is noted that broadly grouping global populations is an over simplification of human diversity and does not capture complex social and cultural identities. As such, geographic grouping pattern with population group identifiers is an important component of knowledge extraction from curated literature [24].

In this study, 18 genetic markers for the CYP2C9 gene were investigated within 101 Jordanian individuals of Arab descent. The minor alleles (T) and (C) of alleles *2 and *3 (defined by rs1799853 and rs1057910) were 0.094 and 0.084, respectively, which are significantly different from frequencies observed across the additional studied populations. These variant genotypes (CT and AC) were correlated with reduced enzyme function, and therefore are associated with PharmGKB recommendations for changes to ibuprofen dosing [7]. Approximately 33% of these individuals were either intermediate (IM) or poor metabolizers (PM) of ibuprofen based on the sequence variant analysis of CYP2C9. The rs1057911 marker was found to be in LD with the variant of CYP2C9*3 (rs1057910; Figure 1A, Table 3), which is consistent with the recently published PharmVar change adding c.1425A>T (rs1057911) to the *3

haplotype definition [44]. The rs67807361 was also significantly different from other populations, and the nucleotide BLAT search revealed a 100% sequence identity with CYP2C19. Further analysis indicated that the array probe used for genotyping was not able to bind specifically with the target SNPs, due to non-specific binding to another genomic region. Awareness of problematic regions is critical during test design and reporting to guide decisions regarding exclusion of regions and/or whether alternative assays must be used. This is particularly the case for CYP2C9*7_5080C>A(L19I) (rs67807361), where both statistical and genetic tests revealed a homologous sequence that may result in false positive or false negative variant calls.

MDS analysis showed that the Jordanian Arab population clustered with multiple regions within European and Near Eastern, particularly with Turkish, Israeli, Caucasian, Italian, Romanian, Iranian and Lebanese populations. Interestingly, this cluster showed three Turkish populations and two Iranian populations. However, all three Turkish populations were the most similar to Jordanian populations based on MDS analysis. MDS results were further validated by the pairwise F_{st} values, where the lowest level of differentiation was observed between the Jordanian Arab population and Saudi Arabian, followed by the Italian and Turkish. Collectively these results showed that the current Jordanian population today falls into two main groups: one sharing more genetic characteristics with modern-day Europeans and Central Asians, and the other with closer genetic affinities to Arabia [45]. In addition, the autosomal analyses are in agreement with recent studies using large-scale genomics that indicated three major genetic events related to Levant populations. During the late neolithic, gene pools across Anatolia and the Southern Caucasus mixed, resulting in an admixture cline [46]. The second event occurred during the Early Bronze Age, where Northern Levant populations, a region flanked by the Middle East and Europe, experienced gene flow in a process that likely involved a yet to-be-sampled neighboring population from Mesopotamia [46]. The most recent event for the modern Levant was largely determined by subsequent repopulations and mass movements associated with multiple cultural changes within the last two millennia. This appeared to have facilitated and maintained admixture between culturally different populations [47]. Conversion of the region's populations to Islam, appeared to have also introduced major rearrangements in the populations genetic relations with an admixture of culturally similar populations [45]. In general, the Jordanian population was not significantly different from their Levantine neighbours, and fit consistently into a Middle East-Anatolia-Balkan-Caucasus geographic and genetic continuum [48].

PharmGKB meta-analysis of several annotation studies supported by a number of clinical trials [7], showed a very strong correlation between CYP2C9 genotypes and plasma levels of ibuprofen [12-16]. PharmGKB level 1 (1A) represents CYP2C9 *2 (rs1799853) and *3 (rs1057910) alleles with a strong evidence of PK/PD alteration [7]. Mapped frequencies of these alleles showed that Central/South Asian, Near Eastern, and European populations were 7.9x more likely to show impaired CYP2C9 metabolism than Sub-Saharan populations, and 4.9x more likely than East Asian ancestry populations (Table 8). Inference that a higher proportion of East Asian and African ancestry populations have normal ibuprofen metabolism, therefore are less susceptible to complications for ibuprofen-based treatment of COVID-19 related symptoms should be done with caution. Figure 5 shows genotype frequencies of global populations of CYP2C9*2 and *3 for more than 100,000 subjects within 412 reports (Table 8). Group

boundaries for the seven geographical groups fall predominantly along national boundaries to aid the assignment of group membership. The two admixed groups of African American/Afro-Caribbean and Latino were not represented on this figure as the geographical grouping based on the location of genetic ancestors pre-diaspora and pre-colonization could not be applied to the two admixed groups.

