

Genetic Landscape of VIP Pharmacogenomic Variants in the Kinh Vietnamese Population

Hang Tong^{1,2}, Vo Thu Nga Phan^{1,2}, Tu Nguyen^{1,2}, Ly Le^{1,2}

¹School of Biotechnology, International University, Ho Chi Minh City, Vietnam; ²Vietnam National University, Ho Chi Minh City, Vietnam

Correspondence: Ly Le; Hang Tong, Email ly.le@hcmiu.edu.vn; tthang@hcmiu.edu.vn

Background: Genetic polymorphisms in the genes encoding enzymes and proteins in drug metabolism, transport, and response can significantly impact enzymatic activity and overall pharmacokinetics and pharmacodynamics, leading to inter-individual and inter-population differences in drug efficacy and safety. The prevalence of pharmacogenomic variants often differs among populations due to unique evolutionary and demographic factors. By studying the genetic variation within 100 pharmacogenes in the Kinh Vietnamese population, a relatively underexplored group in pharmacogenomic research, we aim to provide insights into the population-specific pharmacogenomic landscape.

Materials and Methods: 100 healthy people were recruited for peripheral blood donation after getting consents. Genomic DNA from participants was sequenced at coding regions of 100 pharmacogenes. Sequence reads were qualified, and variants were called using Genome analysis toolkit (GATK) followed with variant processing. Very important variants were characterized. Genetic distance using pairwise fixation index and allele frequencies comparison between the Kinh population and 25 populations of the 1000 Genome Project were analyzed.

Results: 689 variants were called with 652 SNPs and 37 indels including 371 missense-, 266 synonymous-, 30 frameshift-, 14 stop-gain-, 2 stop-lost-, 3 in-frame insertion-, 2 in-frame deletion- and 1 protein variant. There are 59 novel variants (8.6%) present in 39/100 genes in which 13 variants are labeled with damaging phenotype. 28 VIP variants were obtained from these regions. Allele frequencies of variants are mostly similar with East Asians, but different from Africans. Remarkably, variants rs1801280 and rs1208 (NAT2), rs2231142 (ABCG2), rs2306283 (SCLO1B1) and rs4148323 (UGT1A1) distribute significantly between Kinh people and all other populations.

Conclusion: The prevalence of pharmacogenomic variants of 100 pharmacogenes was obtained for Kinh Vietnamese people, in which 28 variants were characterized as very important variants. Kinh Vietnamese shows close genetic distance with East Asians but far from Africans. The variants with distinguished distribution in Kinh people were also highlighted.

Keywords: pharmacogenomics, VIP variants, panel sequencing

Introduction

Drug-response of a patient is influenced by different physiological factors, in which genetic predepositions have been reported to have association with the metabolism and immunological response of a medicine.¹⁻³ Pharmacogenomics is the new field of genetic medicine that investigates how a patient's genetic makeup affects how the body processes medication. Recent research mostly focuses on the variation of genes encoding drug receptors, metabolizing enzymes, transporters, secretion enzymes and immune response proteins,⁴ with the investigation scale from single variant verification to the whole genome study in large population.⁵ Candidate variants are characterized as biomarkers if they have a strong association with the changes in drug metabolism or immune responses. The impact of these variations could then be validated in different populations to establish the level of importance. Pharmacogenomic Knowledge Base (PharmGKB: <https://pharmgkb.org>)⁶ and Clinical Pharmacogenomic Implementation Consortium (CPIC: <https://cpicpgx.org>)⁷ have summarized and organized the evidences of clinical impact of certain genetic changes, therefore, suggesting the implementation action on drug use. The markers having a high level of clinical

evidence are labeled from level 1–4 according to the clinical and variants annotation. Variants with label 1A-1B, indicating a strong association with particular diseases, should be intensively concerned when using the drugs. These variants are symbolized as very important variants (VIP variants). With less evidence collected, variants are labeled 2A-2B, 3 and 4. The US Food and Drug Administration has listed drugs with warning and caution levels for genetic testing prior to prescription.⁸

Several studies have been conducted to explore the prevalence of very important (VIP) variants in different populations. Using next generation sequencing or variant array, target regions have been analyzed to plot the distribution of variant sets. The study by the Pharmacogenomics Research Network had obtained the population-specific variations from a panel of 84 pharmacogenes including those having prior association with the pharmacogenomic traits at various levels.⁹ In the meta-study of 18 European populations, the shared variants were first selected, then the population-specific markers for each participating population were added in to the panel. The total 1931 variants in 231 genes were all-together analyzed using primary and replicated verification. The outcome indicated differences among populations, however, number and types of actionable markers are mostly found in 9 important pharmacogenes.¹⁰ Studies across several different ethnic groups have created the population reference data which turn into valuable recommendations for drug use.^{11–14}

Vietnam is a multicultural country with 54 ethnic groups in which Kinh people account for approximately 86% of the population. Therefore, most of the pharmacogenetic studies in Vietnam focuses on the Kinh group. Number of genetic markers has been verified in replicated analysis associated with drug toxicity and adverse drug reactions.^{15–18} However, there is the need for overall analysis on risk variation for the references of the population and for disease study. In this project, we study the high impact variants among 100 pharmacogenes which are involved in drug metabolism, transportation, oxidation-reduction and others, and their prevalence in Kinh people. We collected DNA from participants during 2019 and sequenced the coding regions of 100 pharmacogenes using next-generation sequencing. Variants were genotyped following the GATK (Genome Analysis ToolKit) best practice framework. We compared the frequencies and genetic differences of these variants with those in 25 populations from the 1000 Genome Project. This result may provide some information for the development of screening methods on adverse drug reactions as well as personalized medicine in the future.

