Molecular Medicine Communications

DOI: 10.12345/mmc.001.01.0016

Research Article
Polymorphisms in CYP2C8 Gene in Pakistani population and their frequencies in various ethnic groups

Rosana Christine¹, Cavalcanti Ximenes¹, Bashir Ahmad², Vahid Nikoui³, Khuseyn Egamnazarov⁴, Muhammad Imran Khan⁵

¹Neuropsychiatry and Behavioral Sciences Graduate Program, Federal University of Pernambuco, Recife, Brazil
²Department of Physiology, Bannu Medical College, Bannu, Pakistan
³Razi Drug Research Center, Iran University of Medical Sciences, Tehran, Iran
⁴Department of environmental health, State Educational Institution “Avicenna Tajik State Medical University”, Dushanbe, the Republic of Tajikistan.
⁵Department of Biomedical Sciences, Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, PKP, Pakistan

*Correspondence: mik9892@gmail.com

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Abstract
Cytochrome P4502C8 represents 7% of the hepatic cytochrome system and metabolizes around 5% of drugs in phase I processes. It also plays a significant role in the metabolism of endogenous compounds. More than 20 single nucleotide polymorphisms (SNPs) have been reported, mainly in exon 3, 5, and 8. Some of the SNP’s lead to decreased enzyme activity and may have impact on drug metabolism. This research study aims to determine the frequencies of the most common SNPs of the CYP2C8 gene (CYP2C8*2, *3, *4) in the Pakistani population. A cross-sectional study consisting of 391 healthy humans was conducted after taking informed consent. DNA was extracted using commercial kits, and allelic and genotype frequencies were determined after PCR amplification, restriction fragment length polymorphism, and gel electrophoresis. The rate of minor allele was found to be 11.64% for CYP2C8*2, 14.71% for CYP2C8*3, and 1.53% for CYP2C8*4. Comparison with the 1000 Genome project reveals that the allelic frequencies of CYP2C8*4 in Pakistani population were similar to other South Asian populations while the frequencies of CYP2C8*2 and CYP2C8*3 as significantly different from other South Asian populations. A significant interethnic variation was also observed among Pakistani ethnicities. The CYP2C8*2 allele was highest in the Sindhi population, CYP2C8*3 in the Pashtoon population and CYP2C8*4 in the Balochi population. These data suggest that the frequency of poor metabolizers of CYP2C8 is high enough in the Pakistani population to warrant further genotype-phenotype correlations studies on individual drugs metabolized by CYP2C8 enzyme.

Keywords: CYP2C8, poor metabolizer, non-responder phenomenon, Pakistan, Genetic polymorphism

Introduction
A significant number of drugs, xenobiotics, and endogenous compounds are metabolized by the cytochrome P450 (CYP) enzyme superfamily (Zanger and Schwab 2013). The enzyme originated about three billion years ago, are found in a large variety of members of the animal kingdom (Danielson 2002). In human beings, the CYP2C subfamily represents approximately 18-30% of the total cytochrome enzymes (Goldstein 2001). Taking up about 6-7% of the entire hepatic cytochrome enzyme content (Achour, Barber, and Rostami-Hodjegan 2014; Inoue et al. 2006; Rostami-Hodjegan and Tucker 2007), the CYP2C8 enzyme is a crucial member of this subfamily and is also responsible for metabolizing over a 100 clinical drugs currently used (Minhas et al. 2013). Moreover, different ethnic groups have shown a distinct spectrum of single nucleotide polymorphisms (SNPs) in the CYP2C8 gene. These variations have been studied in populations...
ranging from the Caucasians to the Chinese to the Africans (Pechandova et al. 2012). Over the past 15-20 years, the drug-drug interactions and pharmacogenetic polymorphisms in the CYP2C8 gene have been given considerable attention. At the start of the millennium, the rhabdomyolysis and deaths caused by the pharmacokinetic drug-drug interactions between the 3-hydroxy, 3-methyl-glutaryl Co-enzyme A reductase inhibitor cerivastatin (a substrate for CYP2C8) and the fibric acid derivative gemfibrozil brought a considerable amount of attention and broader insight towards CYP2C8 enzyme’s activity in drug metabolism (Backman et al. 2002; Staffa, Chang, and Green 2002). This ultimately led to CYP2C8 being recognized as a major drug-metabolizing enzyme by the drug regulatory authorities.

The three most frequently studied CYP2C8 gene variants, CYP2C8*2 (rs11572103 A>T), CYP2C8*3 (rs10509681 T>C), and CYP2C8*4 (rs1058930 G>C), located in the coding region of the enzyme, are responsible for the variation in enzyme activity of the CYP2C8 in humans and their frequencies differ in both intercontinental and intracontinental populations (Daily and Aquilante 2009). The CYP2C8*2 allele is more commonly found in the African population and is responsible for the substitution of Ile269Phe in exon 5, whereas the CYP2C8*3 allele is responsible for the substitution of Arg139Lys in exon 3. During in vitro studies, the defects in the metabolism of a very commonly used anticancer drug paclitaxel were found to be related to the CYP2C8*2 and CYP2C8*3 variants of the CYP2C8 gene (Dai et al. 2001). The CYP2C8*4 allele results in the substitution of Ile264Met in exon 5, and the enzyme formed afterward has approximately 25% activity compared to the wild type enzyme (Daily and Aquilante 2009).

