

# Genetic Polymorphism of Drug Metabolic Gene *CYPs*, *VKORC1*, *NAT2*, *DPYD* and *CHST3* of Five Ethnic Minorities in Heilongjiang Province, Northeast China

Tingting Zhang<sup>1,2,\*</sup>  
 Qiuyan Li<sup>1-3,\*</sup>  
 Bonan Dong<sup>1,2</sup>  
 Xiao Liang<sup>1,2</sup>  
 Mansha Jia<sup>4</sup>  
 Jing Bai<sup>1,2</sup>  
 Jingcui Yu<sup>2,4</sup>  
 Songbin Fu<sup>1,2</sup>

<sup>1</sup>Laboratory of Medical Genetics, Harbin Medical University, Harbin, People's Republic of China; <sup>2</sup>Key Laboratory of Preservation of Human Genetic Resources and Disease Control in China (Harbin Medical University), Ministry of Education, Harbin, People's Republic of China; <sup>3</sup>Editorial Department of International Journal of Genetics, Harbin Medical University, Harbin, People's Republic of China; <sup>4</sup>Scientific Research Centre, The Second Affiliated Hospital of Harbin Medical University, Harbin, People's Republic of China

\*These authors contributed equally to this work

**Introduction:** Genetic variability in genes encoding drug-metabolizing enzymes may contribute to the heterogeneity of drug responses in different populations. Extensive research in pharmacogenomics in major populations around the world provides us with a great deal of information about drug-related genetic polymorphisms.

**Objective:** The purpose of this study was to detect the genetic variation of drug-metabolism-related genes in the five ethnic minorities Daur, Hezhen, Ewenki, Mongolian and Manchu in China, and to analyze the distribution differences among ethnic groups.

**Methods:** We genotyped 32 SNPs of drug metabolism genes in 882 healthy Chinese volunteers from five ethnic groups. The genotype frequency and allele frequency of the five ethnic groups were calculated, and the different variants among the five ethnic groups were compared by chi-square test. Genetic parameters were analyzed using Popgene software. The genetic structure of five ethnic minorities was analyzed by principal component analysis, and compared with 26 populations.

**Results:** We found that SNPs of genes related to drug metabolism existed diversity in different populations. Among them, rs192766 and rs9419082 in *CYP2E1* showed statistical differences between Daur and Manchu, and *NAT2* rs1801280 showed statistical differences between Hezhen and Mongolian. In addition, the five populations we studied had the smallest differences with EAS populations. There was haplotype diversity in *CHST3*, *VKORC1*, *CYP1A2* and *CYP2E1* genes in the five ethnic minorities, and these haplotype polymorphisms were related to the use of corresponding drug doses. Cluster analysis shows that the five ethnic minorities in Heilongjiang Province are clustered together with the EAS populations.

**Conclusion:** These results suggest that understanding the diversity of drug-related genetic markers is critical for individualized drug gene therapy programs in ethnic minorities in China as well as in populations highly mixed with these ethnic groups.

**Keywords:** pharmacogenomics, population genetic, single nucleotide polymorphism, Chinese minorities

## Introduction

In recent years, the safety and efficacy of drugs are important factors of concern in clinical treatment and drug development.<sup>1,2</sup> A study conducted by the World Health Organization shows that about one-third of the deaths of patients worldwide each year are closely related to the irrational use of drugs.<sup>3</sup> In China, unreasonable drug use phenomenon is also quite prominent. It has been reported that at least

Correspondence: Songbin Fu; Jingcui Yu  
 Email fusb@ems.hrbmu.edu.cn;  
 yujingcui@ems.hrbmu.edu.cn



2 million people are hospitalized due to adverse drug reactions every year, and about 200,000 people die as a result.<sup>4</sup> In addition, adverse drug reactions also cause huge costs to the pharmaceutical industry. In recent decades, with the development of the association between pharmacogenomics and population genetics and the completion of the Human Genome Project, many studies have begun to focus on the genetic basis of individual differences in drug response.<sup>5,6</sup> It is gradually recognized that the heterogeneity of drug response in different individuals is mainly due to the differences of drug metabolism-related genes.<sup>2,5,7,8</sup> Therefore, the exploration of drug-metabolism-related gene polymorphism is beneficial to optimize drug selection and dosage, and provides the theoretical basis for personalized medicine.<sup>9</sup>

Drug metabolism involves the interaction of various drug metabolism enzymes, including stage Phase I enzymes, Phase II conjugation enzymes, transporters, cytochrome reductases, and nuclear receptors. One of the most important drug metabolism enzymes is the cytochrome P450 (CYP) family. CYP1A2, CYP2C19, CYP2D6 and CYP3A4/5 are the main human enzymes involved in drug metabolism and metabolize most clinical drugs.<sup>10</sup> However, SNPs of these drug-metabolism-related genes may lead to changes in mRNA activity, gene expression, and even translation or splicing, leading to genetic polymorphisms in drug metabolism among different individuals or populations.<sup>11–14</sup> For example, rs4244285 and rs4986893 of *CYP2C19* gene are the most common major mutations in the Chinese population, which have been shown to profoundly affect the uptake of clopidogrel for the treatment of coronary artery disease, and ultimately affect drug response polymorphism.<sup>6</sup> In addition, individuals with allelic variants may exhibit one of the following metabolic phenotypes: poor metabolizers (PM; A combination of alleles with no function or reduced function and individuals with low or no enzyme activity), intermediate metabolizers (IM; Individuals with a combination of normal and reduced functional alleles that cause reduced enzyme activity); normal metabolizers (NM; Individuals with two normal alleles and full functional enzyme activity) or rapid and ultra-rapid metabolizers (UM; individuals with two increased function alleles or more than two normal function alleles).<sup>15</sup> Studies have shown that due to the differences in genetic backgrounds of different populations, the distribution of some gene metabolic phenotypes presents corresponding population differences.<sup>2,16–18</sup>