Materials and Methods

Study Participants

100 healthy people (aged ≥ 18 years) were randomly recruited from the Vietnamese population from 2019 by the Project GEN2019-28-01/HD-KHCN. Participants were provided a questionnaire about Kinh origin of parental lineages in three generations and medical history. People with chronic diseases were excluded. The procedures were designed to comply with the Declaration of Helsinki for ethical standards in biomedical research. The study was ethically approved by Ethical Committee from Vietnam National University, Ho Chi Minh city (Number 2032/DHQG-KHCN). Before sampling, all participants were provided with a brief of the project and blood sampling protocol along with the informed consents.

DNA Extraction and PGx Gene Panel Sequencing

Peripheral blood was collected in EDTA tubes after obtaining the signed informed consents. Genomic DNA from whole blood was extracted using QIAamp DNA Blood Mini Kit (Qiagen, GmbH, Germany) as per manufacturer instructions. The quality and quantity of genomic DNA were analyzed by spectrometry and agarose gel electrophoresis, yielding an average DNA concentration of 50 ng/ μ L and average A260/A280 ratio of 1.8. The panel of 37 drug transporter genes, 30 cytochrome P450 (CYP) enzyme encoded genes, 10 uridine diphosphate glucuronosyltransferase (UGT) genes, 5 flavin-containing monooxygenase (FMO) genes, 4 glutathione S-transferase (GST) genes, 4 sulfotransferase (SULT) genes, and others^{19,20} was selected for the study. Target sequencing of the coding regions was performed using MiSeq Reagent Kit v2 (Illumina, San Diego, CA, USA) using a bench-top next-generation sequencer.

Variant Genotyping

Sequencing reads were processed to remove adapters using TrimmomaticPE.²¹ The reads were then mapped to human reference sequence (build Hg38) using the Burrows-Wheeler Aligner (0.7.17) after qualification. Genome Analysis

Toolkit (GATK 4.2) was then used to perform variant calling followed with variant recalibration (VQSR) to filter false positives. We calculated the allele frequency of all called sites and determined the co-located variants using 1000 genome global minor allele frequency database. For the functional prediction, called variants were aligned to Ensembl/GENCODE transcript database to determine the transcript biotype and phenotype. Sorting Intolerant from Tolerant (SIFT) and Polymorphism Phenotyping (PolyPhen2) were used to predict the deleterious effect of variant on protein function. Other predictions and verification of allele function were analyzed using embedded tools in Functional effect Predictor from Ensembl (www.ensembl.org).

Genetic Distance

Genetic distance between Kinh Vietnamese (VT) and each of 25 populations in 1000 genome project was calculated using pair-wise fixation index (F_{st}). First, bcftools (1.10.2) and vcftools (0.1.16) were used to manipulate variants stored in variant calling format (vcf) files. Then the estimation of Weir and Cockerham F_{st} mean was performed to obtain the genome wide F_{st} at all sites between 2 populations. Downstream variant annotations and plotting were performed using plink (1.9), PGDSpider-3 and R software (R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org).

VIP Variants Selection

VIP variants from 100 pharmacogenes were selected considering their impact on drug metabolism, dosage and toxicity which were supported by CPIC guidelines²² and met the level 1A and 1B criteria from the Pharmacogenomics Knowledge Base [PharmGKB] <https://www.pharmgkb.org>.⁶ We screened the prevalence of all VIP variants of 100 pharmacogenes in our sequence data. Frequency, phenotypic effect associated with drugs used and recommendation were summarized. Variants of world populations were selected within the targeted regions from resequencing data of 2504 participants in the 1000 Genome Project.²³

Statistical Analysis

Distribution of VIP variants in Kinh people (VT) was statistically analyzed together with 25 world populations from the 1000 Genome Project including: CHB (Han Chinese in Beijing, China), CHS (Southern Han Chinese), CDX (Chinese Dai in Xishuangbanna, China), JPT (Japanese in Tokyo, Japan), CEU (Utah residences with Northern and Western European ancestry), TSI (Toscani in Italia), GBR (British in England and Scotland), FIN (Finnish in Finland), IBS (Iberian population in Spain), YRI (Yoruba in Ibadan, Nigeria), LWK (Luhya in Webuye, Kenya), GWD (Gambian in Western Division, Gambia), MSL (Mende in Sierra Leone), ESN (Esan in Nigeria), ASW (African ancestry in Southwest USA), ACB (African Caribbean in Barbados), MXL (Mexican ancestry in Los Angeles, California, USA), PUR (Puerto Rican in Puerto Rico), CLM (Colombian in Medellin, Colombia), PEL (Peruvian in Lima, Peru), GIH (Gujarati Indian in Houston, Texas, USA), PJI (Punjabi in Lahore, Pakistan), BEB (Bengali in Bangladesh), STU (Sri Lankan Tamil in the UK), ITU (Indian Telugu in the UK). We also compared with data of superpopulations from the GenomeAD database including African/African American (AFR), Latino/Admixed American (AMR), East Asian (EAS), Non-Finnish European (NFE) and South Asian (SAS). Pairwise comparison of allele frequencies for each selected variant was performed in Microsoft Excel and IBM SPSS Statistics using Chi squared test or alternatively Fisher exact test when applicable. The p values are all two-sided and $p < 0.000154$ ($0.05/(26C2)$) and $p < 0.005$ ($0.05/10$) are set as significance with Bonferroni correction for the individual population comparison and for superpopulation comparison, respectively.

