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

Ammar Ali Almarzooq[1]

[1]Galore Consultancy, Flat 12, Building 144, Road 336, Block 314, Manama, Bahrain.

Corresponding author:

Ammar Ali Almarzooq,

Flat 12, Building 144, Road 336, Block 314, Manama, Bahrain.

tel: +973-32171842

email: ammar@galoreconsultancy.com

ABSTRACT

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 treatment with non-steroidal anti-inflammatory drugs (NSAIDs) e.g. ibuprofen, however such medications may increase COVID-19 complications when used in patients with acute viral respiratory infections. 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. This study identified 18 single nucleotide polymorphisms (SNPs) within CYP2C9 in these 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. However, European populations are 7.2x more likely to show impaired ibuprofen metabolism than Sub-Saharan populations, and 4.5x more likely than East Asian ancestry populations.. 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.

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

Introduction

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]. For suspected or confirmed cases of COVID-19, requiring urgent care for conditions such as fever and/or sore throat, pharmacological management may require antibiotics and/or analgesics as 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, fornoxicam, 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 (g.47639A>C; p.(Ile359Leu)) and CYP2C9*3 allele (g.47639A>C; p.(Ile359Leu)) with plasma levels of ibuprofen[12-16]. Based on these results, PharmGKB assigned the highest level of evidence (level 1A) to these associations[9], indicating strong evidence of pharmacokinetic (PK) or pharmacodynamic (PD) alteration[7]. Furthermore, 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 pharmacogenomics 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 on 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 54 were male and 47 were female. After a signed informed consent, a 3ml venous blood sample was

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, Hilden, Germany) according to the manufacturer's instructions. The quality of the purified DNA was determined by using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA USA).

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 1,936 drug metabolism markers consisting of 1,931 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 a call rate of less than 99% were excluded from subsequent analyses. 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 ($p > 0.05$). The workflow applied in this study is summarized in Supplementary material file (Fig. S1).

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 (Table S1). The statistical comparison of allele frequencies on experimental data and reference populations were performed by Pearson's χ^2 test with Bonferroni correction and the negative logarithm of the adjusted significance values [-log10 (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: 1,810 individuals from 22 populations from the 1000 Genomes Project Phase III (1kG-p3) dataset, excluding admixed populations (Table S2); 3,413 individuals from 18 global reports (Table S3); and 31,880 individuals from 118 reports from the European (EUR) and Near Eastern (NEA) populations from the CPIC updated report in March 2020. For a detailed description, refer to the supplementary material.

Fixation Index (Fst) 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 Fst values (0 indicates no divergence, 1 indicates complete separation) between the Jordanian Arab population and the other populations listed above[23].

Pharmacogenetic analyses

The frequencies of the 13 actionable pharmacogenomics biomarkers were assessed cumulatively for the Jordanian Arabs against nine biogeographical groups, consisting of 101,407 individuals from 412 global populations[11,24]. 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 (Supplementary material file). 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] (see Supplemental Material for populations details).

RESULTS

Selection and analysis of CYP2C9 Variations

It is worth mentioning that we have screened all the variants listed in Table S5 for variants fulfilling our selection criteria to ensure that our pipeline had not missed any known causal or candidate variants in this gene. Four CYP2C9 variants across 101 Jordanian individuals of Arab descent associated with reduced enzyme function were selected (Table S5 and S6). The defective allele *2 was the most abundant variant (0.094), followed by allele *3 (0.084). In addition, two rare variants, g.55323A>T; and g.55221C>T; were also detected. These two SNPs had frequencies of less than 0.05. CYP2C9 *2 and *3 together accounted for 0.178 of the allele frequency and about 0.327 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 1).