However, most genetic studies (eg, GWAS) focus on populations of European and African origin, while genetic

studies in China are dominated by Han nationality.<sup>19–22</sup> In China, where many ethnic groups live in compact communities, different ethnic groups have different genetic backgrounds. For example, the Mongolians are a minority ethnic group with a large population in China, and they are significantly different from other ethnic groups in terms of facial features. Similar to the Mongolians, the Daur, Hezhen, Ewenki and Manchu have different genetic backgrounds and the diversity of their environment, including pharmacogenetic difference.<sup>9</sup> Previous studies have compared allele and genotype polymorphisms in ADME in several populations (Tajik, Uygur, Kazak, Kirgiz and Hui).<sup>2,9</sup> In this study, we investigated the genetic diversity of drug metabolism genes of Daur, Hezhen, Ewenki, Mongolian and Manchu, in an attempt to identify highly different SNPs and further explore their clinical significance.<sup>2</sup>

## Materials and Methods

### Study Participants

We recruited 882 volunteers (413 males, 469 females, 45 average age) from Daur (n = 233), Hezhen (n = 250), Ewenki (n = 73), Mongolian (n = 106) and Manchu (n = 220) nationalities in China. Participants from all five ethnic groups were from Heilongjiang province in China. Each participant signed an informed consent form. Each subject has at least three generations of the same ethnic origin, minimizing known environmental and therapeutic factors that affect genetic variation. The study was approved by the Ethics Committee of Harbin Medical University and carried out in strict accordance with the Helsinki Declaration. In addition, we collected and downloaded raw genotype data of 2504 individuals in 26 population groups globally from Ensemble database at [http://grch37.ensembl.org/Homo\\_sapiens/Tools/DataSlicer](http://grch37.ensembl.org/Homo_sapiens/Tools/DataSlicer) for follow-up analysis. Explanations of abbreviations for different groups are shown in [Supplementary Table 1](#).

### DNA Extraction and Concentration Determination

An appropriate amount of peripheral blood samples were taken from the participants. DNA was extracted from the peripheral blood of 882 subjects using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) as recommended by the manufacturer. Concentration and purity of DNA are measured using NanoDrop One spectrophotometer (ThermoFisher, Thermo Fisher Scientific, Madison, USA).

Genotyping was performed using the SNPscan™ Kit (Genesky Biotechnologies Inc., Shanghai, China) according to the manufacturer's instructions.

## Variant Selection and Genotyping

From SNP database, PharmGKB database and published literature,<sup>2,23</sup> relatively important SNPs related to drug metabolism were screened according to the following principles: First, Very Important Pharmacogenes (VIP) variants in PharmGKB database were searched. We screened out clinically relevant variants with a high level of clinical annotation. Then we used dbSNP database to conduct frequency search for the screened variants, and retained the variants with altered alleles greater than 0.05 in the Han population of Beijing, China (CHB), in order to increase the statistical efficacy.<sup>24</sup> Finally, PGx biomarkers that have been reported and recommended for drug dose adjustment by global drug regulatory agencies are selected as candidate SNPs from these SNPs according to published literature.<sup>25</sup> Based on the above criteria and methods, we used SNPscan™ technology to genotype candidate SNPs, and excluded those SNPs that could not be successfully genotyped. Finally, 32 SNPs in 10 genes were reserved as research objects.

## Statistical Analyses

The test of Hardy-Weinberg equilibrium (HWE) and haplotype were estimated using the Haploview 4.2. Chi-square test was used to assess the allele differences between two ethnic groups. We performed multiple corrections using Bonferroni. Subsequently, genetic diversity was analyzed using Population Genetic Analysis (version Popgene32) software. R i386 4.0.2 software was used to plot the principal component analysis (PCA) based on raw genotype and allele frequency, respectively. Phylogenetic neighbor-joining (NJ) tree was reconstructed by MEGA7 software v5.0 using Dendrogram index.

## Results

### The Allele Frequencies of 32 Variants in Different Ethnic Populations

The genotype statistics in overall population is shown in Table 1, and the results show that all SNPs meet the Hardy-Weinberg equilibrium. The allele frequencies of five ethnic minorities from this study and 26 populations from the 1000 Genome Project are shown in [Supplementary Table 2](#).

### Pairwise Comparison of Five Ethnic Populations Screened for Differential SNP

In addition, we compared the SNPs of five ethnic minorities in pairs. After Bonferroni correction ( $0.05/(32*5)$ ), the results showed in [Supplementary Table 3](#) that rs9419082 and rs8192766 of *CYP2E1* gene had statistical difference between Daur and Manchu, and rs1801280 of *NAT2* gene had statistical difference between Hezhen and Mongolian.

### Five Ethnic Populations Were Compared with 26 Populations to Screen the Difference SNP

When Daur was compared with 26 populations ([Supplementary Table 4](#)), except for CHB, JPT and MSL, rs28360521 in *CYP2D6* gene was significantly different after Bonferroni correction ( $0.05/(32*27)$ ). In addition, the allele frequency of Daur population was the least different from EAS populations, but the most different from AFR populations. When Ewenki population is compared with East Asian populations ([Supplementary Table 5](#)), rs28360521 of *CYP2D6* gene is statistically different from KHV, rs8192766 of *CYP2E1* gene is statistically different from CDX after Bonferroni correction ( $0.05/(32*27)$ ). In addition, the alleles of Ewenki were the most different from AFR and EUR, and the least different from EAS population. After Bonferroni correction ( $0.05/(32*27)$ ), rs4148943 and rs730720 in *CHST3* gene in Manchu were statistically different from AFR, AMR, EUR and GIH populations when compared with 26 populations ([Supplementary Table 6](#)). Except EAS and BEB, *CYP1A1* rs4646421 was statistically different from other 20 populations in Manchu after Bonferroni correction ( $0.05/(32*27)$ ). After Bonferroni correction ( $0.05/(32*27)$ ), *CYP2C19* rs12248560 showed statistical difference between Manchu and 20 populations, except EAS and PEL when compared with 26 populations. In addition, as shown in [Supplementary Table 7](#) and [Supplementary Table 8](#), when comparing with the 26 populations, both Hezhen and Mongolian have a large difference with AFR, AMR, EUR and SAS, but a small difference with EAS, after Bonferroni correction ( $0.05/(32*27)$ ).