Results

Characteristics of Study Participants

One hundred healthy people were randomly recruited from Ho Chi Minh City, Vietnam. Participants are from 18 to 47 years old (mean±standard error: 23.6±0.6) with the gender ratio of 1:1. BMI values of participants are from 16.4–29.3 (22.1±0.3), weight from 42–98 kg (59.6±1.2) and height is 1.50–1.83 m (1.64±0.8). Questionnaire reports on personal information and medical history were obtained. Participants have confirmed their origin, their parents and grandparents' origin as Kinh Vietnamese with

confidence. As the research aims to investigate the prevalence of genetic variation in the general population, participants with chronic diseases or using drugs were excluded. All participants signed the informed consents for sample donation.

Variants Genotyping

Panel pharmacogene sequencing from 100 subjects released 11.2 Gbp of raw sequence data in fastq files. The sequence reads were qualified and aligned to the human reference genome assembly hg38. The targeted regions have a read depth of 317x on average, and 98.3% of these regions were covered by at least 20x depth. After recalibration and filtering sites with read depth <20x and p-value less than 1.0×10^{-50} by the Hardy–Weinberg equilibrium test, we obtained 689 variants with 652 SNPs and 37 indels including 371 missense-, 266 synonymous -, 30 frameshift-, 14 stop-gain-, 2 stop-lost-, 3 in-frame insertion-, 2 in-frame deletion- and 1 protein variants. There are 59 novel variants (8.6%) present in 39/100 genes in which 13 variants are labeled with damaging phenotype (Table 1) predicted by two recent prediction tools: Sorting Intolerant from Tolerant (SIFT) and Polymorphism Phenotyping (PolyPhen) embedded in Variant effect Predictor from Ensembl. A total of 689 variants were used for further analysis.

Variation of VIP Pharmacogenomic Variants

Screening was performed to select the very important variants from the coding regions of 100 pharmacogenes. We obtained 29 VIP variants presented in 13 genes including: cytochrome P450s (*CYP2A6*, *CYP2B6*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP4F2*), N-acetyltransferase (*NAT2*), UDP-glucuronosyltransferase (*UGT1A1*), Thiopurine methyltransferase (*TPMT*), Dihydropyrimidine dehydrogenase (*DPYE*), SLC transporter (*SLCO1B1*), ABC transporter (*ABCG2*) and Nudix hydrolase (*NUDT15*). One new variant appears at position chr22:42126660 of chromosome 22 which was previously determined as position of SNP rs1135835 (T>C) of *CYP2D6* gene. However, in this data, there is the deletion of G at chr22:42126661, marking the total change as TG>T. The SNP rs1135835 and deletion variants have not been reported in Kinh Vietnamese before. In addition, the variant at chr22:42126661 was also not observed in other populations, therefore, we excluded it from VIP variant analysis. The characteristics of 28 SNPs are present in Table 2.

Distribution of VIP Variants Among Populations

In 28 important variants from the dataset, we found that variant rs375781227 present in only East Asian with MAF is 0.02%, and rs56337013 is only in South Asian with MAF 0.02%.⁶ There was no corresponding presence of variant in other populations. Therefore, to prevent evaluation bias, we excluded these variants and kept the common sites in the final set of 26 variants for statistical analysis. The pair-wise comparisons have shown that allele frequencies of most VIP variants are not different among East Asian populations including Han Chinese, Dai Chinese and Japanese. However, Kinh people differ significantly from all other populations (Table S1). The largest number of differences is found between Kinh people and Esan people in Nigeria (ESN) and Gambian Western Division in Mandinka (GWD) (10 and 9 differences, respectively). Especially, variants rs1801280 and rs1208 (*NAT2*) distribute significantly differently from all country populations and from selected super populations while rs2231142 (*ABCG2*) differs from all except East Asian. In contrast, variants rs1801160, rs2297595 (*DPYE*), rs1057910 (*CYP2C9*) and rs1809810, rs18192709 (*CYP2A6*) are well conserved. We do not obtain the presence of variants rs72558187 (*CYP2C9*), rs28399447 and rs199916117 (*CYP2A6*) in other populations except in East Asian countries including Kinh people from Vietnam.

Genetic distance between Kinh people and each of 25 populations were also analyzed using pair-wise fixation index (Fst). Mean Fst for each pair of population was calculated from individual Fst values at 29 loci. The lower Fst indicates the stronger genetic relation. Kinh Vietnamese people showed a close genetic distance with east Asian populations including Dai Chinese (Fst 0.005) then Japanese (Fst 0.006) and Han Chinese (CHS Fst 0.0081, CHB Fst 0.0114), but the most distance from African populations consisting of Esan in Nigeria (Fst 0.0638), Luhya in Webuye Kenya (Fst 0.0697) and Gambian in western Division, The Gambia (Fst 0.0704). To visualize the degree of differentiation, we illustrate the matrix of Fst values between populations in the heatmap in Figure 1.

