

ORIGINAL ARTICLE

Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy

Jing Li,¹ Haiyi Lou,¹ Xiong Yang,¹ Dongsheng Lu,¹ Shilin Li,² Li Jin,² Xinwei Pan,² Wenjun Yang,³ Manshu Song,⁴ Dolikun Mamatyusupu,⁵ Shuhua Xu¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2014-102530>).

For numbered affiliations see end of article.

Correspondence to

Professor Shuhua Xu, Max Planck Independent Research Group on Population Genomics, Chinese Academy of Sciences and Max Planck Society (CAS-MPG) Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; xushua@picb.ac.cn

Received 13 May 2014

Revised 7 July 2014

Accepted 7 July 2014

ABSTRACT

Background Drug absorption, distribution, metabolism and excretion (ADME) contribute to the high heterogeneity of drug responses in humans. However, the same standard for drug dosage has been applied to all populations in China although genetic differences in ADME genes are expected to exist in different ethnic groups. In particular, the ethnic minorities in northwestern China with substantial ancestry contribution from Western Eurasian people might violate such a single unified standard.

Methods In this study, we used Affymetrix SNP Array 6.0 to investigate the genetic diversity of 282 ADME genes in five northwestern Chinese minority populations, namely, Tajik, Uyghur, Kazakh, Kirgiz and Hui, and attempted to identify the highly differential SNPs and haplotypes and further explore their clinical implications.

Results We found that genetic diversity of many ADME genes in the five minority groups was substantially different from those in the Han Chinese population. For instance, we identified 10 functional SNPs with substantial allele frequency differences, 14 functional SNPs with highly different heterozygous states and eight genes with significant haplotype differences between these admixed minority populations and the Han Chinese population. We further confirmed that these differences mainly resulted from the European gene flow, that is, this gene flow increased the genetic diversity in the admixed populations.

Conclusions These results suggest that the ADME genes vary substantially among different Chinese ethnic groups. We suggest it could cause potential clinical risk if the same dosage of substances (eg, antitumour drugs) is used without considering population stratification.

INTRODUCTION

Efficacy and safety are among the most important considerations in the discovery and application of drugs. Both are associated with drug metabolism and transportation. To serve their therapeutic effects, drugs in the body must be absorbed, distributed, transported and bound to targets. The toxicity and safety of drugs are determined by their timely metabolism into water-soluble substances and excretion.^{1 2} The diversity of drug absorption, distribution, metabolism and excretion (ADME) genes that encode drug-metabolising enzymes is one of the important factors affecting drug efficacy and safety.

Genetic studies have revealed that different populations have different genetic structures because of their complex demographic histories.³ Genetic differences in ADME genes are also expected to exist in different ethnic groups.⁴ For instance, *CYP3A4* and *CYP3A5*, which metabolise an estimated 50% of the currently used drugs, exhibit significant genetic differences between Africans and non-Africans,⁵ and *CYP1A2*, which mainly metabolises antidepressants, exhibits substantial genetic difference between Europeans and other populations.⁶ It was found the *NAT2* gene, which is involved in detoxification of a large number of chemicals, and its variants associated with slow metabolism are more prevalent in East Asians than in other populations.⁷ We previously investigated the genetic diversity of ADME genes on a worldwide scale;⁸ ancestral information was also found to influence the genetic variations of ADME genes in global populations,⁹ and population specific alleles of ADME genes were discovered to be associated with gene expression level.¹⁰ These results suggest that the population difference of ADME genes could affect the clinic decision-making in different populations. However, most developing countries follow the US FDA/European Medicines Agency guidelines, while using the same therapeutic drug dosages among different populations without considering that the variation of ADME genes could be a potential clinical risk.

The clinical risk of drug response heterogeneity is also present in countries with multiple ethnic groups because of population stratification.¹¹ For instance, the Han Chinese comprises 98% of the total population in China. Population sizes of the other 55 minority groups vary from thousands to millions; some of these ethnic groups differ substantially from the Han Chinese in terms of morphological and genetic characteristics. For example, the Uyghur (UIG) population, the largest minority group in northwestern China, differs significantly from the Han Chinese in terms of facial features and has approximately 55% European genetic components.^{12 13} Similar to UIG, the ethnic groups Tajik (TJK), Kirgiz (KGZ), Kazakh (KZK) and Hui (HUI) are also highly distinguishable from both current Europeans and East Asians due to their multiple ancient genetic sources.^{14 15} The differences in genetic structures between admixed populations and their ancestral source populations (ie, Europeans and East Asians) might also affect the heterogeneity of drug responses.¹⁶

To cite: Li J, Lou H, Yang X, et al. *J Med Genet* Published Online First: [please include Day Month Year] doi:10.1136/jmedgenet-2014-102530

In reality, most people from minority groups are reluctant to participate in clinical tests involving pharmacodynamic and pharmacokinetic studies. Hence, the national dosage standard of China has been mostly determined based on clinical studies in Han Chinese population. This procedure ignores the genetic difference between the minority groups and the Han Chinese. Therefore, it is important and urgent to explore the types of medicines that could elicit different drug responses between the minority groups and the Han Chinese. The genetic differentiation of ADME genes between these populations and the Han Chinese should be considered as a fundamental factor. However, systematical pharmacogenetic studies on these populations are very limited. The present study is the first effort to investigate the diversity patterns of ADME genes between the five northwestern Chinese populations (TJK, UIG, KZK, KGZ and HUI) and Han Chinese in Beijing (CHB). The clinical implications of these genetic variations in drug safety and efficacy are also discussed.

METHODS AND MATERIALS

Genetic variation data

DNA samples of 47 TJK, 42 UIG, 45 KZK and 46 KGZ were collected from the Xinjiang Uyghur Autonomous Region in China, whereas DNA samples of 30 HUI were collected from the Ningxia Hui Autonomous Region in China. All procedures were followed in accordance with the ethical standards of the ethics committee of Fudan University and the Helsinki Declaration of 1975, as revised in 2000.

The genotyping data of five HapMap III¹⁷ population samples (181 Yoruba in Ibadan (YRI), 183 Utah residents with Northern and Western European ancestry (CEU), 91 Toscani in Italia (TSI), 89 CHB and 91 Japanese in Tokyo (JPT)) were included in this study. The full sequencing data of 181 CEU and 89 CHB samples obtained from The 1000 Genomes Project¹⁸ were also included.

Genotyping and data filtering

Genotyping using the Affymetrix Genome-Wide Human SNP Array 6.0 was performed according to the '48 Sample Protocol' (Affymetrix, Genome-Wide Human SNP Nsp/Sty 6.0 User Guide, Rev. 3, 2008, P/N 702504). The raw intensity data were analysed with Birdsuite V1.5.3.¹⁹ After excluding the individuals with genotyping call rate below 90%, SNPs with missing data >10% individuals and SNPs that failed the Hardy-Weinberg equilibrium test ($p < 0.0001$) within each population, we acquired 893 949 autosomal SNPs for further studies.

