Vietnam Journal of Biotechnology **20**(2): 197-212, 2022

IN SILICO STUDY OF CYTOCHROME P450 ALLELES AND PHENOTYPIC DISTRIBUTION IN VIETNAMESE POPULATION

Pham Ngoc Ha[1,*], Nguyen Phan Tuan[1,*], Trinh Thi Xuan[2], Truong Nam Hai[3], Tran Dang Hung[4], Nguyen Cuong[1,✉]

[1]*LOBI Vietnam Limited Liability Company, 27/385 Luong The Vinh Road, Nam Tu Liem District, Hanoi, Vietnam*
[2]*Faculty of Information Technology, Hanoi Open University, B101 Nguyen Hien Street, Hai Ba Trung District, Hanoi, VietNam*
[3]*Institute of Biotechnology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam*
[4]*Faculty of Information Technology, Hanoi National University of Education, 136 Xuan Thuy Road, Cau Giay District, Hanoi, Vietnam*

[*]These authors contributed equally to this work

[✉]To whom correspondence should be addressed. E-mail: juhuvn@gmail.com

Received: 07.07.2021
Accepted: 05.01.2022

SUMMARY

Cytochrome P450 enzymes play an important role in phase I drug metabolism, accounting for approximately 75% of the enzymatic processes. We investigated the allele and phenotypic distributions of five important *CYP* genes (*CYP2B6, CYP3A5, CYP2C9, CYP2C19, CYP2D6*) in the Vietnamese population by using Stargazer and the PharmCAT tool to call star alleles and translating them into phenotypes based on the available dataset of PharmGKB. We compared our computational analysis of the Vietnamese distributions with those of East Asia, Europe, America and other super populations, as well as with previous experimental research. The allele frequencies and phenotypic distributions of the five important *CYP* genes in the Vietnamese population are similar to those in East Asia while significantly different from other populations. In silico analysis also provided consistent results with previous experimental studies. In addition, the resultant data from our research contributes to the database of genetic variations in pharmacogenetics and constructs the fundamentals for future basic and applied research.

Keywords: *CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP3A5*, drug metabolism, Kinh Vietnamese, PharmCAT, pharmacogenomics, star alleles, Stargazer.

INTRODUCTION

Pharmacogenomics is the study and use of genetic variables pertaining to the drug response of individuals. Its applications are of interest to industry and patient care, for instance, increasing drug development efficiency by detecting drug responders and drug non-responders in clinical trials and identifying those at risk of adverse effects. Pharmacogenomics can, particularly, support clinicians with prescription decision-making and determining the best dosage of a medication for patients. Therefore, this is an effective and potentially cost-saving clinical tool (Hockings *et al.*, 2020).

Cytochrome P450 enzymes are involved in approximately 75% of the enzymatic processes in drug metabolism; therefore, the range of disciplines in which P450s are studied has broadened drug development (Guengerich *et al.*, 2016). Multiple medications and genetic polymorphisms that affect drug-metabolizing cytochrome P450 (CYP) enzyme activity are important causes of drug pharmacokinetics and drug response variability, which are important clinical issues among individuals. Dosage guidelines based on CYP genotype would assist doctors in prescribing the optimal medication treatment and desired drug dose for patients (Samer *et al.*, 2013).

Most of the well-known and widely accepted guidelines, such as those published by CPIC (https://cpicpgx.org/), WHO (https://www.who.int/), FDA (https://www.fda.gov/), and ESC (https://www.escardio.org/), are based on the European and American populations. Although lots of studies have demonstrated their efficiency and safety, there are still cases when the recommended prescriptions do not work effectively (Ma, Chan, 2013; Tesar, Hruskova, 2015). One reason for these variations is the difference in genetic factors between ethnicities. Thus, for a Southeast Asian country like Vietnam, some Western guidelines may not be optimal. If the drug metabolizing abilities of Vietnamese people are different than those in Western countries, clinicians should consider adjusting the medication for better responses.

Several studies have reviewed the pharmacogenomics of Vietnamese people. Veiga *et al*. (2009) worked on seven genes related to malaria treatment, *CYP2A6*, *CYP2B6*, *CYP2C19*, *CYP2D6*, *CYP3A4*, *CYP3A5*, *MDR1*, and detected the frequency of common star alleles of these pharmacogenes in the Vietnamese population. Recent studies by many authors have determined the polymorphisms of *CYP2C9*, *CYP2C19*, *CYP3A5*, *CYP2D6* genes using genotyping kits, and identified a number of novel SNPs that appeared in 100 Vietnamese people living in Hanoi (Nguyen *et al.*, 2019; Nhung *et al.*, 2020; Vu *et al.*, 2018; 2019). Our study has a similar aim, but different approach was used for determining the allele prevalence.

In this study, we constructed an allele frequency and phenotypic distribution figure of five essential *CYP* genes (*CYP2B6*, *CYP3A5*, *CYP2C9*, *CYP2C19*, and *CYP2D6*) of 99 Kinh people in the Vietnamese population, thereby comparing them with the distributions of East Asian, European, American super populations, and others. Besides, we also compare our results with the experimental results of previous studies to confirm whether *in silico* analysis is similar to clinical. We expect that our results will determine the need for dose adjustments of the drugs metabolized by the five important *CYP* genes, as well as provide useful information for further study related to the pharmacogenomics field in the Vietnamese population.

MATERIALS AND METHODS

Subjects

The information about variants was collected from the variant call format (VCF) files of the 1000 Genomes Project (Clarke *et al.*, 2017). A total of 2504 individuals are categorized into 5 groups, including 99 Kinh Vietnamese people (KHV), 405 East Asians excluding Vietnamese (EAS), 504 Americans (AMR), 503 Europeans (EUR), and 993 other ethnicities (Others), including both South Asian and African people. In the American group, 157 are African descendants.

Calling star alleles

We identified specific regions and selected the corresponding subsets of the VCF files for genotyping. Haplotypes were identified as star alleles using only the previously generated VCF files with two different tools, Stargazer (Lee *et al.*, 2019) and PharmCAT (Klein, Ritchie, 2018). With Stargazer, the two main candidates, one for each haplotype, were combined to form a diplotype. With PharmCAT, if there are multiple diplotypes predicted, the diplotypes with the

highest frequency in the respective region as reported by PharmGKB are chosen. For the KHV population, the PharmGKB-reported frequencies of the East Asian population were used.

Phenotyping

The diplotype was mapped to phenotypes based on the gene-specific table provided by PharmGKB. If the phenotypes "likely intermediate metabolizer" and "likely poor metabolizer" are found, they will be considered as "intermediate metabolizer" and "poor metabolizer", respectively. For the four genes, *CYP2B6*, *CYP2C9*, *CYP2C19*, and *CYP3A5*, only consensus phenotypes received from the two tools are used for downstream analysis. For the gene *CYP2D6*, as only Stargazer covers this gene, its results are used for analysis.