These findings are supported by recent European reports of potential harm with ibuprofen usage in patients with COVID-19 symptoms [3]. Further supported by multiple reports, the National Agency for the Safety of Medicines and Health Products (ANSM) of France, issued a warning in April 2019 about the use of NSAIDs for patients with infectious diseases based on an analysis of 20 years of safety data of ibuprofen and ketoprofen. Consequently, the French regulatory body was concerned that existing infections might be worsened by the use of NSAIDs [49]. Furthermore, one large case-control study found an association between NSAIDs and respiratory complications, regardless of whether the NSAIDs were taken long term or as a treatment for acute illness, suggesting that the association was not simply a result of increased prescription in response to acute illness [50].

However, additional research is necessary to clarify whether further variants should be incorporated into clinical decision making. Collectively, this work demonstrates the capability and application of large-scale pharmacogenomics studies to elucidate genetic variation effects on NSAID efficacy in COVID-19 patients. Ultimately, the implementation of pharmacogenetics in clinical settings can lead to more efficient, safer and cost-effective treatments.

Declarations

Ethics approval and consent to participate

Signed informed consent were obtained from all participants recruited for this study under the ethical approval from University Review Committee for Research on Humans at Jordan University of Science and Technology (JUST) registered under [67/2/2013] on 4/7/2013.

Consent for publication

Consent for publication of the data obtained in this study were retrospectively approved, following de-identifying information of individuals.

Availability of data and materials

The datasets generated during the current study are included in the supplementary files. Publicly available repositories used in the analyses were previously published under PMID: 32189324 and downloaded from

https://api.pharmgkb.org/v1/download/file/attachment/CYP2C9_frequency_table.xlsx, and https://api.pharmgkb.org/v1/download/file/attachment/CYP2C9_Diplotype_Phenotype_Table.xlsx.

Competing interests

The author declares that there is no conflict of interest.

Funding

No funding was available for this study.

Authors' contributions

The author conceived the research study, performed all the analyses reported and wrote the manuscript.

Acknowledgements

The technical staff from Princess Haya Biotechnology Center at the Jordan University of Science and Technology provided their assistance with this study.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R *et al*: **A Novel Coronavirus from Patients with Pneumonia in China, 2019**. *N Engl J Med* 2020, **382**(8):727-733.
2. Lin L, Li TS: **[Interpretation of "Guidelines for the Diagnosis and Treatment of Novel Coronavirus (2019-nCoV) Infection by the National Health Commission (Trial Version 5)"]**. *Zhonghua Yi Xue Za Zhi* 2020, **100**(11):805-807.
3. Day M: **Covid-19: ibuprofen should not be used for managing symptoms, say doctors and scientists**. *BMJ* 2020, **368**:m1086.
4. Fang L, Karakiulakis G, Roth M: **Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?** *Lancet Respir Med* 2020, **8**(4):e21.
5. Gray JA: **The shift to personalised and population medicine**. *Lancet* 2013, **382**(9888):200-201.
6. Peterson JF, Roden DM, Orlando LA, Ramirez AH, Mensah GA, Williams MS: **Building evidence and measuring clinical outcomes for genomic medicine**. *Lancet* 2019, **394**(10198):604-610.

7. Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, Mountain JL, Perez-Stable EJ, Sheppard D, Risch N: **The importance of race and ethnic background in biomedical research and clinical practice.** *N Engl J Med* 2003, **348**(12):1170-1175.
8. Gemmati D, Tisato V: **Genetic Hypothesis and Pharmacogenetics Side of Renin-Angiotensin-System in COVID-19.** *Genes (Basel)* 2020, **11**(9).
9. Ramamoorthy A, Pacanowski MA, Bull J, Zhang L: **Racial/ethnic differences in drug disposition and response: review of recently approved drugs.** *Clin Pharmacol Ther* 2015, **97**(3):263-273.
10. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, Altman RB, Klein TE: **Pharmacogenomics knowledge for personalized medicine.** *Clin Pharmacol Ther* 2012, **92**(4):414-417.
11. Theken KN, Lee CR, Gong L, Caudle KE, Formea CM, Gaedigk A, Klein TE, Agundez JAG, Grosser T: **Clinical Pharmacogenetics Implementation Consortium Guideline (CPIC) for CYP2C9 and Nonsteroidal Anti-Inflammatory Drugs.** *Clin Pharmacol Ther* 2020, **108**(2):191-200.
12. Garcia-Martin E, Martinez C, Tabares B, Frias J, Agundez JA: **Interindividual variability in ibuprofen pharmacokinetics is related to interaction of cytochrome P450 2C8 and 2C9 amino acid polymorphisms.** *Clin Pharmacol Ther* 2004, **76**(2):119-127.
13. Karazniewicz-Lada M, Luczak M, Glowka F: **Pharmacokinetic studies of enantiomers of ibuprofen and its chiral metabolites in humans with different variants of genes coding CYP2C8 and CYP2C9 isoenzymes.** *Xenobiotica* 2009, **39**(6):476-485.
14. Kirchheiner J, Meineke I, Freytag G, Meisel C, Roots I, Brockmoller J: **Enantiospecific effects of cytochrome P450 2C9 amino acid variants on ibuprofen pharmacokinetics and on the inhibition of cyclooxygenases 1 and 2.** *Clin Pharmacol Ther* 2002, **72**(1):62-75.
15. Lopez-Rodriguez R, Novalbos J, Gallego-Sandin S, Roman-Martinez M, Torrado J, Gisbert JP, Abad-Santos F: **Influence of CYP2C8 and CYP2C9 polymorphisms on pharmacokinetic and pharmacodynamic parameters of racemic and enantiomeric forms of ibuprofen in healthy volunteers.** *Pharmacol Res* 2008, **58**(1):77-84.
16. Ochoa D, Prieto-Perez R, Roman M, Talegon M, Rivas A, Galicia I, Abad-Santos F, Cabaleiro T: **Effect of gender and CYP2C9 and CYP2C8 polymorphisms on the pharmacokinetics of ibuprofen enantiomers.** *Pharmacogenomics* 2015, **16**(9):939-948.
17. Wang Y, Cottman M, Schiffman JD: **Molecular inversion probes: a novel microarray technology and its application in cancer research.** *Cancer Genet* 2012, **205**(7-8):341-355.
18. Rabbee N, Speed TP: **A genotype calling algorithm for affymetrix SNP arrays.** *Bioinformatics* 2006, **22**(1):7-12.
19. Sissung TM, English BC, Venzon D, Figg WD, Deeken JF: **Clinical pharmacology and pharmacogenetics in a genomics era: the DMET platform.** *Pharmacogenomics* 2010, **11**(1):89-103.
20. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA *et al*: **A global reference for human genetic variation.** *Nature* 2015, **526**(7571):68-74.

21. Machiela MJ, Chanock SJ: **LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants.** *Bioinformatics* 2015, **31**(21):3555-3557.
22. Kent WJ: **BLAT—the BLAST-like alignment tool.** *Genome Res* 2002, **12**(4):656-664.
23. R Core Team: **R: A language and environment for statistical computing.** In. Vienna, Austria: R Foundation for Statistical Computing, Vienna, Austria.; 2013.
24. Huddart R, Fohner AE, Whirl-Carrillo M, Wojcik GL, Gignoux CR, Popejoy AB, Bustamante CD, Altman RB, Klein TE: **Standardized Biogeographic Grouping System for Annotating Populations in Pharmacogenetic Research.** *Clin Pharmacol Ther* 2019, **105**(5):1256-1262.
25. Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calo C, De Montis A, Atzori M, Marini M, Tofanelli S, Francalacci P *et al*: **Geographic population structure analysis of worldwide human populations infers their biogeographical origins.** *Nat Commun* 2014, **5**:3513.
26. Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R *et al*: **Genotype, haplotype and copy-number variation in worldwide human populations.** *Nature* 2008, **451**(7181):998-1003.
27. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL *et al*: **Worldwide human relationships inferred from genome-wide patterns of variation.** *Science* 2008, **319**(5866):1100-1104.
28. Djaffar-Jureidini I, Chamseddine N, Keleshian S, Naoufal R, Zahed L, Hakime N: **Pharmacogenetics of coumarin dosing: prevalence of CYP2C9 and VKORC1 polymorphisms in the Lebanese population.** *Genet Test Mol Biomarkers* 2011, **15**(11):827-830.
29. Al-Saikhan FI: **Genetic risk assessment towards warfarin application: Saudi Arabia study with a potential to predict and prevent side effects.** *Saudi J Biol Sci* 2020, **27**(1):456-459.
30. A LL, Dorado P, O'Kirwan F, Jepson R, Licinio J, Wong ML: **Lower frequency of CYP2C9*2 in Mexican-Americans compared to Spaniards.** *Pharmacogenomics J* 2004, **4**(6):403-406.
31. Abdelhedi R, Bouayed NA, Alfadhli S, Abid L, Rebai A, Kharrat N: **Characterization of drug-metabolizing enzymes CYP2C9, CYP2C19 polymorphisms in Tunisian, Kuwaiti and Bahraini populations.** *J Genet* 2015, **94**(4):765-770.
32. Aynacioglu AS, Brockmoller J, Bauer S, Sachse C, Guzelbey P, Ongen Z, Nacak M, Roots I: **Frequency of cytochrome P450 CYP2C9 variants in a Turkish population and functional relevance for phenytoin.** *Br J Clin Pharmacol* 1999, **48**(3):409-415.
33. Peyvandi F, Spreafico M, Karimi M, Zeinali S, Mannucci PM, Bianchi Bonomi A: **Allele frequency of CYP2C9 gene polymorphisms in Iran.** *Thromb Haemost* 2002, **88**(5):874-875.
34. Stubbins MJ, Harries LW, Smith G, Tarbit MH, Wolf CR: **Genetic analysis of the human cytochrome P450 CYP2C9 locus.** *Pharmacogenetics* 1996, **6**(5):429-439.
35. Scordo MG, Aklillu E, Yasar U, Dahl ML, Spina E, Ingelman-Sundberg M: **Genetic polymorphism of cytochrome P450 2C9 in a Caucasian and a black African population.** *Br J Clin Pharmacol* 2001, **52**(4):447-450.