Phasing and merging datasets

The genotyping data were further phased with fastPHASE²⁰ and merged with PLINK²¹ into the haplotype data of the five populations (CEU, TSI, CHB, JPT and YRI) from HapMap III dataset, and non-overlapping SNPs were removed. The merged dataset consisted of haplotypes for 833 658 SNPs in 845 individuals from 10 populations.

ADME genes

The ADME gene lists were obtained from the PharmaADME database (<http://www.pharmaadme.org/>). After excluding the genes located on sex chromosomes and with less than three polymorphic sites in our data, 32 core ADME genes and 250 extended ADME genes were acquired. According to PharmaADME database, ADME core genes are defined as the most important genes in drug metabolism, while ADME extended genes also play roles in drug metabolism but are

relatively not essential as those core genes. Gene coordinate information was obtained from the RefSeq database,²² and 10 kb upstream and downstream of each gene were included.

Functional annotations of SNPs and haplotypes

The functional effects of each SNP from each ADME gene were obtained based on the variance effect prediction tools from the Ensemble database.²³ Here, the functional SNPs were defined as non-synonymous variations or splice sites annotated in Ensemble or variations affecting gene expression annotated in RegulomeDB dataset,²⁴ and the associated SNPs were defined as those that have been reported to be directly associated with drug response/dose as collected and annotated in the PharmGKB database.²⁵ Therefore, we classified the SNPs into four categories: (1) F_{ASS} : functional SNPs with association reported; (2) F_{NAS} : functional SNPs but without association reported yet; (3) NF_{ASS} : having been reported association but not found precise function yet; and (4) NF_{NAS} : no functional and without reported association currently.

Structure and local ancestry inference of admixed populations

STRUCTURE (V2.3)²⁶ was used to infer the population structure. Here, each STRUCTURE analysis was run with 100 000 burn-ins and 100 000 iterations. The local ancestry inference of each admixed population was estimated with HAPMIX.²⁷ The prior ancestry proportions of five admixed populations were obtained from the STRUCTURE results, and other parameters were set as default. We also obtained the local ancestral components of each gene based on the HAPMIX's results.

Imputing allele frequencies of admixed populations

The expected allele frequencies in the five admixed populations could be imputed from their putative ancestral source populations, that is, CEU and CHB, as follows:

$$f^{CNW,exp} = f^{CEU} \cdot \rho^{Eur} + f^{CHB} \cdot (1 - \rho^{Eur})$$

where f^{CEU} and f^{CHB} denote the derived allele frequency of each locus in CEU and CHB, respectively, ρ^{Eur} represents the contribution of the European ancestry in each locus, and CNW represents one of the five admixed populations.

Statistical and population genetic analyses

Principal component analysis (PCA) was performed at the individual level using EIGENSOFT V3.0.²⁸ Analysis of molecular variance (AMOVA) and F_{ST} calculations were performed with Arlequin 3.5.²⁹ Particularly, the F_{ST} of SNPs were calculated following Weir and Cockerham,³⁰ while the F_{ST} of genes were the average F-statistic over loci. The significance of a gene's F_{ST} was measured with permutation test by randomly shuffling individuals among populations. The empirical p values were also calculated for each gene's F_{ST} and each SNP's F_{ST} by comparing the F_{ST} values of 500 randomly selected genes and all autosomal SNPs, respectively.

The populations studied in this work had different sample sizes. Hence, we randomly sampled 30 individuals from each population and calculated the expected heterozygosity (H_e) on all loci of 282 ADME genes for each population. For each ADME core gene, the significance of haplotype diversity (H_d) among populations was assessed with permutation test by randomly resampling individuals among populations.³¹

Selection signal detection

The integrated haplotype score (iHS) and composite likelihood ratio (CLR) tests were used to detect signals of recent positive selection. The unstandardised iHS scores were calculated by iHS.³² The CLR scores were calculated using SweepFinder.³³ We calculated the iHS score and CLR score for each SNP at whole autosomal regions, and identified top SNPs with top 1% highest scores. The selection candidates were identified with at least three top SNPs found in genes including 10 kb upstream and downstream.

RESULTS

Population differences in ADME genes

PCA was used to depict the differences in ADME genes among two European groups (CEU and TSI), two East Asian groups (CHB and JPT) and five minority groups in northwestern China (HUI, KGZ, KZK, UIG and TJK), as shown in figure 1. The principal components (PC) plot based on data of ADME gene regions showed very similar patterns to that of whole genome data (see online supplementary figure S1). This result indicates the presence of clear population stratification among the three groups (Europeans, East Asians and five northwestern Chinese populations). PC1 (accounting for 47.4% of the total genetic variances) clearly showed that the five minority populations were different from Europeans and East Asians. Especially, these populations were intermediate to Europeans and East Asians. PC2 (accounting for 5% of the total genetic variances) also separated the five minority populations from both Europeans and East Asians. The relative positions of the five minority populations in the PC plot reflect their genetic relationship; as shown in figure 1, HUI is very close to CHB and JPT; TJK is closer to CEU and TSI than to East Asians; and KZK, KGZ and UIG are very close to each other and intermediate to the other populations.

F_{ST} value was calculated for each ADME gene to further quantify the differences in ADME genes among different populations. The results of 32 ADME core genes are listed in table 1. Except for *CYP2B6*, *GSTM1* and *TPMT*, the other 29 ADME core genes had extremely different F_{ST} values (permutated p value <0.01) among different populations, suggesting that most of the ADME

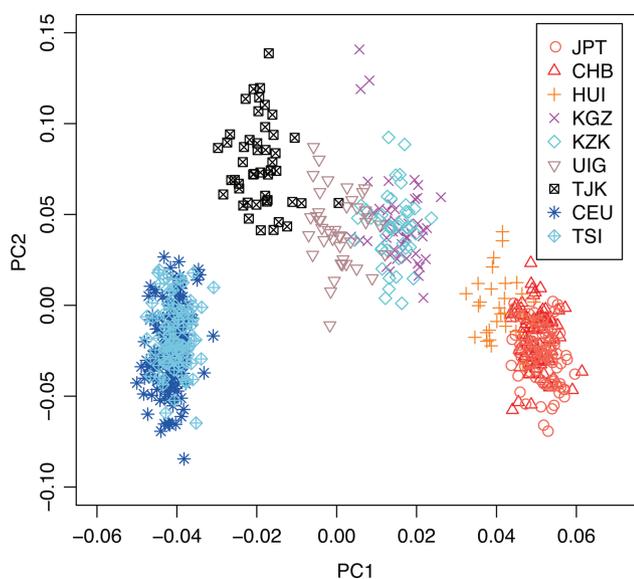


Figure 1 Principal component analysis for 664 individuals from nine populations based on 5462 SNPs from the regions of 282 absorption, distribution, metabolism and excretion genes.