Prior knowledge of the frequencies of mutant alleles found in a population may help revise the prescribing practices of the clinicians, including dose adjustments and alternate medications. However, no such data of CYP2C8 genetic variants are available for the Pakistani population. Therefore, the present study was designed to investigate the frequencies of CYP2C8*2, *3, and *4 alleles in the Pakistani population to find the prevalence of these genetic variants. We specifically investigated the samples of various ethnic populations from Pakistan to examine the frequencies of CYP2C8*2, *3, and *4, and compared them with previous findings in other populations.

Materials and Methods

Study approval, samples, and DNA extraction

This study was approved by the Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Pakistan. Written informed consent forms were obtained from all participating individuals. The study cohort consisted of healthy volunteers from six major ethnicities of Pakistan, including Punjabi, Pashtoon, Sindhi, Balochi, Seraiki, and Urdu Speaking. Ethnicity was self-reported. Five milliliters of venous blood drawn into sterile tubes containing EDTA as an anti-coagulant was stored at 4°C. Genomic DNA was isolated using Gene Jet Genomic DNA extraction Kit (ThermoScientific) and was quantified using 1% agarose gel electrophoresis. Isolated genomic DNA was stored at -20°C until further processing.

Genotyping

CYP2C8*2, CYP2C8*3, and CYP2C8*4 were genotyped using polymerase chain reaction (PCR) followed by restriction fragment length polymorphism (RFLP) with slight modifications in methods previously described (Dai et al. 2001; Nakajima et al. 2003). The products were separated on agarose or polyacrylamide gel and visualized under UV light. PCR for all three SNPs was performed separately in a total reaction volume of 25µl containing 12.5µl of 2X Dream Taq Mastermix (ThermoScientific), 0.18mM of both forward and reverse primers, 7.7µl of sterile PCR water, and 3 µl of template DNA (20-50 ng/µl). Thermal profile was as follows: initial denaturation at 95°C for 2 minutes followed by 35 cycles with denaturation at 95°C for 30 seconds, 30 seconds of primer annealing at 54°C for rs11572103, 58°C for rs10509681 and 50°C for rs1058930, initial extension at 72 °C for 1 minute and a final extension at 72 °C for 5 minutes. Enzymes used for restriction digestion were BclI for CYP2C8*2, BseRI for CYP2C8*3, and TaqI for CYP2C8*4. For visualization, 10µl of PCR product was directly loaded onto agarose/polyacrylamide gel. About one-quarter of the total samples were sequenced to confirm our genotyping results.
Figure 1: Geographical representation of the ethnicities investigated in the study and their respective CYP2C8*2, *3, and *4 allelic frequencies.

Statistical analysis
Data were compiled according to the genotype and allele frequencies estimated from the observed numbers of each specific allele. The frequency of each allele and genotype in our samples is given together with the 95% confidence interval. Data was recorded and analyzed using Statistical Package for Social Sciences (SPSS) version 26. Expected frequency percentages of genotypes are given as per the Hardy-Weinberg equilibrium. Qualitative variables were presented as frequencies and percentages. Chi-square test was applied to compare qualitative variables.
Table 1: CYP2C8*2 allele frequencies in the Pakistani population.

| Ethnicity | n  | % CYP2C8*1 (CI)         | % CYP2C8*2 (CI)         |
|-----------|----|-------------------------|-------------------------|
| Pakistan  | 391| 88.36 (86.1-90.6)       | 11.64 (9.4-13.9)        |
| Punjabi   | 100| 92.50 (88.8-96.2)       | 7.50 (6.9-8.1)          |
| Pashtoon  | 80 | 91.88 (87.6-96.1)       | 8.13 (3.9-12.4)         |
| Sindhi    | 50 | 77.00 (68.8-85.2)       | 23.00 (14.8-31.2)       |
| Seraiki   | 54 | 87.04 (80.7-93.4)       | 12.96 (6.6-19.3)        |
| Balochi   | 58 | 83.62 (76.9-90.4)       | 16.38 (9.6-23.1)        |
| Urdu      | 49 | 92.86 (87.8-98.0)       | 7.14 (62.5-80.3)        |

Figure 2: CYP2C8*2 genotype frequencies in various ethnic groups of Pakistan.

Results

Figure 1 shows geographical representation of the ethnicities investigated in the study and their respective CYP2C8*2, *3, and *4 allelic frequencies. The mean age of the participants of this study was 26±6 years with 215 males (55%) and 176 (45%) females. Frequencies of the CYP2C8*2 allele in the Pakistani population are shown in table 1. The frequency of the major allele was 88.36%, and of minor allele was 11.64%. The major allele was found slightly less prevalent in the Sindhi population at 77% compared to Punjabi, Pashtoon, Baloch, Seraiki, and Urdu populations in which the prevalence of major allele was slightly higher (Table 1). The frequency of wild type genotype (*1*1) was 79.5%, *1*2 was 18.7%, and *2*2 was 1.8% in the Pakistan population. Punjabi, Pashtoon, and Urdu populations showed a slightly higher frequency of wild type genotype at 86%, 85%, and 87.8%, respectively. In contrast, Sindhi, Seraiki, and Baloch populations had a slightly lower prevalence of wild-type genotype. The Sindhi population showed the highest frequency of *1*2 genotype at 34% and of *2*2 genotype at 6% (Figure 2 and supplementary table 1).
Table 2: CYP2C8*3 allele frequencies in the Pakistani population.