Data analysis

Allele frequencies and phenotypic distributions were calculated using the consensus alleles and phenotypes called by Stargazer and PharmCAT. For *CYP2D6*, only Stargazer results were utilized as PharmCAT was not able to call star alleles for this gene. Indeterminate phenotypes were excluded from the analysis. The distribution is visualized by R and Microsoft Excel, and their differences are evaluated using the Chi-squared test and Student's t-test when appropriate.

RESULTS AND DISCUSSION

Allele Frequency

The distributions of allele frequencies between the Vietnamese and East Asian populations are profoundly similar in all five investigated genes. In contrast, other populations show distinctive differences in most allele frequency spectrums, especially with the distributions of *CYP2C9* and *CYP2D6*. The only exception is *CYP3A5*, as allele *CYP3A5**3 is the only variant, besides the reference allele *1, in three groups and the most popular variant in the other two groups. (Fig. 1, Table S1).

Stargazer and PharmCAT have been demonstrated to produce accurate results with the investigated genes (Lee *et al.*, 2019). To predict precisely both the single-nucleotide polymorphism (SNP) and the copy number variation (CNV) of an allele, we would require chromosome structural information, which can be retrieved from BAM files. However, "The 1000 Genome Project" does not provide such a format, and it is impossible to process the pipeline for whole-genome sequencing data of 2504 people due to the limitation of computational resources and time. Therefore, we worked directly with the VCF files, which means we only rely on SNPs to predict genotype and phenotype information of the 5 important pharmacogenes in studied populations. Thus, our result for *CYP2D6* does not show the duplication, deletion, and hybrid alleles which were detected in previous studies. These variants were estimated to make up about 5% over all biogeographical groups (Naranjo *et al.*, 2018; Sistonen *et al*., 2007) and should be taken into consideration for interpretation. Regarding the other four genes, because no large structural variants have been defined by CPIC, the result would be indifferent with or without the information about depth of coverage.

Allele frequencies of the three super populations, ie., American, European, and East Asians, have been reported previously (Zhou *et al.*, 2017) using different algorithms, and their results match ours. The relative frequencies of the three most common *CYP2C9* alleles, namely **1*, **2*, and **3*, between the two studies are deeply consistent. With *CYP2C19*, more than half of its star alleles in the European and American populations is **17*, making it the most dominant variant. For East Asia, *CYP2C19**2 is the major allele that makes up about three-quarters of all variant alleles, followed by *CYP2C19**3 and *CYP2C19**17. For the extremely polymorphic gene *CYP2D6*, there are discrepancies with several uncommon alleles between the two studies; however, the order of the popular variants (**2*, **4*, and **10*) is still the same.

The frequencies of *CYP2B6* alleles also show numerous similarities between the two studies.

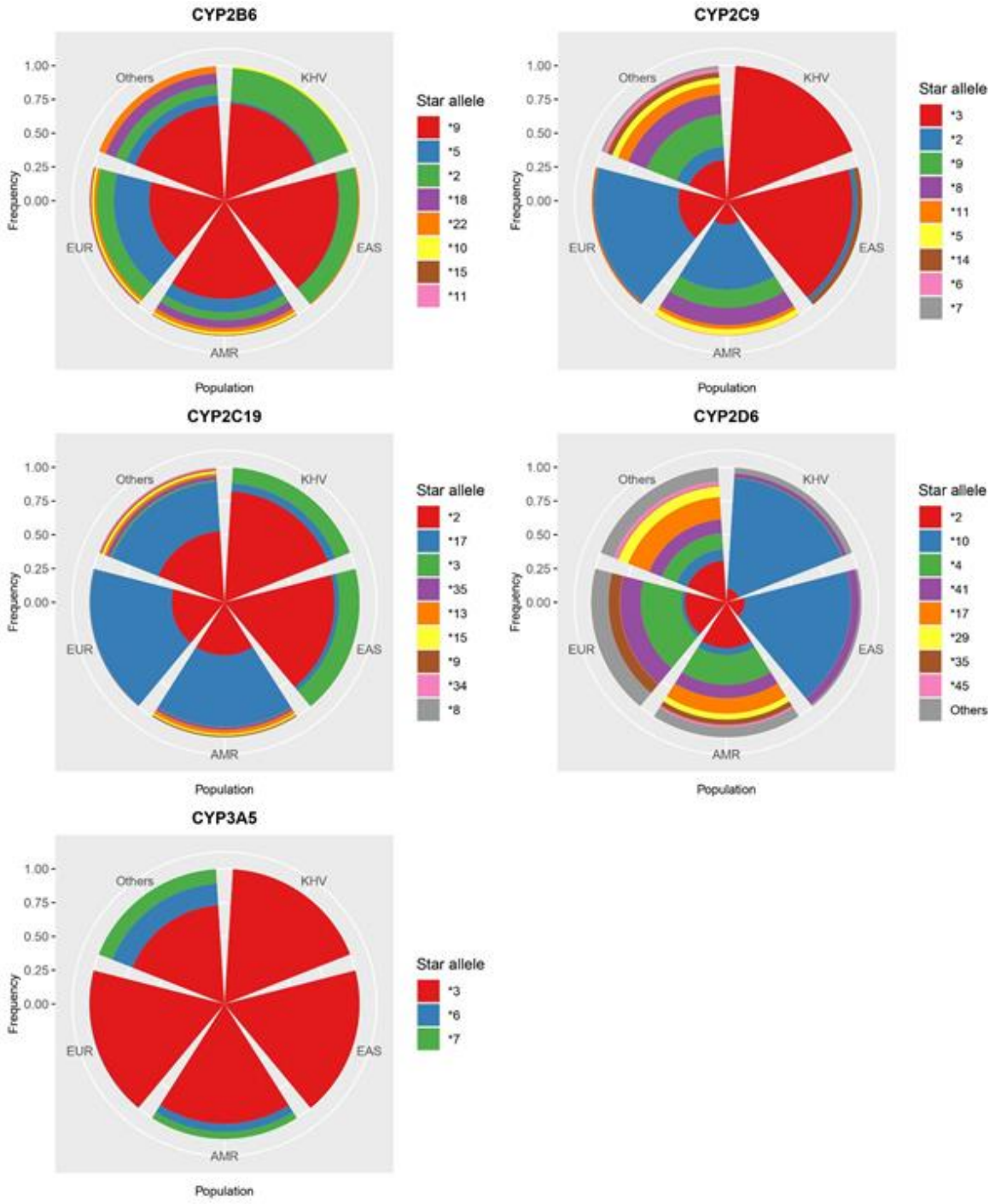

Figure 1. Distribution of variant alleles of five CYP genes in Kinh-Vietnamese (KHV), East Asian (EAS), American (AMR), European (EUR) and other populations. The reference allele **1* is excluded. Each pole is a stacked bar of proportions of the other alleles identified.