Table 1 AMOVA for 32 absorption, distribution, metabolism and excretion core genes based on all SNPs in each gene region in three groups: Europeans (CEU and TSI), East Asians (CHB and JPT) and five ethnic groups in northwestern China (TJK, UIG, KZK, KGZ and HUI)

Type	Gene	Variant (%)			F_{ST}	p Value
		Between group	Within group	Within populations		
Phase I	<i>CYP1A1</i>	20.43	4.70	74.87	0.251	0.0000
	<i>CYP1A2</i>	9.52	1.99	88.49	0.115	0.0000
	<i>CYP2A6</i>	-1.41	2.43	98.98	0.010	0.0020
	<i>CYP2B6</i>	1.18	-0.28	99.10	0.009	0.0176
	<i>CYP2C8</i>	3.70	1.29	95.01	0.050	0.0000
	<i>CYP2C9</i>	2.82	1.05	96.13	0.039	0.0000
	<i>CYP2C19</i>	1.31	1.12	97.57	0.024	0.0000
	<i>CYP2D6</i>	4.49	1.34	94.17	0.058	0.0000
	<i>CYP2E1</i>	5.99	0.47	93.53	0.065	0.0000
	<i>CYP3A4</i>	11.10	-0.46	89.36	0.106	0.0000
	<i>CYP3A5</i>	11.24	-0.53	89.29	0.107	0.0000
	<i>DPYD</i>	1.50	1.90	96.60	0.034	0.0000
	Phase II	<i>GSTM1</i>	-0.53	1.10	99.43	0.006
<i>GSTP1</i>		5.59	1.39	93.03	0.070	0.0000
<i>GSTT1</i>		3.32	4.21	92.47	0.075	0.0000
<i>NAT1</i>		5.42	0.38	94.20	0.058	0.0000
<i>NAT2</i>		7.37	2.90	89.73	0.103	0.0000
<i>SULT1A1</i>		4.55	0.90	94.56	0.054	0.0000
<i>TPMT</i>		1.35	-0.41	99.06	0.009	0.0166
<i>UGT1A1</i>		1.04	3.48	95.48	0.045	0.0000
<i>UGT2B7</i>		3.27	1.04	95.69	0.043	0.0000
<i>UGT2B15</i>		-1.01	2.88	98.13	0.019	0.0000
<i>UGT2B17</i>		3.58	-0.73	97.15	0.029	0.0010
Transporter	<i>ABCB1</i>	6.30	1.35	92.36	0.076	0.0000
	<i>ABCC2</i>	0.85	1.81	97.34	0.027	0.0000
	<i>ABCG2</i>	2.90	1.70	95.40	0.046	0.0000
	<i>SLC15A2</i>	1.21	5.24	93.54	0.065	0.0000
	<i>SLC22A1</i>	3.11	1.90	95.00	0.050	0.0000
	<i>SLC22A2</i>	1.98	0.36	97.66	0.023	0.0000
	<i>SLC22A6</i>	4.31	0.96	94.73	0.053	0.0000
	<i>SLCO1B1</i>	6.81	3.02	90.17	0.098	0.0000
	<i>SLCO1B3</i>	2.48	1.50	96.02	0.040	0.0000

HUI, Hui; KGZ, Kirgiz; KZK, Kazakh; TJK, Tajik; UIG, Uyghur.

core genes differ substantially among the studied populations. Compared with 500 randomly selected genes, though the p values of most of ADME genes and random genes are very small (see online supplementary figure S2A), the distribution of F_{ST} values of ADME core genes was shifted to higher values compared with random genes, as shown in online supplementary figure S2B. Notably, the F_{ST} values of *CYP1A1* and *CYP1A2* are higher than top 1% threshold (0.112), and the F_{ST} of *CYP3A4*, *CYP3A5* and *NAT2* are higher than top 5% threshold (0.101).

AMOVA was used to further investigate the genetic variations of ADME core genes among populations, as shown in table 1 and online supplementary table S1. Compared with 500 randomly selected genes from the whole genome, *CYP1A1*, *CYP1A2*, *CYP3A4*, *CYP3A5* and *NAT2* exhibited higher variations among the three groups (Europeans, East Asians and northwestern Chinese) than top 1% threshold (7.34%), whereas *SLC15A2* exhibited higher variation between populations within group than top 1% threshold (4.97%). The AMOVA results confirm that some ADME genes vary significantly among different populations compared with randomly selected genes.

Identification of highly differentiated functional/associated SNPs in ADME genes

Pharmacogenetic studies focus on the genetic diversity patterns of functional SNPs, that is, the variants that serve important

Table 2 Highly differentiated functional SNPs with F_{ST} ranking top 1% of all autosomal SNPs (empirical $p < 0.01$)

rsID	Gene	F_{ST}	f_CHB	f_HUI	f_KGZ	f_KZK	f UIG	f_TJK	f_CEU	Function	Type	Substrate
rs2235047	<i>ABCB1</i>	0.235	0.500	0.383	0.207	0.278	0.119	0.075	0.008	Intron	NF_{ASS}	Anthracyclines
rs2470890	<i>CYP1A2</i>	0.225	0.118	0.250	0.196	0.267	0.262	0.170	0.656	Synonymous	NF_{ASS}	Paroxetine
rs1799929	<i>NAT2</i>	0.157	0.023	0.033	0.087	0.111	0.333	0.160	0.410	Synonymous	NF_{ASS}	Isoniazid
rs1208	<i>NAT2</i>	0.133	0.023	0.067	0.120	0.100	0.310	0.266	0.399	Non-synonymous	F_{AAS}	Isoniazid
rs4148323	<i>UGT1A9</i>	0.107	0.214	0.183	0.141	0.244	0.143	0.011	0.000	Non-synonymous	NF_{ASS}	Irinotecan
rs17822931	<i>ABCC11</i>	0.479	0.107	0.966	0.900	0.717	0.589	0.255	0.417	Non-synonymous	F_{NAS}	–
rs2066714	<i>ABCA1</i>	0.238	0.869	0.275	0.233	0.576	0.556	0.692	0.548	Non-synonymous	F_{NAS}	–
rs1056836	<i>CYP1B1</i>	0.125	0.522	0.921	0.800	0.826	0.733	0.819	0.810	Non-synonymous	F_{ASS}	Taxanes
rs12418	<i>CHST3</i>	0.112	0.489	0.101	0.183	0.196	0.378	0.223	0.238	3' UTR	NF_{ASS}	Docetaxel
rs4148945	<i>CHST3</i>	0.108	0.486	0.101	0.200	0.196	0.378	0.234	0.238	3' UTR	NF_{ASS}	Docetaxel

HUI, Hui; KGZ, Kirgiz; KZK, Kazakh; TJK, Tajik; UIG, Uyghur.

functions in drug response heterogeneity. In the present study, we used public annotation databases, including PharmGKB, RegulomeDB and Ensemble, to classify all SNPs into four categories as described in Methods and materials section.

