Observational Study

Medicine®

OPEN

Single nucleotide and structural variants of *CYP2D6* gene in Kinh Vietnamese population

Ha Hai Nguyen, PhD[a,b], Thuong Thi Huyen Ma, MSc[a], Nhung Phuong Vu, MSc[a,b], Quynh Thi Nhu Bach, PhD[c], Thang Hong Vu, MD[d], Ton Dang Nguyen, PhD[a,b], Hai Van Nong, PhD[a,b,*]

Abstract

CYP2D6 genetic variations could result in alteration of CYP2D6 enzyme activity, leading to dissimilarity among individuals in regard of drug metabolism.

This study aims to detect all genetic variants, allele, and genotype frequencies of *CYP2D6* gene in 136 unrelated healthy Kinh Vietnamese volunteers. All single nucleotide variants (SNVs) and structural variations (SVs) of *CYP2D6* gene were identified by Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA) assay.

Totally, 30 SNVs and 9 SVs including a whole gene deletion, 8 hybrid structures, and tandem arrangements were identified. Of the 7 novel SNVs detected, the 3157G>T (R329L) substitution was predicted to be deleterious by PROVEAN; the 3851G>A (W358X) variant resulted in a truncated protein; and the 2988G>A variant located in the intron 6 was predicted to be capable of modifying splicing motif by Human Splicing Finder. We determined 29 different genotypes of *CYP2D6* from 136 individuals. The most common alleles were the *CYP2D6**10 (43.75%), *1 (18.75%), and tandem arrangement *36-*10 (12.13%).

This study provides best information on *CYP2D6* polymorphism comprising the newly discovered SNVs, structural variations, and their frequencies in Kinh Vietnamese. These new data would be valuable in view of precise dosing of CYP2D6 metabolized drugs and giving better treatment outcome.

Abbreviations: EM = extensive metabolizer, IM = intermediate metabolizer, LD = linkage disequilibrium, LR-PCR = Longrange PCR, MLPA = multiplex ligation-dependent probe amplification, PM = poor metabolizer, SNVs = single nucleotide variants, SVs = structural variations, UM = ultrarapid metabolizer.

Keywords: *CYP2D6* gene, drug metabolism, Kinh Vietnamese, single nucleotide variants, structural variations

1. Introduction

The *CYP2D6* gene encodes CYP2D6 enzyme that is involved in metabolism of approximately 25% commonly prescribed drugs on the market.[1] Polymorphism of *CYP2D6* can lead to inter-individual difference in clearance of many drugs such as: antidepressant, anticancer, antipsychotics, anti-arrhythmic, and opioid. According to Pharmacogene Variation Consortium (https://www.pharmvar.org), there are over 100 allelic variants of *CYP2D6* have been designated, including single nucleotide polymorphisms (SNPs) and insertions/deletions of a small number of nucleotides. In addition to single nucleotide variants (SNVs), *CYP2D6* has been shown to have SVs consisting of copy number variations (CNVs), conversions (structures carry a small region derived from *CYP2D7*), and structural rearrangements (hybrid genes and tandems) with *CYP2D7* pseudogene. The phenotypes of CYP2D6 were classified according to enzyme activation: poor metabolizers (PMs), intermediate metabolizers (IMs), extensive metabolizers (EMs), and ultrarapid metabolizers (UMs). CYP2D6 enzyme activation alteration can result in variability of drug response, in which PM and UM phenotypes would be at risk of adverse effect or therapeutic failure. There is a wide variability in distribution of *CYP2D6* variants among different ethnic groups all over the world. The PMs (*CYP2D6**3, *4 and *5) were mainly found in Europe whereas percentages of these variants in Asia were very low (0.2%–0.7%).[2] The UMs presented in North Africa and Oceania while IMs primarily located in Asia due to high frequency of *CYP2D6**10.[3] Although *CYP2D6* allele frequencies have been evaluated in many ethnic groups, there was limited information of SVs discrimination of this gene.[4,5] To date, *CYP2D6* genetic polymorphisms in Vietnamese have been reported by 3 studies. Using pyrosequencing, Veiga et al[6] genotyped 78 Vietnamese and obtained frequencies for 4 *CYP2D6* alleles *1, *4, *5, and *10. Kim et al[7] applied multiplex single-base extension to identify 12 *CYP2D6* SNPs and a duplication in 758 Korean, 89 Chinese, and 122 Vietnamese. Love et al[8] used TaqMan genotyping assay to detect frequencies of 8 *CYP2D6* alleles in 138 Filipino and 86 Vietnamese women for tamoxifen therapy in breast cancer. Taken together, these works reported the existence and frequencies of only 14 well-known alleles and lacked

Editor: Nikhil Jain.

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106-YS.02-2014.30.

All authors have no conflicts of interest.

Supplemental Digital Content is available for this article.

[a] Institute of Genome Research, [b] Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Hanoi, [c] Haiphong University of Medicine and Pharmacy, Ngo Quyen, Hai Phong, [d] Hanoi Medical University, Dong Da, Hanoi, Vietnam.

[*] Correspondence: Hai Van Nong, Institute of Genome Research, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam (e-mail: vhnong@igr.ac.vn).

Medicine (2019) 98:22(e15891)

Received: 30 November 2018 / Received in final form: 19 April 2019 / Accepted: 30 April 2019

http://dx.doi.org/10.1097/MD.0000000000015891

information of other known alleles, especially possible novel variants in the whole *CYP2D6* gene for the Kinh Vietnamese population. Therefore, this study aimed to detect all variants of *CYP2D6* gene and their frequencies in the Kinh Vietnamese by Sanger sequencing combined with multiplex ligation-dependent probe amplification (MLPA) assay. The obtained data would provide further understanding of *CYP2D6* pharmacogenetics for Vietnamese population.

2. Materials and methods

2.1. Subjects and DNA extraction

One hundred and thirty six of unrelated healthy Kinh volunteers (85 women and 51 men) from Hanoi Medical University of Vietnam were recruited. All subjects were regarded as healthy according to medical history and physical examination. The study purpose was explained to all individuals and the informed consent was obtained from each study subject before sample collection. This project was approved by the Institutional Review Board (IRB) of the Institute of Genome Research, Vietnam Academy of Science and Technology. For all subjects, 2 mL of whole blood was collected and preserved in ethylenediaminetetraacetic acid containing tubes. Genomic DNA was extracted from the blood samples using E.Z.N.A Blood DNA Mini Kit (Omega-Biotek, Norcross, Georgia, USA) according to manufacturer's protocol.