### Haplotype Analysis of Five Ethnic Populations

The haplotype block was constructed by Haploview software with  $D' > 0.8$ . The results show that the distribution

**Table 1** Genotype Count and HWP of 882 Individuals

SNP	Gene	Chr	Position	Allele		HLJ			HWP val
				A	B	AA	AB	BB	
rs4646421	CYP1A1	15	74,723,851	G	A	318	384	133	>0.05
rs4646422	CYP1A1	15	74,722,964	C	T	686	187	9	>0.05
rs2606345	CYP1A1	15	74,724,835	C	A	735	139	8	>0.05
rs2470890	CYP1A2	15	74,755,085	T	C	11	201	670	>0.05
rs2472304	CYP1A2	15	74,751,897	G	A	668	201	12	>0.05
rs762551	CYP1A2	15	74,749,576	C	A	118	412	352	>0.05
rs12248560	CYP2C19	10	94,761,900	C	T	836	35	0	>0.05
rs4244285	CYP2C19	10	94,781,859	G	A	519	305	57	>0.05
rs4986893	CYP2C19	10	94,780,653	G	A	776	103	3	>0.05
rs1080983	CYP2D6	22	42,132,561	C	T	536	291	51	>0.05
rs28360521	CYP2D6	22	42,132,969	C	T	320	397	156	>0.05
rs2031920	CYP2E1	10	133,526,341	C	T	572	279	29	>0.05
rs2070676	CYP2E1	10	133,537,633	G	C	20	185	676	>0.05
rs8192766	CYP2E1	10	133,525,881	T	G	383	392	105	>0.05
rs9419082	CYP2E1	10	133,520,268	A	G	103	403	375	>0.05
rs4646437	CYP3A4	7	99,767,460	G	A	757	110	8	>0.05
rs2242480	CYP3A4	7	99,763,843	C	T	582	263	32	>0.05
rs2359612	VKORC1	16	31,092,475	A	G	741	132	5	>0.05
rs7294	VKORC1	16	31,091,000	C	T	752	124	5	>0.05
rs8050894	VKORC1	16	31,093,188	C	G	5	133	742	>0.05
rs9934438	VKORC1	16	31,093,557	G	A	5	133	743	>0.05
rs1208	NAT2	8	18,400,806	G	A	1	83	797	>0.05
rs1799929	NAT2	8	18,400,484	C	T	783	66	2	>0.05
rs1801280	NAT2	8	18,400,344	T	C	800	80	2	>0.05
rs17116806	DPYD	1	97,507,696	C	A	484	348	49	>0.05
rs1801159	DPYD	1	97,515,839	T	C	475	350	46	>0.05
rs4970722	DPYD	1	97,886,497	A	T	5	122	754	>0.05
rs1871450	CHST3	10	72,012,256	G	A	634	223	19	>0.05
rs4148943	CHST3	10	72,009,749	C	T	604	238	26	>0.05
rs4148945	CHST3	10	72,009,832	C	T	630	233	18	>0.05
rs4148950	CHST3	10	72,011,948	G	A	638	223	18	>0.05
rs730720	CHST3	10	72,013,004	C	T	599	239	26	>0.05

**Abbreviations:** HWP val, Hardy-Weinberg equilibrium *P* value; SNP, Single nucleotide polymorphism; Chr, Chromosome; A, reference allele; B, altered allele; HLJ, Heilongjiang; AA, wild homozygote; AB, heterozygote; BB mutant homozygote.

of haplotype among the five ethnic groups was different. As shown in Table 2, one block of five SNPs located in the *CHST3* gene was constructed, and the most prevalent haplotype was CCGGC, followed by TTAAT, TCGGT. The results showed that the frequency distributions of CCGGC and TTAAT in Ewenki vs Daur populations and Mongolian vs Ewenki populations were statistically significant ( $P < 0.005$ ). The second block consists of four SNPs in the *VKORC1* gene, the most prevalent haplotype was CAGA, followed by TGCG. The CAGA haplotype frequencies in Manchu vs Mongolian, Ewenki vs Daur, Hezhen vs Daur, Manchu vs Ewenki and Hezhen vs Manchu are statistically in the comparison of the

differences, and TGCG haplotype frequency only in Ewenki vs Daur and Hezhen vs Manchu two comparison differences statistically. The third block includes three SNPs in the *CYP1A2* gene. AGC is the main haplotype in the five ethnic minorities, followed by CGC and AAT. AGC haplotype showed the most statistical difference between Manchu and Mongolian ( $P < 0.005$ ). The haplotype of CGC were statistically different between Manchu and Mongolian, Hezhen and Mongolian, and Mongolian and Daur. The haplotype of AAT were statistically different between Mongolian and Daur and between Hezhen and Ewenki. The fourth block includes three SNPs in the *CYP2E1* gene. The main haplotype in the five nationalities

**Table 2** Composition and Frequency Distribution of Haplotypes of Five Nationalities

Block	Haplotype	Gene	Daur	Ewenki	Manchu	Hezhen	Mongolian
rs4148943, rs4148945, rs4148950, rs1871450, rs730720	CCGGC	<i>CHST3</i>	0.811	0.890	0.822	0.839	0.809
	TTAAT	<i>CHST3</i>	0.165	0.068	0.143	0.149	0.148
	TCGGT	<i>CHST3</i>	0.021	0.021	0.025	0.012	0.019
rs7294, rs2359612, rs8050894, rs9934438	CAGA	<i>VKORC1</i>	0.936	0.877	0.943	0.900	0.901
	TGCG	<i>VKORC1</i>	0.060	0.116	0.057	0.092	0.085
rs762551, rs2472304, rs2470890	AGC	<i>CYP1A2</i>	0.517	0.514	0.452	0.516	0.552
	CGC	<i>CYP1A2</i>	0.393	0.308	0.391	0.372	0.292
	AAT	<i>CYP1A2</i>	0.090	0.178	0.150	0.112	0.156
rs9419082, rs8192766, rs2031920	GTC	<i>CYP2E1</i>	0.712	0.637	0.568	0.626	0.692
	AGT	<i>CYP2E1</i>	0.159	0.199	0.239	0.179	0.165
	AGC	<i>CYP2E1</i>	0.094	0.130	0.168	0.181	0.114

is GTC, followed by AGT and AGC, among which the GTC haplotype showed statistical differences in the comparison of Manchu vs Mongolian, Hezhen vs Daur and Hezhen vs Manchu. The haplotype of AGT were only statistically different between the Manchu and the Mongolian, and the haplotype of AGC were statistically different between Hezhen vs Mongolian and Hezhen vs Daur.