Table 1 Novel Missense Variants with Deleterious Damaging Phenotype from Panel of 100 Pharmacogenes

Location	Allele	GENE	Protein Position	Amino Acid Changes	Codons	SIFT (Score) ^a	PolyPhen (Score) ^a
1:109689097–109689097	C	GSTM1	76	I/T	aTc/aCc	Deleterious_low_confidence(0.01)	Probably_damaging(1)
1:109689228–109689228	A	GSTM1	88	G/E	gGg/gAg	Deleterious_low_confidence(0.02)	Probably_damaging(0.999)
2:38074670–38074670	G	CYP1B1	240	L/P	cTg/cCg	Deleterious(0)	Probably_damaging(1)
2:233682025–233682025	A	UGT1A7	30	V/I	Gta/Ata	Deleterious_low_confidence(0.01)	Probably_damaging(0.982)
4:69096575–69096575	T	UGT2B7	19	S/C	Agc/Tgc	Deleterious_low_confidence(0)	Possibly_damaging(0.9)
10:71362456–71362456	G	SLC29A3	426	N/D	Aac/Gac	Deleterious(0)	Probably_damaging(0.991)
10:95037292–95037292	G	CYP2C8	437	G/R	Gga/Cga	Deleterious_low_confidence(0)	Probably_damaging(0.998)
11:63373715–63373715	G	SLC22A9	193	A/G	gCt/gGt	Deleterious(0)	Probably_damaging(0.974)
12:21176778–21176778	G	SLCO1B1	121	Y/C	tAc/tGc	Deleterious(0)	Probably_damaging(1)
16:16068157–16068157	C	ABCC1	560	V/A	gTg/gCg	Deleterious_low_confidence(0)	Probably_damaging(0.999)
16:16068172–16068172	C	ABCC1	565	F/S	tTt/tCt	Deleterious_low_confidence(0)	Probably_damaging(0.999)
19:15673578–15673578	C	CYP4F12	17	S/P	Tcc/Ccc	Deleterious_low_confidence(0.01)	Probably_damaging(0.958)
19:41016712–41016712	G	CYP2B6	454	L/R	cTc/cGc	Deleterious(0)	Probably_damaging(0.995)

Notes: ^aSIFT and Polyphen cores from 0–1, lower SIFT score (closed to 0) and higher Polyphen score (closed to 1) indicate the significant prediction of phenotype.

Table 2 Characteristics of 28 VIP Variants in 100 Pharmacogenes of Kinh Vietnamese People

CHR	GENE	POS	RS	ALLELE	REF	ALT	AF	Column I	TYPE	Drugs ^a	Phenotypes ^b
1	DPYE	97305364	rs1801160		C	T	C:0.98	T:0.02	SNP	Fluorouracil	Toxicity
		97450068	rs17376848		A	G	A:0.843434	G:0.156566	SNP	Fluorouracil	Other
		97515839	rs1801159		T	C	T:0.777778	C:0.222222	SNP	Capecitabine	Toxicity
		97699535	rs2297595		T	C	T:0.974747	C:0.0252525	SNP	Fluorouracil	Toxicity
		97883329	rs1801265		A	G	A:0.833333	G:0.166667	SNP	Fluorouracil	Toxicity
2	UGT1A1	233760498	rs4148323	*6	G	A	G:0.875	A:0.125	SNP	Atazanavir, atazanavir / ritonavir/FOLFIRI, irinotecan	Toxicity
4	ABCG2	88131171	rs2231142		G	T	G:0.641414	T:0.358586	SNP	Rosuvastatin	Toxicity
6	TPMT	18130687	rs1142345		T	C	T:0.975	C:0.025	SNP	Thioguanin, mercaptopurine, azathioprine	Toxicity, dosage, metabolism
8	NAT2	18400344	rs1801280		T	C	T:0.924242	C:0.0757576	SNP	Drugs for treatment of tuberculosis, isoniazid	Toxicity
		18400593	rs1799930		G	A	G:0.64	A:0.36	SNP	Drugs for treatment of tuberculosis, isoniazid	Toxicity
		18400806	rs1208		G	A	G:0.0909091	A:0.909091	SNP	Drugs for treatment of tuberculosis, isoniazid	Toxicity
		18400860	rs1799931		G	A	G:0.853535	A:0.146465	SNP	Drugs for treatment of tuberculosis, isoniazid	Toxicity
10	CYP2C19	94780653	rs4986893	*19	G	A	G:0.96	A:0.04	SNP	Imipramine, dexlansoprazole, voriconazole, clopidogrel, lansoprazole, pantoprazole, sertraline, escitalopram, clomipramine, amitriptyline	Metabolism, PK
										Clopidogrel, citalopram, omeprazole, pantoprazole	Toxicity, Efficacy
		94781859	rs4244285	*2	G	A	G:0.75	A:0.25	SNP	Voriconazole, lansoprazole, pantoprazole, citalopram, sertraline, doxepin, escitalopram, clomipramine, amitriptyline, imipramine, dexlansoprazole, clopidogrel, trimipramine, omeprazole	Metabolism, PK
										Clopidogrel, pantoprazole, omeprazole, lansoprazole,	Toxicity, Efficacy
		94781944	rs375781227	*26	G	A	G:0.995	A:0.005	SNP	Clopidogrel, omeprazole	Metabolism, PK
		94842866	rs3758581		A	G	A:0.0252525	G:0.974747	SNP	Clopidogrel, citalopram, omeprazole, pantoprazole, lansoprazole	Toxicity, Efficacy
		94852738	rs56337013	*5	C	T	C:0.995	T:0.005	SNP	Clopidogrel	Toxicity, Efficacy
	CYP2C9	94941958	rs72558187	*13	T	C	T:0.995	C:0.005	SNP	Phenytoin, meloxicam, celecoxib, lornoxicam,	Metabolism, PK
		94981296	rs1057910	*3	A	C	A:0.974747	C:0.0252525	SNP	Piroxicam, phenytoin, lornoxicam, bupropfen, tenoxicam, celecoxib, flurbiprofen, fluvastatin, meloxicam, siponimod	Metabolism, PK
										Fluvastatin, warfarin, phenytoin, acenocoumarol	Toxicity, Efficacy
										Warfarin, acenocoumarol	Dosage

