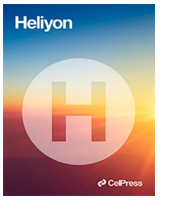
Heliyon 10 (2024) e28952

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Genetic variants, haplotype determination, and function of novel alleles of *CYP2B6* in a Han Chinese population

Li-Qun Zhang [a], Xin-Yue Li [d], Lian-Guo Chen [b], Zhe Chen [c], Ren-Ai Xu [b], Jian-Chang Qian [d], Xiao-yang Zhou [a], Da-Peng Dai [a], Guo-Xin Hu [d], Jian-Ping Cai [a,*]

[a] *The Key Laboratory of Geriatrics, Beijing Institute of Geriatrics, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing Hospital/National Center of Gerontology of National Health Commission, China*
[b] *The First Affiliated Hospital of Wenzhou Medical University, 325000, Wenzhou, Zhejiang, China*
[c] *Pharmacy department, Wenzhou people's hospital, 325000, Wenzhou, Zhejiang, China*
[d] *Institute of Molecular Toxicology and Pharmacology, School of Pharmaceutical Sciences, Wenzhou Medical University, Wenzhou, Zhejiang, China*

ARTICLE INFO

Keywords:
CYP2B6
Genetic variant
Haplotype
Catalytic activity

ABSTRACT

Amino acid variants in protein may result in deleterious effects on enzymatic activity. In this study we investigate the DNA variants on activity of *CYP2B6* gene in a Chinese Han population for potential use in precision medicine. All exons in *CYP2B6* gene from 1483 Chinese Han adults (Zhejiang province) were sequenced using Sanger sequencing. The effects of nonsynonymous variants on recombinant protein catalytic activity were investigated *in vitro* with Sf12 system. The haplotype of novel nonsynonymous variants with other single nucleotide variants in the same allele was determined using Nanopore sequencing. Of 38 alleles listed on the Pharmacogene Variation Consortium, we detected 7 previously reported alleles and 18 novel variants, of which 11 nonsynonymous variants showed lower catalytic activity (0.00–0.60) on bupropion compared to *CYP2B6*1*. Further, these 11 novel star-alleles (*CYP2B6*39–49*) were assigned by the Pharmacogene Variation Consortium, which may be valuable for pharmacogenetic research and personalized medicine.

1. Introduction

Cytochrome P450s (CYPs) are considered as the most important drug metabolizing enzymes in phase I metabolism. Human *CYPs* are categorized into 18 families and 57 genes [1,2]. Fifteen *CYP* enzymes in families 1–4 are responsible for approximately 90% of phase I metabolism of clinical drugs [3,4]. *CYP2B6* activity varied among human microsomes by 25 and 80 fold in the metabolism of S-mephenytoin [5] and bupropion [6], respectively. Such variations in *CYP2B6* activities may lead to variable blood exposure and therapeutic efficacy to drugs mainly metabolized by *CYP2B6*.

CYP2B6 and related *CYP2B7* (pseudogene) are located in *CYP2* gene loci on chromosome 19 [7]. The *CYP2B6* gene is comprised of 9 exons and encodes a 491 amino acids enzyme, which is the only functional isozyme of *CYP2B* [2]. Due to the high genetic variant and inhibitors or inducers, activity of *CYP2B6* is intensively variable among individuals. For clinical drugs such as efavirenz (EFV) and nevirapine, which are used in treatment of human immunodeficiency virus (HIV), *CYP2B6* single nucleotide polymorphisms (SNPs) were shown to be useful in predicting drug responses [8,9]. Data from Chinese HIV patients (n = 166) who had been genotyped for

* Corresponding author.
E-mail address: caijp61@vip.sina.com (J.-P. Cai).

https://doi.org/10.1016/j.heliyon.2024.e28952
Received 18 May 2023; Received in revised form 25 March 2024; Accepted 27 March 2024
Available online 30 March 2024

CYP2B6 and tested for blood EFV levels at steady state showed trough levels for TT, GT and GG genotypes were 9.08, 3.57, and 2.94 mg/L, respectively, for variant of G516T [10]. Different efficacy of EFV-based HIV therapy and other *CYP2B6* substrates may resulted from variations of the *CYP2B6* frequencies due to highly variable genetic differences among racial and ethnic populations [11]. Guan et al. found that the frequencies of common *CYP2B6* variants in a Chinese Han population (n = 193) are similar to those of other Asians including Korean and Japanese, but highly different from those of Caucasians [12]. Herein, we sought to investigate on the DNA variant of *CYP2B6* in a healthy Chinese Han population (n = 1483) and determine the catalytic activity of novel amino acid changes on bupropion, which may be valuable for pharmacogenetic research and personalized medicine.

2. Materials and methods

2.1. Subjects and samples

The study consisted of 1483 unrelated Han Chinese subjects (>18 y) who lived in Wenzhou (Zhejiang Province, China). This study was approved by the Institutional Ethical Committee of Beijing Hospital. All participants provided informed written consent before blood sample collection. DNA was extracted from whole blood cells using magnetic beads (Applied Biosystems, CA, USA).

2.2. DNA sequencing

The 9 exons and flanking introns (over 50 bp) of the *CYP2B6* were amplified by PCR with the primers listed in Supplementary Table 1. After purification, the amplicons were sequenced on an Applied Biosystems 3730XL DNA Analyzer (Applied Biosystems). Acquired chromatograms of the sequences were verified manually by at least two individuals. Amplicons with inconsistent sequences relative to *CYP2B6*1* (NG_007929.1) were confirmed by bidirectional sequencing.

2.3. Data analysis

Chromatogram files were analyzed on Snapgene (version4.3.6, Illinois, USA) using *CYP2B6*1* (NG_007929.1) as the template. Variants that were not listed on the Pharmacogene Variation Consortium (PharmVar) were searched in the SNP database (https://www.ncbi.nlm.nih.gov/snp/) on National Center for Biotechnology Information (NCBI) to confirm they were novel. The effects of novel nonsynonymous variants on the structure and activity of *CYP2B6* were analyzed with online tools PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) and PROVEAN (https://www.jcvi.org/research/provean).