| Ethnicity | no  | % CYP2C8*1 (CI)       | % CYP2C8*3 (CI)       |
|-----------|-----|----------------------|----------------------|
| Pakistan  | 391 | 85.29 (82.8-87.8)    | 14.71 (12.2-17.2)    |
| Punjabi   | 100 | 86.00 (81.2-90.8)    | 14.00 (9.2-18.8)     |
| Pashtoon  | 80  | 81.85 (75.8-87.8)    | 18.15 (12.1-24.1)    |
| Sindhi    | 50  | 84.00 (76.8-91.2)    | 16.00 (8.8-23.2)     |
| Balochi   | 54  | 87.04 (80.7-93.4)    | 12.96 (6.6-19.3)     |
| Seraiki   | 58  | 83.60 (76.9-90.3)    | 16.40 (9.7-23.1)     |
| Urdu      | 49  | 90.85 (83.1-96.5)    | 9.15 (3.4-14.8)      |

Figure 3: CYP2C8*3 genotype frequencies in various ethnic groups of Pakistan.

Frequencies of CYP2C8*3 alleles in the Pakistani population are shown in table 2. The frequency of minor alleles for this polymorphism was higher in the Pakistani population than the CYP2C8*2 allele and was found to be 14.71% (table 2). In Pashtoon, Sindhi, and Seraiki populations, the minor allele was found slightly more prevalent at 18.1%, 16%, and 16.4%, respectively. In the Urdu population, the frequency of minor allele was found to be the lowest at 9.1%. The frequency of wild type genotype (*1*1) was 73.4%, *1*3 was 23.8%, and *3*3 was 2.8% in the Pakistan population. In the Urdu population, wild type genotype (*1*1) was found to be highest at 44%. Urdu population showed the lowest frequency of wild type genotype at 83.7%. Sindhi, Pashtoon, and Punjabi population showed the highest frequencies of heterozygous genotype (*1*3) at 28%, 26.3%, and 26%, respectively. The highest prevalence of homozygous mutant genotype (*3*3) was found in the Pashtoon population at 5%. All other ethnic groups also showed *3*3 genotype, albeit at varying frequencies. (Figure 3 and supplementary table 2).
Table 3: CYP2C8*4 allele frequencies in the Pakistani population.

| Ethnicity | no  | % CYP2C8*1 (CI)   | % CYP2C8*4 (CI)   |
|-----------|-----|-------------------|-------------------|
| Pakistan  | 391 | 98.46 (97.24-99.68) | 1.53 (0.31-2.75)  |
| Punjabi   | 100 | 97.50 (94.44-100)  | 2.50 (0-5.56)     |
| Pashtoon  | 80  | 97.50 (94.08-100)  | 2.50 (0-5.92)     |
| Sindhi    | 50  | 100               |                   |
| Balochi   | 54  | 97.22 (92.84-100)  | 2.77 (0-7.15)     |
| Seraiki   | 58  | 100               |                   |
| Urdu      | 49  | 100               |                   |

Figure 4: CYP2C8*4 genotype frequencies in various ethnic groups of Pakistan.

Frequencies of the CYP2C8*4 allele in the Pakistani population are shown in table 3. The frequency of the major allele was 98.46%, and of minor allele was 1.53%. No minor allele was found in Sindhi, Seraiki, and Urdu populations. At the same time, it was present in the Punjabi population at 2.5%, in the Baloch population at 2.77%, and in the Pashtoon population at 2.5%. The frequency of wild type genotype (*1*1) was 96.93%, heterozygous genotype (*1*4) was 3.06%, while no *4*4 genotype was observed in the Pakistan population. Urdu, Seraiki, and Sindhi populations showed 100 wild type genotypes. Heterozygous genotype was observed at 5% in the Punjabi population, at 5.55% in the Baloch population, and 5% in the Pashtoon population. As mentioned earlier, the homozygous mutant (*4*4) was not observed in any ethnicity (Figure 4 and supplementary table 3).
Table 4: Comparison of CYP2C8*2 allele in Pakistani population with various worldwide populations reported in 1000 Genome Project.

| Population | CYP2C8*1 | CYP2C8*2 | Chi-square statistic | p-value  |
|------------|----------|----------|---------------------|----------|
| PAK        | 691      | 91       | 19.52               | P<0.00001|
| PEL        | 169      | 1        | 12.60               | P<0.000385|
| PUR        | 201      | 7        | 21.02               | 0.00001  |
| GBR        | 181      | 1        | 20.59               | 0.00001  |
| IBS        | 211      | 3        | 24.01               | 0.00001  |
| GIH        | 205      | 1        | 21.50               | 0.00001  |
| ITU        | 202      | 2        | 15.98               | 0.00006  |
| PJL        | 188      | 4        | 15.53               | 0.00008  |
| STU        | 199      | 5        | 15.53               | 0.00008  |

CLM-Colombian in Medellin, Colombia, MXL-Mexican ancestry in Los Angeles, California, PEL-Peruvian in Lima, Peru, PUR-Puerto Rican in Puerto Rico, CDX-Chinese Dai in Xishuangbanna, China CHB-Han Chinese in Beijing, China, CHS-Southern Han Chinese, China, JPT-Japanese in Tokyo, Japan, KHV-Kinh in Ho Chi Minh City, Vietnam, CEU-Utah residents with northern and western European ancestry, FIN-Finnish in Finland, GBR-British in England and Scotland, IBS-Iberian population in Spain, TSI-Toscani in Italy, BEB-Bengali in Bangladesh, GIH-Gujrati Indian in Houston, TX, IBS-Indian Telugu in UK, PJL-Punjabi in Lahore, Pakistan, STU-Sri Lankan Tamil in the UK.