12	SLCO1B1	21176804	rs2306283		A	G	A:0.272727	G:0.727273	SNP	Simvastatin, simvastatin acid, rosuvastatin, atorvastatin, fluvastatin, pitavastatin, pravastatin	Metabolism, PK
										Lovastatin, fluvastatin, rosuvastatin, atorvastatin, simvastatin, pravastatin	Toxicity, Efficacy
		21178615	rs4149056		T	C	T:0.905	C:0.095	SNP	Atorvastatin, simvastatin, simvastatin acid, pitavastatin, fluvastatin, rosuvastatin, pravastatin	Metabolism, PK
										Rosuvastatin, fluvastatin, atorvastatin, pravastatin, lovastatin, simvastatin,	Toxicity, Efficacy
13	NUDT15	48045719	rs116855232	*2, *3	C	T	C:0.954545	T:0.0454545	SNP	Mercaptopurine, azathioprine	Toxicity, Dosage
19	CYP4F2	15879621	rs2108622		C	T	C:0.835	T:0.165	SNP	Warfarin	Dosage
	CYP2A6	40844759	rs1809810		T	A	T:0.979592	A:0.0204082	SNP	Nicotine	Metabolism
		40847036	rs28399447	*11	A	G	A:0.985	G:0.015	SNP	Nicotine	Metabolism
		40848293	rs199916117	*15	T	C	T:0.994949	C:0.00505051	SNP	Nicotine	Metabolism
	CYP2B6	40991369	rs8192709	*2	C	T	C:0.944444	T:0.0555556	SNP	Efavirenz	Toxicity
		41006936	rs3745274		G	T	G:0.767677	T:0.232323	SNP	Efavirenz, sertraline,	Metabolism
										Efavirenz	Toxicity, Dosage
22	CYP2D6	42126660	NA		TG	T	TG:0.993151	T:0.00684932	INDEL	Position reported with rs1135835 but this variant is not available in VN genome	

Notes: ^{a,b} Associated drugs and their impact on drug response are obtained from CPIC and PharmGKB. * allele determined in haplotype database.

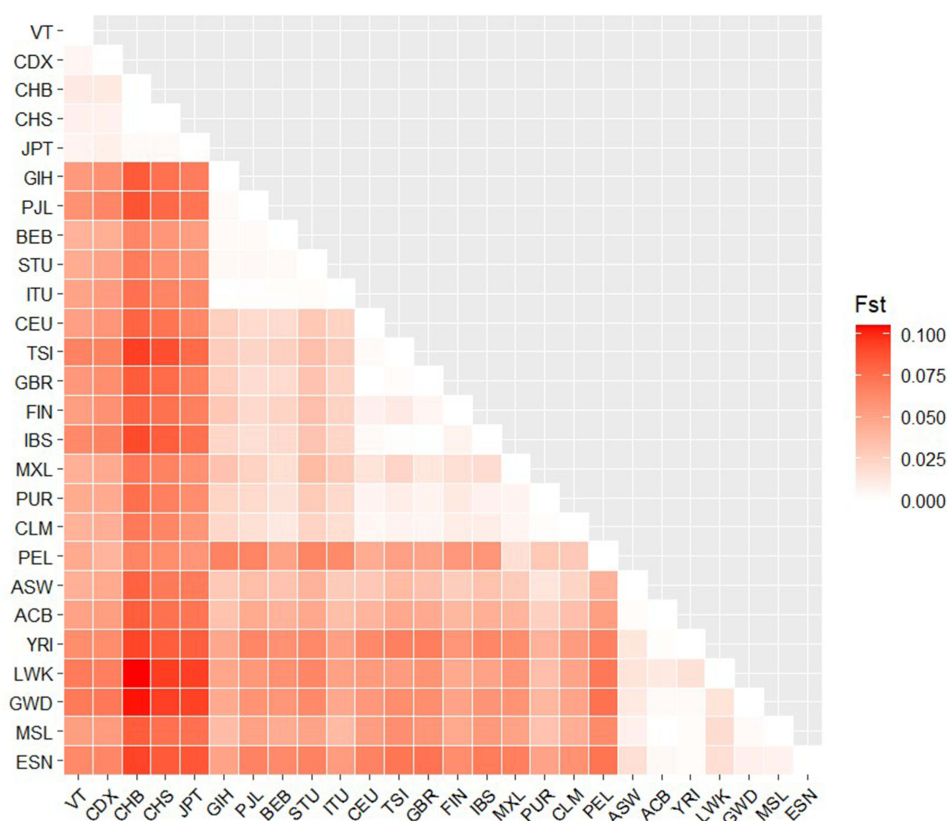


Figure 1 Heatmap of pair-wise F_{st} values between Kinh Vietnamese people (VT) and 25 populations from the 1000 Genome Project. Lighter color indicates the lower genetic distance.