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 (indicating complete LD), whereas, r^2 values across the same region, show a LD block between the *7 allele (g.5080C>A) at exon 1 and the *14 allele (g.8577G>A) at exon 3. A clear LD block was also observed between CYP2C9*3 (g.47639A>C) at exon 3 and between g.55323A>T at exon 9, crossing an approximately 8-kb region (Fig. 1A and Table S7). The experimental allele of CYP2C9*7 (g.5080C>A; p.(Leu19Ile)) was significantly different from the other sampled populations ($p = 4.9x10^{22}$; Table S8). However, the nucleotide BLAT search showed that the DNA sequence obtained from the flanking region of this SNP (124 bp) had 100% sequence identity with the CYP2C19 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 ~40 bp. 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 14 variants shared between the Jordanian Arab population and the 22 global populations from the 1000 Genomes Project Phase III (1kG-p3) dataset (Fig. 2A), captured 60.63% and 21.38% of the variance respectively, showing a well-defined separation between the Jordanian Arab population and African, East Asian, and South Asian super populations. The Jordanian Arab population had a close affinity with European populations, and validated by pairwise Fst analyses (Table S9). The lowest level of differentiation was observed between the Jordanian Arab population and British in England and Scotland (Fst = 5.97×10^{-3}), followed by Iberian (Fst = 6.39×10^{-3}) and Finnish in Finland (Fst = 6.69×10^{-3}) populations, whereas the greatest divergence was observed with Gambian in western divisions in the Gambia (Fst = 8.54×10^{-2}).

Lack of genomic data for additional ethnic groups in the 1000 Genomes Project such as near eastern populations, can reduce robustness and potentially result in biased geographic-based genomic analysis. Therefore, a secondary analysis was performed to include under-represented populations. The two leading principal components shared between the Jordanian Arab population, and the 18 global reports including the Near Eastern population for *1, *2 and *3 captured 98.35% and 1.62% of the variance, respectively, suggesting a well-defined genetic separation between Jordanian Arabs and African and East Asian populations (Fig. 2B). In addition, defined clusters of European and Near Eastern populations were found, which were further validated using pairwise Fst analyses (Table S10). The lowest level of differentiation was observed between the Jordanian Arab population and Saudi Arabian population (Fst = 7.4×10^{-4}), followed by Italian (Fst = $1.62 \times 10^{-}$

^{3}) and Turkish populations (Fst = $1.8x10^{-3}$), whereas the greatest divergence was observed with the Korean population (Fst = $1.13x10^{-1}$).

These findings were investigated further by increasing the coverage of the variant analysis to include populations from 118 reports across European and Near Eastern groups. MDS analysis showed that Jordanian Arabs cluster with Turkish, Israeli, Caucasian, Italian, Romanian, Iranian and Lebanese populations (Fig. 2C).

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 S11). Distinct differences were found among these populations, with direct impacts on ibuprofen clinical outcomes (Fig. 3B). The CYP2C9 *2 and *3 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 S12). These significant variant alleles and genotypes were classified as PharmGKB Level 1A evidence with reduced enzyme function, 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 European populations are 7.2x more likely to show impaired CYP2C9 metabolism than Sub-Saharan populations, and 4.5x more likely than East Asian populations (Tables S13 and S14).

Interestingly, a large number of generally less common alleles were also identified (Table S15). Allele *9 (g.15560A>G) was significantly over-represented in Sub-Saharan Africans (0.13), but was not detected in other global populations (Fig. 3A). Alleles *5 (g.47644C>G), *6 (g.15626delA), *8 (g.3260T>C; g.8652G>A; g.8652G>T) and *11 (g.47567C>T) 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 (g.8573C>T) and *55 (g.47645C>A) (Tables S16 and S17).

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. Having diverse population genetic information and genetic databases, could help clinicians avoid additional risks for treating COVID-19 patients. 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. Moreover, controversy arose on the use of ibuprofen due to the possibility of a worse COVID-19 prognosis[3,4]. 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.

Approximately 33% within 101 individuals of Jordanian Arab descent were either intermediate or poor metabolizers of ibuprofen based on the sequence variant analysis of CYP2C9. The g.55323A>T marker was also found to be in LD with the variant of CYP2C9*3 (Fig. 1A, Table S7), which is consistent with the recently published PharmVar

change adding g.55323A>T to the *3 haplotype definition[25]. The CYP2C9*7 was 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 SNP, due to non-specific binding to another genomic region. Awareness of problematic regions is critical during test design and reporting to guide decision regarding exclusion of regions and/or whether alternative assays must be used. This is particularly the case for CYP2C9*7, 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. MDS results were further validated by the pairwise Fst 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[26]. These autosomal analyses are in agreement with recent studies using large-scale genomics that indicated three major genetic events related to Levant populations, including Jordanian. During the late neolithic, gene pools across Anatolia and the Southern Caucasus mixed, resulting in an admixture cline[27]. 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[27]. 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[28]. 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[26]. 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[29].