36. Dorado P, Berez R, Norberto MJ, Yasar U, Dahl ML, A LL: **CYP2C9 genotypes and diclofenac metabolism in Spanish healthy volunteers.** *Eur J Clin Pharmacol* 2003, **59**(3):221-225.
37. Sullivan-Klose TH, Ghanayem BI, Bell DA, Zhang ZY, Kaminsky LS, Shenfield GM, Miners JO, Birkett DJ, Goldstein JA: **The role of the CYP2C9-Leu359 allelic variant in the tolbutamide polymorphism.** *Pharmacogenetics* 1996, **6**(4):341-349.
38. Dickmann LJ, Rettie AE, Kneller MB, Kim RB, Wood AJ, Stein CM, Wilkinson GR, Schwarz UI: **Identification and functional characterization of a new CYP2C9 variant (CYP2C9*5) expressed among African Americans.** *Mol Pharmacol* 2001, **60**(2):382-387.
39. Wang SL, Huang J, Lai MD, Tsai JJ: **Detection of CYP2C9 polymorphism based on the polymerase chain reaction in Chinese.** *Pharmacogenetics* 1995, **5**(1):37-42.
40. Gaedigk A, Casley WL, Tyndale RF, Sellers EM, Jurima-Romet M, Leeder JS: **Cytochrome P4502C9 (CYP2C9) allele frequencies in Canadian Native Indian and Inuit populations.** *Can J Physiol Pharmacol* 2001, **79**(10):841-847.
41. Kimura M, Ieiri I, Mamiya K, Urae A, Higuchi S: **Genetic polymorphism of cytochrome P450s, CYP2C19, and CYP2C9 in a Japanese population.** *Ther Drug Monit* 1998, **20**(3):243-247.
42. Nasu K, Kubota T, Ishizaki T: **Genetic analysis of CYP2C9 polymorphism in a Japanese population.** *Pharmacogenetics* 1997, **7**(5):405-409.
43. Yoon YR, Shon JH, Kim MK, Lim YC, Lee HR, Park JY, Cha IJ, Shin JG: **Frequency of cytochrome P450 2C9 mutant alleles in a Korean population.** *Br J Clin Pharmacol* 2001, **51**(3):277-280.
44. Gaedigk A, Ingelman-Sundberg M, Miller NA, Leeder JS, Whirl-Carrillo M, Klein TE, PharmVar Steering C: **The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database.** *Clin Pharmacol Ther* 2018, **103**(3):399-401.
45. Haber M, Gauguier D, Youhanna S, Patterson N, Moorjani P, Botigue LR, Platt DE, Matisoo-Smith E, Soria-Hernanz DF, Wells RS *et al*: **Genome-wide diversity in the levant reveals recent structuring by culture.** *PLoS Genet* 2013, **9**(2):e1003316.
46. Skourtanioti E, Erdal YS, Frangipane M, Balossi Restelli F, Yener KA, Pinnock F, Matthiae P, Ozbal R, Schoop UD, Guliyev F *et al*: **Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus.** *Cell* 2020, **181**(5):1158-1175 e1128.
47. Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, Latini V, Varesi L, Memmi MM, Romano V *et al*: **Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region.** *Hum Immunol* 2001, **62**(9):871-884.
48. Flores C, Maca-Meyer N, Larruga JM, Cabrera VM, Karadsheh N, Gonzalez AM: **Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan.** *J Hum Genet* 2005, **50**(9):435-441.
49. **EU Regulatory Roundup: France's ANSM Warns About NSAIDs Following Safety Review**
[<https://www.raps.org/news-and-articles/news-articles/2019/4/eu-regulatory-roundup-frances-an-sm-warns-about-n>]

50. Basille D, Thomsen RW, Madsen M, Duhaut P, Andrejak C, Jounieaux V, Sorensen HT: **Nonsteroidal Antiinflammatory Drug Use and Clinical Outcomes of Community-acquired Pneumonia.** *Am J Respir Crit Care Med* 2018, **198**(1):128-131.