## Significant Difference SNP and Drug Related Information

In addition, from the PharmGKB database, we found that the basic drug-related information for loci with statistical differences is summarized in [Supplementary Table 9](#). Different genotype compositions of rs8192766 and rs4646421 did not affect the increased risk of bladder tumors associated with caffeine exposure. Rs1801280 was associated with cisplatin or cyclophosphamide, but CC genotype was not associated with an increased risk of anemia in women treated with cisplatin and cyclophosphamide compared with the TT genotype. Rs4148943 and rs730720 in *CHST3* gene are associated with docetaxel and thalidomide, and allele changes of these two variants can affect the treatment risk of prostate patients. For rs12248560, located on *CYP2C19*, the T allele was associated with an increased risk of bleeding in patients with coronary artery disease treated with clopidogrel when compared with the C allele. So far, the relationship between *CYP2E1* rs9419082 and drugs has not been reported in literature.

Further, we made a linear distribution of the frequency of rs8050894 G allele and rs9934438 G allele in *VKORC1* gene in five populations and 26 populations, as shown in [Supplementary Figure 1](#). The results showed that rs8050894

G allele frequency in EAS population was higher than that in other regions of the world, while rs9934438 G allele frequency in East Asian population was lower than that in other regions of the world. The distribution of these two alleles in five different populations we studied was consistent with that in EAS populations.

## Genetic Parameters of Five Populations

Popgene software was used to calculate the genetic parameters among five ethnic groups in Heilongjiang Province. First of all, the heterozygosity statistics of all loci in the five ethnic groups are shown in [Table 3](#). The average heterozygotic value of 32 SNPs in five populations was 0.2494. Rs28360521 located on *CYP2D6* gene has a relatively large heterozygosity, while rs12248560 located on *CYP2C19* gene has the smallest heterozygosity. As shown in [Table 4](#), the genetic distances of Nei's for pair-for-pair-comparison of the five ethnic minorities ranged from 0.0018 to 0.0051. The upper part of the diagonal is the genetic consistency. The larger the value is, the higher the consistency is between the two populations. The consistency value of Hezhen and Manchu is 0.9982, indicating that the genetic consistency of Hezhen and Manchu is higher. The lower part of the diagonal is the genetic distance, the higher the value is, the farther the relationship is. The maximum value of the genetic distance between Ewenki and Manchu is 0.0051.

## Analysis of Genetic Structure and Population Membership

Here, we conducted a population genetic analysis of the five ethnic minorities. We performed a series of principal

**Table 3** Summary of Heterozygosity Statistics for All Loci in Five Ethnic Populations

Locus	Sample Size	Obs_Hom	Obs_Het	Exp_Hom	Exp_Het	Nei	Ave_Het
rs4646421	1670	0.5401	0.4599	0.5243	0.4757	0.4755	0.4770
rs4646422	1762	0.7877	0.2123	0.7960	0.2040	0.2039	0.1963
rs2606345	1764	0.8424	0.1576	0.8396	0.1604	0.1603	0.1485
rs2470890	1764	0.7721	0.2279	0.7790	0.2210	0.2209	0.2347
rs2472304	1762	0.7719	0.2281	0.7771	0.2229	0.2228	0.2357
rs762551	1764	0.5329	0.4671	0.5349	0.4651	0.4648	0.4521
rs12248560	1742	0.9598	0.0402	0.9606	0.0394	0.0394	0.0461
rs4244285	1762	0.6538	0.3462	0.6373	0.3627	0.3625	0.3437
rs4986893	1764	0.8832	0.1168	0.8840	0.1160	0.1159	0.1059
rs1080983	1756	0.6686	0.3314	0.6524	0.3476	0.3474	0.3525
rs2836052	1746	0.5452	0.4548	0.5174	0.4826	0.4824	0.4811
rs2031920	1760	0.6830	0.3170	0.6902	0.3098	0.3096	0.3078
rs2070676	1762	0.7900	0.2100	0.7771	0.2229	0.2228	0.2164
rs8192766	1760	0.5545	0.4455	0.5496	0.4504	0.4501	0.4394
rs9419082	1762	0.5426	0.4574	0.5474	0.4526	0.4523	0.4449
rs4646437	1750	0.8743	0.1257	0.8663	0.1337	0.1336	0.1397
rs2242480	1754	0.7001	0.2999	0.6965	0.3035	0.3033	0.2921
rs2359612	1756	0.8497	0.1503	0.8513	0.1487	0.1487	0.1596
rs7294	1762	0.8593	0.1407	0.8594	0.1406	0.1405	0.1499
rs8050894	1760	0.8489	0.1511	0.8506	0.1494	0.1493	0.1600
rs9934438	1762	0.8490	0.1510	0.8508	0.1492	0.1491	0.1592
rs1208	1762	0.9058	0.0942	0.9081	0.0919	0.0918	0.1040
rs1799929	1702	0.9224	0.0776	0.9211	0.0789	0.0789	0.0849
rs1801280	1764	0.9093	0.0907	0.9092	0.0908	0.0907	0.1037
rs1711680	1762	0.6050	0.3950	0.6217	0.3783	0.3781	0.3835
rs1801159	1742	0.5982	0.4018	0.6211	0.3789	0.3787	0.3832
rs4970722	1762	0.8615	0.1385	0.8613	0.1387	0.1386	0.1243
rs1871450	1752	0.7454	0.2546	0.7463	0.2537	0.2536	0.2384
rs4148943	1736	0.7258	0.2742	0.7216	0.2784	0.2783	0.2663
rs4148945	1762	0.7355	0.2645	0.7411	0.2589	0.2587	0.2476
rs4148950	1758	0.7463	0.2537	0.7486	0.2514	0.2512	0.2336
rs730720	1728	0.7234	0.2766	0.7198	0.2802	0.2801	0.2690
Mean	1752	0.7496	0.2504	0.7488	0.2512	0.2511	0.2494
St.Dev	–	0.1277	0.1277	0.1294	0.1294	0.1293	0.1258

**Abbreviations:** Obs\_Hom, observed homozygosity; Obs\_Het, observed heterozygosity; Exp\_Hom, expected homozygosity; Exp\_Het, expected heterozygosity; Ave\_Het, Average heterozygosity.