2.4. Haplotype sequencing

For nonsynonymous variants, the entire *CYP2B6* (30,957 bp) gene was first amplified with KOD OneTM PCR Master Mix (TOYOBO, Japan) and then purified with 0.6 × AMPure XP beads (Beckman Coulter, High Wycombe, UK). A DNA A-tailing Kit (Takara, Japan) was used for dA-tailing for 20 min at 72 °C. The A-tailed DNA amplicons were purified by 0.6 × AMPure XP beads and then ligated to barcodes (EXP-NBD104, ONT, UK). After purification, adapters were ligated to the DNA ends with the SQK-LSK110 ligation kit (ONT, Oxford, UK) and the samples were purified again with 0.6 × volume of AMPure XP beads. The adapted library was quantified with a Qubit fluorometer and 50–100 fmol DNA library was loaded onto a GridION flow-cell (R9.4.1) according to manufacturer instructions and run for 17.5 h. The raw sequencing data was base called using the MinKNOW software in default mode. The reads, in fastq format, were mapped to the human reference genome GRCh38 using Minimap2. The alignment data were transferred to bam format and sorted using samtools. The sorted and indexed bam data were viewed with IGV_2.14.0 and the haplotype of each SNP was determined manually (Supplementary Fig. 1). The novel haplotypes were submitted to DNA Data Bank of Japan and PharmVar for nomenclature.

2.5. Expression of CYP2B6 novel variant proteins in Sf21 cells and in vitro function analysis with UPLC-MS/MS

Briefly, the whole length of *CYP2B6* ORF was synthesized and constructed into pFastbac (Thermo) by BGI laboratory (Beijing, China). Pointed nucleotide variant was introduced with double-reversed PCR and confirmed by Sanger sequencing. Recombinant baculoviruses with high expression of both *CYP2B6* and *CYP oxidoreductase (OR)* proteins were obtained from infection of Sf21 cells as previously described method for study on *CYP2C19* variants [13]. Microsomal fractions (2 μg) were separated on 10% (w/v) sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gels and western blotting was carried out for quantification of *CYP2B6* proteins. Antibodies for *CYP2B6* and *OR* were purchased from Abcam (Cambridge, UK).

Bupropion hydrochloride, midazolam and hydroxybupropion were purchased from J&K Scientific Ltd. (Hong Kong, China). Briefly, 200 μL incubation buffer containing 100 mM Tris-HCl (pH 7.4), 0.1 mg/mL *CYP2B6*1* or other novel variants, 2.7 mg/mL CYB5 and 100 mM bupropion was gently mixed and incubated for 5 min at 37 °C. Next, 1 mM NADPH was added in the solution and incubated for 40 min before transferring to −80 °C to terminate the reaction. Subsequently, 200 μL acetonitrile and 20 μL Midazolam (500 ng/mL, as internal standard) were added and vortexed for 2 min. Samples were then centrifuged at 13,000×*g* for 10 min at 4 °C and the supernatants were analyzed with UPLC-MS/MS as previously described [14]. The relative ratio of hydroxybupropion to bupropion hydrochloride was calculated to assess the enzyme activities of *CYP2B6*1* and 11 novel variants. These experiments were performed in triplicate.

For Vmax/Km analysis, 1, 2, 5, 10, 20 and 50 μM of efavirenz were added in 200 μL incubation buffer containing 1 × PBS buffer, 20

μg *CYP2B6*1* or other *CYP2B6* variants, 25 μg/ml cytochrome *b*5, 1 mM NADPH. The mixture without NADPH was pre-incubated at 37 °C for 5 min; then, 1 mM of NADPH was added to initiate the reaction. After incubation for 40 min, the reaction was immediately terminated by cooling to −80 °C. Then, 200 μl of acetonitrile and 20 μl of telmisartan (10 μg/ml) as an internal standard, were added to the mixture. After being vortexed for 2 min and centrifuged at 13,000×*g* for 10 min, the supernatant was analyzed by UPLC-MS/MS to detect efavirenz and 8-OH efavirenz. The analytes were firstly separated on a BEH C18 column (2.1 × 100 mm,1.7 μm; Waters Corp., Millipore, Bedford, MA, USA), which incubated at 40 °C. The mobile phase was consisted of 0.1% formic acid-water (A) and methanol (B), and elution at 0.40 mL/min for 4.5 min with a gradient condition. The program was set as 10–90% methanol (0–2.5 min), 90-10% methanol (2.5–2.8 min), and 10% methanol (2.8–4.5 min). The monitoring transitions were *m/z* 329.9 → 257.9 for 8-OH efavirenz.

2.6. Statistical analyses

One-way analysis of variance (ANOVA) test was performed for multiple comparisons. Student's *t*-test was performed for two group comparisons. Significant level was $P < 0.05$.

3. Results

3.1. Allele frequencies

To investigate the overall variants of the *CYP2B6* gene in the Chinese population, 1483 healthy subjects in the south of China were recruited and 9 exons of their *CYP2B6* gene were sequenced. The results showed that there were 7 previously reported allelic variants in PharmVar and 18 novel variants identified in the present study. These include 11 nonsynonymous and 7 synonymous variants (Table 1). Additionally, most variants, except for c.12776A > T and c.18008A > C, have been reported in the library of Reference SNPs.

As illustrated in Table 1, the first three common variants were c.18053A > G (26.57%), c.15631G > T (18.41% compared to 31.6% in the 1000 Genomes Project) and c.64C > T (3.67% comparing with 4.8% in the 1000 Genomes Project). Allele frequency analysis showed that most individuals carried the wild-type *CYP2B6* allele (*CYP2B6*1*, 74.68%), and the first three common alleles were *CYP2B6*6* (combination of c.15631G > T and c.18053A > G, 12.37%), *CYP2B6*4* (c.18053A > G, 6.54%) and *CYP2B6*2* (c.64C > T, 4.28%) as shown in Table 2. Additionally, four other alleles, *CYP2B6*5* (c.25505C > T), *CYP2B6*9* (c.15631G > T), *CYP2B6*23* (c.25421A > G) and *CYP2B6*26* (c.15614C > G, c.15631G > T, c.18053A > G) were also found in this study, with frequencies of no more than 0.4%. Moreover, 11 other novel alleles were reported for the first time and assigned star-allele names by PharmVar (*CYP2B6*39–49*).