Tables

Table 5. The negative logarithm of adjusted significance values.

SNP ID	p-value	Heatmap (-log p-value)
rs1799853	3.70E-01	0.0
rs1057910	8.04E-01	0.0
rs56165452	8.17E-01	0.0
rs28371686	1.0	0.0
rs9332131	1.0	0.0
rs67807361	4.93E-22	18.71
rs2256871	N/A	N/A
rs9332130	N/A	N/A
rs28371685	1.0	0.0
rs9332239	1.19E-01	0.0
rs72558187	N/A	N/A
rs72558189	1.19E-01	0.0
rs72558190	N/A	N/A
rs72558192	N/A	N/A
rs869277704	N/A	N/A
rs2017319	N/A	N/A
rs1057909	N/A	N/A
rs1057911	N/A	N/A

Table 6. Global phenotype frequencies of CYP2C9 core alleles in biogeographical groups against the experimental data.

EHR Priority Result Notation A	Priority/ High Risk	Normal	None	Total
Central/South Asian	0.40		0.00	0.40
Near Eastern	0.39		0.00	0.39
European	0.37		0.00	0.37
Jordanian Arabs*	0.33		0.00	0.33
Sub-Saharan African	0.27	0.02	0.00	0.29
Latino	0.25		0.00	0.25
African American/Afro-Caribbean	0.24		0.00	0.24
American	0.17		0.00	0.17
East Asian	0.15		0.00	0.15
Oceanian	0.09		0.00	0.09

Table 7. Global allele frequencies of CYP2C9 level A alleles in biogeographical groups against the experimental data.

CYP2C9 allele	*2	*3
Jordanian Arabs*	0.094	0.084
African American/ Afro-Caribbean	0.022	0.014
American	0.033	0.030
Central/South Asian	0.114	0.110
East Asian	0.002	0.038
European	0.127	0.076
Latino	0.076	0.040
Near Eastern	0.130	0.083
Oceanian	0.029	0.016
Sub-Saharan African	0.013	0.011

Table 8. Global phenotype frequencies of CYP2C9 level A alleles in biographical groups against the experimental data.

EHR Priority Result Notation A	Priority /	Normal	None	Total
	High Risk			
Central/South Asian	0.40	0.00	0.00	0.40
Near Eastern	0.38	0.00	0.00	0.38
European	0.36		0.00	0.36
Jordanian Arabs*	0.33		0.00	0.33
Latino	0.22	0.00	0.00	0.22
American	0.12	0.00	0.00	0.12
Oceanian	0.09		0.00	0.09
East Asian	0.08		0.00	0.08
African American/Afro-Caribbean	0.07	0.00	0.00	0.07
Sub-Saharan African	0.05		0.00	0.05