Both studies demonstrate that allele *CYP2B6*9* is the most common variant in all three groups, *CYP2B6*5* is the second most common allele in Europe and America, and *CYP2B6*2* is the second most common in East Asia. However, a discordant observation in our study is the surprisingly high frequency of *CYP2B6*9* and the absence of *CYP2B6*6*, which were estimated to occupy 7–10% of the variants (excluding *CYP2B6*1*) (Zhou *et al.*, 2017). The allele *CYP2B6**6 consists of two SNPs, rs3745274 and rs2279343, and these SNPs alone correspond to alleles *CYP2B6*9* and *CYP2B6*4*, respectively. The 1000 Genome Project did not call the variant rs2279343 at all, so allele *CYP2B6*6* might be counted as *CYP2B6*9* in the present study and lead to an addition in the frequency of *CYP2B6*9*.

Our result is also in accordance with several experimental studies conducted on Vietnamese people (see Table 1) (Nguyen *et al.*, 2019; Nhung *et al.*, 2020; Veiga *et al.*, 2009; Vu *et al.*, 2018; 2019). In particular, Vu *et al.* (2018) have studied the polymorphism of multiple *CYP* genes in 99 Vietnamese people of the Kinh ethnic group. The reported allele frequencies of *CYP2C9**1 and *CYP2C9**3 were identical to ours, 96.5% and 3.5%, respectively (Vu *et al*., 2018). *CYP2C9*3* differs from the reference allele by one single nucleotide (rs1057910, 42614A > C). Based on a genetic database constructed from 206 Vietnamese individuals, this substitution occurs with a frequency of approximately 3% (Le *et al*., 2019). Similarly, *CYP3A5*3* is the only identified variant of the respective gene with a prevalence of about 70% in all the relevant studies and the reference allele *CYP3A5**1 occupies the other 30% (Nhung *et al*., 2020).

Table 1. Frequency of the most common star alleles in Kinh – Vietnamese (KHV) population in different studies.

Allele	KHV population in this study (n = 99)	Vu *et al*. (2020) (n=100)	Veiga *et al*. (2009) (n=78)
CYP2B6			
*2	7.58	-	-
*6	-	-	27.1
*9	22.22	-	-
CYP2C9			
*3	3.54	3.5	-
CYP2C19			
*2	28.28	20.5	30.6
*3	4.04	2.5	6.3
*17	2.02	1	-
CYP2D6			
*2	8.08	7.35	-
*4	0.51	0.74	1.4
*10	65.66	43.75	43.5
CYP3A5			
*3	71.21	67.5	66.7

*CYP2C19*2* is the most common *CYP2C19* variant in all studies, but its frequency varies from 20 to 30%. Though both studies by Veiga *et al*. (2009) and Vu *et al*. (2018) showed that the allele frequencies of *CYP2C19* were in Hardy-Weinberg equilibrium, it is still possible that a gradual decrease in the allele frequency has occurred. As the present study utilizes "The 1000 Genome Project" data collected from 2008 to 2015 (Clarke *et al*., 2017), it is reasonable that

our result is closer to that of the Veiga *et al*. (2009)'s study. Allele *CYP2C19*17* is another noteworthy allele, as it is the major allele in other populations but almost absent in Vietnamese and East Asians. This difference may lead to crucial clinical implementation because *CYP2C19*2* and *CYP2C19*3* are alleles with no function, while *CYP2C19*17* has increased metabolic activity. Thus, the metabolism of relevant drugs may be distinctively different between these populations.

In our results, for *CYP2D6*, there is probably an overestimation of the allele *CYP2D6*10* frequency. The frequency of many common alleles (**1, *2, *4*) was nearly the same as demonstrated in a previous study (Nguyen *et al*., 2019). However, in this study, the frequency of *CYP2D6*10* was reported to be about 44%, while that of ours is vastly greater (66%). A possible explanation is the absence of structural variants in our results, as duplications of *CYP2D6*10* (**10xN*) as well as hybrids containing *CYP2D6*10* (e.g., **36 + *10*) would all be classified as *CYP2D6*10* in the present study. However, as most structural variants containing *CYP2D6**10 are considered to have decreased function, our phenotypic distribution should not be affected drastically.

We experience a similar discordance with the distribution of *CYP2B6* in Vietnam as we did with the previous three super populations. Veiga *et al.* (2009)'s study reported that the frequency of *CYP2B6*6* was 27.1%, but the data from "The 1000 Genome Project" as well as the database provided by Le *et al*. (2019) did not identify the SNP rs2279343, and consequently, no *CYP2B6*6* were found. Therefore, the prevalence of allele *CYP2B6*9* might be overestimated while *CYP2B6*6* might be underestimated. Fortunately, both have decreased function, and substituting one by another would not interfere with phenotype interpretation. Furthermore, our study also identified the frequency of *CYP2B6*2*, which was not included in the study of Veiga et al. (2009), and the result is consistent with the variation analysis in Le *et al.* (2019).

Phenotypic Distribution

The comparison of the *CYP2B6*, *CYP2C9*, *CYP2C19*, and *CYP3A5* phenotypic distributions obtained by Stargazer and PharmCAT is shown in Figs. 2 to 6. Since PharmCAT cannot call star alleles for CYP2D6, the comparison for this gene is excluded from the figure. The PharmCAT algorithm provides multiple genotypes but does not score the most reliable. Hence, we select the proper genotypes for downstream phenotype matching using the population allele frequency database provided by PharmGKB. In contrast, Stargazer predicts the most likely genotypes based on the given variants of each individual. Though both tools provide reliable results, the simple setup process, short analysis time, and extensive gene coverage make Stargazer the better genotyping software to call star alleles of these five *CYP* genes as well as other pharmacogenes in the future.

Phenotypic distribution of the *CYP2B6* gene

The *CYP2B6* gene is found on chromosome 19's long arm along with the closely related pseudogene *CYP2B7* and numerous other members of the *CYP2* gene family. The *CYP2B6* gene has at least 38 allelic variations, 25 of which are deemed significant and eight of which are prevalent in at least one racial/ethnic community. The CYP2B6 enzyme metabolizes a broad spectrum of substrates, accounting for roughly 8% of all commercially available medicines (Wang *et al*., 2019). Therefore, *CYP2B6* genetic testing should be considered before prescribing.