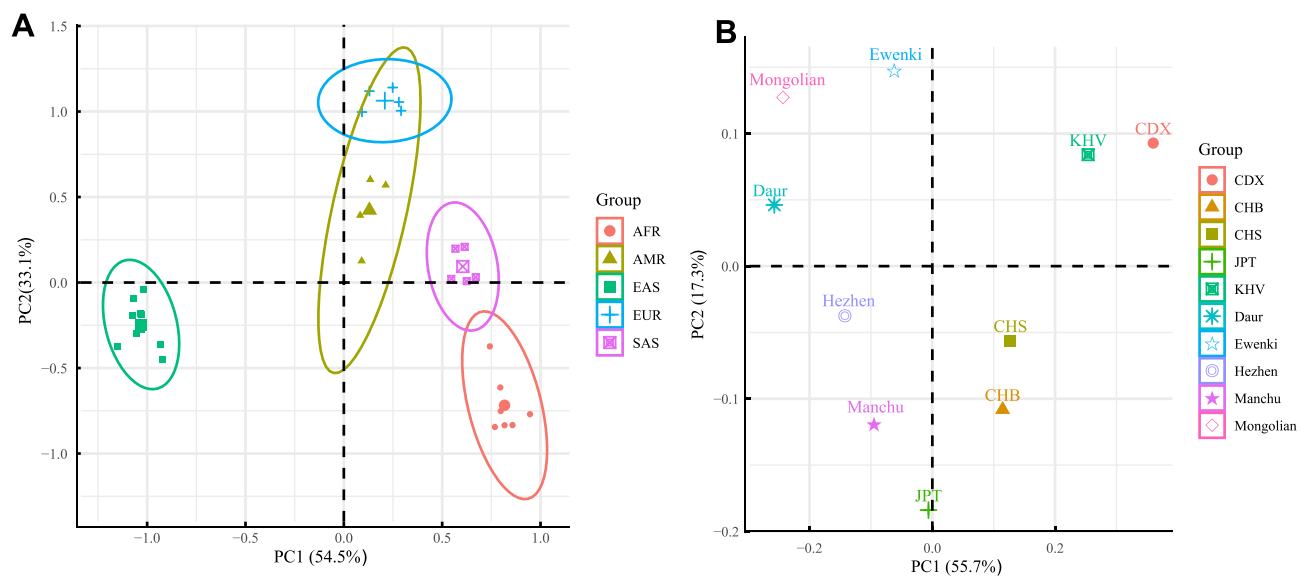
**Table 4** Nei's Unbiased Measures of Genetic Identity and Genetic Distance

Population	Daur	Ewenki	Manchu	Hezhen	Mongolian
Daur	/	0.9952	0.9967	0.9980	0.9981
Ewenki	0.0048	/	0.9949	0.9961	0.9962
Manchu	0.0033	0.0051	/	0.9982	0.9961
Hezhen	0.0020	0.0039	0.0018	/	0.9973
Mongolian	0.0019	0.0039	0.0039	0.0027	/

**Notes:** Nei's genetic identity (above diagonal) and genetic distance (below diagonal); /means no value is required.

component analyses (PCA) based on allele frequencies of 32 SNPs in the Daur, Hezhen, Ewenki, Mongolian, and Manchu populations in the Heilongjiang province and in

26 populations from 1000 Genome Project. **Figure 1A** shows the PCA plot of five ethnic groups from Heilongjiang Province and 26 populations divided into



**Figure 1** Principal component analysis. **(A)** Principal component analysis was performed on 3386 individuals using 32 SNPs from 31 populations. **(B)** Principal component analysis was performed using 32 SNPs from 10 populations in East Asia.

**Abbreviations:** AFR, African; AMR, American; EAS, East Asian; CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese, China; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; EUR, European; SAS, South Asian; PC1, principal component 1; PC2, principal component 2.

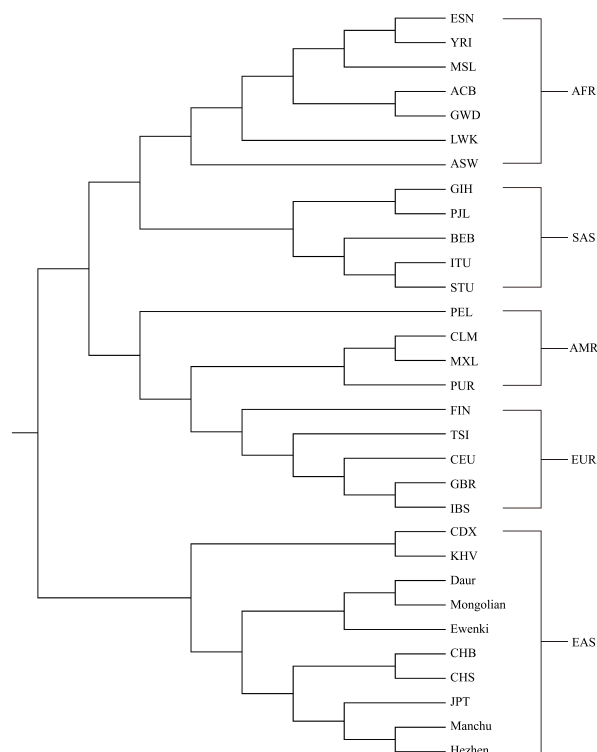
AFR, AMR, EAS, EUR and SAS groups. PC1 and PC2 accounted for 54.5% and 33.1% of the total genetic variation. The results showed that the population of different continents is clearly stratified. The five ethnic minorities in Heilongjiang Province were clustered together with EAS. In addition, [Figure 1B](#) shows the results of principal component analysis of 5 ethnic minorities in Heilongjiang Province and 5 populations in EAS. PC1 accounted for 55.7% of the total genetic variation, which differentiated the ten populations with a north-south trend. PC2 accounted for 17.3% of the total genetic variation. The results showed that the genetic distance between Daur and Mongolian was close, the genetic distance between Manchu and Hezhen was close, CHS was closely related to CHB, and CDX was closely related to KHV.