Mapped frequencies of CYP2C9 genotypes of *2 and *3 alleles based on both the PharmGKB meta-analysis[7,12-16] and the CPIC updated report[11], showed that European populations were 7.2x more likely to show impaired CYP2C9 metabolism than Sub-Saharan populations, and 4.5x more likely than East Asian ancestry populations (Fig. 4, Table S13). Inference that a higher proportion of East Asian and African ancestry populations have normal ibuprofen metabolism, and therefore are less susceptible to complications for ibuprofen-based treatment.

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 these two NSAIDs[30]. Following this analysis, the European Committee in charge of risk assessment and pharmacovigilance (PRAC) concluded in April 2020 that taking ibuprofen or ketoprofen (oral, rectal or injectable) can lead, in certain infections, to mask symptoms such as fever or pain, leading to a delay in the management of the patient with the consequence of a risk of complications of the infection. The PRAC also concluded that this risk has been observed for bacterial infections in the context of chickenpox and pneumonia[30]. Furthermore, one large case-control study found a clear

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[31].

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 leads to more efficient, safer, and cost-effective treatments.

ACKNOWLEDGEMENTS

Microarray wet- and dry-lab procedures were all performed by the Princess Haya Biotechnology Centre (PHBC) at Jordan University of Science and Technology (JUST). This work used data from an approved project by JUST (project 67/2/2013).

CONFLICT OF INTEREST

The author declares that there is no conflict of interest.

FUNDING

No funding was available for this study.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, *et al*. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; **382:** 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:** 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:** e21.
5. Gray JA. The shift to personalised and population medicine. *Lancet* 2013; **382:** 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:** 604-610.
7. Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, *et al*. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* 2003; **348:** 1170-1175.
8. Gemmati D, Tisato V. Genetic Hypothesis and Pharmacogenetics Side of Renin-Angiotensin-System in COVID-19. *Genes (Basel)* 2020; **11**: 1044.
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:** 263-273.

10. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF *et al*. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 2012; **92:** 414-417.

11. Theken KN, Lee CR, Gong L, Caudle KE, Formea CM, Gaedigk A, *et al*. Clinical Pharmacogenetics Implementation Consortium Guideline (CPIC) for CYP2C9 and Nonsteroidal Anti-Inflammatory Drugs. *Clin Pharmacol Ther* 2020; **108:** 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:** 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:** 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:** 62-75.

15. López-Rodríguez R, Novalbos J, Gallego-Sandín S, Román-Martínez M, Torrado J, Gisbert JP, *et al*. 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:** 77-84.

16. Ochoa D, Prieto-Pérez R, Román M, Talegón M, Rivas A, Galicia I, *et al*. Effect of gender and CYP2C9 and CYP2C8 polymorphisms on the pharmacokinetics of ibuprofen enantiomers. *Pharmacogenomics* 2015; **16:** 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:** 341-355.

18. Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 2006; **22:** 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:** 89-103.

20. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al*. A global reference for human genetic variation. *Nature* 2015; **526:** 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:** 3555-3557.

22. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res* 2002; **12:** 656-664.

23. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2013. https://www.r-project.org

24. Huddart R, Fohner AE, Whirl-Carrillo M, Wojcik GL, Gignoux CR, Popejoy AB, *et al*. Standardized Biogeographic Grouping System for Annotating Populations in Pharmacogenetic Research. *Clin Pharmacol Ther* 2019; **105:** 1256-1262.

25. Gaedigk A, Ingelman-Sundberg M, Miller NA, Leeder JS, Whirl-Carrillo M, Klein TE, *et al*. The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clin Pharmacol Ther* 2018; **103:** 399-401.