Among people suffering from HIV, the frequency of reduced or loss-of-function alleles of the *CYP2B6* gene was highest in African ancestry patients (Klein *et al*., 2005). Our study with healthy individuals showed a similar result; in particular, the two groups with the highest intermediate and poor metabolizer phenotype percentage are the American population, which includes 157/504 African Americans, and the Others group, in which half are African.

The *CYP2B6* gene polymorphism significantly affects the pharmacokinetics of

efavirenz, an important antiretroviral agent used to treat HIV. For individuals with a poor metabolic phenotype, plasma efavirenz concentrations are often elevated (the likelihood ratio is 35) and strongly correlated with an increased risk of suicide in patients receiving the drug (Mollan *et al*., 2017; Rotger *et al*., 2007). According to the results, 42.9% of Vietnamese people have a poor metabolic phenotype (95% CI 33-52.8) and should use the reduced starting dose when treated with efavirenz. This ratio highlights the importance of individualized treatment for the Vietnamese population.

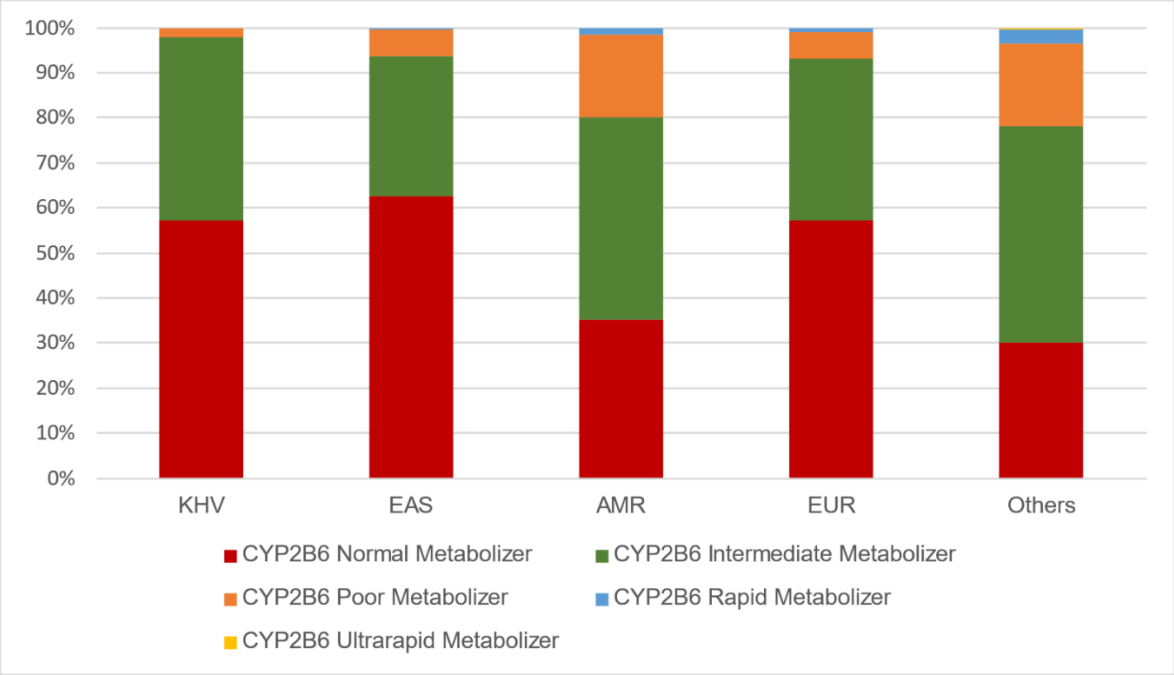

Figure 2. Phenotypic distribution of the CYP2B6 gene in five populations. The normal metabolizer accounted for the highest percentage of the Vietnamese population, followed by the intermediate metabolizer and the poor metabolizer. The phenotypic distribution of the Vietnamese population is compatible with the East Asian super population. The others have witnessed the presence of rapid and ultrarapid metabolizers ata very low rate.

Phenotypic distribution of the *CYP2C9* gene

The *CYP2C9* gene is found on chromosome 10q24.1, and there are around 60 different *CYP2C9* alleles (Cavallari, Momary, 2019). Numerous medications, such as nonsteroidal anti-inflammatory drugs, losartan, tolbutamide, warfarin, phenytoin, or carbamazepine, are metabolized by the *CYP2C9* gene (Lazar *et al*., 2004). The majority of East Asians are normal Metabolizer, which partially explains the high tolerability of celecoxib in Asians (Essex *et al*., 2016). Due to the differences in phenotypic distribution, clinical drugs applicable to American or European populations might not be suitable to East Asians, especially Vietnamese people. Therefore, the results above support the strict control of over-the-counter NSAIDs in Vietnam.

In addition, CYP2C9 is the main enzyme responsible for the elimination of various drugs with a narrow therapeutic range, such as warfarin or phenytoin, so the phenotype of *CYP2C9* gene has a considerable influence on the efficacy and safety of the drug (Daly *et al*., 2017). The metabolism of these drugs also depends on other genes which were not analyzed in this study (such as *VKORC1, HLA-B*), so the phenotypic distribution of the *CYP2C9* gene might not accurately reflect the differences in the risk of adverse reactions between populations.

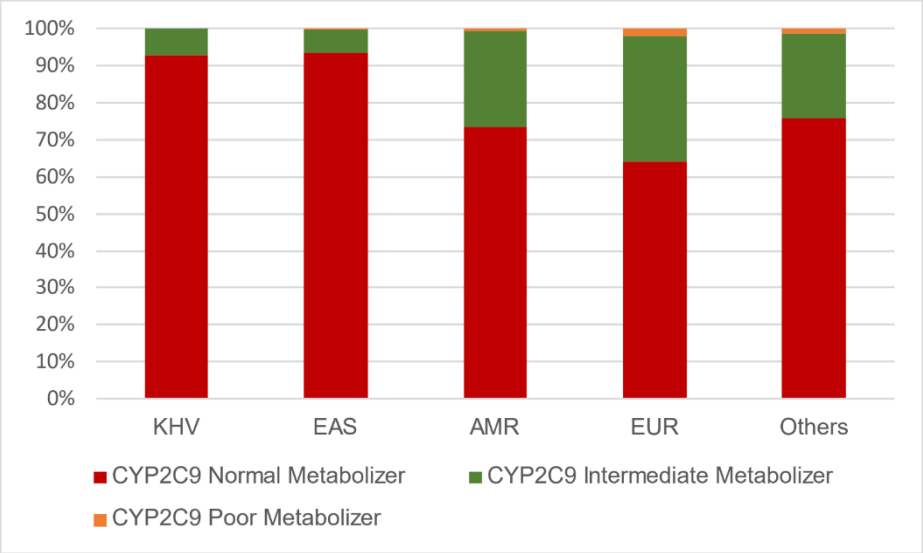

Figure 3. Phenotypic distribution of the CYP2C9 gene in the five populations. For the Vietnamese population, the Normal Metabolizer accounted for the highest percentage, followed by the Intermediate Metabolizer. This phenotypic distribution is compatible with the East Asian super population. The others had witnessed the presence of Poor Metabolizer, at a very low rate.