2.2. DNA sequencing

Sanger sequencing method was used to identify nucleotide variants in the promoter, entire 9 coding exons and their flanking regions of *CYP2D6* gene. Primers provided by PHUSA Biochem Company (Can Tho, Vietnam) were listed in Supplementary Table 1, http://links.lww.com/MD/D12. Polymerase chain reaction (PCR) reaction was performed with a total volume of 20 μL containing: 10 ng total genomic DNA, 1× DreamTaq Master mix (Thermo Fisher Scientific, Waltham, Massachusetts), 10 pmole each primer, 1 μL DMSO (Bioworld, Dublin, Ohio), and 16.5 μL deionized water. The thermo cycle was as following: denaturation at 95 °C for 5 minutes, following by 40 cycles of 95 °C for 15 seconds, 58 °C for 30 seconds, 72 °C for 40 seconds to 2 minutes, and a final extension at 72 °C for 5 minutes. For exon 5 to exon 8, the thermo cycle was as following: denaturation at 95 °C for 2 minutes, following by 35 cycles of 95 °C for 30 seconds, 59 °C for 30 seconds, 72 °C for 2 minutes, and a final extension at 72 °C for 5 minutes. PCR products were purified using Multiscreen PCR 96 Filter Plate (Merck-Millipore, Burlington, Massachusetts) and were then sequenced using ABI Prism BigDye Terminator Cycle Sequencing Kit Version 3.1 (Applied Biosystems), on an ABI genetic analyzer 3500 (Applied Biosystems).

2.3. Multiplex ligation-dependent probe amplification (MLPA) assay

To identify deletions/duplications on *CYP2D6* gene, MPLA technique was performed using the commercial SALSA MLPA P128-C1 Cytochrome P450 Probemix kit (MRC-Holland, Amsterdam, Netherlands) following the manufacturer's protocol and collected data were analyzed by Coffalyzer.net software. The MLPA probes were specifically designed to amplify exons 1, 5, 6, and 9. For each genomic DNA sample, 50 ng was used for this assay.

2.4. Longrange (LR) PCR

Primers used in LR-PCR were described in previous study.[9] For detection of *CYP2D6*5*, a 25 μL long PCR reaction was performed. The reaction mixture consisted of: 100 ng template DNA, 1X Gotaq Long Mastermix (Promega, Fitchburg, Wisconsin), 0.16 μM of each primer Dup and Dlow, 0.32 μM of each primer DPKup and DPKlow,10.5 μL double-distilled water. The thermo cycle was as following: denaturation at 94 °C for 2 minutes, following by 35 cycles of 94 °C for 20 seconds, 70 ° C for 5 minutes, and a final extension at 72 °C for 10 minutes. All amplified fragments were subsequently analyzed on 0.8% agarose gel.

2.5. Data analysis

Bioedit software was used for initial analysis of the sequences. The *CYP2D6* variants were named based on the nucleotide reference sequence (NCBI reference sequence: M33388.1) and CYP Allele Nomenclature (https://www.pharmvar.org/). Variants not reported in database of CYP Allele Nomenclature Committee were subsequently searched in NCBI SNP database (dbSNP) and 1000 Genomes Data (http://www.internationalgenome.org) were assigned as novel variants. Assessment of Linkage Disequilibrium (LD) for SNP variants was accomplished using HAPLOVIEW 4.2 (https://www.broadinstitute.org/haploview/version-42-15-september-2009).

2.6. Statistical analysis

Inter-ethnic comparison of *CYP2D6* allele frequencies and difference of allele frequencies between Kinh Vietnamese and global populations were assessed by Chi-square (χ^2) and Fisher exact test. $P < .05$ was considered as statistically significant.

2.7. In silico analysis of novel variants

The online tools as MATCH, PROVEAN (http://provean.jcvi.org/index.php)/Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/), and Human Splicing Finder (http://www.umd.be/HSF3/) were used to search potential transcriptional factor binding sites in *CYP2D6* promoter,[10] predict possible impact of non-synonymous SNPs in the coding regions of *CYP2D6*[11,12] and predict possible impact of SNPs in introns of CYP2D6.[13]

3. Results

3.1. Genetic variants and linkage disequilibrium analysis

The Sanger sequencing analysis identified totally 30 SNVs including 7 novel and 23 previously reported on the human *CYP2D6* Pharmacogene Variation data (https://www.pharmvar.org/). Of the 7 novel variants, there were 3 in the promoter (−498C>A, −184A>T, −175A>T), 2 in the intron 4 and 6 (2137G>C, 2988G>A), 2 in the exon 7 and 8 (3157G>T, 3851G>A) of the *CYP2D6* gene. For the 23 known SNPs, 2 were located in the promoter, 12 in introns, and 9 in exons (Table 1).

Linkage Disequilibrium analysis was performed to examine the SNPs relationship by Haploview 4.2 software. We identified only one LD block spanning a region from 214 in intron1 (marker 214G>C) to 232 in intron 1 (marker 232G>C) and these markers showed tight correlation with LOD #2 and $D' \leq 1$ (Fig. 1).