26. Haber M, Gauguier D, Youhanna S, Patterson N, Moorjani P, Botigué LR, *et al*. Genome-wide diversity in the levant reveals recent structuring by culture. *PLoS Genet* 2013; **9:** e1003316.

27. Skourtanioti E, Erdal YS, Frangipane M, Balossi Restelli F, Yener KA, Pinnock F, *et al*. Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell* 2020; **181:** 1158-1175.

28. Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, *et al*. Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 2001; **62:** 871-884.

29. 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:** 435-441.

30. ANSM. Nonsteroidal anti-inflammatory drugs (NSAIDs) and serious infectious complications - Information point updated on 20/05/2020. 2020. https://www.ansm.sante.fr/S-informer/Points-d-information-Points-d-information/Anti-inflammatoires-non-steroidiens-AINS-et-complications-infectieuses-graves-Point-d-Information-actualise-le-20-05-2020.

31. Basille D, Thomsen RW, Madsen M, Duhaut P, Andrejak C, Jounieaux V, *et al*. Nonsteroidal Antiinflammatory Drug Use and Clinical Outcomes of Community-acquired Pneumonia. *Am J Respir Crit Care Med* 2018; **198:** 128-131.

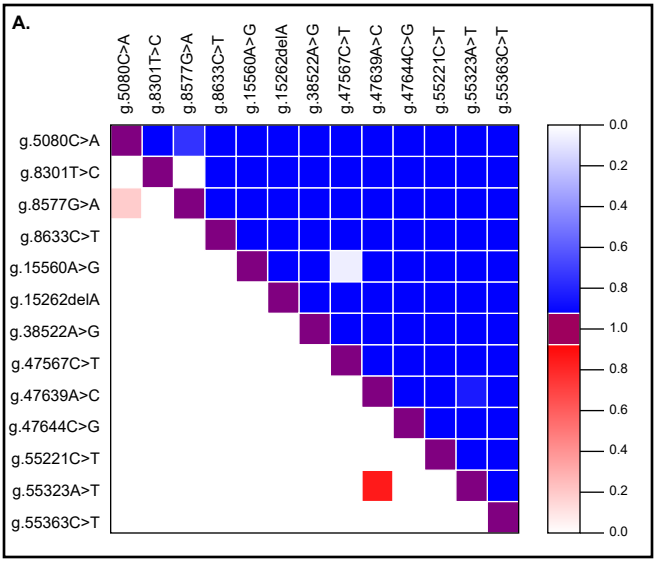
FIGURE LEGENDS

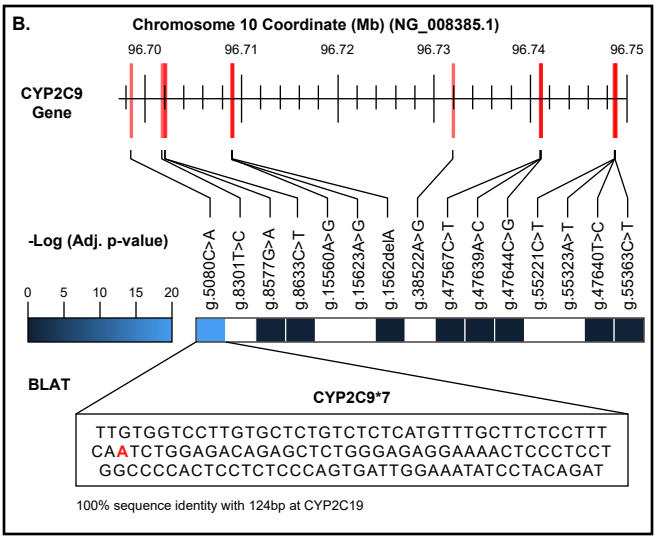
Figure 1: CYP2C9 genetic interpretation in the context of Linkage Disequilibrium (LD) and high homology sequence. **A.** LD (D' and r^2) between 13 single nucleotide polymorphisms (SNPs) in the CYP2C9 gene. **B.** Statistical comparison of allele frequencies on experimental data and reference populations. Color intensity indicates a more significant difference between the given experimental data and the population.