Tables 4, 5, and 6 show the results of comparisons between allelic frequencies of CYP2C8*2, *3 and *4 and different world-wide populations as reported in the 1000 Genome Project. These results indicate that allelic frequencies of CYP2C8*2 in our study are different than various South American and South Asian populations (table 4). Allelic frequencies of CYP2C8*3 were statistically insignificant in comparison with European populations but different than South Asian populations (table 5). The frequency of CYP2C8*4 in our study was found similar to South Asian ans South American populations but different from European populations (table 6).

Table 5: Comparison of CYP2C8*3 allele in Pakistani population with various worldwide populations reported in 1000 Genome Project.

| Population | CYP2C8*1 | CYP2C8*3 | Chi-square statistic | p-value  |
|------------|----------|----------|---------------------|----------|
| PAK        | 667      | 115      | 1.12                | 0.28831  |
| CLM        | 166      | 22       | 1.88                | 0.16992  |
| MXL        | 115      | 13       | 19.48               | 0.00001  |
| PEL        | 166      | 4        | 32.46               | 0.00001  |
| PUR        | 178      | 30       | 0.01                | 0.91834  |
| CHS        | 209      | 1        | 32.46               | 0.00001  |
| CEU        | 172      | 26       | 5.98                | 0.01440  |
| FIN        | 182      | 16       | 3.59                | 0.05791  |
| GBR        | 165      | 17       | 3.59                | 0.05791  |
| IBS        | 182      | 32       | 0.0082              | 0.92796  |
| TSI        | 186      | 28       | 0.35                | 0.54884  |
| BEB        | 169      | 3        | 21.85               | 0.00001  |
| GIH        | 198      | 8        | 17.52               | 0.00002  |
| ITU        | 199      | 5        | 22.73               | 0.00001  |
| PJL        | 183      | 9        | 13.92               | 0.00019  |
| STU        | 200      | 4        | 28.27               | 0.00001  |
Discussion
According to its Statistics Bureau, Pakistan, with an estimated population of over 220 million, is the sixth most populous country in the world (“Block Wise Provisional Summary Results of 6th Population & Housing Census-2017 [As on January 03, 2018] | Pakistan Bureau of Statistics” n.d.). The country has a young, multi-ethnic, and multi-cultural society. Despite being home to a vast population, pharmacogenetic studies on how its population responds to various pharmaceutical drugs are rare. The largest ethnic group in Pakistan is Punjabis, which make up about 38.78% of the population, followed by Pashtuns (18.24%), Sindhis (14.57%), Seraikis (10.53%), Urdu speaking (7.57%), and Baloch (3.57%) (Taus-Bolstad 2003). These ethnic groups represent about 94% of the Pakistani population. Genetic variations in CYP genes affecting the metabolism of xenobiotics and drug response have not been investigated in these ethnic groups. Our study partly addresses this issue by reporting frequencies of three of the most important single nucleotide polymorphisms in the CYP2C8 gene (figure 1).

The frequency of CYP2C8*2 minor alleles in the Pakistani population was similar to the one found in the Zanzibar population (Cavaco et al. 2005). The lowest frequency of the minor allele has previously been reported from various European populations. Literature search shows that in the African-American population, minor allele is reported at the highest frequency and is around double the frequency observed in the Pakistani population. No CYP2C8*2 allele is reported from the Japanese population (Nakajima et al. 2003). Allelic frequency of CYP2C8*2, as reported in the 1000 Genome Project, was significantly lower in Peruvian (PEL), Puerto Rican (PUR), British (GBR), and Iberian (IBS) populations. This allele was absent in the Chinese and East Asian populations and was rare in European populations. In the regional populations, this allele was found in lower frequencies in Gujrati Indians (GIH), Indian Telegu (ITU), Sri Lankan Tamils (STU), and even in Pakistanis from Lahore (PJL), while it was absent in the Bengali population (BEB) (Table 4).

Table 6: Comparison of CYP2C8*4 allele in Pakistani population with various worldwide populations reported in 1000 Genome Project.

| Population | CYP2C8*1 | CYP2C8*4 | Chi-square statistic | p-value   |
|------------|----------|----------|----------------------|-----------|
| PAK        | 770      | 12       | 0.0037               | 0.95129   |
| CLM        | 185      | 3        | 0.44                 | 0.50505   |
| MXL        | 125      | 3        | 0.73                 | 0.39104   |
| PEL        | 168      | 2        | 0.11                 | 0.72520   |
| PUR        | 203      | 5        | 3.32                 | 0.06815   |
| CEU        | 191      | 7        | 16.57                | 0.00001   |
| FIN        | 182      | 16       | 12.89                | 0.00033   |
| GBR        | 171      | 11       | 24.39                | 0.00004   |
| IBS        | 200      | 14       | 0.37                 | 0.54255   |
| TSI        | 204      | 10       | 0.35                 | 0.55133   |
| BEB        | 171      | 2        | 0.26                 | 0.60717   |
| GIH        | 204      | 2        | 0.95                 | 0.32897   |
| ITU        | 202      | 2        | 0.37                 | 0.54255   |
| PJL        | 190      | 2        | 0.37                 | 0.54255   |