Table 1

***CYP2D6* variants in Kinh Vietnamese population identified by Sanger sequencing.**

Number	Region	Nucleotide change	Position	SNP ID Genbank	Effects	Number of detected subjects
1	Promoter	C>T	(−740)	rs28624811	Not translated	26/136
2		G>A	(−678)	rs28633410	Not translated	32/136
3		C>A	(−498)	Novel	Not translated	16/136
4		A>T	(−184)	Novel	Not translated	1/136
5		A>T	(−175)	Novel	Not translated	10/136
6	Exon 1	C>T	100	rs1065852	P34S	110/136
7		insT	137–138	rs774671100	47 frameshift	1/136
8	Intron 1	G>C	214	rs1080995	Not translated	25/136
9		C>A	221	rs1080996	Not translated	26/136
10		C>G	223	rs1080997	Not translated	22/136
11		T>C	227	rs1080998	Not translated	27/136
12		G>C	232	rs29001518	Not translated	34/136
13		A>C	233	rs28695233	Not translated	25/136
14		A>G	245	rs1081000	Not translated	21/136
15		G>T	310	rs28371699	Not translated	116/136
16	Exon 2	C>T	1039	rs1081003	None	114/136
17	Exon 3	G>C	1661	rs1058164	None	125/136
18		G>A	1759	rs5030865	G169R	2/136
19	Intron 3	G>A	1846	rs3892097	Not translated	3/136
20	Intron 4	A>G	2097	rs2267447	Not translated	115/136
21	Intron 4	G>C	2137	Novel	Not translated	1/136
22	Exon 5	G>A	2607	rs77913725	E278K	5/136
23		T>A	2611	rs1135828	M279K	1/136
24	Exon 6	C>T	2851	rs16947	R269C	35/136
25	Intron 6	G>A	2988	Novel	Not translated	6/136
26	Exon 7	G>T	3157	Novel	R329L	1/136
27	Intron 7	A>C	3384	rs1985842	Not translated	66/136
28		C>T	3790	rs4987144	Not translated	33/136
29	Exon 8	G>A	3851	Novel	W358*	2/136
30	Exon 9	G>C	4181	rs1135840	S486T	82/136

The nucleotide position is according to reference sequence M33388.1 in NCBI, nucleotides were numbered according to base A in the initiation codon ATG as +1.

To acquiring copy number information of the *CYP2D6* gene, the MLPA assay measuring the dosage of the exon 1, exon 5, exon 6, and exon 9 was applied to all 136 study subjects. The dosage of the exon 1, 5, 6, and 9 based on the ratio of the peak height of the reference versus the subjects was summarized in the Table 2. There were total 12 different types of copy number chances, in which the subjects with reduced copy number of all exon 1, 5, 6, and 9 were subsequently validated by LR-PCR (Supplementary Fig 1, http://links.lww.com/MD/D12).

3.2. Functional assessment of CYP2D6 novel variants

To predict where the 3 novel variants in the promoter region of *CYP2D6* have ability to regulate translation of gene, the promoter sequences containing each of novel variants (−498C>A, −184A>T, −175A>T) was analyzed using MATCH bioinformatic tool. However, none of them shown to be overlapped with the transcriptional factor binding sites in *CYP2D6* promoter.

Of the 2 novel variants located in the coding regions, the non-synonymous mutation 3851G>A (exon 8) resulted in a premature stop codon instead of tryptophan at position 385 (W385X) and a truncated, usually reduced or no functional protein product. The other one, 3157G>T (exon 7) causing the amino acid substitution from arginine to leucine at the position 329 (R329L), was evaluated for the probable effect by using Polyphen-2 and PROVEAN online tools. As a consequence, both HumDiv and HumVar trained Polyphen-2 models predicted that this variant was benign. In contrast, PROVEAN tool indicated that the R329L mutation would be deleterious (score = −4.326) (Fig. 2A and B).

Finally, the possible impact of novel mutations detected in intron 4 and 6 was performed using Human Splicing Finder online tool. The Position Weight Matrices were developed in order to assess the strength of 5′ and 3′ splice site and branch points. As a result, the novel variant 2988G>A located in the intron 6 was predicted to affect branching points and was capable of modifying splicing motif (Fig. 2C). Meanwhile, no impact was found when examining the remaining novel variant 2137G>C in the intron 4.

3.3. Genotype frequency and allelic frequencies

To determine the genotypes of *CYP2D6*, we combined the sequencing-based genotypes and copy number alterations identified by MLPA of each individual. Consequently, 9 SVs presenting in Kinh population containing an entire gene deletion and 8 hybrid/tandem rearrangements were explored as shown in the Fig. 3. In 136 subjects, 29 different genotypes with the frequency varied from 0.7% to 22.8% were assigned (Table 2). Among these, the most abundant genotypes were the homozygous *10/*10 (22.8%) followed by the heterozygous *1/*10 (15.4%) and *10/*36-*10 (11%). Other genotypes detected in over 5% of participants were *2/*10 (6.6%) and *1/*1 (5.9%).