Table 1
Single nucleotide variation (SNV) found in coding region of *CYP2B6* and their frequencies in 1483 Chinese Han subjects.

Exon	GRCh38.p13 position	Coding DNA position	Complementary DNA position	Effect on protein	Hit	Frequency (%)	Frequency in 1000 Genomes Project (%)	Reference SNP	Recorded on PharmVar
1	40991369	64C > T	64C > T	R22C	109	3.67	4.80	rs8192709	yes
2	41004045	12740G > C	216G > C	/	76	2.56	5.20	rs2279341	yes
2	41004081	12776A > T	252A > T	/	1	0.03	–	NA	no
2	41004133	12828G > A	304G > A	A102T	1	0.03	–	rs148009906	no
3	41004380	13075C > T	418C > T	R140W	3	0.10	<0.10	rs535039125	no
4	41006919	15614C > G	499C > G	P167A	10	0.34	<0.10	rs3826711	yes
4	41006936	15631G > T	516G > T	Q172H	546	18.41	31.60	rs3745274	yes
4	41006967	15662G > A	547G > A	V183I	1	0.03	0.10	rs58871670	no
5	41009313	18008A > C	740A > C	H247P	1	0.03	–	NA	no
5	41009358	18053A > G	785A > G	K262R	788	26.57	–	rs2279343	yes
7	41012393	21088T > C	1060T > C	Y354H	1	0.03	–	rs754621576	no
7	41012394	21089A > G	1061A > G	Y354C	1	0.03	–	rs780991919	no
7	41012466	21161G > A	1133G > A	R378Q	2	0.07	–	rs200458614	no
7	41012471	21166T > C	1138T > C	Y380H	12	0.40	<0.10	rs201500445	no
7	41012473	21168C > T	1140C > T	/	1	0.03	–	rs374509631	no
7	41012478	21173T > A	1145T > A	I382 N	6	0.20	<0.10	rs200238771	no
8	41012739	21434C > T	1218C > T	/	1	0.03	<0.10	rs35661880	no
9	41016652	25347G > A	1301G > A	R434Q	1	0.03	–	rs764288403	no
9	41016677	25372C > T	1326C > T	/	1	0.03	–	rs1476466408	no
9	41016679	25374G > A	1328G > A	R443H	1	0.03	–	rs374099483	no
9	41016686	25381A > G	1335A > G	/	1	0.03	<0.10	rs146956370	no
9	41016726	25421A > G	1375A > G	M459V	8	0.27	<0.10	rs3211369	yes
9	41016741	25436G > A	1390G > A	A464T	5	0.17	<0.10	rs117872433	no
9	41016743	25438C > T	1392C > T	/	8	0.27	–	rs200993638	no
9	41016746	25441A > T	1395A > T	/	8	0.27	–	rs3211370	no
9	41016810	25505C > T	1459C > T	R487C	5	0.17	5.40	rs3211371	yes

Table 2
Allele frequencies of *CYP2B6* variants detected in 1483 Chinese Han subjects.

Allele	Position (coding DNA)	Effect on protein	Hit	Frequency
*CYP2B6*1*	/	/	2215	0.7468
*CYP2B6*2*	64C > T	R22C	127	0.0428
*CYP2B6*4*	18053A > G	K262R	194	0.0654
*CYP2B6*5*	25505C > T	R487C	5	0.0017
*CYP2B6*6*	15631G > T and 18053A > G	Q172H and K262R	367	0.1237
*CYP2B6*9*	15631G > T	Q172H	4	0.0013
*CYP2B6*23*	25421A > G	M459V	8	0.0027
*CYP2B6*26*	15614C > G, 15631G > T, 18053A > G	P167A, Q172H and K262R	11	0.0037
*CYP2B6*39*	13075C > T, 15631G > T and 18053A > G	R140W, Q172H and K262R	3	0.0010
*CYP2B6*40*	21161G > A, 15631G > T and 18053A > G	R378Q, Q172H and K262R	2	0.0007
*CYP2B6*41*	21166T > C, 15631G > T and 18053A > G	Y380H, Q172H and K262R	12	0.0040
*CYP2B6*42*	25374G > A, 15631G > T and 18053A > G	R443H, Q172H and K262R	1	0.0003
*CYP2B6*43*	25436G > A, 15631G > T and 18053A > G	A464T, Q172H and K262R	5	0.0017
*CYP2B6*44*	12828G > A	A102T	1	0.0003
*CYP2B6*45*	15662G > A	V183I	1	0.0003
*CYP2B6*46*	18008A > C	H247P	1	0.0003
*CYP2B6*47*	21088T > C and 21089A > G	Y354R	1	0.0003
*CYP2B6*48*	21173T > A	I382 N	7	0.0024
*CYP2B6*49*	25347G > A	R434Q	1	0.0003

3.2. Comparison of allele frequency in different biogeographical groups

The allele frequency of *CYP*2B6 in the Han Chinese population was compared to that of different biogeographical groups as illustrated in Table 3. The allele frequencies of *CYP2B6* are similar to those of other Asian populations, such as Central/South Asian and East Asian, but quite different from that of European, Latino, African American/Afro-Caribbean, Near Eastern, Oceanian and Sub-Saharan African. However, *CYP2B6*1* and *CYP2B6*6* are the most prevalent alleles for all the groups.

3.3. Genotype frequencies

Genotype analyses (Table 4) showed that the first three common genotypes were *CYP2B6*1/*1* (56.04%), *CYP2B6*1/*6* (18.41%) and *CYP2B6*1/*4* (10.92%). Additionally, *CYP2B6*1/*2*, *CYP2B6*2/*2*, *CYP2B6*2/*4*, *CYP2B6*4/*4*, *CYP2B6*1/*5*, *CYP2B6*2/*6*, *CYP2B6*6/*6*, *CYP2B6*1/*9*, *CYP2B6*1/*23* and *CYP2B6*1/*26* were also detected with frequencies of no more than 9.0%. 35 individuals had genotypes of *CYP2B6*1/39–49* (3.98%).