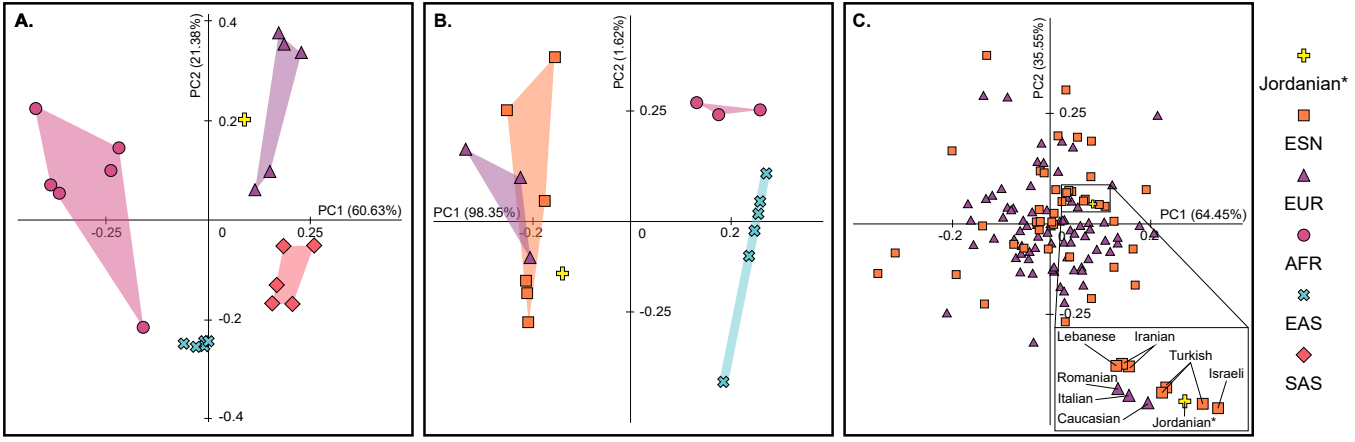
Figure 2: Population structure analysis of the CYP2C9 *1, *2 and *3 alleles. **A.** MDS plot of the 14 variants shared between the Jordanian Arab population and 22 global populations from the 1000 Genomes Project Phase III (1kG-p3) dataset. **B.** MDS plot of the Jordanian Arab population and 18 public worldwide populations. **C.** MDS plot of the Jordanian Arab population and 118 populations from the Near Eastern (ESN) and European (EUR) groups.

Figure 3: Biogeographical groups and pharmacogenetic analyses for CYP2C9 and ibuprofen associations. **A.** Frequencies of 13 actionable pharmacogenomics (PGx) alleles and phenotypes assessed cumulatively for the Jordanian Arab population against the nine biogeographical groups. **B.** Mapping of the total allele and phenotype frequency of the CYP2C9 alleles *2 and *3.