Abbreviations: CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese; JPT, Japanese in Tokyo, Japan; GIH, Gujarati Indian in Houston, Texas, USA; PJL, Punjabi in Lahore, Pakistan; BEB, Bengali in Bangladesh; STU, Srilankan Tamil in the UK; ITU, Indian Telugu in Tthe UK; CEU, Utah residences with Northern and Western European ancestry; TSI, Toscani in Italia; GBR, British in England and Scotland; FIN, Finnish in Finland; IBS, Iberian population in Spain; MXL, Mexican ancestry in Los Angeles, California, USA; PUR, Puerto Rican in Puerto Rico; CLM, Colombian in Medellin, Colombia; PEL, Peruvian in Lima, Peru; ASW, African ancestry in Southwest USA; ACB, African Caribbean in Barbados; YRI, Yoruba in Ibadan, Nigeria; LWK, Luhya in Webuye, Kenya; GWD, Gambian in Western Division, Gambia; MSL, Mende in Sierra Leone; ESN, Esan in Nigeria.

Discussion

With the increasing of adverse drug events reported yearly in pharmacovigilance programs, exploring genetic markers associated with certain adverse reactions is crucial for preventing and predicting these outcomes of drug consumption. The Kinh ethnics (VT) is the major group in Vietnam with more than 86% of the population. However, this group is relatively underexplored in pharmacogenomic research. In this study, we investigated the variation in well-presented genes involved in several pathways of drug metabolism, transport, and secretion in the general Kinh population. Variants present in the coding regions could effectively alter enzymatic activity, protein expression, the pharmacokinetics and pharmacodynamics, which finally lead to inter-individual and inter-population differences in drug efficacy and safety. Distribution of the variants is populationspecific. We found some variants presented in this dataset but not obtained in the previously released Vietnamese variation database (genomes.vn, data not shown), suggesting that more extensive research for Kinh people should be conducted in overall country. Genotypic frequencies of 28 very important pharmacogenetic (VIP) variants were similar between VT and CDX, CHB, CHS and JPT. In addition, pair-wise F_{st} values indicate that VIP variants in VT have a close relationship with above-mentioned populations. We highlighted the variants with significant differences in VT including *rs1801280*, *rs1799931* and *rs1208* (*NAT2*), *rs2231142* (*ABCG2*), *rs17376848* (*DPYE*), *rs4148323* (*UGT1A1*) and *r2306283* (*SLCO1B1*) for further discussion of their impact on enzyme activity and drug metabolism.

N-acetyltransferase (*NAT2*) enzyme bio-transforms several drugs and chemicals with primary aromatic amine or hydrazine groups. Mutations in this gene were reported as common and significant effects have been observed in drug

acetylation.²⁴ Variants *rs1801280* (T>C), *rs1208* (G>A), *rs1799930* (G>A) and *rs1799931* (G>A) lead to the slow acetylation of medicines and are strongly linked with metabolism of drugs for treatments of tuberculosis and isoniazid. Slow acetylators (slow acetylation carriers) are associated with increased risk of hepatotoxicity due to anti-TB drugs²⁵ in Brazilian people. *Rs1208* was found associated with phenytoin intoxication on isoniazid prescription.²⁶ Therefore, the use of isoniazid for the treatment of tuberculosis should be carefully considered in patients carrying *rs1801280* (T>C), *rs1799930* (G>A) and *rs1799931* (G>A). Frequencies of these variants are similar with that in East Asians reported in genomeAD database, but significantly different from the rest of the world including African, European, South Asian and American. Similar distribution of certain variants should suggest the same potential in enzyme activity, hence conversion and drug metabolism. Allele frequency in different Kinh group samples is not significantly different²⁷ in which disease variants presented in the population with frequencies from 3–9%.

The other high variable SNP from our data is *rs2231142* of *ABCG2* gene which encodes ATP Binding Cassette Subfamily G Member 2 enzyme. The enzyme is highly involved in excretion of drug metabolism products. Reduced enzyme activity might lead to the accumulation of metabolites and cause cellular toxicity. The locus is multiallelic with either G or C or T, in which G is normal allele. We observed in Kinh people only conversion G>T with allele frequency (T) of 35.9%. Allele T is associated with increased plasma concentrations of rosuvastatin as compared with allele G. Genotypes GT + TT decrease transporter efflux function, therefore associated with increased exposure to rosuvastatin in people with diabetes mellitus and hypercholesterolemia as compared with genotype GG.²⁸ The variant is strongly associated with metabolism, dosage, and toxicity of rosuvastatin treatments in Chinese people.²⁹ In addition, the variant affects the efficacy and dosage of allopurinol. *Rs2231142* presented with high frequencies in East Asian (30.97%) compared with other population (3–14%), therefore, suggestion for implementation of genetic testing is recommended.

In our dataset, thiopurine methyltransferase (TPMT) has one VIP variant *rs1142345*. This enzyme methylates thiopurine compounds using S-adenosyl-L-methionine as methyl donor. TPMT participates actively in metabolism of several drugs such as azathioprine, mercaptopurine and thioguanine. The SNP *rs1142345* (T>C) is present in inactivating alleles *3A and *3B making enzyme carriers as intermediate metabolizer-IM (*1/*3) or poor metabolizer-PM (*3/*3). IM and PM are significantly associated with hepatotoxicity following azathioprine therapy.³⁰ The variant is important for categorizing patient groups prior to azathioprine use, therefore pre-emptive genetic testing was suggested. Cheng reported high variation of *rs1142345* (TPMT) between Dai population from Yunnan and Chinese Dai in Xishuangbanna, China (CDX).¹³ The variant frequency between Dai from Yunnan is also different from all other continental populations in the 1000 genome project. The study particularly noticed the high variation of this variant within Chinese population including Han Chinese in Beijing and in the South (CHB, CHS). However, in our study, we obtained no difference in genetic variation among Kinh people (including KHV from 1000 genome data, Le's database²⁷ and this study) and between Kinh populations (VT) and other populations in the world. The result suggests the homogeneity of this variant in different samplings from Kinh groups and could imply the well-conserved enzyme activity in the population.