To further explore the genetic clustering of each ethnic group, we used Popgene to calculate a Dendrogram index and clustered the 5 ethnic populations and 26 populations from the 1000 Genome Project via MEGA. As shown in [Figure 2](#), when the phylogenetic tree was constructed based on 31 populations, five main clusters were identified: African cluster, American cluster, European cluster, South Asian cluster and East Asian cluster. Five ethnic minority population from Heilongjiang Province in China were dispersed in the East Asian population, which was consistent with the results of principal component analysis. In the East Asian branch, on the one hand, the Manchu and

Hezhen ethnic groups cluster together, which then converges with JPT, while CHB and CHS cluster together, on the other hand, the Daur and Mongolian cluster together, which then converges with Ewenki, and CDX and KHV cluster together.

## Discussion

In recent decades, with the completion of three major projects, namely the Human Genome Project, the International Human Genome Haplogram Project and the International 1000 Genome Project, many studies of human genetic diversity around the world have inferred the global distribution of pharmacogenomic variation and defined its population structure, which is consistent with the geographical location of the population.<sup>26–29</sup> However, previous studies in China have been limited to the Han population and a small minority of ethnic groups. In the present study, genetic polymorphisms of drug-metabolism-related genes were analyzed in 5 ethnic groups in northern China and 26 populations (from 1000 genomic databases) from around the world. The results showed that the allele frequencies of 32 SNPs in 10 genes were different among these populations. This study not only provides useful genomic diversity data for the further study of pharmacogenomics among these populations, but also reveals the reasons for heterogeneity of drug responses in populations of different ethnic origins.



**Figure 2** Phylogenetic trees of 31 populations from around the world.

**Abbreviations:** AFR, African; ACB, African Caribbean in Barbados; ASW, African ancestry in southwestern USA; ESN, Esan in Nigeria; GWD, Gambian in Western Division, The Gambia; LWK, Luhya people in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria; AMR, American; CLM, Colombian in Medellin, Colombia; MXL, Mexican Ancestry in Los Angeles, California; PEL, Peruvian in Lima, Peru; PUR, Puerto Rican in Puerto Rico; EAS, East Asian; CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese, China; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; EUR, European; CEU, Utah residents with Northern and Western European ancestry; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian populations in Spain; TSI, Tuscans in Italy; SAS, South Asian; BEB, Bengali in Bangladesh; GIH, Gujarati Indians in Houston, Texas, USA; ITU, Indian Telugu in the UK; PJI, Punjabi in Lahore, Pakistan; STU, Sri Lankan Tamil in the UK.

Among the statistically significant drug-metabolism-related gene polymorphisms we analyzed, most of them were located in *CYP450* gene, the most important first-stage drug metabolism gene.<sup>30</sup> Many studies have proven that genetic variation of *CYP450* gene was associated with drug metabolism in different diseases.<sup>10,31,32</sup> In 2013, P. Maisano Delser et al showed that the locus encoding cytochrome P450s showed a significantly higher level of genetic diversity among populations than other gene categories.<sup>26</sup> In the present study, SNPs rs8192766 and rs9419082, located on *CYP2E1* gene, had different distribution among populations although after multiple comparisons. The *CYP2E1* gene encodes enzymes that are involved in the metabolism and biological activation of

a large number of low molecular compounds. Studies have shown that the loci of rs8192766 of *CYP2E1* gene polymorphisms in Chinese Han population have a significant statistical relationship with schizophrenia susceptibility.<sup>33</sup> Therefore, it is reasonable to speculate that the five ethnic groups may have different levels of schizophrenia susceptibility due to rs8192766 locus polymorphism. In addition, the enzyme encoded by *CYP2C19* gene is involved in the metabolism of clopidogrel, a commonly used drug for the treatment of coronary artery disease in clinical practice. Previous results showed that more than 50% of the carriers of the IM type of *CYP2C19* were in the Han, Tibetan and Miao ethnic groups, while more than 30% of the UM type of *CYP2C19* were in the Uyghur and Kazak ethnic groups. In this study, SNPs of *CYP2C19* gene showed significant differences in allele frequencies among different populations, so *CYP2C19* gene testing should be carried out prior to corresponding drug treatment in accordance with FDA's warning.<sup>34</sup> Therefore, the heterogeneity of metabolic types should be considered when different populations take drugs to prevent the occurrence of adverse drug reactions. Based on the results of previous pharmacogenetic observation, further studies on the genetic variants related to important drug metabolism in different regions of the population may have new findings, which may help to predict the general drug response of individuals or the trend of drug response of the population.

Widespread diversity is also shown when the haplotypes of the five ethnic groups are considered. We mainly found four haplotypes of diversity, which were located on *CHST3*, *VKORC1*, *CYP1A2* and *CYP2E1* genes. Five SNPs in *CHST3* gene (rs4148943, rs1871450, rs4148950, rs730720, rs4148945) were associated with the sensitivity of Docetaxel and Thalidomide to the clinical treatment of prostate cancer. Both drugs were most effective when the haplotype composition of the five polymorphic loci was "CTTCC".<sup>14</sup> Although there were five minority loci without such haplotype, alleles at each locus of the haplotype affected the metabolism of Docetaxel and Thalidomide. Therefore, we strongly recommend that different groups of prostate cancer patients conduct genetic testing before using these two drugs, so as to prevent the occurrence of adverse reactions and optimize the efficacy of the drugs according to the actual situation of the patients. The SNPs of rs7294, rs2359612, rs8050894 and rs9934438 on *VKORC1* gene are all correlated with the sensitivity of warfarin, a drug used in the clinical treatment of thromboembolic diseases.