Figures

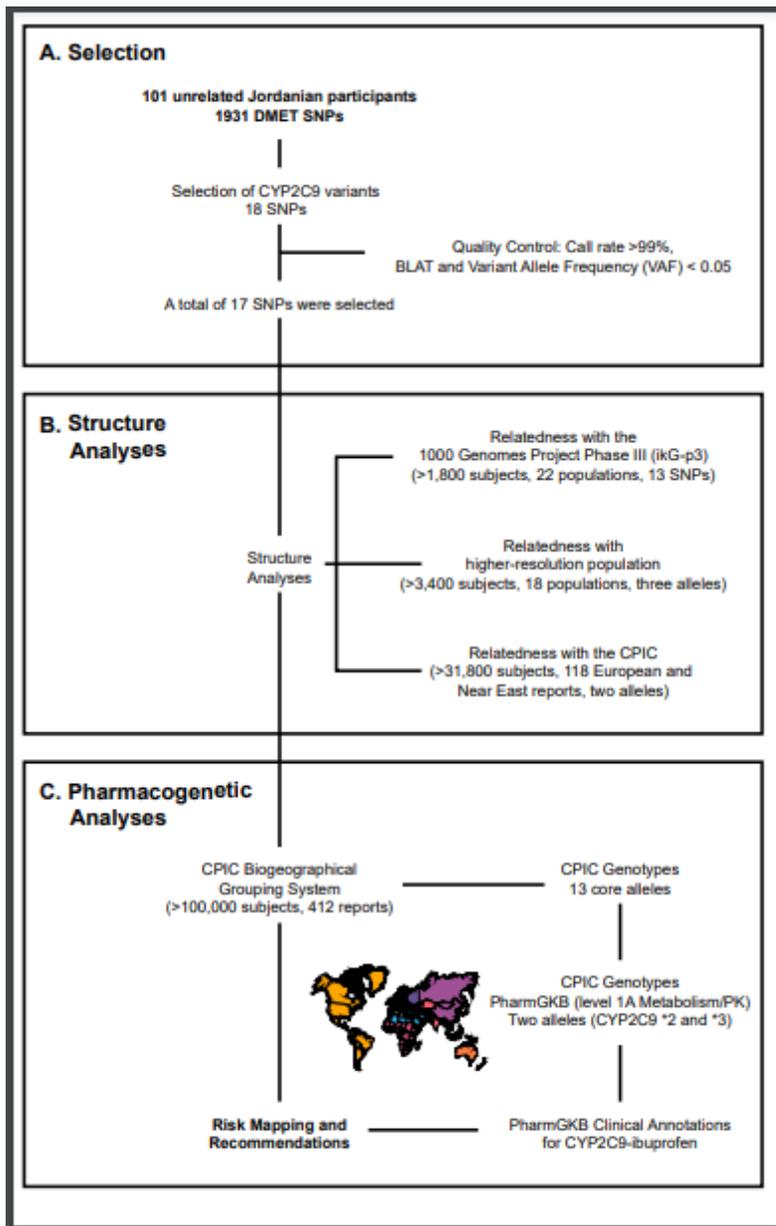


Figure 1

Flow chart depicting various steps involved in variant selection, analysis, and annotation. A. Selection of CYP2C9 Variations by several statistical and genetic analyses. B. Structure Analysis by multidimensional scaling (MDS) plots between the Jordanian Arab population and three defined datasets: 1,810 individuals from 22 populations from the 1000 Genomes Project Phase III (1kG-p3) dataset (Table S5), 3,413 individuals from 18 global reports (Table 4), and 31,880 individuals from 118 reports of the European (EUR) and Near Eastern (NEA) populations from the CPIC updated report (Additional file 3). Fixation Index (F_{st}) was used to quantify population differentiation from genetic structure using SNP allele frequencies. C. Finally, the findings were adapted with nine geographically-defined groups that have been compiled for CPIC guidelines (>100,000 subjects, 412 reports) in order to estimate genotype-predicted phenotype status across world populations.