Very important variants of 100 pharmacogenes have been analyzed to establish the overall variation for further investigation and application in Kinh people. While most of the variants show the relation with that in East Asians which may correspond to the similar proportions of enzymatic activity in the population and similar drug metabolism, some variants present with particular frequencies. We noted all conserved and variable distributions to add the discussion for how the variant might be more investigated toward implementation. Our data may be the source of reference for further pharmacogenomic study in Kinh people. However, there are limitations in the study due to the small sample size. We would suggest extended research with larger sample sizes to capture more variants with smaller MAF, as well as including more pharmacogenes in the analysis such as HLA genes.

Conclusion

We identified 689 variants in the coding regions of 100 pharmacogenes in the Kinh population. Among these, 59 are novel alleles, with 13 predicted to have damaging effects. From this dataset, 28 key variants were highlighted. Kinh people have close genetic distance with Dai Chinese, Han Chinese and Japanese, but far distance from African populations. The distribution of VIP variants was compared across 25 other populations, revealing significant differences at loci *rs4148323* (UGT1A1), *rs2231142* (ABCG2), *rs1801280*, *rs1208*, and *rs1799931* (NAT2), as well as *rs2306283*

(SLCO1B1). These findings provide valuable insights into genetic differentiation and hold potential for advancing personalized medicine tailored to the Kinh population in Vietnam.

Abbreviations

VIP, very important pharmacogenomic; PharmGKB, Pharmacogenomic Knowledge Base; CPIC, Clinical Pharmacogenomic Implementation Consortium; PGx, pharmacogenomic; EDTA, Ethylenediaminetetraacetic acid; CYP, cytochrome P450; UGT, uridine diphosphate glucuronosyltransferase; FMO, 5 flavin-containing monooxygenase; GST, 4 glutathione S-transferase; SULT, 4 sulfotransferase; Hg38, Homo sapiens (human) genome assembly GRCh38; GATK, Genome Analysis Toolkits; VQSR, variant quality score recalibration; SIFT, Sorting Intolerant from Tolerant; Polyphen, Polymorphism Phenotyping; BMI, body mass index; SNP, single nucleotide polymorphism; ATP: adenosine triphosphate; G: guanine; T, thymine; C, cytosine.

Acknowledgments

We thank Dr. Koya Fukunaga and Dr. Taishi Mushiroda at Laboratory for Pharmacogenomics, Riken Center for Integrative Medical Sciences (Yokohama, Japan) for providing sequencing service under SEAPharm project. We also thank all participants for their sample donation.

Disclosure

The authors report no conflicts of interest in this work.

References

- Gaebler AJ, Schneider KL, Stingl JC, Paulzen M. Subtherapeutic bupropion and hydroxybupropion serum concentrations in a patient with CYP2C19*1/*17 genotype suggesting a rapid metabolizer status. *Pharmacogenomics J.* 2020;20(6):840–844. doi:10.1038/s41397-020-0169-y
- Kwan PK, Ng MH, Lo SV. Association between HLA-B*15:02 allele and antiepileptic drug-induced severe cutaneous reactions in Hong Kong Chinese: a population-based study. *Hong Kong Med J.* 2014;20(Suppl 7):16–18.
- Man CB, Kwan P, Baum L, et al. Association between HLA-B*1502 allele and antiepileptic drug-induced cutaneous reactions in Han Chinese. *Epilepsia.* 2007;48(5):1015–1018. doi:10.1111/j.1528-1167.2007.01022.x
- Cecchin E, Stocco G. Pharmacogenomics and Personalized Medicine. *Genes.* 2020;11(6):679. doi:10.3390/genes11060679
- Tong H, Phan NVT, Nguyen TT, Nguyen DV, Vo NS, Le L. Review on Databases and Bioinformatic Approaches on Pharmacogenomics of Adverse Drug Reactions. *Pharmgenomics Pers Med.* 2021;14:61–75. doi:10.2147/PGPM.S290781
- PharmGKB. Pharmacogenomic knowledge Base. Available from: <https://www.pharmgkb.org/>. Accessed August 23, 2025.
- CPIC. Clinical Pharmacogenomic Implementation Consortium. Available from <https://cpicpgx.org/>. Accessed August 23, 2025.
- FDA UFA DA. *Table of Pharmacogenomic Biomarkers in Drug Labeling.* 2016.
- Gordon AS, Fulton RS, Qin X, Mardis ER, Nickerson DA, Scherer S. PGRNseq: a targeted capture sequencing panel for pharmacogenetic research and implementation. *Pharmacogenet Genomics.* 2016;26(4):161–168. doi:10.1097/FPC.0000000000000202
- Mizzi C, Dalabira E, Kumuthini J, et al. A European Spectrum of Pharmacogenomic Biomarkers: implications for Clinical Pharmacogenomics. *PLoS One.* 2016;11(9):e0162866. doi:10.1371/journal.pone.0162866
- Gonzalez-Covarrubias V, Morales-Franco M, Cruz-Correa OF, et al. Variation in Actionable Pharmacogenetic Markers in Natives and Mestizos From Mexico. *Front Pharmacol.* 2019;10:1169. doi:10.3389/fphar.2019.01169
- Wen YF, Culhane-Pera KA, Thyagarajan B, et al. Potential Clinical Relevance of Differences in Allele Frequencies Found within Very Important Pharmacogenes between Hmong and East Asian Populations. *Pharmacotherapy.* 2020;40(2):142–152. doi:10.1002/phar.2360
- Cheng Y, Dai R, Chen W, Li Q, Zhang C, Yang T. Genetic polymorphisms of pharmacogenomic VIP variants in the Dai population from Yunnan province. *Mol Genet Genomic Med.* 2020;8(7):e1231. doi:10.1002/mgg3.1231
- Tafazoli A, van der Lee M, Swen JJ, et al. Development of an extensive workflow for comprehensive clinical pharmacogenomic profiling: lessons from a pilot study on 100 whole exome sequencing data. *Pharmacogenomics J.* 2022;22(5–6):276–283. doi:10.1038/s41397-022-00286-4
- Veiga MI, Asimus S, Ferreira PE, et al. Pharmacogenomics of CYP2A6, CYP2B6, CYP2C19, CYP2D6, CYP3A4, CYP3A5 and MDR1 in Vietnam. *Eur J Clin Pharmacol.* 2009;65(4):355–363. doi:10.1007/s00228-008-0573-8
- Nguyen TVA, Le BH, Nguyen MT, et al. Pharmacogenomic Analysis of CYP3A5*3 and Tacrolimus Trough Concentrations in Vietnamese Renal Transplant Outcomes. *Pharmgenomics Pers Med.* 2024;17:53–64. doi:10.2147/PGPM.S439400
- van Nguyen D, Chu HC, Vidal C, et al. Genetic susceptibilities and prediction modeling of carbamazepine and allopurinol-induced severe cutaneous adverse reactions in Vietnamese. *Pharmacogenomics.* 2021;22(1):1–12. doi:10.2217/pgs-2019-0146
- Vu NP, Tth M, Tran NTB, et al. Polymorphic analysis of CYP2C9 gene in Vietnamese population. *Mol Biol Rep.* 2018;45(5):893–900. doi:10.1007/s11033-018-4235-3
- Runchaen C, Fukunaga K, Sensorn I, et al. Prevalence of pharmacogenomic variants in 100 pharmacogenes among Southeast Asian populations under the collaboration of the Southeast Asian Pharmacogenomics Research Network (SEAPharm). *Hum Genome Var.* 2021;8(1):7. doi:10.1038/s41439-021-00135-z
- Yoshihama T, Fukunaga K, Hirasawa A, et al. GSTP1 rs1695 is associated with both hematological toxicity and prognosis of ovarian cancer treated with paclitaxel plus carboplatin combination chemotherapy: a comprehensive analysis using targeted resequencing of 100 pharmacogenes. *Oncotarget.* 2018;9(51):29789–29800. doi:10.18632/oncotarget.25712