The frequency of CYP2C8*3 minor alleles in the Pakistani population was found in a similar range to West Germany (Weise et al. 2004). The highest CYP2C8*3 allelic frequency is reported in the Portuguese population at 19.8% (Cavaco et al. 2005). Studies from the Czech Republic and North East England have also reported CYP2C8*3 frequencies in double digits (Pechandova et al. 2012) and (Bahadur et al. 2002). Malaysian, Indian, and African-American populations reported low single-digit frequencies of this polymorphism (Dai et al. 2001). In contrast, studies from Japan and Ghana did not report any CYP2C8*3 in their population (Nakajima et al. 2003). Comparison of our results with the 1000 Genome Project revealed that...
among South American populations, the frequency of this allele in Colombian (CLM), Mexican (MXL), and PUR populations was similar to the Pakistani population observed in this study while the Peruvian population displayed significantly lower prevalence of this allele (Table 5). This allele is not present in many of the East Asian and Chinese populations except in Southern Chinese Hans (CHS), which showed a significantly lower frequency of this allele compared to the Pakistani population (Table 5). Among the European populations, the prevalence of this allele was similar to the Pakistani population except in the Finnish population, which showed a significantly lower prevalence. Among the South Asian populations, all the populations investigated, including BEB, GIH, ITU, PJL, and STU, showed a significantly lower frequency of this allele. This analysis indicates that the distribution pattern of CYP2C8*3 allele in Pakistani population resembles more closely with the European populations than with South Asian populations.

In the Pakistani population, the frequency of the CYP2C8*4 minor allele was low. A study from the British population reported the prevalence of CYP2C8*4 at 7.5% (Bahadur et al. 2002) while the same variant was reported at a frequency of 4% from the North Indian population (Minhas et al. 2013). However, this gene variant was absent in Malaysian Indian, Ghanian, and Japanese populations (Kudzi, Dodoo, and Mills 2009; Muthiah et al. 2005; Nakajima et al. 2003). Comparing our results with the 1000 Genome project showed that the frequency of this allele in South American populations was not significantly different from the Pakistani population reported in the current study. However, this allele was absent from the Chinese and East Asian populations (Table 6). In the European populations, this allele was found significantly higher than Pakistani populations, while no significant differences were observed between Pakistani and various South Asian populations except for STU where the CYP2C8*4 allele was not found.

The Pakistani population is a heterogeneous mixture of Asian, Middle Eastern, and European populations partly because of the Arab invasion of the 8th century and British invasions of the 18th and 19th centuries, and partly owing to its high geographic and ethnic diversity (Bhatti et al. 2017). The genetic structures of various Pakistani populations have been analyzed and several distinct variants identified among different ethnicities by global projects such as the 1000 Genome Project and Human Genome Diversity Project (Auton et al. 2015; Bergström et al. 2020). Some studies indicate that the genetic structure of these ethnicities is closely related to both South Indian and European populations (Metspalu et al. 2011), while others suggest Pakistani ethnicities to be similar to European populations (Ayub et al. 2003; Mansoor et al. 2004). The differences observed with other world-wise populations as well as within the various ethnic populations of Pakistan may be due to their diverse ancestry belonging to Aryan, Arab, Persian, Turkish, Kurdish, Dravidian, Sewais, and black African lineages (Ahmed and Khan, n.d.).

To our knowledge, this is the first study to report frequencies of CYP2C8 gene polymorphisms in various ethnicities of the Pakistani population. Genetic information about patients’ CYP2C8 gene is likely to help physicians prescribe to patients the most suitable and safest drug based on their genetic make-up. With roughly 5% clinically available drugs metabolized by the CYP2C8 enzyme (Zanger and Schwab 2013) and a non-trivial fraction of the Pakistani population having at least one minor allele, the number of patients affected by these genetic variations is significant. We propose carrying out further studies with individual drugs metabolized by CYP2C8 to shed more light on genotype-phenotype relations.

Conflict of interest
The authors declare that they have no competing interests.

Funding
The authors did not receive any specific funding for this project.

Ethics approval
Yes. The study was approved by the Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Pakistan.

Consent forms
Yes. Consent forms were obtained from the participating patients.

Authors contribution
MIK, and RCCX conceptualized the study, MIK, and BA, collected samples, BA, VN, and KE performed experiments and data analysis, MIK and RCCX wrote the final draft. All the authors have read and approved the final version of the manuscript.
Acknowledgements
Not applicable.

References
Achour, Brahim, Jill Barber, and Amin Rostami-Hodjegan. 2014. “Expression of Hepatic Drug-Metabolizing Cytochrome P450 Enzymes and Their Intercorrelations: A Meta-Analysis.” Drug Metabolism and Disposition: The Biological Fate of Chemicals 42 (8): 1349–56. https://doi.org/10.1124/dmd.114.058834.

Ahmed, Manzoor, and Gulawar Khan. n.d. “The History of Baloch and Balochistan: A Critical Appraisal,” 14.

Auton, Adam, Gonçalo R. Abecasis, David M. Altshuler, Richard M. Durbin, Gonçalo R. Abecasis, David R. Bentley, Aravinda Chakravarti, et al. 2015. “A Global Reference for Human Genetic Variation.” Nature 526 (7571): 68–74. https://doi.org/10.1038/nature15393.