From our dataset, we found 70 functional/associated (F_{ASS} , F_{NAS} and NF_{ASS}) SNPs in the ADME core genes and 543 functional/associated SNPs in the ADME extended genes. Among those SNPs, four highly differentiated functional/associated SNPs (one F_{AAS} and three NF_{ASS} SNPs) were identified in the ADME core genes, and another six (one F_{ASS} , two F_{NAS} and three NF_{ASS} SNPs) were identified in the ADME extended genes, as shown in table 2. These SNPs with high F_{ST} values might lead to heterogeneity in drug response. For instance, compared with ancestral allele (A), the derived allele (C) of the rs2235047 SNP (NF_{ASS}) from the *ABCB1* gene has higher correlation with increased likelihood of cardiotoxicity upon exposure to anthracyclines.³⁴ The derived allele (also the risk allele C) of this SNP is in the highest frequency in CHB (0.500) and the lowest frequency in CEU (0.008). Accordingly, in the five northwestern Chinese groups with different ancestry contributions from eastern (CHB) and western (CEU) populations, the allele frequency of rs2235047 varies from HUI (0.383) to TJK (0.075). These results suggested that the metabolism of *ABCB1*

substrates, such as anthracyclines, leads to different distribution of phenotypes among those populations.

Of the 10 functional/associated SNPs with substantial frequency differences (table 2), the F_{ASS} SNPs (rs1208 and rs1056838) could be directly considered as markers for the heterogeneity of drug response, and the six NF_{ASS} SNPs could be tagged with undetected functional SNPs, although they could also be regarded as potential biomarker because of the reported associations. Furthermore, the two F_{NAS} SNPs might be informative for future clinic verifications since they could affect the protein structures of related enzymes.

Functional/associated SNPs with different heterozygous states among populations

Apart from the SNPs with extremely different frequencies among different populations, variants with highly different heterozygous states are also important in clinical studies. As shown in table 3, 14 functional/associated SNPs (five F_{ASS} , four F_{NAS} and five NF_{ASS} SNPs) with low heterozygous states were present in the Han Chinese population ($H_e < 5\%$); however, high heterozygous states were present in the other populations ($H_e > 15\%$). For instance, with respect to the intronic mutation rs7089580 (NF_{ASS}) in *CYP2C9* (11809A>T), patients with both

Table 3 Highly differentiated heterozygous state of functional SNPs. Listed below are the SNPs with low heterozygosity in Han Chinese population (<5%) but high heterozygosity (>15%) in at least one of the five northwestern Chinese populations

rsID	Gene	He_CHB (%)	He_HUI (%)	He_KGZ (%)	He_KZK (%)	He UIG (%)	He_TJK (%)	He_CEU (%)	Function	Type	Substrate
rs9282564	<i>ABCB1</i>	2.20	6.4	0.0	4.3	11.2	13.8	18.6	Non-synonymous	F_{ASS}	Bisantrene
rs2235040	<i>ABCB1</i>	6.50	15.3	4.3	8.5	15.3	19.0	26.3	Intron_variant	NF_{ASS}	Amitriptyline
rs7089580	<i>CYP2C9</i>	4.40	6.4	14.1	14.3	19.1	19.0	34.8	Intron_variant	NF_{ASS}	Warfarin
rs1799929	<i>NAT2</i>	4.40	6.4	15.9	19.8	44.4	26.8	48.4	Synonymous	NF_{ASS}	Isoniazid
rs1208	<i>NAT2</i>	4.40	12.4	21.1	18.0	42.7	39.0	48.0	Non-synonymous	F_{AAS}	Isoniazid
rs1344279	<i>SLC15A2</i>	9.60	15.3	14.1	23.1	35.0	31.0	32.6	Intron	F_{NAS}	–
rs11045819	<i>SLCO1B1</i>	1.10	0.0	0.0	8.5	9.1	20.7	30.3	Non-synonymous	F_{ASS}	Fluvastatin
rs7435335	<i>UGT2B7</i>	5.50	3.3	12.2	6.4	21.0	32.3	28.1	Intron_variant	F_{NAS}	–
rs17822931	<i>ABCC11</i>	6.50	18.0	40.5	48.4	48.6	38.0	19.0	Non-synonymous	F_{NAS}	–
rs4773866	<i>ABCC4</i>	8.60	6.4	2.2	10.5	21.0	25.4	15.0	Intron	NF_{ASS}	–
rs1800440	<i>CYP1B1</i>	2.20	3.3	15.9	23.1	36.3	45.6	30.6	Non-synonymous	F_{ASS}	Capecitabine
rs41507953	<i>EPHX2</i>	1.10	0.0	2.2	6.4	9.1	22.3	15.5	Non-synonymous	F_{NAS}	–
rs4253778	<i>PPARA</i>	0.00	3.3	12.2	4.3	21.0	13.8	31.3	Intron_variant	NF_{ASS}	Fenofibrate
rs1801282	<i>PPARG</i>	8.60	15.3	15.9	19.8	13.3	32.3	19.0	Non-synonymous	F_{ASS}	Pioglitazone

HUI, Hui; KGZ, Kirgiz; KZK, Kazakh; TJK, Tajik; UIG, Uyghur.

Table 4 Haplotype diversity analysis of absorption, distribution, metabolism and excretion genes. Shown here are genes with at least three functional SNPs defined by PharmaGKB

Gene	Number of SNPs	Hd_CHB	Hd_HUI	Hd_KGZ	Hd_KZK	Hd UIG	Hd_TJK	Hd_CEU	p Value
ABCA1	3	0.806	0.799	0.760	0.714	0.760	0.666	0.479	0.001
ABCB1	6	0.633	0.718	0.635	0.687	0.667	0.688	0.691	0.028
ABCC1	4	0.701	0.647	0.630	0.594	0.610	0.648	0.558	0.001
ABCC2	7	0.780	0.701	0.821	0.796	0.820	0.828	0.797	0.070
ABCC4	4	0.679	0.698	0.648	0.674	0.713	0.723	0.657	0.144
ABCG2	4	0.765	0.776	0.655	0.644	0.665	0.622	0.531	0.001
CYP1A2	3	0.676	0.701	0.676	0.665	0.706	0.626	0.501	0.001
CYP2C8	3	0.479	0.452	0.418	0.510	0.475	0.567	0.494	0.082
CYP39A1	3	0.364	0.490	0.390	0.485	0.617	0.690	0.530	0.001
NAT2	3	0.353	0.476	0.552	0.552	0.689	0.701	0.690	0.001
SLC10A2	3	0.793	0.719	0.799	0.808	0.776	0.761	0.740	0.001
SLC22A14	3	0.478	0.442	0.347	0.445	0.489	0.408	0.418	0.104
SLCO1B1	4	0.620	0.459	0.440	0.558	0.421	0.424	0.501	0.000
UGT1A10	3	0.425	0.453	0.381	0.535	0.536	0.530	0.495	0.070
UGT1A6	3	0.425	0.453	0.381	0.535	0.536	0.530	0.495	0.024
UGT1A7	3	0.425	0.453	0.381	0.535	0.536	0.530	0.495	0.018
UGT1A8	3	0.425	0.453	0.381	0.535	0.536	0.530	0.495	0.024
UGT1A9	3	0.425	0.453	0.381	0.535	0.536	0.530	0.495	0.028

HUI, Hui; KGZ, Kirgiz; KZK, Kazakh; TJK, Tajik; UIG, Uyghur.

the TA and TT genotypes require higher doses of warfarin than patients with the AA genotype.³⁵ At this variation, the expected heterozygosity in CHB is 4.4%, whereas in both UIG and TJK it is $\geq 19.0\%$. The results of F_{ST} screening indicate that the allele frequency differentiation of these SNPs among different Chinese ethnic groups is not significant compared with the entire genome ($F_{ST}=0.061$). However, the great difference of heterozygosity of rs7089580 might indicate more complexity drug responses among the Chinese minority groups compared with Han Chinese, which could lead to unpredicted clinical risk for the drugs listed in table 3.