21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170
22. CPIC. Guidelines on gene-drug use. Available from: <https://cpicpgx.org/genes-drugs/>. Accessed Aug 23, 2025.
23. Byrska-Bishop M, Evani US, Zhao X, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell*. 2022;185(18):3426–3440e3419. doi:10.1016/j.cell.2022.08.004
24. Furet Y, Bechtel Y, Le Guellec C, Bechtel PR, Autret-Leca E, Piantaud G. Clinical relevance of N-acetyltransferase type 2 (NAT2) genetic polymorphism. *Therapie*. 2002;57(5):427–431.
25. Possuelo LG, Castelan JA, de Brito TC, et al. Association of slow N-acetyltransferase 2 profile and anti-TB drug-induced hepatotoxicity in patients from Southern Brazil. *Eur J Clin Pharmacol*. 2008;64(7):673–681. doi:10.1007/s00228-008-0484-8
26. Adole PS, Kharbanda PS, Sharma S. N-acetyltransferase 2 (NAT2) gene polymorphism as a predisposing factor for phenytoin intoxication in tuberculous meningitis or tuberculoma patients having seizures - A pilot study. *Indian J Med Res*. 2016;143(5):581–590. doi:10.4103/0971-5916.187106
27. Le VS, Tran KT, Bui HTP, et al. A Vietnamese human genetic variation database. *Hum Mutat*. 2019;40(10):1664–1675. doi:10.1002/humu.23835
28. Zhang D, Ding Y, Wang X, et al. Effects of ABCG2 and SLCO1B1 gene variants on inflammation markers in patients with hypercholesterolemia and diabetes mellitus treated with rosuvastatin. *Eur J Clin Pharmacol*. 2020;76(7):939–946. doi:10.1007/s00228-020-02882-4
29. Zhou Q, Ruan ZR, Yuan H, Xu DH, Zeng S. ABCB1 gene polymorphisms, ABCB1 haplotypes and ABCG2 c.421c > A are determinants of inter-subject variability in rosuvastatin pharmacokinetics. *Pharmazie*. 2013;68(2):129–134.
30. Sheu HS, Chen YM, Liao YJ, et al. Thiopurine S-Methyltransferase Polymorphisms Predict Hepatotoxicity in Azathioprine-Treated Patients with Autoimmune Diseases. *J Pers Med*. 2022;12(9):1399. doi:10.3390/jpm12091399

Pharmacogenomics and Personalized Medicine

Publish your work in this journal

Pharmacogenomics and Personalized Medicine is an international, peer-reviewed, open access journal characterizing the influence of genotype on pharmacology leading to the development of personalized treatment programs and individualized drug selection for improved safety, efficacy and sustainability. This journal is indexed on the American Chemical Society's Chemical Abstracts Service (CAS). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/pharmacogenomics-and-personalized-medicine-journal>

Dovepress
Taylor & Francis Group