Ayub, Qasim, Atika Mansoor, Muhammad Ismail, Shagufta Khaliq, Aisha Mohyuddin, Abdul Hameed, Kehkashan Mazhar, et al. 2003. “Reconstruction of Human Evolutionary Tree Using Polymorphic Autosomal Microsatellites.” American Journal of Physical Anthropology 122 (3): 259–68. https://doi.org/10.1002/ajpa.10234.

Backman, Janne T., Carl Kyrkklund, Mikko Neuvonen, and Perni J. Neuvonen. 2002. “Gemfibrozil Greatly Increases Plasma Concentrations of Cerivastatin.” Clinical Pharmacology and Therapeutics 72 (6): 685–91. https://doi.org/10.1006/cpth.2002.128469.

Bahadur, Namrata, Julian B. S. Leathart, Elaine Mutch, Dorothy Steimel-Crespi, Stuart A. Dunn, Ron Gilissen, Jos Van Houdt, et al. 2002. “CYP2C8 Polymorphisms in Caucasians and Their Relationship with Paclitaxel 6alpha-Hydroxylase Activity in Human Liver Microsomes.” Biochemical Pharmacology 64 (11): 1579–89. https://doi.org/10.1016/s0006-2952(02)01354-0.

Bergström, Anders, Shane A. McCarthy, Ruoyun Hui, Mohamed A. Almarri, Qasim Ayub, Petr Daneeek, Yuan Chen, et al. 2020. “Insights into Human Genetic Variation and Population History from 929 Diverse Genomes.” Science 367 (6484). https://doi.org/10.1126/science.aay5012.

Bhatti, Shahzad, Muhammad Aslamkhan, Marcella Attimonelli, Sana Abbas, and Hikmet Hakan Aydin. 2017. “Mitochondrial DNA Variation in the Sindh Population of Pakistan.” https://doi.org/10.1080/00450618.2016.1144788.

“Block Wise Provisional Summary Results of 6th Population & Housing Census-2017 [As on January 03, 2018] | Pakistan Bureau of Statistics.” n.d. Accessed February 11, 2020. http://www.pbs.gov.pk/content/block-wise-provisional-summary-results-6th-population-housing-census-2017-january-03-2018.

Cavaco, L, J. Strömberg-Nörklit, A. Kaneko, M. I. Msellem, M. Dahoma, V. L. Ribeiro, A. Bjorkman, and J. P. Gil. 2005. “CYP2C8 Polymorphism Frequencies among Malaria Patients in Zanzibar.” European Journal of Clinical Pharmacology 61 (1): 15–18. https://doi.org/10.1007/s00228-004-0871-8.

Dai, D., D. C. Zeldin, J. A. Blaisdell, B. Chanas, S. J. Coulter, B. I. Ghanayem, and J. A. Goldstein. 2001. “Polymorphisms in Human CYP2C8 Decrease Metabolism of the Anticancer Drug Paclitaxel and Arachidonic Acid.” Pharmacogenetics 11 (7): 597–607. https://doi.org/10.1097/00008571-200110000-00006.

Daily, Elizabeth B., and Christina L. Aquilante. 2009. “Cytochrome P450 2C8 Pharmacogenetics: A Review of Clinical Studies.” Pharmacogenomics 10 (9): 1489–1510. https://doi.org/10.2217/pgs.09.82.

Danielson, P. B. 2002. “The Cytochrome P450 Superfamily: Biochemistry, Evolution and Drug Metabolism in Humans.” Current Drug Metabolism 3 (6): 561–97. https://doi.org/10.2174/13892000023337054.

Goldstein, J. A. 2001. “Clinical Relevance of Genetic Polymorphisms in the Human CYP2C Subfamily.” British Journal of Clinical Pharmacology 52 (4): 349–55. https://doi.org/10.1046/j.0306-5251.2001.01499.x.

Inoue, S., E. M. Howgate, K. Rowland-Yeo, T. Shimada, H. Yamazaki, G. T. Tucker, and A. Rostami-Hodjegan. 2006. “Prediction of in Vivo Drug Clearance from in Vitro Data. II: Potential Inter-Ethnic Differences.” Xenobiotica; the Fate of Foreign Compounds in Biological Systems 36 (6): 499–513. https://doi.org/10.1080/0098250600683262.

Kudzi, William, Alexander No Dodoo, and Jeremy J. Mills. 2009. “Characterisation of CYP2C8, CYP2C9 and CYP2C19 Polymorphisms in a Ghanaian Population.” BMC Medical Genetics 10 (December): 124. https://doi.org/10.1186/1471-2350-10-124.

Mansoor, Atika, Kehkashan Mazhar, Shagufta Khaliq, Abdul Hameed, Sadia Rehman, Saima Siddiqi, Myrto Papaioannou, L. L. Cavalli-Sforza, S. Qasim Mehdi, and Qasim Ayub. 2004. “Investigation of the Greek Ancestry of
Populations from Northern Pakistan.” Human Genetics 114 (5): 484–90. https://doi.org/10.1007/s00439-004-1094-x.

Metspalu, Mait, Irene Gallego Romero, Bayazit Yunusbayev, Gyaneshwer Chaubey, Chandana Basu Mallick, Georgi Hudjashov, Mari Nelis, et al. 2011. “Shared and Unique Components of Human Population Structure and Genome-Wide Signals of Positive Selection in South Asia.” The American Journal of Human Genetics 89 (6): 731–44. https://doi.org/10.1016/j.ajhg.2011.10.002.