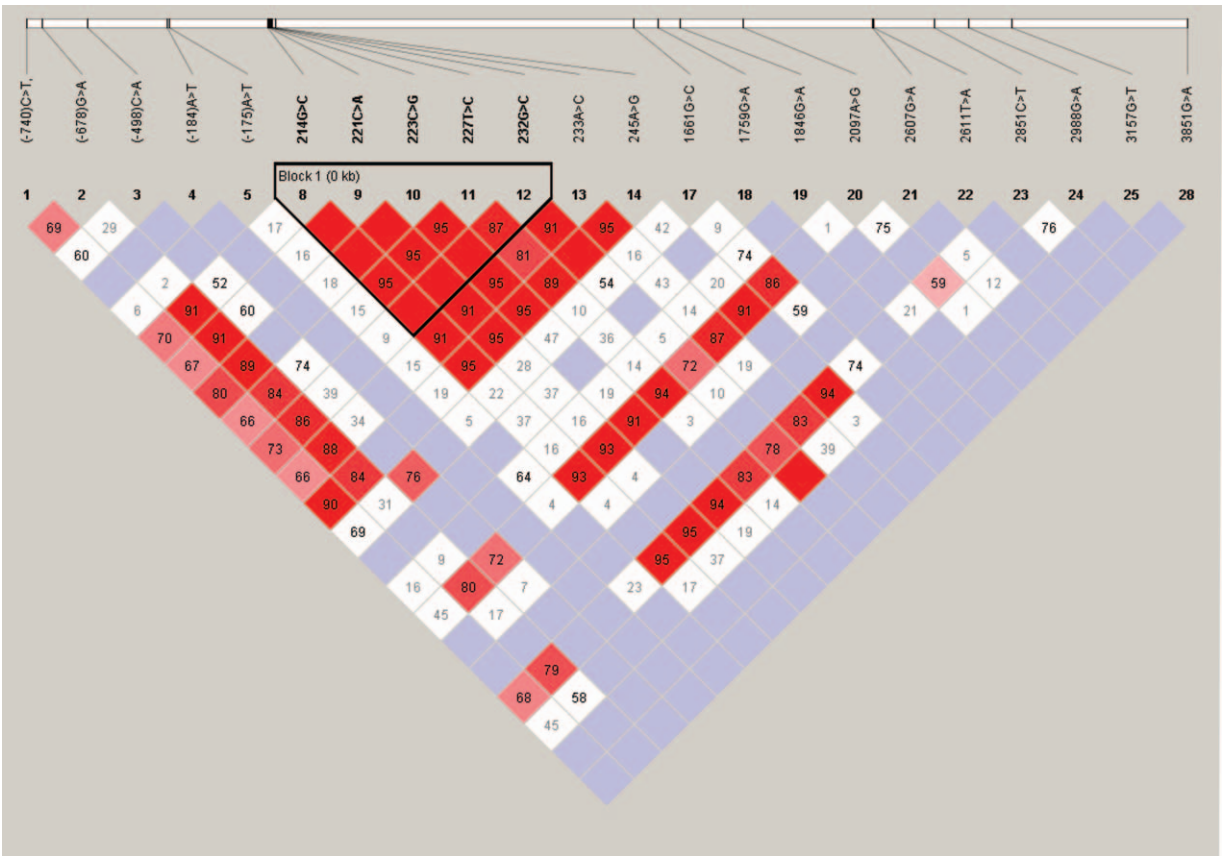

Figure 1. Linkage disequilibrium analysis of *CYP2D6* in Kinh Vietnamese. LD plots show distance and pair-wise LD between markers. It is displayed by standard color schemes with bright red for very strong LD (LOD #2, $D' = 1$), shades of pink/red (LOD#2, $D' < 1$), blue (LOD <2, $D' = 1$) for intermediate LD and white in otherwise. LD = linkage disequilibrium.

Except the wild type genotype, *10/*10, *36-*10/*36-*10 and *60/*60, all remaining genotypes were found as heterozygous.

For all population, there are total 9 *CYP2D6* variant alleles, 8 hybrid structures or tandem arrangements, and a whole gene deletion were identified which accounted for 75.37%, 16.54%, and 8.09% of 136 individuals, respectively (Table 3). Among 9 variant alleles, the *CYP2D6*10, known for decreased function, showed highest frequency (43.75%) while, the normal function alleles *CYP2D6*1 and *CYP2D6*2 accounted for 18.75% and 7.35%, respectively. Additionally, we identified 5 non-functional alleles, of which percentage varied from 0.37% to 8.09% (*4, *5, *14, *15, *36). Three alleles of unknown function were *60, *65, and *86 accounting for 0.74%, 2.94%, and 0.37%. Of the detected tandem arrangements, the most prominent variant was *36-*10 (12.13%) —a structure that the *36 hybrid was located upstream of *10 copy. Other tandem arrangements had frequencies ranged from 0.37% to 1.47%. The tandem types *13-*1 and *13-*2 had normal function while *36-*10 and *36-*36-*10 had reduced function. Both of *13 and *68-*4 types had no function (Table 3). We did not find any individual with homozygous deletion (*5/*5) as well as duplication (increased function) of *CYP2D6* in this study.

3.4. Comparison of CYP2D6 allele frequencies among global populations

We further compared frequency of *CYP2D6* alleles associated with enzyme activity alteration between Kinh Vietnamese and other populations from various geographical areas in the world (Table 4). Among nonfunctional alleles, the *CYP2D6*4 frequency was significantly lower than those in other populations located in Europe and America ($P < .0001$). In Asian countries, the null allele *CYP2D6*5* prevalence was found almost similar except Thailand ($P < .05$). Other alleles having no enzyme activity (*13, *14, and *15) showed no significant difference compared with populations in global. Among decreased functional alleles, it was shown that *CYP2D6*10 frequency statistically higher than that found in Europe and America ($P < .001$) as well as in other remaining parts of the world. Concerning *CYP2D6*36 percentage, this allele was shown to be of lower significance than that seen in Thai ($P < .0001$) but equal to those in Han Chinese, Papua New Guineans, and South Africa.

4. Discussion

To our knowledge, this is the first study applied both the Sanger sequencing and MLPA technique to identify all the *CYP2D6* variants presenting in Kinh Vietnamse population. Therefore, we could newly uncovered not only all SNVs located in the 5′ upstream region, 9 exons and their flanking regions but also the recurrent SVs of the *CYP2D6* gene. In fact, more than half (40/76) of *CYP2D6* genotypes carring *10 allele, which were assigned by only sequencing, were reassigned to genotypes consisted of hybrid structure (*13, *36, *67) and tandem hybrids (*36-*10, *36x2-*10, *68-*4) after copy number was analyzed

Table 2

Assignment of *CYP2D6* genotype in Kinh Vietnamese population.

Number	Sanger sequencing	MLPA				Long range PCR	Final genotype	Frequency (%)
		CN (E1)	CN (E5)	CN (E6)	CN (E9)			
1	*1/*1	2	2	2	2	–	*1/*1	5.9
2	*1/*2	2	2	2	2	–	*1/*2	4.4
3	*1/*10	2	2	2	2	–	*1/*10	15.4
4	*1/*86	2	2	2	2	–	*1/*86	0.7
5	*1/*14	2	2	2	2	–	*1/*14	0.7
6	*2/*10	2	2	2	2	–	*2/*10	6.6
7	*2/*65	2	2	2	2	–	*2/*65	2.2
8	*4/*10	2	2	2	2	–	*4/*10	0.7
9	*60/*60	2	2	2	2	–	*60/*60	0.7
10	*10/*10	2	2	2	2	–	*10/*10	22.8
11	*10/*15	2	2	2	2	–	*10/*15	0.7
12	*10/*65	2	2	2	2	–	*10/*65	2.2
13	*1/*10	3	3	3	2	–	*1/*36-*10	1.5
14	*10/*10	3	3	3	2	–	*10/*36-*10	11.0
15	*10/*65	3	3	3	2	–	*65/*36-*10	0.7
16	*10/*10	4	4	4	2	–	*36-*10/*36-*10	3.7
17	*2/*10	3	2	2	2	–	*2/*68-*4	0.7
18	*1/*4	2	2	2	3	–	*4/*13-*1	0.7
19	*10/*10	3	3	3	1	Deletion	*5/*36-*36-*10	0.7
20	*10/*10	2	2	2	1	Deletion	*5/*36-*10	3.7
21	*10/*10	1	2	2	2	–	*13/*10	1.5
22	*2/*10	1	1	2	2	–	*10/*67	0.7
23	*1/*2	1	1	1	2	Deletion	*5/*13-*2	1.5
24	*1/*1	1	1	1	2	Deletion	*5/*13-*1	0.7
25	*1/*1	1	1	1	1	Deletion	*1/*5	2.9
26	*2/*2	1	1	1	1	Deletion	*2/*5	0.7
27	*10/*10	1	1	1	1	Deletion	*5/*10	4.4
28	*2/*10	1	1	1	1	Deletion	*5/*65	0.7
29	*10/*10	1	1	1	0	Deletion	*5/*36	0.7

CN = copy number, E1 = exon 1, E5 = exon 5, E6 = exon 6, E9 = exon 9.

by MLPA. Furthermore, there were 22/136 individuals harboring *5, *13-*1, and *13-*2 alleles, which were only indentified after MLPA result obtained.