3.4. Haplotype determination of SNP

The haplotype of the novel nonsynonymous variant and all other variants in promoter, intron and untranslated region (UTR) in each sample was not clear because of the limitation of the Sanger sequencing method. Nanopore sequencing was then utilized to distinguish the haplotype of the novel variants both in coding and non-coding regions. The entire *CYP2B6* (30957 bp) gene was amplified with long-fragment PCR and finally sequenced with GridION flow-cell (R9.4.1). The haplotypes of the 11 novel non-synonymous variants and variants in non-coding regions were determined finally (Supplementary Fig. 2), and the star allele names of these haplotypes were assigned as *CYP2B6*39*–49 by PharmVar. These results indicated that long-PCR followed by long-read sequencing was a useful tool to distinguish the haplotype of long length genes such as *CYP2B6*.

Table 3
Allele frequencies of *CYP2B6* variants in this study and other biogeographical populations (from PharmVar).

Allele	Position (coding DNA)	Frequency								
		Current study	African American/Afro-Caribbean	Central/ South Asian	East Asian	European	Latino	Near Eastern	Oceanian	Sub-Saharan African
*CYP2B6*1*	/	0.7468	0.4135	0.6085	0.6400	0.4907	0.4974	0.4760	0.3400	0.3147
*CYP2B6*2*	64C > T	0.0428	0.0310	0.0410	0.0458	0.0492	0.0440	0.0390	0.0000	0.0310
*CYP2B6*4*	18053A > G	0.0654	0.0103	0.0990	0.0842	0.0409	0.1062	0.1040	0.0000	0.0000
*CYP2B6*5*	25505C > T	0.0017	0.0621	NA	0.0110	0.1155	0.0381	0.0240	0.0200	0.0200
*CYP2B6*6*	15631G > T , 18053A > G	0.1237	0.3170	0.1850	0.1745	0.2330	0.2116	0.2310	0.6200	0.3749
*CYP2B6*9*	15631G > T	0.0013	0.0465	0.0590	0.0344	0.0147	0.0727	0.1020	0.0100	NA
*CYP2B6*23*	25421A > G	0.0027	0.0000	0.0000	0.0021	0.0001	0.0000	NA	NA	0.0000

Table 4
Genotype frequencies and phenotypes of CYP2B6 in 1483 Chinese Han subjects.

Genotype	Hit	Frequency (%)	Phenotype Summary	Electronic Health Record (HER) Priority Notation
CYP2B6*1/*1	776	52.33	Normal Metabolizer	Normal/Routine/Low Risk
CYP2B6*1/*2	125	8.43	Normal Metabolizer	Normal/Routine/Low Risk
CYP2B6*2/*2	1	0.07	Normal Metabolizer	Normal/Routine/Low Risk
CYP2B6*1/*4	162	10.92	Rapid Metabolizer	Normal/Routine/Low Risk
CYP2B6*2/*4	7	0.47	Rapid Metabolizer	Normal/Routine/Low Risk
CYP2B6*4/*4	16	1.08	Ultrarapid Metabolizer	Normal/Routine/Low Risk
CYP2B6*1/*5	5	0.34	Normal Metabolizer	Normal/Routine/Low Risk
CYP2B6*1/*6	273	18.41	Intermediate Metabolizer	Abnormal/Priority/High Risk
CYP2B6*2/*6	13	0.88	Intermediate Metabolizer	Abnormal/Priority/High Risk
CYP2B6*6/*6	47	3.17	Poor Metabolizer	Abnormal/Priority/High Risk
CYP2B6*1/*9	4	0.27	Intermediate Metabolizer	Abnormal/Priority/High Risk
CYP2B6*1/*23	8	0.54	Indeterminate	none
CYP2B6*1/*26	11	0.74	Intermediate Metabolizer	Abnormal/Priority/High Risk
CYP2B6*1/*39	3	0.20	NA	NA
CYP2B6*1/*40	2	0.13	NA	NA
CYP2B6*1/*41	12	0.81	NA	NA
CYP2B6*1/*42	1	0.07	NA	NA
CYP2B6*1/*43	5	0.34	NA	NA
CYP2B6*1/*44	1	0.07	NA	NA
CYP2B6*1/*45	1	0.07	NA	NA
CYP2B6*2/*46	1	0.07	NA	NA
CYP2B6*1/*47	1	0.07	NA	NA
CYP2B6*1/*48	7	0.47	NA	NA
CYP2B6*1/*49	1	0.07	NA	NA

3.5. Functional prediction of novel variants with online PROVEAN and PolyPhen-2

Each of the 11 novel nonsynonymous variants in the *CYP2B6* gene was analyzed using the online tools PROVEAN and PolyPhen-2. Based on the analysis, it was found that the function of variants p.A102T, p.H247P, p.Y380H, p.I382 N, p.R434Q, and p.R443H was predicted to be damaged by both PROVEAN and PolyPhen-2. On the other hand, variants p.V183I, p.Y354H, and p.A464T were predicted to have a "Tolerated" or "Benign" function. The function of variants p.R140W and p.R378Q showed inconsistent predictions between PROVEAN and PolyPhen-2 (Supplementary Table 3).