Phenotypic distribution of the *CYP2C19* gene

The *CYP2C19* gene is found on chromosome 10q24 and 35 variants are presently known. CYP2C19 is the most important enzyme in the hepatic metabolism of drugs such as antimalarials (proguanil), oral anticoagulants (R-warfarin), chemotherapeutic agents (cyclophosphamide), anti-epileptics (S-mephenytoin, diazepam, phenobarbitone), antiplatelets (clopidogrel), proton pump inhibitors (omeprazole, pantoprazole, lansoprazole, rabeprazole), antivirals (nelfinavir), and antidepressants (amitriptyline, clomipramine) (Gurusamy, Shewade, 2014).

It is worth noting that intermediate and poor metabolizer phenotypes dominate in the East Asian super population as well as the Vietnamese population. In contrast, the percentage of rapid metabolizers in these two super populations is much higher than in the East Asian super population, which is coupled with the presence of the ultrarapid metabolizer phenotype. Doctors should consider genetic testing for the *CYP2C19* gene to provide optimal drug doses for patients or even use an alternative therapy.

Phenotypic distribution of the *CYP2D6* gene

The *CYP2D6* gene, which codes for the CYP2D6 enzyme, is found on chromosome 19. *CYP2D6* is one of the most polymorphic *CYP* genes in humans, with about 80 distinct allelic variants and 130 genetic variations documented (Demkow, 2016). Antidepressants, antipsychotics, beta-blockers, antiviral medicines, antiarrhythmics, morphine derivatives, and tamoxifen are among the medications metabolized by this enzyme, which has a restricted therapeutic window (Vuppalanchi, 2011). *CYP2D6* exhibits extraordinary variability, sometimes with complete gene duplication, with over 90 confirmed allelic variations identified. More than 50 drug substrates are known to be metabolized by this route, which accounts for 20% to 30% of all medicines. The *CYP2D6* gene has been widely investigated because of these critical characteristics (Schaffenburg *et al*., 2021).

Previous studies have determined that *CYP2D6*∗10 was determined to be the main variant causing decreased CYP2D6 enzyme activity in the Vietnamese population,

constituting an intermediate metabolizer phenotype (Nguyen *et al*., 2019; Veiga *et al*., 2009). Therefore, physicians should consider genetic testing for allele *CYP2D6**10 to provide an appropriate drug dose for each patient. Our method could not identify CNVs, so the proportion of *CYP2D6* ultrarapid metabolizers was underestimated. In future studies, the identification of *CYP2D6* CNVs by algorithms should be considered.

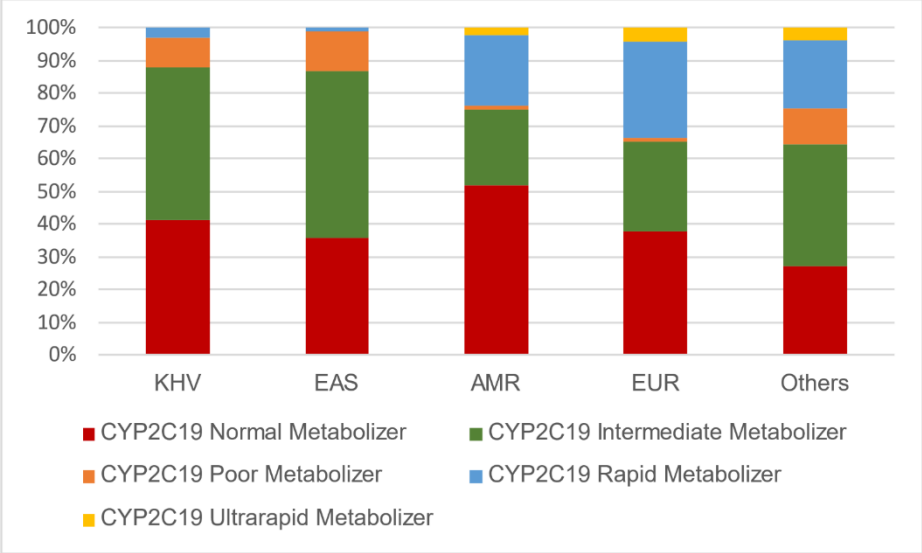

Figure 4. Phenotypic distribution of the CYP2C19 gene in the five populations. For the Vietnamese population, intermediate metabolizers and normal metabolizers accounted for nearly 90% of the population, which is compatible with the East Asian super population. Rapid metabolizers took up a small percentage in the Vietnamese population, whereas there was no presence of ultrarapid metabolizers. By contrast, rapid metabolizers made up a significant proportion in the American and European super populations.

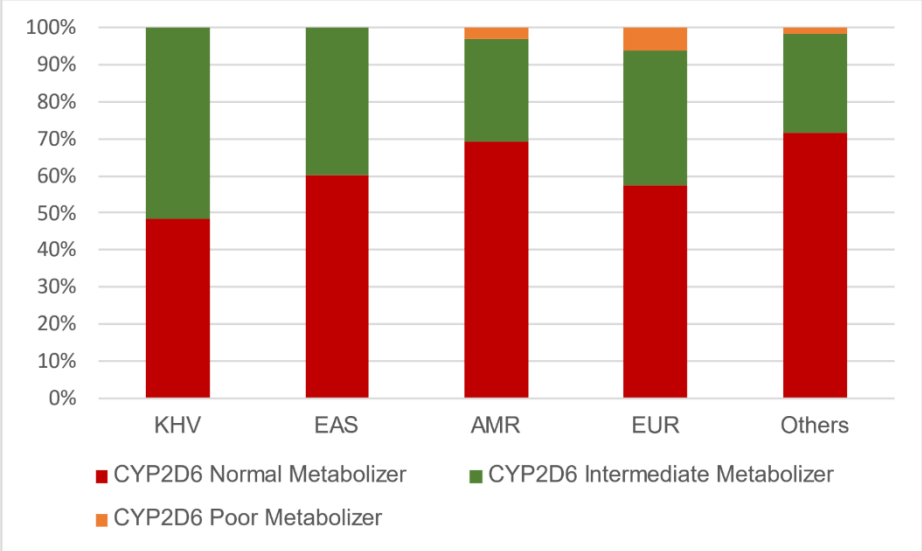

Figure 5. Phenotypic distribution of CYP2D6 gene in the five populations. Normal metabolizer and intermediate metabolizer are predominant in all populations. The normal metabolizer percentage in Vietnamese people is significantly different from the American and European super population (p-value after Bonferroni correction < 0.05).