According to our examination, the percentage of *CYP2D6*10 assigned in the previous studies possibly involved structural variants. Currently, available methodologies for CNV detection comprised of conventional methods (Long distance and multiplex PCR, Realtime PCR) and microarrays hybridization based methods (array comparative genotmic hybridization, Roche Amplichip CYP450 system, Affymetrix DMET plus microarray), but their limitations are low throughput, low reproductivity, and costly.[38] MLPA technique has been shown to be high-throughput, cost effective, and accurate, and was widely applied in genetic testing in various inherited diseases of human.[39]

By sequencing we found 7 novel SNVs and it was interesting to predict whether they might have impact on gene regulation or function. Among 5 new SNVs identified in the noncoding regions, only variant 2988G>A (intron 6) was shown the ability to have harmful effect on branching point thereby may cause of abnormal pre-mRNA splicing. Both novel variants found in coding regions might have effect on protein activity because the variant 3851G>A (W385X) in the exon 8 would be translated into a truncated protein while the other one, 3157G>T (R329L) in the exon 7, was predicted as deleterious by PROVEAN. In fact, the truncated protein resulted from 3851G>A would lead to a nonfunctional product or the enzyme thereby would have no activity. However, the impact of 3157G>T mutation should be further clarified by functional analysis, in which the metabolizer status of mutant protein will be investigated in vitro and/or in vivo animal model or cell culture systems. Based on the activation of encoded proteins, we can determine whether their functions are increased/decreased or not affected. Until very recently, 2 volunteers carrying 3157G>T (R329L) and another with 3851G>A (W385X) have informed us that they have no health problem.

The prevalent of the allele *CYP2D6*10 in Kinh Vietnamese from this study is in consistent with previous findings in Kinh Vietnamese[6] as well as in other Asian populations.[15,16,40] Statistical analysis also demonstrated significant higher frequency of *CYP2D6*10 in Kinh Vietnamese compared with other worldwide regions except for Asian countries.[20,27,31,33,36] Our work further approved the high prevalence of *CYP2D6*10 retricted to Asian populations. Meanwhile, the nonfunction allele *CYP2D6*4 was very rare in our study subjects as well as in other Asian populations. This was significant difference compared with figure observed in Europe, America, and Middle East, where *CYP2D6*4 was reported with high frequency.[20,22,29,36] This result is in accordance with previous works, in which a gradual decrease of *CYP2D6*4 from Europe to Asia countries was reported.[41] Although *CYP2D6*36 frequency (0.37%) was shown significantly lower than that in Thailand (16.4%),[18] however, similar figures were reported in various populations in the

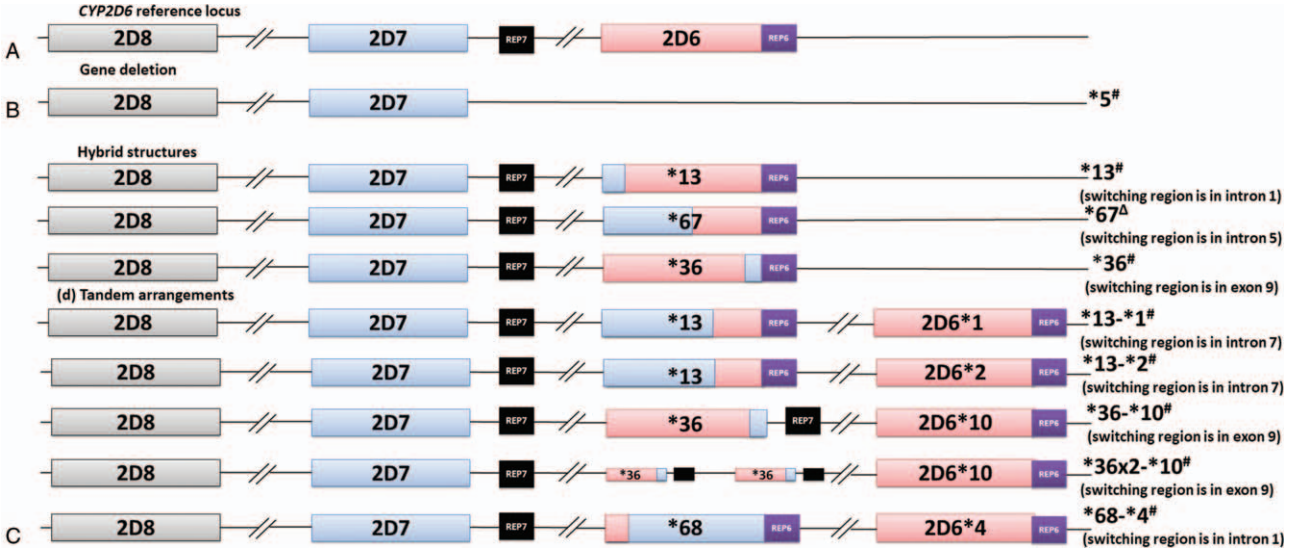

Figure 2. Functional prediction of non-synonymous mutation 3157G>T by Polyphen-2 and Provean (A, B) and splicing effect prediction of intronic variant 2988G>A (C).

world containing Chinese Han, Papua New Guineans, and South African.[15,32,34] Likewise, numerous studies demonstrated the infrequent existence of *CYP2D6**36 in Asian populations as well as other worldwide communities.[15,32,33,34,42]