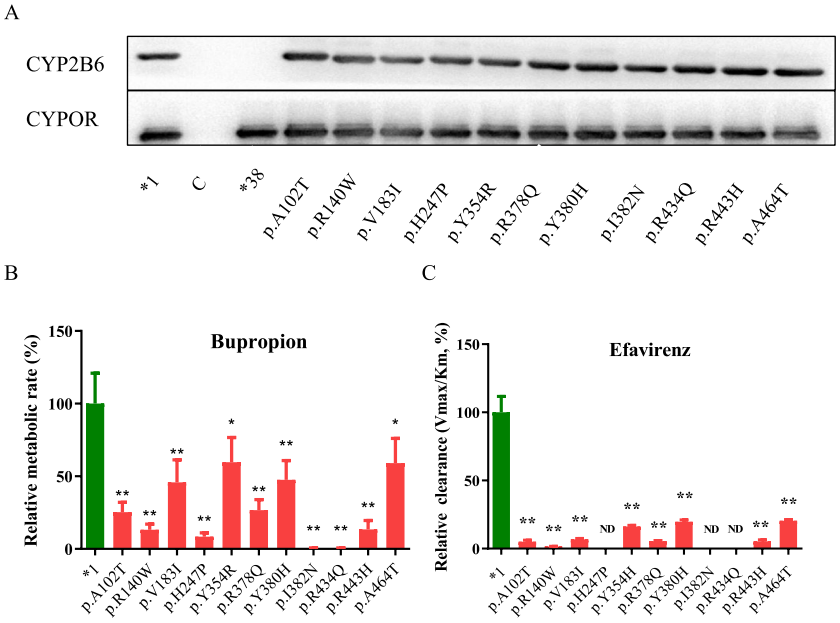


Fig. 1. *In vitro* enzymatic activities of *CYP2B6* variants. A. Protein expression of *CYPOR* and *CYP2B6* in each expression vector assessed by western blotting (Original images of gels and blots were in Supplementary Fig. 3). *1, wild type enzyme. C, control (empty vector). **38*, a truncated enzyme. p.A102T, p.R140W, p.V183I, p.H247P, p.Y354H, p.R378Q, p.Y380H, p.I382 N, p.R434Q, p.R443H and p.A464T were the novel variants found in our study. B. Relative metabolic rate (hydroxybupropion to bupropion hydrochloride) of each variant comparing with wide type **1*. C. Relative clearance rate (Vmax/Km) of each variant on efavirenz comparing with wide type **1*. ND, not detectable. Student's *t*-test, *P < 0.05, **P < 0.01.

3.6. In vitro enzymatic activities of CYP2B6 variants

A total of 11 novel nonsynonymous variants in the coding regions of the *CYP2B6* gene were identified in this study. The enzyme characteristics of these novel variants expressed in Sf21 cells were assessed with UPLC-MS/MS using bupropion and efavirenz respectively as the substrate. The relative ratio of hydroxybupropion to bupropion hydrochloride and relative clearance rate (Vmax/Km) on efavirenz were calculated to assess the enzyme activities of *CYP2B6*1* (*in vitro* activity was set as "1") and 11 novel variants. The results showed that all the novel variants (p.A102T in **44*, p.R140W in **39*, p.V183I in **45*, p.H247P in **46*, p.Y354R in **47*, p.R378Q in **40*, p.Y380H in **41*, p.I382 N in **48*, p.R434Q in **49*, p.R443H in **42*, and p.A464T in **43*) exhibited significant lower activities than the reference *CYP2B6*1* (Fig. 1. A, B and C). As a result, patients with these variants may require a lower oral dose of bupropion to achieve the desired concentration in the bloodstream and therapeutic effects.

4. Discussion

CYP2B6 accounts for 2–5% of total *CYP* content in liver and is thought to metabolize about 8% of clinical drugs, including bupropion, EFV, nevirapine, meperidine, methadone, ketamine, propofol, mephobarbital, and cyclophosphamide [15–17]. Owing to genetic variants, inducibility or inhibition by many compounds, *CYP2B6* activity is highly variable between individuals. Some variants of *CYP2B6* affect transcriptional regulation, splicing, protein expression and thus catalytic activity.

The PharmVar website lists 38 star-alleles, each of which has a variant amino acid sequence or demonstrates functional effect (Last Updated: Jun 1st, 2022). According to the Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for *CYP2B6* and EFV-containing antiretroviral therapy, the *CYP2B6* phenotype can be classified into 5 categories as ultrarapid, rapid, normal, intermediate and poor metabolizer [18]. According to CPIC Guidelines for *CYP2B6* variants, 63.53 % of Han Chinese population are normal metabolizers (Normal/Routine/Low Risk), 3.17% are poor metabolizers (Abnormal/Priority/High Risk), 20.84% are intermediate metabolizers (Abnormal/Priority/High Risk), 11.39% are rapid metabolizers (Normal/Routine/Low Risk), and 1.08% are ultrarapid metabolizers (Normal/Routine/Low Risk). The remaining 2.37% remained undefined (Table 4). Importantly, the "Low Risk" assigned to rapid and ultrarapid metabolizers indicates that the blood concentration of the drug may not be higher than normal, but there is a real risk of therapy failure and drug resistance because of lower drug concentrations in circulation.

Dose adjustment based on genotype demonstrated that the dose of EFV could be reduced successfully and side effects could also be decreased [19,20]. A prior dose reduction in homozygous *CYP2B6*6* (c.516G > T and c.785A > G) patients could maintain the drug exposure in the therapeutic range [21]. An *in vitro* study in mammalian cells showed that the c.516G > T variant resulted in erroneous splicing and lower expression of mRNA [22]. One hypothesis suggested that the binding of splice factors to exon 4 were affected by this variant [23]. All haplotypes harboring the variant c.516G > T were designated as decreased enzymatic activity (*CYP2B6*6*, **7*, **9*, **19*, **20*, **26*, **34*, and **36*) or no activity owing to the presence of other variants such as *CYP2B6*13*, **37*, and *38 by the CPIC.

*CYP2B6*6* functional differences were also affected by variants-related inhibitors. The c.785A > G variant resulted in lower susceptibility to inhibitors such as sertraline and clopidogrel compared to *CYP2B6*1* [24]. Methadone dose requirements for effective treatment of opioid addiction was shown to be significantly reduced in carriers of this genotype [25]. Further *in vivo* investigation is needed to confirm the role of genetic variants in drug–drug interaction of *CYP2B6*. Increased *in vitro* transcription [26] and increased activity *in vivo* [27] were observed in *CYP2B6*22* allele (c.-82T > C), which is a gain-of-function variant. Previous reports [26] suggest that a c.-82T > C variant could alter the TATA-box into CCAAT, an enhancer-binding site to increase transcription from an alternative initiation site.