When the alleles of rs7294, rs2359612, rs805894 and rs9934438 are “G”, “T”, “G” and “C” respectively, the effective warfarin dose required by thromboembolism patients is lower.<sup>9,12,35,36</sup> The FDA has specified that the use of warfarin in clinical trials should be subject to genetic testing for *VKORC1*.<sup>34</sup> Our research data made up for the lack of pharmacogenomic data among Chinese ethnic minorities, and provided negligible help for the clinical medication of five ethnic minorities in Heilongjiang Province, China.

Through population genetic clustering analysis, the distribution of different populations was further clarified. Most populations are clustered according to their geographic locations in Asia, Europe, Africa, and the Americas. In this study, PCA results show that five ethnic minorities in Heilongjiang Province, China are associated with East Asian groups, which may be due to the geographical location of Heilongjiang Province in Northern China and East Asia resulting in a possible genetic fusion. This fusion increases the genetic diversity of genes. In addition, the principal component analysis of the five ethnic minorities found that the Daur ethnic group was close to the Mongolian ethnic group, which was consistent with the results of phylogenetic tree. Interestingly, the distribution of drug-metabolism-related genes is consistent with the human linguistic language population and its origins. Five ethnic groups in this study show clear language clustering tendency, Daur and Mongolian belong to the Altaic language family - The Mongolian languages, Manchu and Hezhen nationality belong to the Altaic - Manchu-Tungusic languages, the Ewenki people also belong to the Altaic language family with the Tungusic languages but its branch of Bei language.<sup>37</sup> Interestingly, the grouping of this linguistic family is consistent with the clustering results of this study. Consequently, our data may provide a basis for future studies on drug metabolism and clinical treatment of ethnic groups with different population structures.

Nevertheless, there are still many defects in this study, the first of which is the limitation of sample size. Secondly, the target variants of this study were not comprehensive enough, and only a small number of pharmacogenetic variants were studied. Last but not least, the relationship between genetic variants of different populations and complex diseases still needs to be explored in depth, so as to improve drug efficacy in clinical drug therapy, reduce adverse drug reactions, and promote drug development and clinical rational drug use and precision medicine.<sup>38</sup>

## Conclusion

In conclusion, our results demonstrate that SNPs in drug metabolism-related genes are similar and heterogeneous in five ethnic groups. At the pharmacogenomic level, five ethnic groups in Heilongjiang province clustered with East Asians. The genetic distance between Daur and Mongolian is the closest, followed by Ewenki, and the genetic distance between Hezhen and Manchu is also close. We believe that pharmacogenomic differences between different populations should be paid attention to, and genetic testing should be conducted before the treatment of corresponding drugs, so as to improve the availability of drugs.

## Ethics Approval and Informed Consent

This study was approved by the Ethics Committee of Harbin Medical University. All volunteers signed informed consent prior to participating.

## Consent for Publication

The authors have declared that they agreed to publish.

## Acknowledgments

All participants in this study are appreciated for their valuable contributions.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Disclosure

The authors declare that they have no competing interests in this work.

## References

1. Zhou Y, Lauschke VM. Comprehensive overview of the pharmacogenetic diversity in Ashkenazi Jews. *J Med Genet*. 2018;55(9):617–627. doi:10.1136/jmedgenet-2018-105429
2. Li J, Lou H, Yang X, et al. Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. *J Med Genet*. 2014;51(9):614–622. doi:10.1136/jmedgenet-2014-102530