Phenotypic distribution of the *CYP3A5* gene

The *CYP3A5* gene, which codes for the CYP3A5 enzyme, is located on chromosome 7q21.1. and involved in the metabolism of medicines used to treat three of the most common infectious diseases: malaria (artemether, lumefantrine, mefloquine, primaquine, chloroquine), HIV/AIDS (efavirenz, saquinavir, maraviroc, indinavir, nelfinavir, ritonavir, lopinavir), and tuberculosis (ritonavir, rifampicin) (Masimirembwa *et al*., 2014). Kuehl *et al*. (2001) demonstrated that individuals need to carry at least one *CYP3A5**1 allele to express the CYP3A5 protein, whereas *CYP3A5**3 (A to G at 6986) in the intron 3 region results in the absence of the CYP3A5 protein.

The proportion of Vietnamese people carrying genotypes **1/*3* and **3/*3* is very high, so the distribution of *CYP3A5* genotypes of the Kinh population mainly belongs to intermediate and poor metabolizers, with 40% and 50%. However, drug dose adjustment is not necessary for individuals who are CYP3A5 poor metabolizers because most drugs have been developed from CYP3A5 poor metabolizers. In contrast, individuals with intermediate and normal metabolizer phenotypes required dose adjustment to get effective treatment (Birdwell *et al*., 2015).

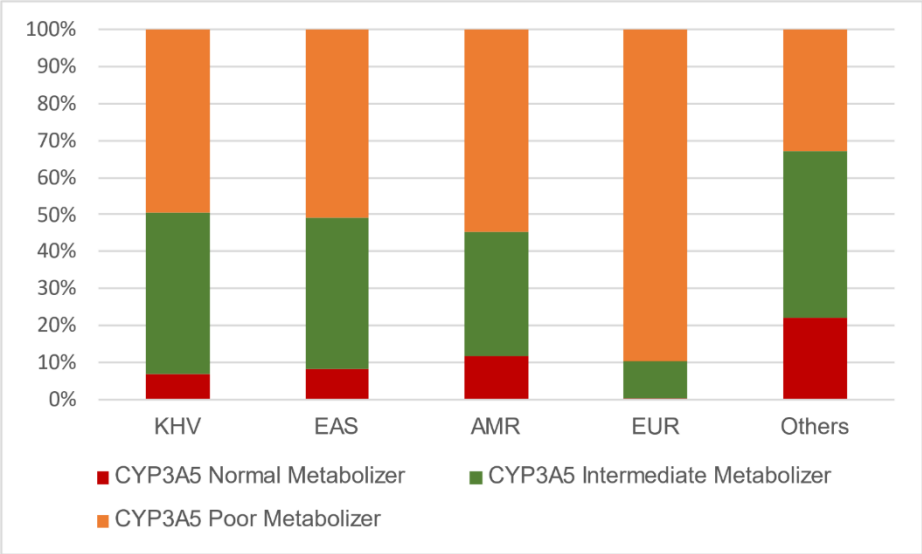

Figure 6. Phenotypic distribution of the CYP3A5 gene in the five populations. Intermediate and poor metabolizers were predominant in all populations. The distribution of intermediate and poor metabolizer phenotypes in the Vietnamese population is similar to other groups, except the Europeans. No drug dose adjustment is required for patients who have the CYP3A5 poor metabolizer phenotype as most drugs are developed from CYP3A5 poor metabolizers.

CONCLUSION

The genotypic and phenotypic distributions of five important pharmacogenes provided useful information about the Kinh population in particular. This result has a high similarity when compared with experimental as well as previous research.

The phenotypic distribution of the Kinh population is significantly different from other populations in the world. These results will provide directions for researchers to target the key points that should be exploited in the field of pharmacogenomics of the Vietnamese population.

However, studies on genetic variation of important pharmacogenes are mainly being carried out with the genomes of the Kinh ethnic

group, while the ethnic minorities of Vietnam appear only in a few previous studies with a small population size: 3 Thai people (Veiga *et al*., 2009), 275 people from four minority groups (Vu *et al*., 2019), and 279 people from four other ethnic groups (Vu *et al*., 2018). Therefore, it is necessary to collect more genomic data from ethnic minorities on a larger scale to have a better overview of the phenotypic distribution of important pharmacological genes of the Vietnamese ethnic groups.

Acknowledgement: *This research was supported by LOBI Vietnam's research and development fund. LOBI Vietnam is a bioinfomatics company.*

REFERENCES

Birdwell K, Decker B, Barbarino J, Peterson J, Stein C, Sadee W, Wang D, Vinks A, He Y, Swen J, Leeder J, van Schaik R, Thummel K, Klein T, Caudle K, MacPhee I (2015) Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. *Clin Pharm Therap* 98(1): 19–24.

Cavallari LH, Momary KM (2019) *Pharmacogenomics (Second Edition)*, Academic Press.

Clarke L, Fairley S, Zheng-Bradley X, Streeter I, Perry E, Lowy E, Tassé A-M, Flicek P (2017) The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res* 45(D1): D854–D859.

Daly AK, Rettie AE, Fowler DM, Miners JO (2017) Pharmacogenomics of CYP2C9: Functional and Clinical Considerations. *J Pers Med* 8(1): 1.

Demkow U (2016) *Clinical Applications for Next-Generation Sequencing*, Academic Press.

Essex MN, O'Connell MA, Behar R, Bao W (2016) Efficacy and safety of nonsteroidal anti-inflammatory drugs in Asian patients with knee osteoarthritis: Summary of a randomized, placebo-controlled study. *Int J Rheum Dis* 19(3): 262–270.

Gurusamy U, Shewade DG (2014) *Handbook of Pharmacogenomics and Stratified Medicine*, Academic Press.

Hockings JK, Pasternak AL, Erwin AL, Mason NT, Eng C, Hicks JK (2020) Pharmacogenomics: An evolving clinical tool for precision medicine. *Clev Clin J Med* 87(2): 91–99.

Klein TE, Ritchie MD (2018) PharmCAT: A Pharmacogenomics Clinical Annotation Tool. *Clin Pharmacol Ther* 104(1): 19–22.

Klein K, Lang T, Saussele T, Barbosa-Sicard E, Schunck W-H, Eichelbaum M, Schwab M, Zanger UM (2005) Genetic variability of CYP2B6 in populations of African and Asian origin: Allele frequencies, novel functional variants, and possible implications for anti-HIV therapy with efavirenz. *Pharmacogenet Genom* 15(12): 861–873.

Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, Watkins PB, Daly A, Wrighton SA, Hall SD, Maurel P, Relling M, Brimer C, Yasuda K, Venkataramanan R, Strom S, Thummel K, Boguski MS, Schuetz E (2001) Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat Genet* 27(4): 383–391.

Lazar A, Tomalik-Scharte D, Fuhr U (2004) *Handbook of Analytical Separations*, Elsevier Science BV.

Le VS, Tran KT, Bui HTP, Le HTT, Nguyen CD, Do DH, Ly HTT, Pham LTD, Dao LTM, Nguyen LT (2019) A Vietnamese human genetic variation database. *Hum Mutat* 40(10): 1664–1675.

Lee S, Wheeler MM, Thummel KE, Nickerson DA. (2019). Calling Star Alleles With Stargazer in 28 Pharmacogenes With Whole Genome Sequences. *Clin Pharmacol Ther* 106(6): 1328–1337.

Ma RC, Chan JC (2013) Type 2 diabetes in East Asians: Similarities and differences with populations in Europe and the United States. *Ann NY Acad Sci* 1281(1): 64–91.

Masimirembwa C, Dandara C, Hasler J (2014). *Handbook of Pharmacogenomics and Stratified Medicine*, Academic Press.

Mollan KR, Tierney C, Hellwege JN, Eron JJ, Hudgens MG, Gulick RM, Haubrich R, Sax PE, Campbell TB, Daar ES, Robertson KR, Ventura D, Ma Q, Edwards DRV, Haas DW (2017) Race/Ethnicity and the Pharmacogenetics of Reported Suicidality With Efavirenz Among Clinical Trials Participants. *J Infect Dis* 216(5): 554–564.

Naranjo MG, Rodrigues-Soares F, Peñas-Lledó EM, Tarazona-Santos E, Fariñas H, Rodeiro I, Terán E, Grazina M, Moya GE, López-López M, Sarmiento AP, Calzadilla LR, Ramírez-Roa R, Ortiz-López R, Estévez-Carrizo FE, Sosa-Macías M, Barrantes R, LLerena A; CEIBA-Consortium of the Ibero-American Network of Pharmacogenetics and Pharmacogenomics RIBEF (2018). Interethnic Variability in CYP2D6, CYP2C9, and CYP2C19 Genes and Predicted Drug Metabolism Phenotypes Among 6060 Ibero- and Native Americans: RIBEF-CEIBA Consortium Report on Population Pharmacogenomics. *Omics* 22(9): 575–588.

Nguyen HH, Ma TTH, Vu NP, Bach QTN, Vu TH, Nguyen TD, Nong HV (2019) Single nucleotide and structural variants of CYP2D6 gene in Kinh Vietnamese population. *Medicine*: 98(22).

Nhung V, Nguyen DT, Nguyen H (2020) Study of CYP3A5 genetic polymorphism in Vietnamese Kinh ethnic group. *TAP CHI SINH HOC* 42.

Guengerich FP, Waterman MR, Egli M (2016) Recent Structural Insights into Cytochrome P450 Function. *Trends Pharmacol Sci* 37(8): 625–640.

Rotger M, Tegude H, Colombo S, Cavassini M, Furrer H, Décosterd L, Blievernicht J, Saussele T, Günthard HF, Schwab M, Eichelbaum M, Telenti A, Zanger UM (2007). Predictive value of known and novel alleles of CYP2B6 for efavirenz plasma concentrations in HIV-infected individuals. *Clin Pharm Ther* 81(4): 557–566.

Samer CF, Lorenzini KI, Rollason V, Daali Y, Desmeules JA (2013). Applications of CYP450 testing in the clinical setting. *Mol Diagn Ther* 17(3): 165–184.

Schaffenburg WC, Lockshin BN, DeKlotz CMC (2021) *Comprehensive Dermatologic Drug Therapy (Fourth Edition)*, Elsevier.

Sistonen J, Sajantila A, Lao O, Corander J, Barbujani G, Fuselli S (2007) CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. *Pharmacogenet Genom* 17(2): 93–101.

Tesar V, Hruskova Z (2015) Lupus Nephritis: A Different Disease in European Patients? *Kidney Dis (Basel)* 1(2): 110–118.

Veiga MI, Asimus S, Ferreira PE, Martins JP, Cavaco I, Ribeiro V, Hai TN, Petzold MG, Björkman A, Ashton M, Gil JP (2009) Pharmacogenomics of CYP2A6, CYP2B6, CYP2C19, CYP2D6, CYP3A4, CYP3A5 and MDR1 in Vietnam. *Eur J Clin Pharmacol* 65(4): 355–363.

Vu NP, Ma TTH, Tran NTB, Huynh HTT, Nguyen TD, Nguyen DT, Van Nong H, Lee MTM, Nguyen HH (2018) Polymorphic analysis of CYP2C9 gene in Vietnamese population. *Mol Biol Rep* 45(5): 893–900.

Vu NP, Nguyen HTT, Tran NTB, Nguyen TD, Huynh HTT, Nguyen XT, Nguyen DT, Nong HV, Nguyen HH (2019) CYP2C19 genetic polymorphism in the Vietnamese population. *Ann Hum Biol* 46(6): 491–497.

Vuppalanchi R (2011) *Practical Hepatic Pathology: A Diagnostic Approach*, WB Saunders.

Wang PF, Neiner A, Kharasch ED (2019) Efavirenz Metabolism: Influence of Polymorphic CYP2B6 Variants and Stereochemistry. *Drug Metab Dispos* 47(10): 1195–1205.

Zhou Y, Ingelman-Sundberg M, Lauschke V (2017) Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. *Clin Pharmacol Ther* 102(4): 688–700.

SUPPLEMENTS

Table S1. All Allele Frequency of five CYP450 genes in five populations.