Some genetic studies on Han Chinese population have been performed on genetic variants of *CYP2B6* in Guangdong (unrelated healthy subjects, n = 193) [12], Shanghai (HIV-infected outpatients, n = 376) [28], Beijing (unrelated healthy subjects, n = 223) [29] and Shandong (unrelated healthy subjects, n = 179) [30]. Most of those studies did not focus on the effects of the variants on the function of *CYP2B6* and its metabolism of specific drugs. In this study, based on the Sanger sequencing method and a large number of subjects (n = 1483), a total of 11 nonsynonymous (c.12828G > A, 0.03%; c.13075C > T, 0.1%; c.15662G > A, 0.03%; c.18008A > C, 0.03%; c.21088T > C and c.21089A > G, 0.03%; c.21161G > A, 0.07%; c.21166T > C, 0.4%; c.21173T > A, 0.24%; c.25347G > A, 0.03%; c. 25374G > A, 0.03%; c. 25436G > A, 0.17%) and 7 synonymous (c. 12776A > T, 0.03%; c. 21168C > T, 0.03%; c. 21434C > T, 0.03%; c. 25372C > T, 0.03%; c. 25381A > G, 0.03%; c. 25438C > T, 0.27%; c. 25441A > T, 0.27%) variants in the coding region of the *CYP2B6* gene were detected and most of them are rare variants. To further investigate the enzymatic activity of these novel variants *in vitro*, wild-type *CYP2B6*1* and 11 novel nonsynonymous variants were expressed in Sf12 cells, and the enzymatic activities of these variants towards bupropion and efavirenz were investigated. As showed in Fig. 1B and C, all 11 variants (p.A102T in **44*, p.R140W in **39*, p.V183I in **45*, p.H247P in **46*, p.Y354R in **47*, p.R378Q in **40*, p.Y380H in **41*, p.I382 N in **48*, p.R434Q in **49*, p.R443H in **42*, and p.A464T in **43*) exhibited significantly lower activity compared to *CYP2B6*1* on both bupropion and efavirenz ($P < 0.05$). Additionally, when using the online prediction tools PROVEAN and PolyPhen-2, it was observed that over half of these novel variants (p.A102T, p.H247P, p.Y380H, p.I382 N, p.R434Q, and p.R443H) were predicted to have damaged functions by both tools (Supplementary Table 3).

According to CPIC, the single variant c.18053A > G (**4*) leads to increased function, while the single variant c.15631G > T (**9*) results in decreased function. When these variants are combined, such as in the case of c.18053A > G (p.K262R) and c.15631G > T (p.Q172H), the overall function is decreased, indicating that c.15631G > T plays a major role in enzymatic activity. All alleles carrying variants c.15631G > T and c.18053A > G exhibit lower activity (**6*, **7*, **9*, **19*, **20*, **26*, **34*, and **36*), except for **13*, **37*, and **38*, which show no activity. In our study, we identified five novel variants (**39-*43*) that carry these two variants along with the novel base variant. Each of these individual novel variants in **39-*43* demonstrated lower enzymatic activity *in vitro* experiments. It is likely that

**39-*43* would exhibit even lower or no activity when expressed as triple variants, as compared to **1*. Since c.15631G > T (18.41%) and c.18053A > G (26.57%) have a high frequency and play a significant role in determining the function of *CYP2B6*, we recommend testing these two sites in a clinical setting. It would be beneficial if they could be distinguished in trans or in cis simultaneously using advanced methods such as nanopore sequencing.

Due to differences among expression systems and substrates used *in vitro* enzymatic activity analysis may produce inconsistent results [17,31,32]. For example, the *CYP2B6*6* allele is associated with enhanced cyclophosphamide metabolism, but appears to confer decreased EFV metabolism [17,33]. (S)-methadone plasma levels in *CYP2B6*6/*6* carriers were increased and resulted in methadone associated deaths or higher risk of severe cardiac arrhythmias [34,35]. Thus, functional data should be extrapolated with caution except for true no function alleles such as *CYP2B6*38* which has a premature stop codon and results in a truncated protein. The *in vivo* enzymatic activities of these 11 variants should be further investigated in future studies.

Because *CYP2B6* variants have been shown to affect catalytic activity, splicing, transcriptional regulation and protein expression, sequencing of the entire gene (including promoter, 5′UTR, intron and 3′UTR) was needed to investigate the relationship between different variants in different positions (haplotype and genotype). In this study, we employed the Sanger method to perform bidirectional sequencing of the promoter region, all exonic regions, and the 3′UTR region of the CYP2B6 gene in 1483 study subjects. This is, to our knowledge, the largest-scale fine sequencing of the CYP2B6 single gene conducted in China. More importantly, for the newly discovered variants, we used nanopore sequencing to further distinguish the haplotypes of these new variants, achieving, for the first time, full-length haplotype sequencing of the CYP2B6 gene at >30 kb level, which was almost unattainable with previous sequencing methods. The whole *CYP2B6* gene (30957 bp) was amplified with long-fragment PCR and finally sequenced using a GridION flow-cell (R9.4.1). In this way, we have determined the haplotypes of the 11 novel nonsynonymous variants and variants in non-coding regions (Supplementary Fig. 2 and Supplementary Table 2). The results show that the relationship between different variants in the whole gene were distinguished, including both haplotype and genotype, which may provide robust evidence for metabolizer level prediction and decrease the gap between gene diagnosis and clinical phenotype.

5. Ethics declaration

This study was reviewed and approved by Institutional Review Boards of Beijing Hospital, with the approval number: 2019BJYYEC-054-02.

All participants provided informed consent to participate in the study.

Funding

This work was supported by Beijing Hospital Project (BJ-2019-186) and National Key R&D Program of China (2020YFC2008300).

Data availability statement

Sequencing data of novel variants in this study has been deposited into DNA Data Bank of Japan and National Center for Biotechnology Information with accession numbers of LC761190, LC761191, LC761192, LC761193, LC761194, LC761195, LC761196, LC761197, LC761198, LC761199 and LC761200. Further experimental details and data are available upon request.