Minhas, S., N. Setia, S. Pandita, R. Saxena, I. C. Verma, and S. Aggarwal. 2013. “Prevalence of CYP2C8 Polymorphisms in a North Indian Population.” Genetics and Molecular Research: GMR 12 (3): 2260–66. https://doi.org/10.4238/2013.July.8.7.

Muthiah, Y. D., W. L. Lee, L. K. Teh, C. E. Ong, and R. Ismail. 2005. “Genetic Polymorphism of CYP2C8 in Three Malaysian Ethnicities: CYP2C8*2 and CYP2C8*3 Are Found in Malaysian Indians.” Journal of Clinical Pharmacy and Therapeutics 30 (5): 487–90. https://doi.org/10.1111/j.1365-2710.2005.00653.x.

Nakajima, Miki, Yuto Fujiki, Kumiko Noda, Hiroki Ohtsuka, Hisashi Ohkuni, Satoru Kyo, Masaki Inoue, Yukio Kuroiwa, and Tsuyoshi Yokoi. 2003. “Genetic Polymorphisms of CYP2C8 in Japanese Population.” Drug Metabolism and Disposition: The Biological Fate of Chemicals 31 (6): 687–90. https://doi.org/10.1124/dmd.31.6.687.

Pechandova, Kristina, Helena Buzkova, Olga Matouskova, Frantisek Perlik, and Ondrej Slanar. 2012. “Genetic Polymorphisms of CYP2C8 in the Czech Republic.” Genetic Testing and Molecular Biomarkers 16 (7): 812–16. https://doi.org/10.1089/gtmb.2011.0275.

Rostami-Hodjegan, Amin, and Geoffrey T. Tucker. 2007. “Simulation and Prediction of in Vivo Drug Metabolism in Human Populations from in Vitro Data.” Nature Reviews. Drug Discovery 6 (2): 140–48. https://doi.org/10.1038/nrd2173.

Staffa, Judy A., Jennie Chang, and Larh Green. 2002. “Cerivastatin and Reports of Fatal Rhabdomyolysis.” The New England Journal of Medicine 346 (7): 539–40. https://doi.org/10.1056/NEJM200202143460721.

Taus-Bolstad, Stacy. 2003. Pakistan in Pictures. Revised, Expanded edition. Minneapolis: Lerner Pub Group.

Weise, A., S. Grundler, D. Zaumsegel, M. Klotzek, B. Gröndahl, T. Forst, and A. Pfützner. 2004. “Development and Evaluation of a Rapid and Reliable Method for Cytochrome P450 2C8 Genotyping.” Clinical Laboratory 50 (3–4): 141–48.

Zanger, Ulrich M., and Matthias Schwab. 2013. “Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation.” Pharmacology & Therapeutics 138 (1): 103–41. https://doi.org/10.1016/j.pharmthera.2012.12.007.
## Supplementary Table 1: CYP2C8*2 Genotype Frequencies in the Pakistani Population

| Genotype  | n  | Observed genotype frequency (CI) | Expected genotype Count | Chi-Square statistic | p-value |
|-----------|----|----------------------------------|-------------------------|---------------------|---------|
| Pakistani |    |                                   |                         |                     |         |
| *1*1      | 309| 79.5 (75.5-83.5)                  | 305.29                  | 3.32                | p>0.05  |
| *1*2      | 73 | 18.7 (14.8-22.5)                  | 80.41                   |                     |         |
| *2*2      | 9  | 1.8 (0.5-3.1)                     | 5.29                    |                     |         |
| Punjabi   |    |                                   |                         |                     |         |
| *1*1      | 86 | 86 (79.2-92.8)                    | 85.56                   | 0.39                | p>0.05  |
| *1*2      | 13 | 13 (6.4-19.6)                     | 13.87                   |                     |         |
| *2*2      | 1  | 1 (0-3.0)                         | 0.56                    |                     |         |
| Pashtoon  |    |                                   |                         |                     |         |
| *1*1      | 68 | 85 (77.2-92.8)                    | 67.52                   | 0.49                | p>0.05  |
| *1*2      | 11 | 13.8 (6.2-21.3)                   | 11.94                   |                     |         |
| *2*2      | 1  | 1.2 (0-3.7)                       | 0.52                    |                     |         |
| Urdu      |    |                                   |                         |                     |         |
| *1*1      | 43 | 87.8 (78.6-96.9)                  | 43.24                   | 2.69                | p>0.05  |
| *1*2      | 5  | 10.2 (1.7-18.7)                   | 6.51                    |                     |         |
| *2*2      | 1  | 2.0 (0-6.0)                       | 0.24                    |                     |         |
| Seraiki   |    |                                   |                         |                     |         |
| *1*1      | 41 | 75.9 (64.5-87.3)                  | 40.90                   | 0.01                | p>0.05  |
| *1*2      | 12 | 22.2 (11.1-33.3)                  | 12.15                   |                     |         |
| *2*2      | 1  | 1.9 (0-5.4)                       | 0.90                    |                     |         |
| Balochi   |    |                                   |                         |                     |         |
| *1*1      | 41 | 70.7 (59.0-82.4)                  | 40.55                   | 0.18                | p>0.05  |
| *1*2      | 15 | 25.9 (14.6-37.1)                  | 15.88                   |                     |         |
| *2*2      | 2  | 3.4 (0-8.1)                       | 1.55                    |                     |         |
| Sindhi    |    |                                   |                         |                     |         |
| *1*1      | 30 | 60 (46.4-73.6)                    | 29.64                   | 0.08                | p>0.05  |
| *1*2      | 17 | 34 (20.9-47.1)                    | 17.71                   |                     |         |
| *2*2      | 3  | 6 (0-12.6)                        | 2.64                    |                     |         |
### Supplementary table 2: CYP2C8*3 genotype frequencies in the Pakistani population.