The 14 functional/associated SNPs with extremely different heterozygous states listed in table 3 could affect response heterogeneity of a broad range of drugs. Examples of drugs that could be affected include the antitumour drugs bisantrene and capecitabine, the antituberculosis drug isoniazid, the anticoagulant drug warfarin, and the hypercholesterolaemia treatment drugs fluvastatin and fenofibrate. Therefore, different dosages should be considered for these drugs and their related substrates due to population stratification.

Identification of highly differentiated functional/associated haplotypes in ADME genes

In pharmacogenetic studies, the clinical phenotypes of drug metabolism are more likely to be determined by haplotypes which are composed of functional/associated variants rather than by single independent SNPs. Table 4 shows the diversity of clinical haplotypes of 18 genes with at least three functional/associated SNPs from our genotyping data among the five minority populations and their ancestral source populations (CHB and CEU). Of these 18 genes, eight showed significantly different haplotype diversity patterns among different populations with permuted p value less than 0.01. For instance, figure 2 shows that NAT2 haplotypes encompass rs1799929 (NF_{ASS}), rs1799930 (F_{ASS}) and rs1208 (F_{ASS}), while CHB was dominated by the haplotype CGA (78%); the five minority groups demonstrated broader haplotype distributions. We could

not directly infer the detailed metabolic phenotype of NAT2 from only three SNPs. Nevertheless, rs1799929 (481C>T), rs1799930 (590G>A) and rs1208 (803G>A) are reported signature alleles for the NAT2 *6, NAT2 *11 and NAT2 *12 haplogroups, which are all associated with either the slow or the rapid acetylator phenotype. Therefore, the different haplotype distributions between Han Chinese and the minority groups suggest that they also differ substantially in metabolism phenotypes for NAT2 substrates.

Inference of ancestral origins of alleles in ADME genes

The above analysis clearly indicates that substantial differences in the ADME genes exist between Han Chinese and the five minority groups. We further determined the factors that may have caused these differences. In our previous study,¹² we reported that UIG is an admixed population with both European and East Asian ancestries. Figure 3A shows the

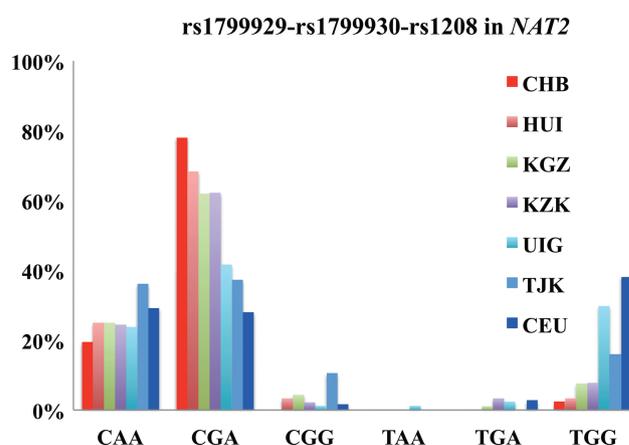


Figure 2 Distribution of NAT2 haplotypes, which were composed of three functional SNPs, namely, rs1799929, rs1799930 and rs1208, in five minority Chinese ethnic groups and their putative ancestral source populations.

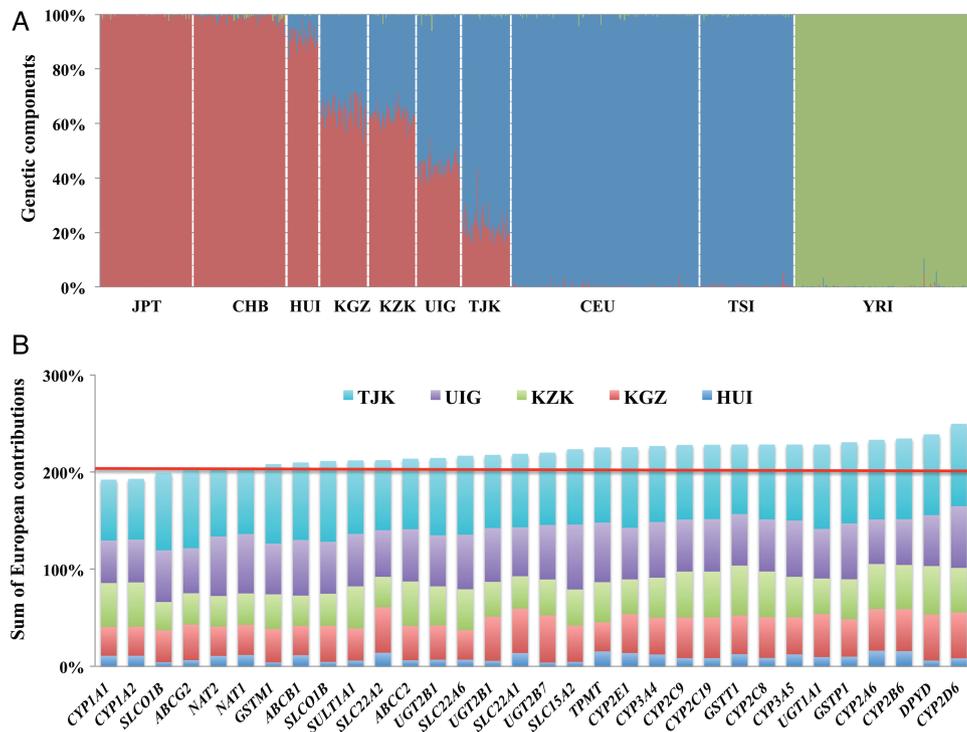


Figure 3 Ancestry inferences for five northwestern Chinese ethnic groups: (A) Global ancestry inferences for 10 populations with STRUCTURE, wherein each column represents the average proportion of ancestral component of each individual. (B) Local ancestral inference of 32 absorption, distribution, metabolism and excretion core genes with HAPMIX, wherein each column represents the average proportion of ancestral European component in each population.

genetic structure analysis of the five populations compared with Africans (YRI), Europeans (CEU and TSI), and East Asians (CHB and JPT). Results showed that the genetic components of the five minority groups were admixed with different proportions of Europeans and East Asians. In detail, the genetic contribution of Europeans was 76.9% to TJK (SD=0.049), 53.5% to UIG (SD=0.053), 38.4% to KZK (SD=0.050), 35.6% to KGZ (SD=0.059) and 9.1% to HUI (SD=0.035).