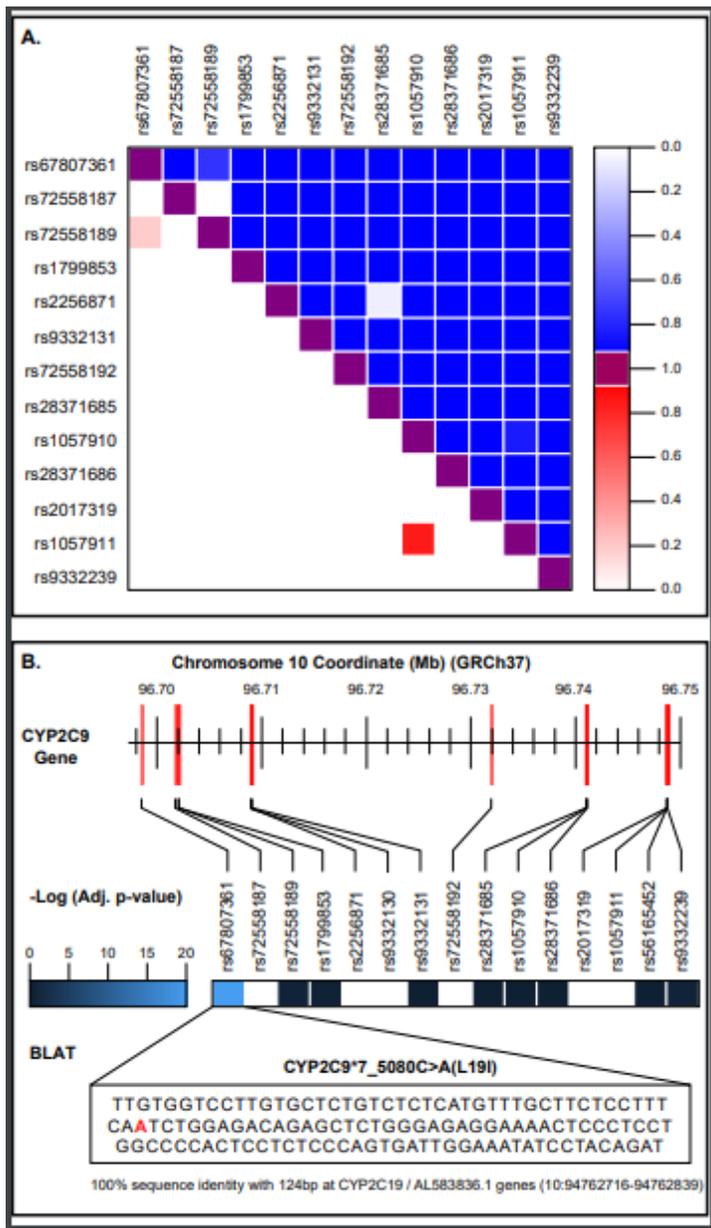


Figure 2

CYP2C9 genetic interpretation in the context of Linkage Disequilibrium (LD) and high homology sequence. A. LD (D' and r^2) between 13 single nucleotide polymorphisms (SNPs) in the CYP2C9 gene. B. Statistical comparison of allele frequencies. Grey areas indicate non-significant results due to absence or low sample size in the population data. Color intensity indicates a more significant difference between the given experimental data and the population. The DNA sequence obtained from the flanking region of this SNP (124 bp) showed a 100% sequence identity with the CYP2C19 and AL583836.1 genes at the region of 10:94762716-94762839.

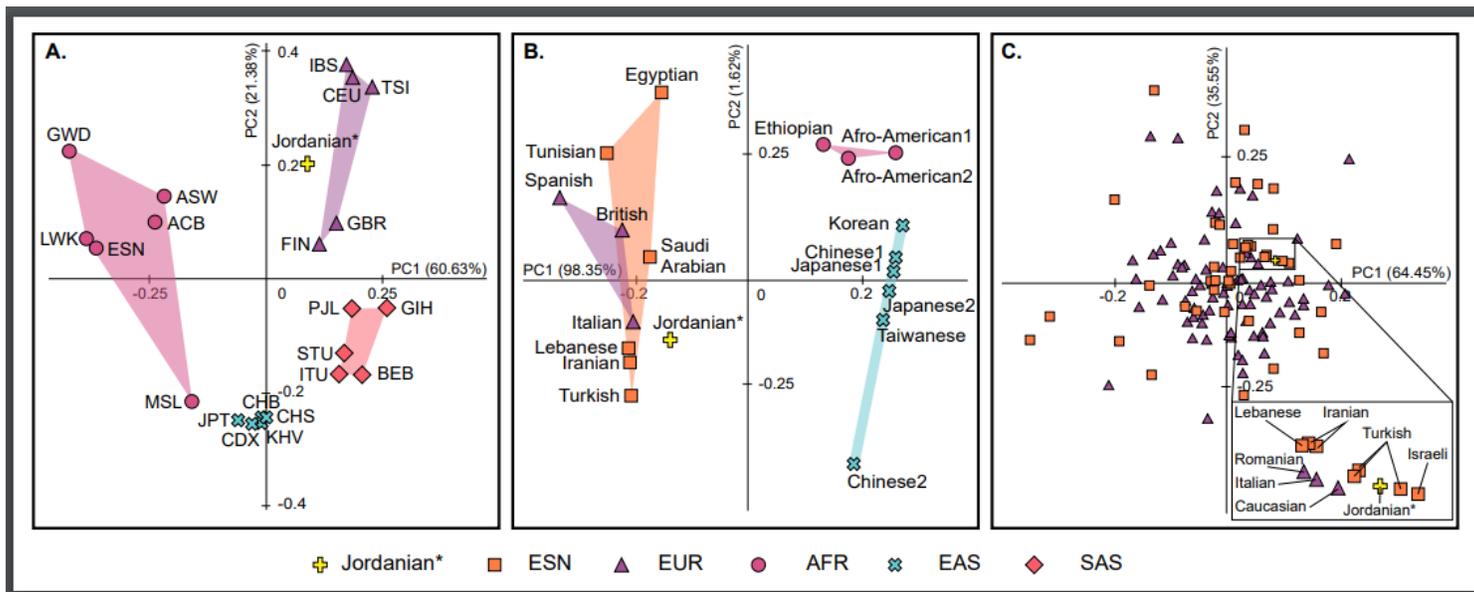


Figure 3

Population structure analysis of the CYP2C9 *1, *2 (defined by rs1799853) and *3 (defined by rs1057910) alleles. A. MDS plot of the 13 variants shared between the Jordanian Arab population and 22 global populations from the 1000 Genomes Project Phase III (1kG-p3) dataset. B. MDS plot of the Jordanian Arab population and 18 public worldwide populations. C. MDS plot of the Jordanian Arab population and 118 populations from the Near Eastern (ESN) and European (EUR) groups.