| Genotype | n  | Observed genotype frequency (CI) | Expected genotype Count HWE | Chi-Square statistic | p-value |
|----------|----|----------------------------------|----------------------------|----------------------|---------|
| **Pakistani** | | | | | |
| *1*1     | 287 | 73.4 (69.0-77.8) | 284.45 | 1.05 | p>0.05 |
| *1*3     | 93  | 23.8 (19.6-28.0) | 98.08 |     |         |
| *3*3     | 11  | 2.8 (1.2-4.5)    | 8.45  |     |         |
| **Punjabi** | | | | | |
| *1*1     | 73  | 73 (64.3-81.7)    | 73.96 | 0.63 | p>0.05 |
| *1*3     | 26  | 26 (17.4-34.6)    | 24.08 |     |         |
| *3*3     | 1   | 1 (0-3.0)         | 1.96  |     |         |
| **Pashtoon** | | | | | |
| *1*1     | 55  | 68.8 (58.6-78.9)  | 53.62 | 1.06 | p>0.05 |
| *1*3     | 21  | 26.3 (16.6-35.9)  | 23.74 |     |         |
| *3*3     | 4   | 5.0 (0.2-9.8)     | 2.62  |     |         |
| **Urdu** | | | | | |
| *1*1     | 41  | 83.7 (73.3-94.0)  | 40.41 | 1.01 | p>0.05 |
| *1*3     | 7   | 14.3 (4.5-24.12.0)| 8.17  |     |         |
| *3*3     | 1   | 2 (0-6.0)         | 0.41  |     |         |
| **Seraiki** | | | | | |
| *1*1     | 41  | 70.7 (59.0-82.4)  | 40.55 | 0.18 | p>0.05 |
| *1*3     | 15  | 25.9 (14.6-37.1)  | 15.88 |     |         |
| *3*3     | 2   | 3.4 (0-8.1)       | 1.55  |     |         |
| **Balochi** | | | | | |
| *1*1     | 42  | 77.8 (66.7-88.9)  | 40.90 | 1.73 | p>0.05 |
| *1*3     | 10  | 18.5 (8.2-28.9)   | 12.18 |     |         |
| *3*3     | 2   | 3.7 (0-8.7)       | 0.90  |     |         |
| **Sindhi** | | | | | |
| *1*1     | 35  | 70 (57.3-82.7)    | 35.28 | 0.08 | p>0.05 |
| *1*3     | 14  | 28 (15.6-40.4)    | 13.44 |     |         |
| *3*3     | 1   | 2 (0-5.9)         | 1.28  |     |         |
**Supplementary table 3:** CYP2C8*4 genotype frequencies in the Pakistani population.

| Genotype | n  | Observed genotype frequency (CI) | Expected genotype count by HWE | Chi-Square Statistic | p-value |
|----------|----|----------------------------------|--------------------------------|----------------------|---------|
| Pakistani |    |                                   |                                |                      |         |
| *1*1     | 379| 96.93 (95.22-98.64)               | 379.0921                       | 0.095                | P>0.05  |
| *1*4     | 12 | 3.06 (1.35-4.77)                 | 11.8159                        |                      |         |
| *4*4     | 0  | 0                                | 0.0921                         |                      |         |
| Punjabi  |    |                                   |                                |                      |         |
| *1*1     | 95 | 95 (90.73-99.27)                 | 95.0625                        | 0.06                 | P>0.05  |
| *1*4     | 5  | 5 (0.73-9.27)                    | 4.875                          |                      |         |
| *4*4     | 0  | 0                                | 0.0625                         |                      |         |
| Pashtoon |    |                                   |                                |                      |         |
| *1*1     | 76 | 95 (90.22-99.78)                 | 76.05                          | 0.05                 | P>0.05  |
| *1*4     | 4  | 5 (0.22-9.78)                    | 3.9                            |                      |         |
| *4*4     | 0  | 0                                | 0.05                           |                      |         |
| Urdu     |    |                                   |                                |                      |         |
| *1*1     | 49 | 100                              | 49                             | 0                    | P>0.05  |
| *1*4     | 0  | 0                                | 0                              |                      |         |
| *4*4     | 0  | 0                                | 0                              |                      |         |
| Seraiki  |    |                                   |                                |                      |         |
| *1*1     | 58 | 100                              | 58                             | 0                    | P>0.05  |
| *1*4     | 0  | 0                                | 0                              |                      |         |
| *4*4     | 0  | 0                                | 0                              |                      |         |
| Balochi  |    |                                   |                                |                      |         |
| *1*1     | 51 | 94.44 (88.33-100)               | 51.0417                        | 0.04                 | P>0.05  |
| *1*4     | 3  | 5.55 (0-11.66)                  | 2.9167                         |                      |         |
| *4*4     | 0  | 0                                | 0.0417                         |                      |         |
| Sindhi   |    |                                   |                                |                      |         |
| *1*1     | 50 | 100                              | 50                             | 0                    | P>0.05  |
| *1*4     | 0  | 0                                | 0                              |                      |         |
| *4*4     | 0  | 0                                | 0                              |                      |         |