We used Hapmix to investigate the local ancestral genetic origins of the five admixed populations (see online supplementary figure S3). Figure 3B shows the sum of European ancestry contributions to the 32 ADME core genes in five admixed populations. Overall, the CYP1A2 shows the lowest European contributions (192.1% in total), while the CYP2D6 shows the highest European components (249.6% in total). The red line in figure 3B represents the average value (204.7%) based on whole genome data; thus, totally there are 26 of 32 ADME genes showing higher European contribution than average.

Imputation of functional/associated SNPs from sequence data using admixture analysis

Minority ethnic groups have rarely been sampled for deep sequencing to explore the polymorphic state of functional variants. We explored the possibility to estimate the frequencies of functional SNPs in the admixed populations from their putative ancestral source populations, which have been sequenced and with data available in public database such as The 1000 Genomes Project.

Figure 4 shows acceptable correlation between the observed and expected frequencies of the functional SNPs ($r=0.93$, $p<10^{-15}$) in the five admixed populations. Therefore, we could

roughly infer the frequencies of the functional SNPs for the admixed populations to identify the high-risk alleles that might cause drug response heterogeneity among populations.

Given the significant correlation between the actual and predicted frequencies in the admixed populations, we found 17 and 46 additional functional/associated SNPs from the ADME core and extended genes (table 5 and see online supplementary table S2), respectively, indicating high differentiations between

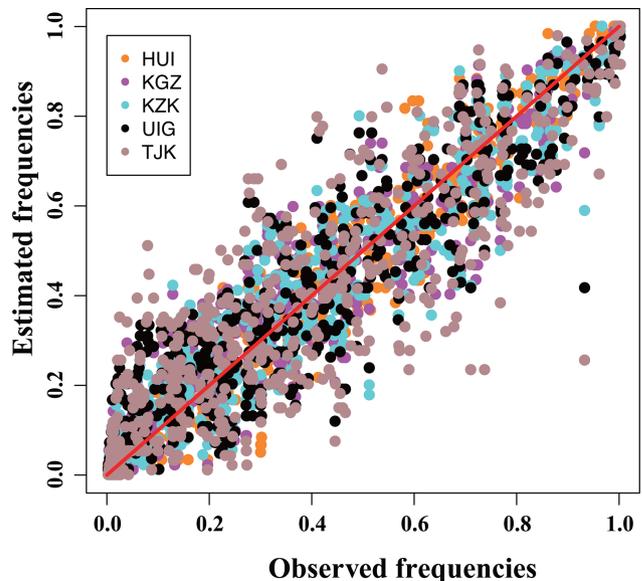


Figure 4 Correlation of observed allele frequencies versus estimated frequencies from CHB and CEU in five northwestern Chinese ethnic groups.

Table 5 Highly differentiated functional SNPs in absorption, distribution, metabolism and excretion core genes between CEU and CHB (empirical $p < 0.01$). Shown here are SNPs based on the 1000 Genomes Project, but data are not available in our genotyping data or in the HapMap III dataset

rsID	Gene	F _{ST}	f_CEU	f_CHB	Function	Type	Substrate
rs2231164	<i>ABCG2</i>	0.194	0.912	0.505	Intron	N _F ASS	Antineoplastic
rs13120400	<i>ABCG2</i>	0.161	0.265	0.000	Intron	N _F ASS	Methotrexate
rs4646421	<i>CYP1A1</i>	0.142	0.088	0.423	Intron	F _{ASS}	Caffeine
rs2606345	<i>CYP1A1</i>	0.438	0.682	0.052	Intron	N _F ASS	Carbamazepine
rs3826041	<i>CYP1A1</i>	0.140	0.877	0.531	Upstream	F _{NAS}	–
rs2470893	<i>CYP1A1</i>	0.225	0.353	0.000	Upstream	N _F ASS	Caffeine
rs2472304	<i>CYP1A2</i>	0.318	0.659	0.113	Intron	N _F ASS	Caffeine
rs146991580	<i>CYP2D6</i>	0.142	0.229	0.603	Non-synonymous	F _N	–
rs9419082	<i>CYP2E1</i>	0.171	0.894	0.510	Intron	F _{NAS}	–
rs8192766	<i>CYP2E1</i>	0.149	0.894	0.541	Upstream	N _F ASS	Caffeine
rs4646458	<i>CYP3A5</i>	0.146	0.988	0.706	Downstream	N _F ASS	Tacrolimus
rs1801280	<i>NAT2</i>	0.245	0.424	0.021	Non-synonymous	F _{NAS}	–
rs2306283	<i>SLCO1B1</i>	0.154	0.588	0.206	Non-synonymous	F _{ASS}	Mycophenolate
rs2417966	<i>SLCO1B1</i>	0.320	0.788	0.222	Intron	F _{NAS}	–
rs10034124	<i>UGT2B17</i>	0.150	0.071	0.407	Intergenic	F _{NAS}	–
rs7435827	<i>UGT2B17</i>	0.130	0.706	0.345	Intron	F _{NAS}	–
rs7434408	<i>UGT2B17</i>	0.215	0.665	0.206	Intron	F _{NAS}	–

CHB and CEU from the full sequencing dataset of The 1000 Genomes Project. These SNPs were still enriched in the functional variants associated with heterogeneous responses to caffeine and antitumour, immunosuppressant and antidepressant drugs. Although these functional/associated SNPs were not found in our dataset, they were supposed to have great difference of frequencies between CHB and the five minority populations because of European gene flow. Thus, the influence of these variants on drug safety and efficacy among different populations must also be considered.

Natural selection on ADME genes in different populations

In the admixed populations, most of the allele frequencies inferred from their ancestral populations were generally fit to their actual frequencies. However, some SNPs still obviously deviated from the expected allele frequency (figure 4). This deviation may be attributed to natural selection. In the present study, we used two methods to detect natural selection signals among the five minority groups and their putative ancestral source populations.

The natural selection signals were enriched in ADME genes (figures 5). For instance, five phase I ADME core genes (*CYP2B6*, *CYP2E1*, *CYP3A4*, *CYP3A5* and *DPYD*) and four core transporter genes (*ABCB1*, *SLC15A1*, *SLC22A1* and *SLCO1B3*) showed significant selection signals in at least two populations. Genes with selective signals were detected in the two ancestral source populations. Selection signals were consistently present in the admixed populations, such as in most of the genes mentioned above. In particular, *CYP3A4* and *CYP3A5* exhibited prevalently selective signals in CEU and most of the admixed populations but not CHB. By contrast, *SLC22A1* and *SLCO1B3* only had selective signals in the admixed populations but not in either CHB or CEU. These results indicate that the adaptation of ADME genes could occur before or after the admixture of ancestral European and Asian populations.