CYP2B6	**KHV**	**EAS**	**AMR**	**EUR**	**Others**
*1	69.19	74.69	48.81	57.65	46.58
*2	7.58	3.58	3.08	5.37	4.38
*5	0.51	0.12	5.06	11.03	3.93
*9	22.22	21.36	37.20	23.56	37.87
*10	0.51	0.00	0.89	0.70	0.05
*11	0.00	0.00	0.00	0.40	0.10
*15	0.00	0.00	0.40	0.40	0.00
*18	0.00	0.00	2.98	0.00	4.33
*22	0.00	0.25	1.59	0.89	2.77
CYP2C9	**KHV**	**EAS**	**AMR**	**EUR**	**Others**
*1	96.46	96.54	83.73	79.82	82.23
*2	0.00	0.12	7.84	12.43	1.76
*3	3.54	3.21	2.88	7.26	5.39
*5	0.00	0.00	0.69	0.00	0.81
*6	0.00	0.00	0.10	0.00	0.50
*7	0.00	0.00	0.00	0.00	0.35
*8	0.00	0.00	2.08	0.20	2.57
*9	0.00	0.00	2.18	0.10	4.23
*11	0.00	0.00	0.50	0.20	1.46
*14	0.00	0.12	0.00	0.00	0.70
CYP2C19	**KHV**	**EAS**	**AMR**	**EUR**	**Others**
*1	65.66	60.74	70.14	62.82	49.45
*2	28.28	31.98	11.71	14.51	26.64
*3	4.04	5.93	0.10	0.00	0.70
*8	0.00	0.00	0.10	0.30	0.05
*9	0.00	0.00	0.30	0.00	0.55
*13	0.00	0.00	0.89	0.00	0.96
*15	0.00	0.00	0.40	0.00	0.96
*17	2.02	1.36	15.77	22.37	18.53
*34	0.00	0.00	0.00	0.00	0.65
*35	0.00	0.00	0.60	0.00	1.51
CYP2D6	**KHV**	**EAS**	**AMR**	**EUR**	**Others**
*1	19.70	30.12	45.34	38.37	34.34
*2	8.08	9.51	18.55	18.89	20.44
*3	0.00	0.00	0.69	1.89	0.10
*4	0.51	0.12	12.10	18.59	7.80
*6	0.00	0.00	0.30	1.99	0.05

*7	0.00	0.00	0.00	0.00	0.45
*9	0.00	0.00	0.99	2.58	0.00
*10	65.66	54.81	2.78	1.59	5.44
*14	2.53	0.62	0.00	0.00	0.00
*17	0.00	0.00	6.25	0.20	10.98
*28	0.00	0.00	0.10	0.50	0.00
*29	0.00	0.00	2.38	0.00	5.19
*31	0.00	0.00	0.40	0.20	0.00
*33	0.00	0.12	0.20	0.60	0.10
*34	0.00	0.00	0.00	0.00	0.05
*35	0.00	0.00	2.18	5.07	0.30
*39	0.00	0.12	0.20	0.00	0.91
*40	0.00	0.00	0.10	0.00	0.55
*41	2.53	4.07	5.36	9.34	6.65
*43	0.00	0.00	0.50	0.20	1.31
*45	0.00	0.00	0.89	0.00	2.17
*46	0.00	0.00	0.50	0.00	0.35
*71	1.01	0.49	0.00	0.00	0.00
*86	0.00	0.00	0.00	0.00	1.01
*106	0.00	0.00	0.10	0.00	0.81
*111	0.00	0.00	0.00	0.00	0.40
*113	0.00	0.00	0.00	0.00	0.40
*125	0.00	0.00	0.10	0.00	0.20
CYP3A5	**KHV**	**EAS**	**AMR**	**EUR**	**Others**
*1	28.79	28.64	28.47	5.37	44.56
*3	71.21	71.36	63.39	94.33	40.53
*6	0.00	0.00	4.37	0.30	8.86
*7	0.00	0.00	3.77	0.00	6.04

KHV: Kinh Vietnamese population; EAS: East Asian population (without Vietnameses); AMR American population; EUR: European population; Others: Other population included in 1000 Genome Project (African and South Asian populations)

Table S2. Absolute frequency of common alleles of five CYP450 genes in five populations.

		KHV	**EAS**	**AMR**	**EUR**	**Others**
CYP2B6	*1	137	605	492	580	924
	*9	44	173	375	237	751
	*5	1	1	51	111	78
	*2	15	29	31	54	87
	Others	1	2	59	24	144
			χ^2 = 412.2, p-value < 2.2e-16			
		KHV	**EAS**	**AMR**	**EUR**	**Others**
CYP2C9	*1	191	782	844	803	1631

	*3	7	26	29	73	107
	*2	0	1	79	125	35
	Others	0	1	56	5	211
			χ^2 = 469.49, p-value < 2.2e-16			
		KHV	**EAS**	**AMR**	**EUR**	**Others**
CYP2C19	*1	130	492	707	632	981
	*2	56	259	118	146	529
	*17	4	11	159	225	367
	Others	8	48	24	3	107
			χ^2 = 422.29, p-value < 2.2e-16			
		KHV	**EAS**	**AMR**	**EUR**	**Others**
CYP2D6	*1	39	244	457	386	682
	*2	16	77	187	190	406
	*10	130	444	28	16	108
	*4	1	1	122	187	155
	*41	5	33	54	94	132
	*17	0	0	63	2	218
	Others	7	11	97	131	285
			χ^2 = 2231.3, p-value < 2.2e-16			
		KHV	**EAS**	**AMR**	**EUR**	**Others**
CYP3A5	*3	141	578	639	949	805
	*1	57	232	287	54	885
	*6	0	0	44	3	176
	*7	0	0	38	0	120
			χ^2 = 961.02, p-value < 2.2e-16			

KHV: Kinh Vietnamese population; EAS: East Asian population (without Vietnameses); AMR American population; EUR: European population; Others: Other population included in 1000 Genome Project (African and South Asian populations)

Table S3. Absolute phenotypic frequencies of five CYP450 genes in five populations.

2		**KHV**	**EAS**	**AMR**	**EUR**	**Others**
CYP2B6	NM	56	251	169	279	287
	IM	40	124	221	172	460
	PM	2	22	87	27	180
	RM	0	2	7	5	26
	UM	0	0	0	0	2
	Indet	1	0	12	15	3
		KHV	**EAS**	**AMR**	**EUR**	**Others**
CYP2C9	NM	91	371	369	318	738
	IM	6	26	128	169	212
	PM	0	1	3	11	11
	Indet	0	0	0	0	5
		KHV	**EAS**	**AMR**	**EUR**	**Others**

CYP2C19	NM	41	143	257	189	262
	IM	46	207	108	137	330
	PM	9	50	5	6	101
	RM	3	4	102	148	200
	UM	0	0	12	22	40
	Indet	0	0	0	0	13
		KHV	**EAS**	**AMR**	**EUR**	**Others**
CYP2D6	NM	47	242	343	285	657
	IM	50	159	138	180	242
	PM	0	0	15	31	31
	UM	0	0	0	0	0
	Indet	2	4	8	7	77
		KHV	**EAS**	**AMR**	**EUR**	**Others**
CYP3A5	NM	7	33	59	2	220
	IM	43	166	155	50	378
	PM	49	206	257	451	280
	Indet	0	0	0	0	0

KHV: Kinh Vietnamese population; EAS: East Asian population (without Vietnameses); AMR American population; EUR: European population; Others: Other population included in 1000 Genome Project (African and South Asian populations)

NM: Normal metabolizer; IM: Intermediate Metabolizer; PM: Poor Metabolizer; RM: Rapid Metabolizer; UM: Ultrarapid Metabolizer; Indet: Indeterminate