CRediT authorship contribution statement

Li-Qun Zhang: Writing – original draft, Formal analysis. **Xin-Yue Li:** Methodology. **Lian-Guo Chen:** Resources, Data curation. **Zhe Chen:** Resources, Methodology. **Ren-Ai Xu:** Data curation. **Jian-Chang Qian:** Validation, Methodology. **Xiao-yang Zhou:** Investigation, Data curation. **Da-Peng Dai:** Funding acquisition. **Guo-Xin Hu:** Supervision. **Jian-Ping Cai:** Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Da-peng Dai reports financial support was provided by Ministry of Science and Technology of the People's Republic of China. Da-Peng Dai reports a relationship with Ministry of Science and Technology of the People's Republic of China that includes: funding grants.

5. Acknowledgements

Thanks a lot for PhD. Gao-Yuan Sun and professor Fei Xiao from Beijing Hospital in providing the platform of nanopore sequencing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e28952.

References

[1] D.R. Nelson, L. Koymans, T. Kamataki, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estabrook, I.C. Gunsalus, D. W. Nebert, P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature, Pharmacogenetics 6 (1996) 1–42.

[2] D.R. Nelson, D.C. Zeldin, S.M. Hoffman, L.J. Maltais, H.M. Wain, D.W. Nebert, Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants, Pharmacogenetics 14 (2004) 1–18.

[3] M. Hidestrand, M. Oscarson, J.S. Salonen, L. Nyman, O. Pelkonen, M. Turpeinen, M. Ingelman-Sundberg, CYP2B6 and CYP2C19 as the major enzymes responsible for the metabolism of selegiline, a drug used in the treatment of Parkinson's disease, as revealed from experiments with recombinant enzymes, Drug metabolism and disposition: the biological fate of chemicals 29 (2001) 1480–1484.

[4] K. Purnapatre, S.K. Khattar, K.S. Saini, Cytochrome P450s in the development of target-based anticancer drugs, Cancer letters 259 (2008) 1–15.

[5] S. Ekins, M. Vandenbranden, B.J. Ring, J.S. Gillespie, T.J. Yang, H.V. Gelboin, S.A. Wrighton, Further characterization of the expression in liver and catalytic activity of CYP2B6, J. Pharmacol. Exp. Therapeut. 286 (1998) 1253–1259.

[6] S.R. Faucette, R.L. Hawke, E.L. Lecluyse, S.S. Shord, B. Yan, R.M. Laethem, C.M. Lindley, Validation of bupropion hydroxylation as a selective marker of human cytochrome P450 2B6 catalytic activity, Drug metabolism and disposition: the biological fate of chemicals 28 (2000) 1222–1230.

[7] S.M. Hoffman, D.R. Nelson, D.S. Keeney, Organization, structure and evolution of the CYP2 gene cluster on human chromosome 19, Pharmacogenetics 11 (2001) 687–698.

[8] A. Telenti, U.M. Zanger, Pharmacogenetics of anti-HIV drugs, Annu. Rev. Pharmacol. Toxicol. 48 (2008) 227–256.

[9] N.Y. Rakhmanina, J.N. van den Anker, Efavirenz in the therapy of HIV infection, Expet Opin. Drug Metabol. Toxicol. 6 (2010) 95–103.

[10] T.N. Lam, K.H. Hui, D.P. Chan, S.S. Lee, Genotype-guided dose adjustment for the use of efavirenz in HIV treatment, J. Infect. 71 (2015) 607–609.

[11] Z. Desta, A. El-Boraie, L. Gong, A.A. Somogyi, V.M. Lauschke, C. Dandara, K. Klein, N.A. Miller, T.E. Klein, R.F. Tyndale, M. Whirl-Carrillo, A. Gaedigk, PharmVar GeneFocus: CYP2B6, Clinical pharmacology and therapeutics 110 (2021) 82–97.

[12] S. Guan, M. Huang, E. Chan, X. Chen, W. Duan, S.F. Zhou, Genetic polymorphisms of cytochrome P450 2B6 gene in Han Chinese, European journal of pharmaceutical sciences, official journal of the European Federation for Pharmaceutical Sciences 29 (2006) 14–21.

[13] X.Y. Zhou, X.X. Hu, M.F. Li, H. Wang, L.Q. Zhang, G.X. Hu, J.P. Cai, Functional characterization of CYP2C19 variants in nebivolol 4-hydroxlation in vitro, Drug Test. Anal. 10 (2018) 807–813.

[14] X. Hu, T. Lan, D. Dai, R.A. Xu, L. Yuan, Q. Zhou, Y. Li, J. Cai, G. Hu, Evaluation of 24 CYP2D6 Variants on the Metabolism of Nebivolol In Vitro, Drug metabolism and disposition: the biological fate of chemicals 44 (2016) 1828–1831.

[15] Z. Desta, T. Saussele, B. Ward, J. Blievernicht, L. Li, K. Klein, D.A. Flockhart, U.M. Zanger, Impact of CYP2B6 polymorphism on hepatic efavirenz metabolism in vitro, Pharmacogenomics 8 (2007) 547–558.

[16] M. Nakajima, S. Komagata, Y. Fujiki, Y. Kanada, H. Ebi, K. Itoh, H. Mukai, T. Yokoi, H. Minami, Genetic polymorphisms of CYP2B6 affect the pharmacokinetics/pharmacodynamics of cyclophosphamide in Japanese cancer patients, Pharmacogenetics Genom. 17 (2007) 431–445.

[17] U.M. Zanger, K. Klein, Pharmacogenetics of cytochrome P450 2B6 (CYP2B6): advances on polymorphisms, mechanisms, and clinical relevance, Front. Genet. 4 (2013) 24.

[18] Z. Desta, R.S. Gammal, L. Gong, M. Whirl-Carrillo, A.H. Gaur, C. Sukasem, J. Hockings, A. Myers, M. Swart, R.F. Tyndale, C. Masimirembwa, O.F. Iwuchukwu, S. Chirwa, J. Lennox, A. Gaedigk, T.E. Klein, D.W. Haas, Clinical pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2B6 and efavirenz-containing antiretroviral therapy, Clinical pharmacology and therapeutics 106 (2019) 726–733.