DISCUSSION

In this work, we studied five minority groups in northwestern China that are admixed populations with both eastern and

western Eurasian ancestries. Generally, admixture increases genetic diversity of genomes and genes, including intermediate allele frequencies, heterozygous allele states and haplotype combinations. Consequently, the distribution of different metabolic phenotypes in the admixed populations is expected to be broader than that in non-admixed populations. Thus, more clinical patient samples are needed to determine appropriate drug dosages in those admixed populations. However, people of minority groups usually refuse to participate in clinical studies. Our study should advance the understanding of the genetic basis of the drug response heterogeneity among populations in China, and have significant implications for evaluating potential heterogeneity of drug responses within multiple ethnic groups of Chinese. It is noteworthy that there could be further subpopulation structure within each ethnic group studied here. For example, considerable genetic differences between south and north Han Chinese were reported previously.^{36 37} Those substructures could have unpredictable but substantial influences on evaluation of drug response and great implications for future pharmacogenomic studies. Therefore, we suggest further extensive investigations should be conducted.

It is noteworthy that the genetic variants are only one of the factors affecting drug responses and most of explicit consequences of genetic variants are not yet fully understood. Thus, the phenotypic consequences of population differentiations of ADME genes should be carefully validated in future studies. On the other hand, although the role of ethnicity in pharmacogenomics studies is still debateable, there are essential ethnic consequences of different drug dose requirements among different populations.³⁸ Therefore, individual genotyping/sequencing is necessary in future pharmacogenomic studies because higher heterogeneity of drug responses is also expected in admixed populations and any oversimplified ethnic medicine standards might be inappropriate.

We systemically investigated the influence of admixture to the diversity of ADME genes. The results of this study could be used as reference in future pharmacogenetic studies on admixed populations. TJK, UIG, KZK, KGZ and HUI exhibit different ancestral European components ranging from high to low. The

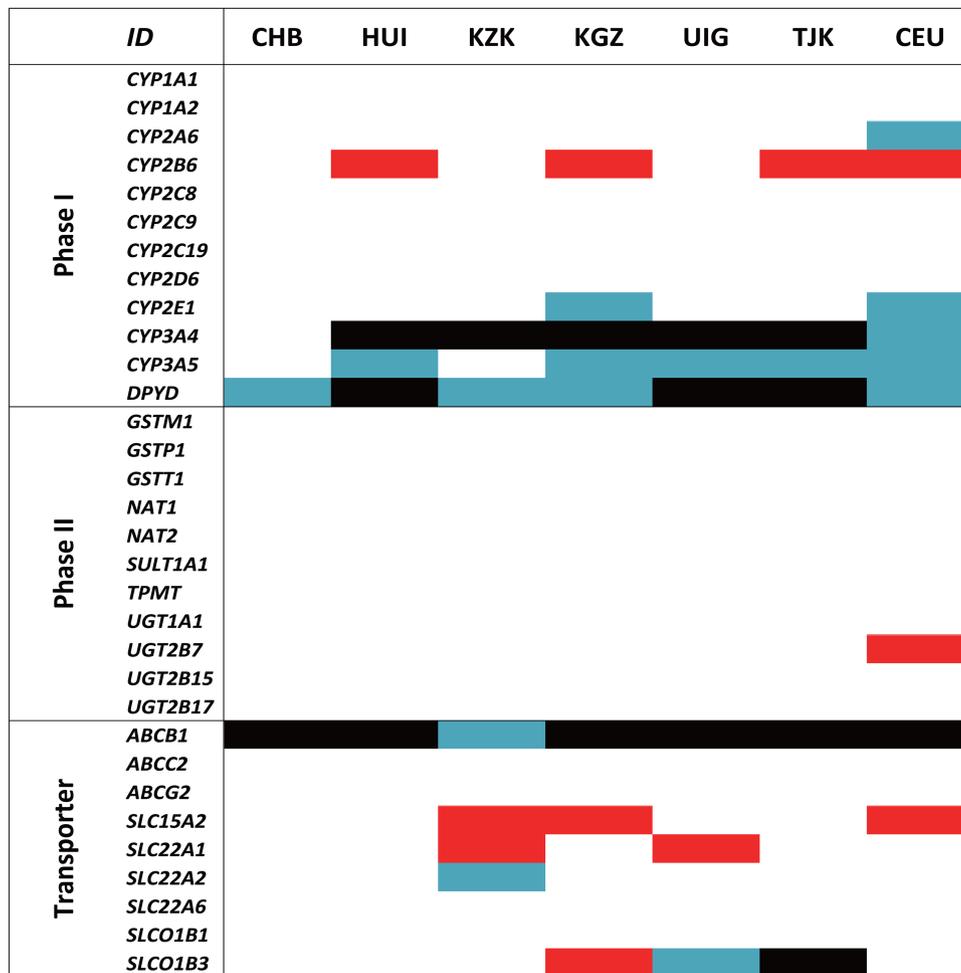


Figure 5 Natural selective signals among 32 absorption, distribution, metabolism and excretion core genes detected in five northwestern Chinese ethnic groups and their putative ancestral source populations. Red blocks represent the signals found by the CLR approach, blue blocks represent the signals found by the iHS approach and black blocks represent the signals detected by both methods.

intermediate admixed proportions exhibit the highest genetic diversity at the entire genome and then at the ADME genes. That is, UIG, KZK and KGZ (average European components are 53.5%, 38.4% and 35.6%) have substantially higher genetic diversity than the other populations, as shown in online supplementary figure S4. These results suggest that ancestral sources of admixed populations are equivalent. All of them exhibit high diversity at the genetic architecture. Conversely, our genotype dataset shows that Europeans exhibit slightly higher genetic diversity than East Asians. Thus, the European components are deemed to have higher contribution to the increase in genetic diversity than the East Asian components. For instance, TJK (76.9% European component) exhibit substantially higher genetic diversity than HUI (9.1% European component) (see online supplementary figure S4). However, despite being unlikely to affect our main results and conclusions, the contribution of the ancestral genetic component could be misleading due to possible ascertainment bias of genetic variants using genotyping approach. Therefore, full sequencing is worthy for future studies.

Aside from demographic histories, such as admixture events, natural selection is another important factor that affects the diversity of genes. Enriched signals of positive selection were observed in the ADME core genes, which are consistent with our previous results.⁸ Interestingly, we found that some ADME genes (*CYP3A4* and *CYP3A5*) showed positively selective signals

before population admixture, while some other genes such as *SLC22A1* and *SLCO1B3* exhibited the selection signals after the admixture events.

CONCLUSIONS

Our results show that the northwestern Chinese populations exhibited substantial differences at some ADME genes compared with the Han Chinese. We suggest population differences should be carefully considered before related drugs are introduced to the Chinese market.