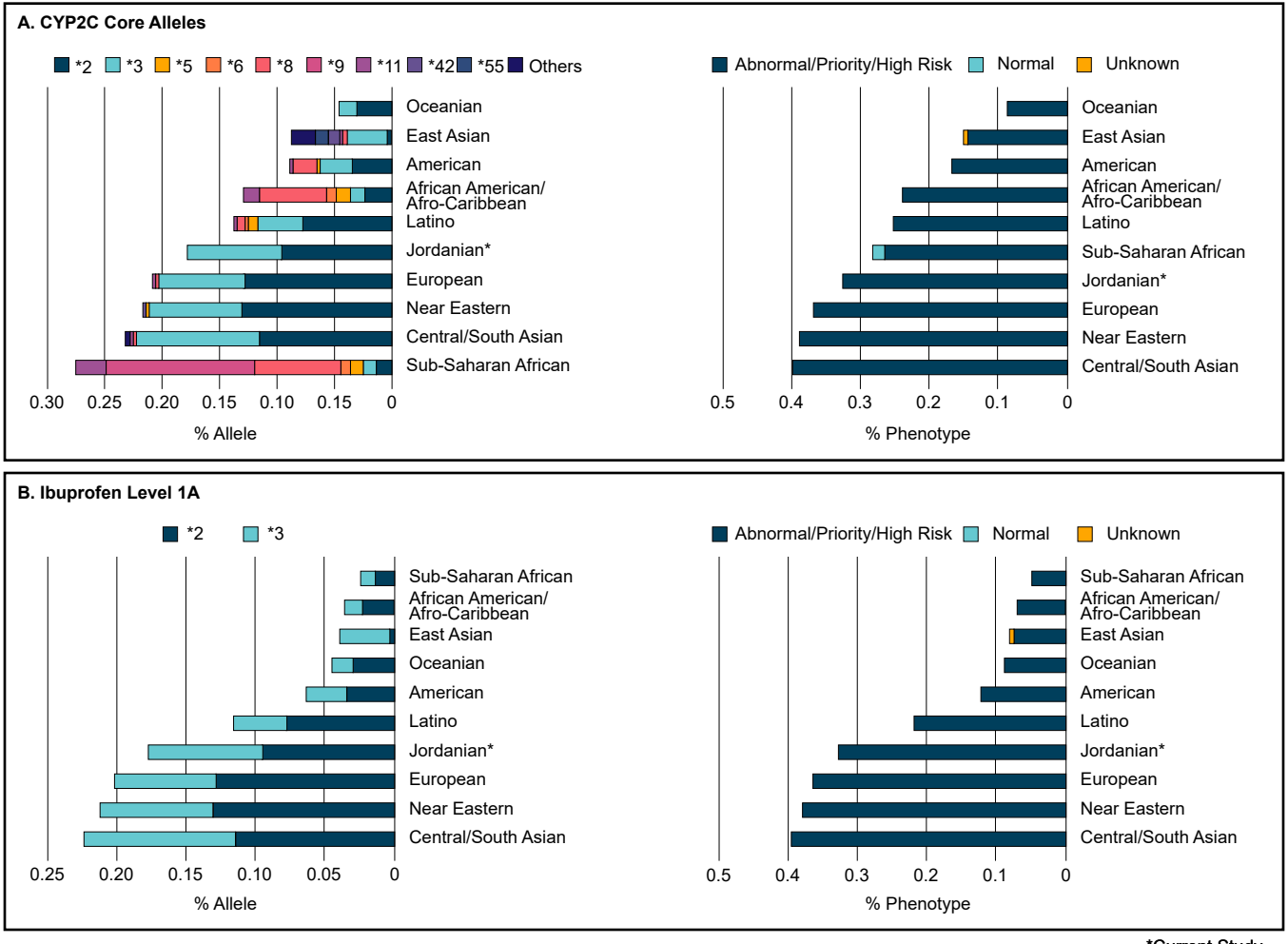
Figure 4: Global frequency map of variability in of CYP2C9 level A alleles and impact on ibuprofen response. Phenotype prevalence shown for global populations of CYP2C9*2 and *3. The two admixed groups of African American/Afro-Caribbean and Latino were not represented on this figure as the geographical grouping based on the location of genetic ancestors pre-diaspora and pre-colonization could not be applied to the two admixed groups.


A.
g.5080C>A
g.8301T>C
g.8577G>A
g.8633C>T
g.15560A>G
g.15262delA
g.38522A>G
g.47567C>T
g.47639A>C
g.47644C>G
g.55221C>T
g.55323A>T
g.55363C>T
0.0
0.2
0.4
0.6
0.8
1.0
0.8
0.6
0.4
0.2
0.0


B.
Chromosome 10 Coordinate (Mb) (NG_008385.1)
96.70
96.71
96.72
96.73
96.74
96.75
CYP2C9 Gene
-Log (Adj. p-value)
0
5
10
15
20
g.5080C>A
g.8301T>C
g.8577G>A
g.8633C>T
g.15560A>G
g.15623A>G
g.1562delA
g.38522A>G
g.47567C>T
g.47639A>C
g.47644C>G
g.55221C>T
g.55323A>T
g.47640T>C
g.55363C>T
BLAT
CYP2C9*7
TTGTGGTCCTTGTGCTCTGTCTCTCATGTTTGCTTCTCCTTT
CAATCTGGAGACAGAGCTCTGGGAGAGGAAAACTCCCTCCT
GGCCCCACTCCTCTCCCAGTGATTGGAAATATCCTACAGAT
100% sequence identity with 124bp at CYP2C19