Because *CYP2D6**10 was shown to be main variant with decreased function found in Kinh Vietnamese, then this is the major factor causing diminished enzyme activity of CYP2D6 in this population. Owing to this evidence, physicians in Vietnam should pay attention on the underlying genetic determinant in

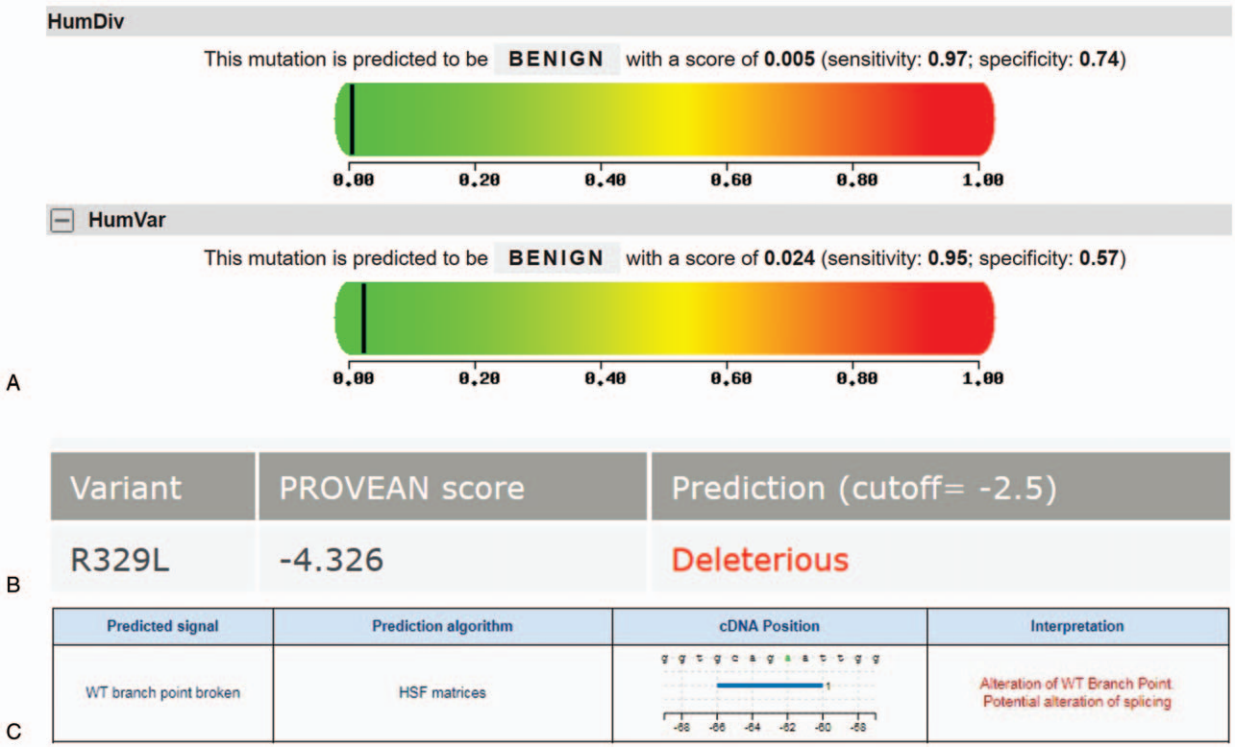

Figure 3. Graphic view of *CYP2D6* structural variants determined in present study. (A) Reference gene locus of *CYP2D6* (red boxes) and pseudo genes *CYP2D7* (blue boxes), *CYP2D8* (grey boxes). REP6 (dark purple boxes) and REP7 (black boxes) are repetitive sequences located downstream of *CYP2D6* and *CYP2D7*, respectively. (B) Deletion of entire *CYP2D6* gene. (C) Hybrid structures of *CYP2D6* containing *CYP2D6-2D7* (the 5′ portion is derived from *CYP2D6* and the 3′ portion is derived from *CYP2D7*) and *CYP2D7-2D6* (the 5′ portion is derived from *CYP2D7* and the 3′ portion is derived from *CYP2D6*). (D) *CYP2D6* tandem arrangements including 2 or more gene copies that are different with "duplications/multiplications." # Details of structural variants were reported in Pharmacogene Variation (PharmVar) Consortium (https://www.pharmvar.org/gene/CYP2D6). Δ Detail of structural variants were reported in Andrea Gaedigk work.[14]

Table 3

Allelic frequencies of *CYP2D6* in the Kinh Vietnamese population.

Variants	Number	Allele	Frequency (%)	Function
Variant alleles	1	*1	18.75	Normal
	2	*2	7.35	Normal
	3	*4	0.74	No function
	4	*10	43.75	Decreased
	5	*14	0.37	No function
	6	*15	0.37	No function
	7	*60	0.74	Unknown
	8	*65	2.94	Unknown
	9	*86	0.37	Unknown
Copy number variant	1	*5	8.09	No function
Hybrids and Tandem arrangement	1	*13	1.547	None
	2	*13-*1	0.74	Normal
	3	*13-*2	0.74	Normal
	4	*36	0.37	No function
	5	*36-*10	12.13	Decreased
	6	*36-*36-*10	0.37	Decreased
	7	*67	0.37	Unknown
	8	*68-*4	0.37	None

regard to drugs response, especially *CYP2D6**10 during treatment for Vietnamese subjects having Kinh genetic background. Individuals who carry >2 copies of *CYP2D6* normal alleles (classified as ultra rapid metabolizer) will metabolize CYP2D6 substrates more rapidly, leading to reduction of drug treatment efficacy. No duplication or multiplication of *CYP2D6* normal function alleles were detected in present study subjects. However, the influence of *CYP2D6* duplication and multiplication on corresponding phenotypes in Vietnam should be considered in future studies.

5. Conclusions

We present for the first time all genetic variants of *CYP2D6* including SNVs and SVs in 136 Kinh Vietnamese subjects. Our newly obtained data provide important knowledge for determin-

Table 4

Comparison of *CYP2D6* allele frequencies responsible for functional alteration of protein between Kinh Vietnamese and populations in worldwide.