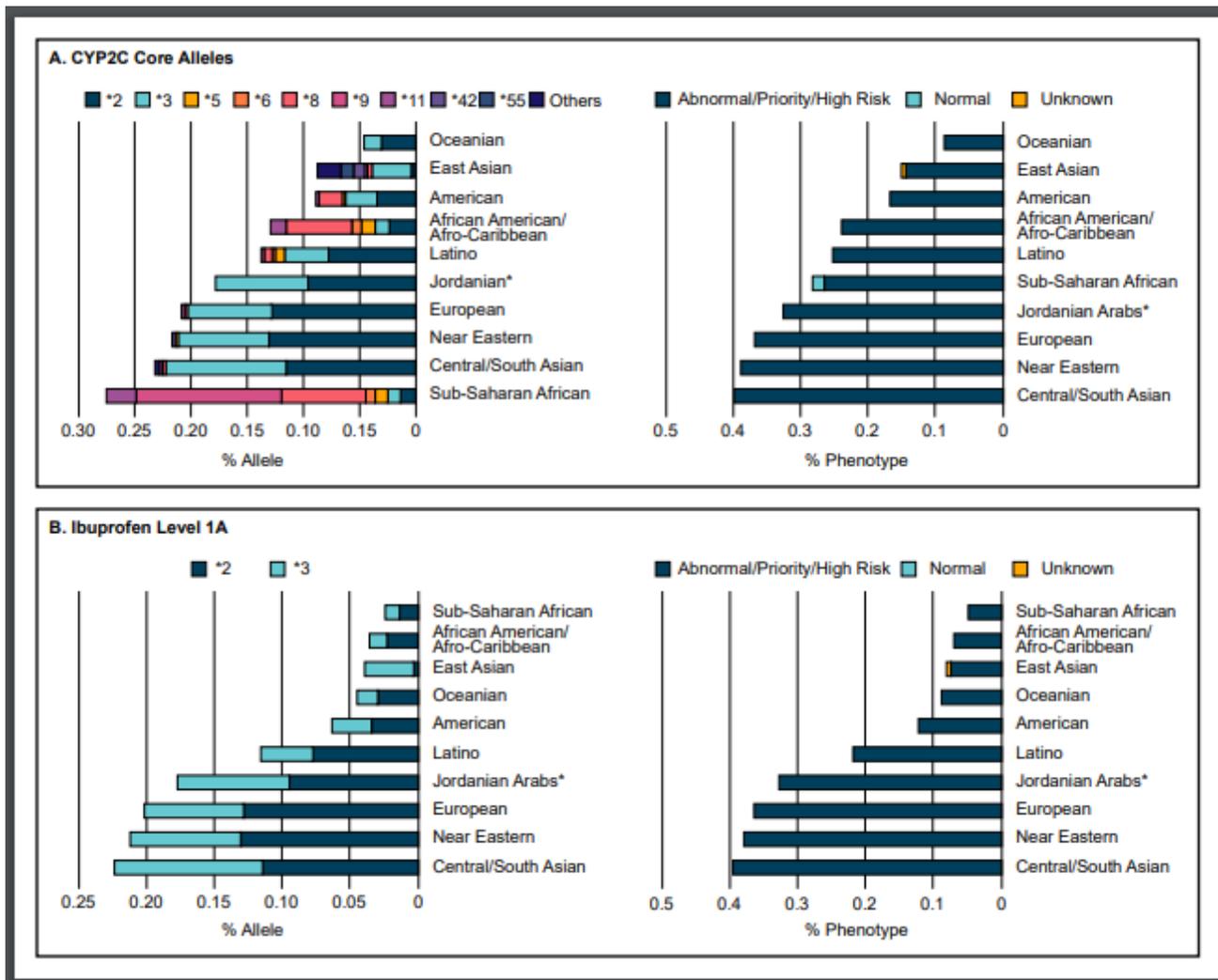


Figure 4

Biogeographical groups and pharmacogenetic analyses for CYP2C9 and ibuprofen associations. A. Frequencies of 14 actionable pharmacogenomics (PGx) alleles and phenotypes assessed cumulatively for the Jordanian Arab population against the nine biogeographical groups (Table 6, Additional files 9 & 10). B. Mapping of the total allele and phenotype frequency of the CYP2C9 alleles *2 (defined by rs1799853) and *3 (defined by rs1057910), strongly associated with Level 1A PharmGKB (Table 7, 8 & Additional file 8).



Figure 5

Global frequency map of variability in of CYP2C9 level A alleles and impact on ibuprofen response. Genotype frequencies shown for global populations of CYP2C9*2 and *3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfiles.xlsx](#)