3. Donaldson LJ, Kelley ET, Dhingra-Kumar N, et al. Medication without harm: WHO's Third global patient safety challenge. *Lancet*. 2017;389(10080):1680–1681. doi:10.1016/S0140-6736(17)31047-4
4. Wester K, Jonsson AK, Spigset O, et al. Incidence of fatal adverse drug reactions: a population based study. *Br J Clin Pharmacol*. 2008;65(4):573–579. doi:10.1111/j.1365-2125.2007.03064.x
5. Chowbay B, Zhou S, Lee EJD. An interethnic comparison of polymorphisms of the genes encoding drug-metabolizing enzymes and drug transporters: experience in Singapore. *Drug Metab Rev*. 2005;37(2):327–378. doi:10.1081/DMR-28805
6. Zuo LJ, Guo T, Xia DY, et al. Allele and genotype frequencies of CYP3A4, CYP2C19, and CYP2D6 in Han, Uighur, Hui, and Mongolian Chinese populations. *Genet Test Mol Biomarkers*. 2012;16(2):102–108. doi:10.1089/gtmb.2011.0084
7. Kuehl P, Zhang J, Lin Y, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat Genet*. 2001;27(4):383–391. doi:10.1038/86882
8. Zhang C, Jiang X, Chen W, et al. Population genetic difference of pharmacogenomic VIP gene variants in the Lisu population from Yunnan Province. *Medicine*. 2018;97(52):e13674. doi:10.1097/MD.00000000000013674
9. Jin T, Zhao R, Shi X, et al. Genetic polymorphisms study of pharmacogenomic VIP variants in Han ethnic of China's Shaanxi province. *Environ Toxicol Pharmacol*. 2016;46:27–35. doi:10.1016/j.etap.2016.06.026
10. McGraw J, Waller D. Cytochrome P450 variations in different ethnic populations. *Expert Opin Drug Metab Toxicol*. 2012;8(3):371–382. doi:10.1517/17425255.2012.657626
11. Magalon H, Patin E, Austerlitz F, et al. Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. *Eur J Hum Genet*. 2008;16(2):243–251. doi:10.1038/sj.ejhg.5201963
12. Sun B, Wen YF, Culhane-Pera KA, et al. Differences in predicted warfarin dosing requirements between Hmong and East Asians using genotype-based dosing algorithms. *Pharmacotherapy*. 2021;41(3):265–76. doi:10.1002/phar.2487
13. Yan M, Li D, Zhao G, et al. Genetic polymorphisms of pharmacogenomic VIP variants in the Yi population from China. *Gene*. 2018;648:54–62. doi:10.1016/j.gene.2018.01.040
14. Deeken JF, Cormier T, Price DK, et al. A pharmacogenetic study of docetaxel and thalidomide in patients with castration-resistant prostate cancer using the DMET genotyping platform. *Pharmacogenomics J*. 2010;10(3):191–199. doi:10.1038/tpj.2009.57
15. Rajman I, Knapp L, Morgan T, et al. African genetic diversity: implications for cytochrome P450-mediated drug metabolism and drug development. *EBioMedicine*. 2017;17:67–74. doi:10.1016/j.ebiom.2017.02.017
16. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Clin Pharmacol Ther*. 2013;138(1):103–141.
17. Jin T, Aikemu A, Zhang M, et al. Genetic polymorphisms analysis of pharmacogenomic VIP variants in Miao ethnic group of Southwest China. *Med Sci Monit*. 2015;21:3769–3776. doi:10.12659/MSM.895191
18. Zhang J, Jin T, Yunus Z, et al. Genetic polymorphisms of VIP variants in the Tajik ethnic group of northwest China. *BMC Genet*. 2014;15(1):102. doi:10.1186/s12863-014-0102-y
19. Gudbjartsson DF, Helgason H, Gudjonsson SA, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet*. 2015;47(5):435–444. doi:10.1038/ng.3247
20. Hehir-Kwa JY, Marschall T, Kloosterman WP, et al. A high-quality human reference panel reveals the complexity and distribution of genomic structural variants. *Nat Commun*. 2016;7(1):12989. doi:10.1038/ncomms12989
21. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017;45(D1):D896–D901. doi:10.1093/nar/gkw1133
22. Gao Y, Zhang C, Yuan L, et al. PGG.Han: the Han Chinese genome database and analysis platform. *Nucleic Acids Res*. 2020;48(D1):D971–D6. doi:10.1093/nar/gkz829
23. Zhang J, Wang H, Niu G, et al. Deciphering DMET genetic data: comprehensive assessment of Northwestern Han, Tibetan, Uyghur populations and their comparison to eleven 1000 genome populations. *Artif Cell Nanomed B*. 2019;46(sup3):S1176–S85. doi:10.1080/21691401.2018.1533849
24. Chen W, Ding H, Cheng Y, et al. Genetic polymorphisms analysis of pharmacogenomic VIP variants in Bai ethnic group from China. *Mol Genet Genomic Med*. 2019;7(9):e884. doi:10.1002/mgg3.884
25. Rodrigues JCG, Fernandes MR, Guerreiro JF, et al. Polymorphisms of ADME-related genes and their implications for drug safety and efficacy in Amazonian Amerindians. *Sci Rep*. 2019;9(1):1–8.
26. Maisano Delser P, Fuselli S. Human loci involved in drug biotransformation: worldwide genetic variation, population structure, and pharmacogenetic implications. *Hum Genet*. 2013;132(5):563–577. doi:10.1007/s00439-013-1268-5
27. Zhou Y, Ingelman-Sundberg M, Lauschke VM. Worldwide distribution of cytochrome P450 alleles: a meta-analysis of population-scale sequencing projects. *Clin Pharmacol Ther*. 2017;102(4):688–700. doi:10.1002/cpt.690
28. Kozyra M, Ingelman-Sundberg M, Lauschke VM. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet Med*. 2017;19(1):20–29. doi:10.1038/gim.2016.33
29. Abecasis GR, Auton A, Brooks LD, et al.; Genomes Project C. An integrated map of genetic variation from 1092 human genomes. *Nature*. 2012;491(7422):56–65.
30. Zhou S-F, Liu J-P, Chowbay B. Polymorphism of human cytochrome P450 enzymes and its clinical impact. *Drug Metab Rev*. 2009;41(2):89–295.
31. Al-Eitan LN, Rababa'h DM, Hakooz NM, et al. Genetic polymorphisms of pharmacogenes among the genetically isolated Circassian subpopulation from Jordan. *J Pers Med*. 2020;10(1):2. doi:10.3390/jpm10010002
32. Jaja C, Burke W, Thummel K, et al. Cytochrome p450 enzyme polymorphism frequency in indigenous and native American populations: a systematic review. *Community Genet*. 2008;11(3):141–149.
33. Huo R, Tang K, Wei Z, et al. Genetic polymorphisms in CYP2E1: association with schizophrenia susceptibility and risperidone response in the Chinese Han population. *PLoS One*. 2012;7(5):e34809. doi:10.1371/journal.pone.0034809
34. Relling MV, Klein TE. CPIC: Clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. *Clin Pharmacol Ther*. 2011;89(3):464–467. doi:10.1038/clpt.2010.279
35. Geisen C, Watzka M, Sittlinger K, et al. VKORC1 haplotypes and their impact on the inter-individual and inter-ethnic variability of oral anticoagulation. *Thromb Haemost*. 2005;94(4):773–779.
36. Askari B, Khaleqsefat E, Khalafkhani D, Khalaj-Kondori M, Khademvatan K, Soraya H. Study on a novel polymorphism in the VKORC1 promoter region using bioinformatic tools and warfarin dosing data. *Thromb Res*. 2017;158:76–78. doi:10.1016/j.thromres.2017.08.012
37. Peng L. An overview of the seed origin myths of the Chinese Altaic language family ethnic group. *J Hulunbuir Univ*. 2021;29(3):143–148.
38. Zhang H, De T, Zhong Y, Perera MA. The advantages and challenges of diversity in pharmacogenomics: can minority populations bring us closer to implementation? *Clin Pharmacol Ther*. 2019;106(2):338–349. doi:10.1002/cpt.1491

## Pharmacogenomics and Personalized Medicine

Dovepress

### 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>