			Allele frequencies (%)									
Regions	Population	Number of subjects	*4	*5	*10	*13	*14	*15	*36	*36-*10	*36x2-*10	Reference
Asia	Kinh Vietnamese	136	0.74	8.09	43.75	1.547	0.37	0.37	0.37	12.13	0.37	Present study
	Vietnamese	74	1.4	8.0	43.5	–	–	–	–	–	–	[6]
	Chinese Han	1954	0.4[b]	6.0	42.9	–	0.9	0.0	0.6	–	–	[15]
	Japanese	1017	–	–	42.7	–	–	–	–	–	–	[16]
	Korean	448	0.8[b]	6.2	44.1	–	0.5	–	–	–	–	[17]
	Thai	288	0.7	4.3[a]	44.6	–	1.0	–	16.4[d]	–	–	[18]
	Indian	881	9.3	–	–	–	–	–	–	–	–	[19]
Europe	Hungarians	431	18.1[d]	–	2.4[c]	–	–	–	–	–	–	[20]
	Italian	218	17.7[d]	0.0[d]	–	–	–	–	–	–	–	[21]
	Macedonia	184	18.7[d]	9.1	2.7[c]	–	–	–	–	–	–	[22]
	Greek	283	17.8[d]	–	–	–	–	–	–	–	–	[23]
	Danish	244	19.7[d]	5.3	1.6[c]	–	–	0.0	–	–	–	[24]
	Czech	223	22.9[d]	3.1[b]	–	–	–	–	–	–	–	[25]
America	American	264	10.0[d]	1.7	2.8[c]	–	0.0	–	–	–	–	[26]
	Brazilian	1020	9.4[d]	4.6[b]	2.1[c]	–	–	–	–	–	–	[27]
	Canadian	90	8.6[d]	–	2.3[c]	–	–	–	–	–	–	[28]
	Chilean	321	12.0[d]	–	–	–	–	–	–	–	–	[29]
	Mexican	154	10.4[d]	0.0[d]	0.6[c]	–	–	–	–	–	–	[30]
Oceania	Oceania	39	0.0	1.3[a]	2.6	–	–	–	–	–	–	[31]
	Papua New Guineans	88	0.0	3.0[a]	0.0[c]	–	0.0	0.0	0.0	–	–	[32]
Africa	African American	251	5.4	6.6[b]	3.6[c]	–	0.0	0.0	–	–	–	[33]
	South Africa	100	3.1[c]	8.7	5.1[c]	0.0	0.0	0.0	0.0	–	–	[34]
	Zimbabwean	114	2.0[b]	4.0[a]	–	–	–	–	–	–	–	[35]
Middle East	Iranian	100	12.5[d]	3.0[a]	9.0[c]	–	–	–	–	–	–	[36]
	Saudi Arabians	192	–	–	–	–	0.3	–	–	–	–	[37]

[a] $P < .05$ compared with data of current study.
[b] $P < .01$ compared with data of current study.
[c] $P < .001$ compared with data of current study.
[d] $P < .0001$ compared with data of current study.

ing metabolic phenotypes of CYP2CD6 substrates and optimizing drug therapies in Vietnam. Based on certain *CYP2D6* genotype identified, individuals would receive an appropriate medication therapy and avoid the risk of side effect caused by improper dose of drugs.

Author contributions

Conceptualization: Hai Van Nong, Ha Hai Nguyen.
Data curation: Thuong Thi Huyen Ma, Nhung Phuong Vu, Thang Hong Vu, Quynh Thi Nhu Bach, Ton Dang Nguyen.
Formal analysis: Nhung Phuong Vu.
Funding acquisition: Ha Hai Nguyen.
Investigation: Ha Hai Nguyen, Thuong Thi Huyen Ma, Nhung Phuong Vu, Thang Hong Vu, Quynh Thi Nhu Bach, Ton Dang Nguyen.
Methodology: Ha Hai Nguyen.
Project administration: Ha Hai Nguyen.
Supervision: Hai Van Nong.
Validation: Thuong Thi Huyen Ma, Nhung Phuong Vu.
Writing – original draft: Ha Hai Nguyen, Nhung Phuong Vu.
Writing – review & editing: Hai Van Nong, Ha Hai Nguyen.

Ha Hai Nguyen: orcid 0000-0002-5431-5935
Nhung Phuong Vu: orcid 0000-0002-0889-4534
Ton Dang Nguyen: orcid 0000-0003-0182-8996
Hai Van Nong: orcid 0000-0003-2939-8430
Van Hai Nong orcid: 0000-0003-2939-8430.