Author affiliations

¹Max Planck Independent Research Group on Population Genomics, Chinese Academy of Sciences and Max Planck Society (CAS-MPG) Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

²State Key Laboratory of Genetic Engineering and Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China

³Key Laboratory of Reproduction and Heredity of Ningxia Region, Ningxia Medical University, Yinchuan, Ningxia, China

⁴School of Public Health, Capital Medical University and Municipal Key Laboratory of Clinical Epidemiology, Beijing, China

⁵College of the Life Sciences and Technology, Xinjiang University, Urumqi, China

Acknowledgements We thank all volunteers for participating in this study. SX also gratefully acknowledges the support of the National Program for Top-notch

Young Innovative Talents of The 'Ten-Thousand-Talents' Project and the support of K.C. Wong Education Foundation, Hong Kong.

Contributors SX conceived the study. SX and JL designed the study. SL, LJ, XP, WY, MS and DM collected samples. JL, HL, XY and DL performed the analyses. SX and JL interpreted the data and wrote the paper. All authors have read and approved the final version of the manuscript. SX is Max-Planck Independent Research Group Leader. SX and LJ are members of CAS Youth Innovation Promotion Association.

Funding This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040100), by the National Science Foundation of China (NSFC) grants (91331204, 31370505, 31171218, 81160249, 30901238 and 30971577), by the Knowledge Innovation Program of Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences (2013KIP108), and by Natural Science Foundation of Xinjiang Uyghur Autonomous Region (2013211A016). Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests None.

Patient consent Obtained.

Ethics approval The ethics committee of Fudan University.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Shashty BS. Pharmacogenetics and the concept of individualized medicine. *Pharmacogenomics J* 2006;6:16–21.
- Sim SC, Kacevska M, Ingelman-Sundberg M. Pharmacogenomics of drug-metabolizing enzymes: a recent update on clinical implications and endogenous effects. *Pharmacogenomics J* 2013;13:1–11.
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008;319:1100–4.
- Delsler PM, Fuselli S. Human loci involved in drug biotransformation: worldwide genetic variation, population structure, and pharmacogenetic implications. *Hum Genet* 2013;132:563–77.
- Chen X, Wang H, Zhou G, Zhang X, Dong X, Zhi L, Jin L, He F. Molecular population genetics of human CYP3A locus: signatures of positive selection and implications for evolutionary environmental medicine. *Environ Health Perspect* 2009;117:1541–8.
- Wooding SP, Watkins WS, Bamshad MJ, Dunn DM, Weiss RB, Jorde LB. DNA sequence variation in a 3.7-kb noncoding sequence 5' of the CYP1A2 gene: implications for human population history and natural selection. *Am J Hum Genet* 2002;71:528–42.
- Sabbagh A, Darlu P, Crouau-Roy B, Poloni ES. Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey. *PLoS ONE* 2011;6:e18507.
- Li J, Zhang L, Zhou H, Stoneking M, Tang K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum Mol Genet* 2011;20:528–40.
- Ramos E, Doumatey A, Elkahlon AG, Shriner D, Huang H, Chen G, Zhou J, McLeod H, Adeyemo A, Rotimi CN. Pharmacogenomics, ancestry and clinical decision making for global populations. *Pharmacogenomics J* 2014;14:217–22.
- Yang HC, Lin CW, Chen CW, Chen JJ. Applying genome-wide gene-based expression quantitative trait locus mapping to study population ancestry and pharmacogenetics. *BMC Genomics* 2014;15:319.
- Suarez-Kurtz G, Paula DP, Struchiner CJ. Pharmacogenomic implications of population admixture: Brazil as a model case. *Pharmacogenomics* 2014;15:209–19.
- Xu S, Huang W, Qian J, Jin L. Analysis of genomic admixture in Uyghur and its implication in mapping strategy. *Am J Hum Genet* 2008;82:883–94.
- Xu S, Jin L. A genome-wide analysis of admixture in Uyghurs and a high-density admixture map for disease-gene discovery. *Am J Hum Genet* 2008;83:322–36.
- Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet* 2004;12:495–504.
- Abilimit A, Qin W, Shan W, Wu W, Ling F, Ling KH, Zhao C, Zhang F, Ma Z, Zheng X. Genetic diversities of cytochrome B in Xinjiang Uyghur unveiled its origin and migration history. *BMC Genet* 2013;14:100.
- Magalon H, Patin E, Austerlitz F, Hegay T, Aldashev A, Quintana-Murci L, Heyer E. 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;16:243–51.
- Alshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu FL, Bonnen PE, de Bakker PIW, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inoué M, Jia XM, Palotie A, Parkin M, Whittaker P, Chang K, Hawes A, Lewis LR, Ren YR, Wheeler D, Muzny DM, Barnes C, Darvishi K, Hurler M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesi J, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Gonzaga-Jauregui C, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Zhang QR, Ghorji MJR, McGinnis R, McLaren W, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng CQ, Brooks LD, McEwen JE, Consortium IH. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52–8.
- Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
- Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008;40:1253–60.
- Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* 2006;78:629–44.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
- Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2007;35(Database issue):D61–5.
- McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010;26:2069–70.
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karzewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–7.
- Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, Altman RB, Klein TE. Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther* 2012;92:414–17.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 2009;9:1322–32.
- Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet* 2009;5:e1000519.
- Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2:e190.
- Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010;10:564–7.
- Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* 1984;38:1358–70.
- Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB. Founding mothers of Jewish communities: Geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* 2002;70:1411–20.
- Voight BF, Kudaravalli S, Wen XQ, Pritchard JK. A map of recent positive selection in the human genome (vol 4, pg 154, 2006). *PLoS Biol* 2006;4:659.
- Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Res* 2005;15:1566–75.
- Visscher H, Ross CJ, Rassekh SR, Barhadi A, Dube MP, Al-Salooos H, Sandor GS, Caron HN, van Dalen EC, Kremer LC, van der Pal HJ, Brown AM, Rogers PC, Phillips MS, Rieder MJ, Carleton BC, Hayden MR, Canadian Pharmacogenomics Network for Drug Safety C. Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. *J Clin Oncol* 2012;30:1422–8.
- Perera MA, Gamazon E, Cavallari LH, Patel SR, Poindexter S, Kittles RA, Nicolae D, Cox NJ. The Missing Association: Sequencing-Based Discovery of Novel SNPs in VKORC1 and CYP2C9 That Affect Warfarin Dose in African Americans. *Clin Pharmacol Ther* 2011;89:408–15.
- Chen J, Zheng H, Bei JX, Sun L, Jia WH, Li T, Zhang F, Seielstad M, Zeng YX, Zhang X, Liu J. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *Am J Hum Genet* 2009;85:775–85.
- Xu S, Yin X, Li S, Jin W, Lou H, Yang L, Gong X, Wang H, Shen Y, Pan X, He Y, Yang Y, Wang Y, Fu W, An Y, Wang J, Tan J, Qian J, Chen X, Zhang X, Sun Y, Zhang X, Wu B, Jin L. Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am J Hum Genet* 2009;85:762–74.
- Pena SD. The fallacy of racial pharmacogenomics. *Braz J Med Biol Res* 2011;44:268–75.