[19] H. Gatanaga, T. Hayashida, K. Tsuchiya, M. Yoshino, T. Kuwahara, H. Tsukada, K. Fujimoto, I. Sato, M. Ueda, M. Horiba, M. Hamaguchi, M. Yamamoto, N. Takata, A. Kimura, T. Koike, F. Gejyo, S. Matsushita, T. Shirasaka, S. Kimura, S. Oka, Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 2B6 *6 and *26, Clinical infectious diseases, an official publication of the Infectious Diseases Society of America 45 (2007) 1230–1237.

[20] H. Gatanaga, S. Oka, Successful genotype-tailored treatment with small-dose efavirenz, Aids 23 (2009) 433–434.

[21] C. Nyakutira, D. Roshammar, E. Chigutsa, P. Chonzi, M. Ashton, C. Nhachi, C. Masimirembwa, High prevalence of the CYP2B6 516G–>T(*6) variant and effect on the population pharmacokinetics of efavirenz in HIV/AIDS outpatients in Zimbabwe, Eur. J. Clin. Pharmacol. 64 (2008) 357–365.

[22] M.H. Hofmann, J.K. Blievernicht, K. Klein, T. Saussele, E. Schaeffeler, M. Schwab, U.M. Zanger, Aberrant splicing caused by single nucleotide polymorphism c.516G>T [Q172H], a marker of CYP2B6*6, is responsible for decreased expression and activity of CYP2B6 in liver, J. Pharmacol. Exp. Therapeut. 325 (2008) 284–292.

[23] W. Sadee, D. Wang, A.C. Papp, J.K. Pinsonneault, R.M. Smith, R.A. Moyer, A.D. Johnson, Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy, Clinical pharmacology and therapeutics 89 (2011) 355–365.

[24] J.C. Talakad, S. Kumar, J.R. Halpert, Decreased susceptibility of the cytochrome P450 2B6 variant K262R to inhibition by several clinically important drugs, Drug metabolism and disposition: the biological fate of chemicals 37 (2009) 644–650.

[25] O. Levran, E. Peles, S. Hamon, M. Randesi, M. Adelson, M.J. Kreek, CYP2B6 SNPs are associated with methadone dose required for effective treatment of opioid addiction, Addiction Biol. 18 (2013) 709–716.

[26] J. Zukunft, T. Lang, T. Richter, K.I. Hirsch-Ernst, A.K. Nussler, K. Klein, M. Schwab, M. Eichelbaum, U.M. Zanger, A natural CYP2B6 TATA box polymorphism (-82T–> C) leading to enhanced transcription and relocation of the transcriptional start site,, Mol. Pharmacol. 67 (2005) 1772–1782.

[27] M. Rotger, H. Tegude, S. Colombo, M. Cavassini, H. Furrer, L. Decosterd, J. Blievernicht, T. Saussele, H.F. Gunthard, M. Schwab, M. Eichelbaum, A. Telenti, U. M. Zanger, Predictive value of known and novel alleles of CYP2B6 for efavirenz plasma concentrations in HIV-infected individuals, Clinical pharmacology and therapeutics 81 (2007) 557–566.

[28] X.M. Meng, Z.R. Li, X.Y. Zheng, Y.X. Liu, W.J. Niu, X.Y. Qiu, H.Z. Lu, Effect of albumin and CYP2B6 polymorphisms on exposure of efavirenz: a population pharmacokinetic analysis in Chinese HIV-infected adults, European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences 167 (2021) 105986.

[29] G.Z. Qi, Z.Y. Zhang, X. Wang, S.J. Yin, Y.Q. Lou, G.L. Zhang, Functional allele and genotype frequencies of CYP1A2, CYP2B6 and iNOS among mainland Chinese Tibetan, Mongolian, Uygur and Han populations, J. Clin. Pharm. Therapeut. 41 (2016) 84–91.

[30] L. Tan, J.T. Yu, Y.P. Sun, J.R. Ou, J.H. Song, Y. Yu, The influence of cytochrome oxidase CYP2A6, CYP2B6, and CYP2C9 polymorphisms on the plasma concentrations of valproic acid in epileptic patients,, Clin. Neurol. Neurosurg. 112 (2010) 320–323.

[31] T. Watanabe, T. Saito, E.M.G. Rico, E. Hishinuma, M. Kumondai, M. Maekawa, A. Oda, D. Saigusa, S. Saito, J. Yasuda, M. Nagasaki, N. Minegishi, M. Yamamoto, H. Yamaguchi, N. Mano, N. Hirasawa, M. Hiratsuka, Functional characterization of 40 CYP2B6 allelic variants by assessing efavirenz 8-hydroxylation, Biochem. Pharmacol. 156 (2018) 420–430.

[32] T. Watanabe, K. Sakuyama, T. Sasaki, Y. Ishii, M. Ishikawa, N. Hirasawa, M. Hiratsuka, Functional characterization of 26 CYP2B6 allelic variants (CYP2B6.2-CYP2B6.28, except CYP2B6.22), Pharmacogenetics Genom. 20 (2010) 459–462.

[33] N. Ariyoshi, M. Ohara, M. Kaneko, S. Afuso, T. Kumamoto, H. Nakamura, I. Ishii, T. Ishikawa, M. Kitada, Q172H replacement overcomes effects on the metabolism of cyclophosphamide and efavirenz caused by CYP2B6 variant with Arg262, Drug metabolism and disposition: the biological fate of chemicals 39 (2011) 2045–2048.

[34] S. Crettol, J.J. Deglon, J. Besson, M. Croquette-Krokkar, I. Gothuey, R. Hammig, M. Monnat, H. Huttemann, P. Baumann, C.B. Eap, Methadone enantiomer plasma levels, CYP2B6, CYP2C19, and CYP2C9 genotypes, and response to treatment, Clinical pharmacology and therapeutics 78 (2005) 593–604.

[35] C.B. Eap, S. Crettol, J.S. Rougier, J. Schlapfer, L. Sintra Grilo, J.J. Deglon, J. Besson, M. Croquette-Krokar, P.A. Carrupt, H. Abriel, Stereoselective block of hERG channel by (S)-methadone and QT interval prolongation in CYP2B6 slow metabolizers, Clinical pharmacology and therapeutics 81 (2007) 719–728.