*Current Study



*Current Study

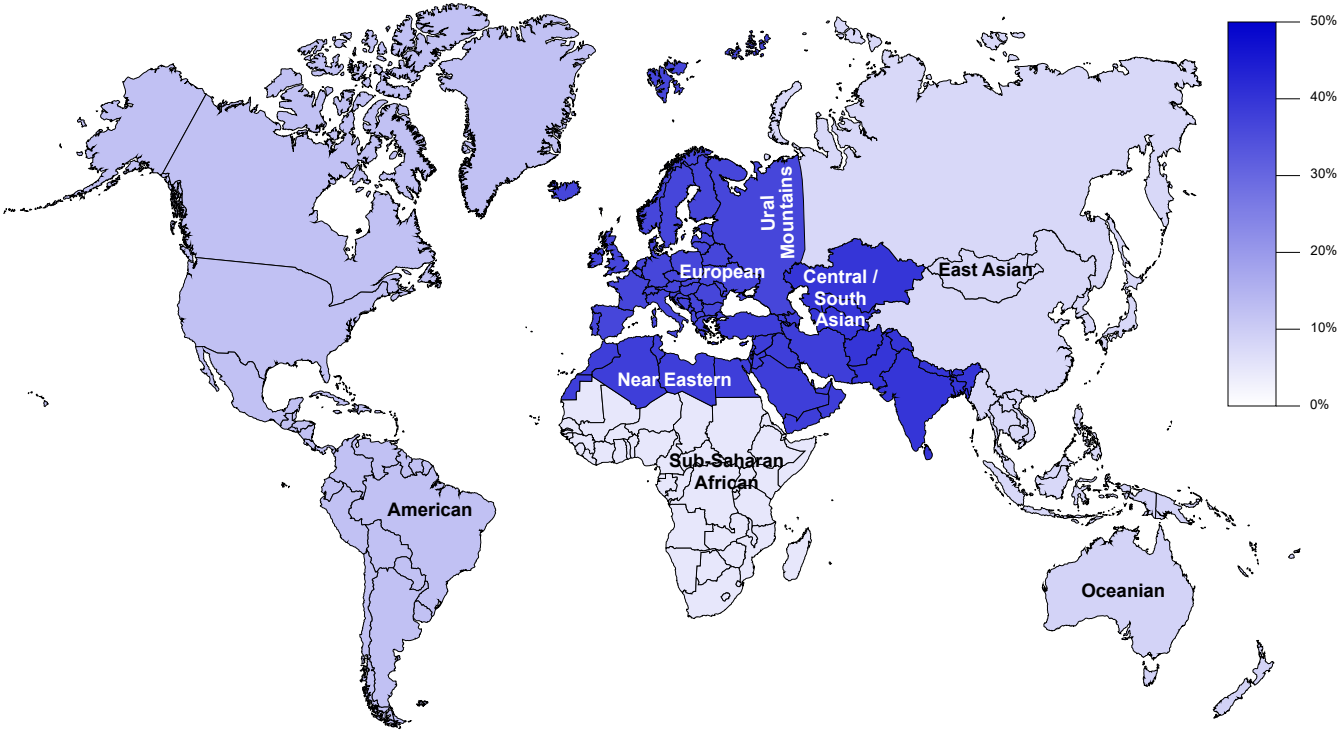

Ural Mountains
European
Central / South Asian
East Asian
Near Eastern
Sub-Saharan African
American
Oceanian
50%
40%
30%
20%
10%
0%

Table 1. Distribution of CYP2C9 alleles and genotypes in Jordanian Arabs (NG_008385.1; NP_000762.2).

Allele	Variant	Substitution	Allele Count	Allelic Frequency	Genotype	Genotypic Frequency	Observed Genotypes	Expected Genotypes	χ^2	*p*-value
*2	g.8633C>T; p.(Arg144Cys)	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	0	0.89		
*3	g.47639A>C; p.(Ile359Leu)	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	0	0		