References

[1] Eichelbaum M, Ingelman-Sundberg M, Evans WE. Pharmacogenomics and individualized drug therapy. Annu Rev Med 2006;57:119–37.
[2] Kurose K, Sugiyama E, Saito Y. Population differences in major functional polymorphisms of pharmacokinetics/pharmacodynamics-related genes in Eastern Asians and Europeans: implications in the clinical trials for novel drug development. Drug Metab Pharmacokinet 2012;27:9–54.
[3] Ingelman-Sundberg M, Sim SC, Gomez A, et al. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. Pharmacol Ther 2007;116:496–526.
[4] Beoris M, Amos Wilson J, Garces JA, et al. CYP2D6 copy number distribution in the US population. Pharmacogenet Genomics 2016;26:96–9.
[5] Gaedigk A, Ndjountché L, Divakaran K, et al. Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. Clin Pharmacol Ther 2007;81:242–51.
[6] 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;65:355–63.
[7] Kim EY, Lee SS, Jung HJ, et al. Robust CYP2D6 genotype assay including copy number variation using multiplex single-base extension for Asian populations. Clin Chim Acta 2010;411:2043–8.
[8] Love RR, Desta Z, Flockhart D, et al. CYP2D6 genotypes, endoxifen levels, and disease recurrence in 224 Filipino and Vietnamese women receiving adjuvant tamoxifen for operable breast cancer. SpringerPlus 2013;2:52.
[9] Hersberger M, Marti-Jaun J, Rentsch K, et al. Rapid detection of the CYP2D6*4 and CYP2D6*6 alleles by Tetra-Primer PCR and of the CYP2D6*5 allele by multiplex long PCR. Clin Chem 2000;46:1072–7.
[10] de Vooght KMK, van Wijk R, van Solinge WW. Management of gene promoter mutations in molecular diagnostics. Clin Chem 2009;55:698–708.
[11] Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;76:7.20.1–41.
[12] Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015;31:2745–7.
[13] Desmet FO, Hamroun D, Lalande M, et al. Human splicing finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009;37:e67.
[14] Gaedigk A, Jaime LKM, Bertino JS, et al. Identification of Novel CYP2D7-2D6 hybrids: non-functional and functional variants. Front Pharmacol 2010;1:121.
[15] Qian JC, Xu XM, Hu GX, et al. Genetic variations of human CYP2D6 in the Chinese Han population. Pharmacogenomics 2013;14:1731–43.
[16] Ota T, Kamada Y, Hayashida M, et al. Combination analysis in genetic polymorphisms of drug-metabolizing enzymes CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A5 in the Japanese population. Int J Med Sci 2015;12:78–82.
[17] Man M, Farmen M, Dumaual C, et al. Genetic variation in metabolizing enzyme and transporter genes: comprehensive assessment in 3 major East Asian subpopulations with comparison to Caucasians and Africans. J Clin Pharmacol 2010;50:929–40.
[18] Suwannasri P, Thongnoppakhun W, Pramyothin P, et al. Combination of multiplex PCR and DHPLC-based strategy for CYP2D6 genotyping scheme in Thais. Clin Biochem 2011;44:1144–52.
[19] Buch S, Kotekar A, Kawle D, et al. Polymorphisms at CYP and GST gene loci. Eur J Clin Pharmacol 2001;57:553–5.
[20] Weber A, Szalai R, Sipeky C, et al. Increased prevalence of functional minor allele variants of drug metabolizing CYP2B6 and CYP2D6 genes in Roma population samples. Pharmacol Rep 2015;67:460–4.
[21] De Luca C, Scordo MG, Cesareo E, et al. Biological definition of multiple chemical sensitivity from redox state and cytokine profiling and not from polymorphisms of xenobiotic-metabolizing enzymes. Toxicol Appl Pharmacol 2010;248:285–92.
[22] Kapedanovska Nestorovska A, Jakovski K, Naumovska Z, et al. Distribution of the most common genetic variants associated with a variable drug response in the population of the Republic of Macedonia. Balkan J Med Genet 2015;17:5–14.
[23] Arvanitidis K, Ragia G, Iordanidou M, et al. Genetic polymorphisms of drug-metabolizing enzymes CYP2D6, CYP2C9, CYP2C19 and CYP3A5 in the Greek population. Fundam Clin Pharmacol 2007;21:419–26.
[24] Rasmussen JO, Christensen M, Svendsen JM, et al. CYP2D6 gene test in psychiatric patients and healthy volunteers. Scand J Clin Lab Invest Suppl 2006;66:129–36.
[25] Buzková H, Pechandová K, Slanar O, et al. Frequency of single nucleotide polymorphisms of CYP2D6 in the Czech population. Cell Biochem Funct 2008;26:76–81.
[26] Luo HR, Gaedigk A, Aloumanis V, et al. Identification of CYP2D6 impaired functional alleles in Mexican Americans. Eur J Clin Pharmacol 2005;61:797–802.
[27] Friedrich DC, Genro JP, Sortica VA, et al. Distribution of CYP2D6 alleles and phenotypes in the Brazilian Population. PLoS One 2014;9:e110691.
[28] Jurima-Romet M, Foster BC, Casley WL, et al. CYP2D6-related oxidation polymorphism in a Canadian Inuit population. Can J Physiol Pharmacol 1997;75:165–72.
[29] Varela N, Quiñones L, Stojanova J, et al. Characterization of the CYP2D6 drug metabolizing phenotypes of the Chilean mestizo population through polymorphism analyses. Pharmacol Res 2016;110:276.
[30] López-López M, Peñas-Lledó E, Dorado P, et al. CYP2D6 genetic polymorphisms in Southern Mexican Mayan Lacandones and Mestizos from Chiapas. Pharmacogenomics 2014;15:1859–65.
[31] Sistonen J, Sajantila A, Lao O, et al. CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. Pharmacogenet Genomics 2007;17:93–101.
[32] von Ahsen N, Tzvetkov M, Karunajeewa HA, et al. CYP2D6 and CYP2C19 in Papua New Guinea: high frequency of previously uncharacterized CYP2D6 alleles and heterozygote excess. Int J Mol Epidemiol Genet 2010;1:310–9.
[33] Gaedigk A, Bradford LD, Marcucci KA, et al. Unique CYP2D6 activity distribution and genotype-phenotype discordance in black Americans. Clin Pharmacol Ther 2002;72:76–89.
[34] Dodgen TM, Labuschagne CDJ, van Schalkwyk A, et al. Pharmacogenetic comparison of CYP2D6 predictive and measured phenotypes in a South African cohort. Pharmacogenomics J 2016;16:566–72.
[35] Dandara C, Mutowembwa Masimirembwa C, Magimba A, et al. Genetic polymorphism of CYP2D6 and CYP2C19 in East- and Southern African populations including psychiatric patients. Eur J Clin Pharmacol 2001;57:11–7.
[36] Hashemi-Soteh MB, Sarzare F, Merat F, et al. Frequencies of three CYP2D6 nonfunctional alleles (CYP2D6*3, *4, and *6) within an

Iranian population (Mazandaran). Genet Test Mol Biomarkers 2011;15:821–5.
[37] Al-Dosari MS, Al-Jenoobi FI, Alkharfy KM, et al. High prevalence of CYP2D6*41 (G2988A) allele in Saudi Arabians. Environ Toxicol Pharmacol 2013;36:1063–7.
[38] He Y, Hoskins JM, McLeod HL. Copy Number Variants in pharmacogenetic genes. Trends Mol Med 2011;17:244–51.
[39] Stuppia L, Antonucci I, Palka G, et al. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. Int J Mol Sci 2012;13:3245–76.
[40] Yi S, An H, Lee H, et al. Korean, Japanese, and Chinese populations featured similar genes encoding drug-metabolizing enzymes and transporters: a DMET Plus microarray assessment. Pharmacogenet Genomics 2014;24:477–85.
[41] Sistonen J, Fuselli S, Palo JU, et al. Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. Pharmacogenet Genomics 2009;19:170–9.
[42] Rasmussen JO, Christensen M, Svendsen JM, et al. CYP2D6 gene test in psychiatric patients and healthy volunteers. Scand J Clin Lab Invest 2006;66:129–36.